

Generalizable neuromarker for autism spectrum disorder across imaging sites and developmental stages: A multi-site study

Takashi Itahashi

Medical Institute of Developmental Disabilities Research, Showa University

Ayumu Yamashita

Brain Information Communication Research Laboratory Group, Advanced Telecommunications Research Institute International

Yuji Takahara

Brain Information Communication Research Laboratory Group, Advanced Telecommunications Research Institute International

Noriaki Yahata

National Institutes for Quantum and Radiological Science and Technology

Motoaki Nakamura

Medical Institute of Developmental Disabilities Research, Showa University

Yuta Aoki

Medical Institute of Developmental Disabilities Research, Showa University

Ryuta Aoki

Medical Institute of Developmental Disabilities Research, Showa University

Junya Fujino

Medical Institute of Developmental Disabilities Research, Showa University

Haruhisa Ohta

Medical Institute of Developmental Disabilities Research, Showa University

Yujiro Yoshihara

Department of Psychiatry, Graduate School of Medicine, Kyoto University

Yuki Sakai

Brain Information Communication Research Laboratory Group, Advanced Telecommunications Research Institute International <https://orcid.org/0000-0003-2475-8548>

Masahiro Takamura

Department of Psychiatry and Neurosciences, Hiroshima University

Naho Ichikawa

Department of Psychiatry and Neurosciences, Hiroshima University

Go Okada

Department of Psychiatry and Neurosciences, Hiroshima University

Naohiro Okada

Department of Neuropsychiatry, Graduate School of Medicine, The University of Tokyo

Kiyoto Kasai

Department of Neuropsychiatry, Graduate School of Medicine, The University of Tokyo

Saori Tanaka

Brain Information Communication Research Laboratory Group, Advanced Telecommunications
Research Institute International

Hiroshi Imamizu

Brain Information Communication Research Laboratory Group, Advanced Telecommunications
Research Institute International

Nobumasa Kato

Medical Institute of Developmental Disabilities Research, Showa University

Yasumasa Okamoto

Department of Psychiatry and Neurosciences, Hiroshima University

Hidehiko Takahashi

Department of Psychiatry and Behavioral Sciences, Graduate School of Medical and Dental Sciences
Tokyo Medical and Dental University

Mitsuo Kawato

ATR Computational Neurosciences Laboratories

Okito Yamashita

Brain Information Communication Research Laboratory Group, Advanced Telecommunications
Research Institute International

Ryuichiro Hashimoto (✉ dbridges50@gmail.com)

Showa University <https://orcid.org/0000-0002-9661-3412>

Article**Keywords:**

Posted Date: May 15th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-2853362/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.

**Title: Generalizable neuromarker for autism spectrum disorder across
imaging sites and developmental stages: A multi-site study**

Authors:

Takashi Itahashi^{1†}, Ayumu Yamashita^{2,3†}, Yuji Takahara^{2,4}, Noriaki Yahata^{2,5,6}, Yuta Y. Aoki^{1,7}, Junya Fujino^{1,8}, Yujiro Yoshihara⁹, Motoaki Nakamura¹, Ryuta Aoki^{1,10}, Haruhisa Ohta¹, Yuki Sakai², Masahiro Takamura^{11,12}, Naho Ichikawa¹¹, Go Okada¹¹, Naohiro Okada^{6,13}, Kiyoto Kasai^{6,13,14}, Saori C. Tanaka^{2,15}, Hiroshi Imamizu^{2,16}, Nobumasa Kato¹, Yasumasa Okamoto¹¹, Hidehiko Takahashi^{8,9,17}, Mitsuo Kawato^{2,18}, Okito Yamashita^{2,19}, Ryu-ichiro Hashimoto^{1,9*}

Affiliations:

¹Medical Institute of Developmental Disabilities Research, Showa University, Tokyo, Japan.

²Brain Information Communication Research Laboratory Group, Advanced Telecommunications Research Institute International, Kyoto, Japan.

³Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan.

⁴Laboratory for Drug Discovery and Disease Research, SHIONOGI & CO., LTD, Osaka, Japan

⁵Institute for Quantum Life Science, National Institutes for Quantum Science and Technology, Chiba, Japan.

⁶Department of Neuropsychiatry, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan.

⁷Department of Psychiatry, Aoki Clinic, Tokyo, Japan.

⁸Department of Psychiatry and Behavioral Sciences, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan.

⁹Department of Psychiatry, Graduate School of Medicine, Kyoto University, Kyoto, Japan

¹⁰Department of Language Sciences, Tokyo Metropolitan University, Tokyo, Japan.

¹¹Department of Psychiatry and Neurosciences, Hiroshima University, Hiroshima, Japan.

¹²Department of Neurology, Shimane University, Shimane, Japan.

¹³The International Research Center for Neurointelligence (WPI-IRCN) at The University of Tokyo Institutes for Advanced Study (UTIAS), The University of Tokyo, Tokyo, Japan.

¹⁴UTokyo Institute for Diversity and Adaptation of Human Mind (UTIDAHM), The University of Tokyo, Tokyo, Japan.

¹⁵Division of Information Science, Nara Institute of Science and Technology, Nara, Japan.

¹⁶Department of Psychology, Graduate School of Humanities and Sociology, The University of Tokyo, Tokyo, Japan.

¹⁷Center for Brain Integration Research, Tokyo Medical and Dental University, Tokyo, Japan

¹⁸XNef Incorporation, Kyoto, Japan.

¹⁹RIKEN, Center for Advanced Intelligence Project, Tokyo, Japan.

*Corresponding author. E-mail: dbridges50@gmail.com (R.H.)

†These authors contributed equally to this work.

Abstract

Autism spectrum disorder (ASD) is a lifelong condition, and its underlying biological mechanisms remain elusive. The complexity of various factors, including inter-site and development-related differences, makes it challenging to develop generalizable neuroimaging-based biomarkers for ASD. This study used a large-scale, multi-site dataset of 730 Japanese adults to develop a generalizable neuromarker for ASD across independent sites and different developmental stages. Our adult ASD neuromarker achieved successful generalization for the US and Belgium adults and Japanese adults. The neuromarker demonstrated significant generalization for children and adolescents. We identified 141 functional connections (FCs) important for discriminating individuals with ASD from TDCs. Finally, we mapped schizophrenia (SCZ) and major depressive disorder (MDD) onto the biological axis defined by the neuromarker and explored the biological continuity of ASD with SCZ and MDD. We observed that SCZ, but not MDD, was located proximate to ASD on the biological dimension defined by the ASD neuromarker. The successful generalization in multifarious datasets and the observed relations of ASD with SCZ on the biological dimensions provide new insights for a deeper understanding of ASD.

Introduction

Establishing robust biomarkers for autism spectrum disorder (ASD) is essential for understanding the pathophysiological mechanisms of this disorder and for early diagnosis and appropriate interventions. A plethora of modalities, including genetics, electrophysiology, and neuroimaging, are employed to identify biomarkers for ASD¹. Yet, no reliable biomarker has been established² because of the complicated relationships of various factors, such as genetic and environmental factors^{3,4}, biological sex⁵, cultural factors⁶, and developmental factors^{7,8}, all of which form the heterogeneity of ASD. Neuroimaging-based biomarkers, here we call neuromarkers, hold promise in their potential to achieve greater classification accuracy with fewer participants than genetic biomarkers, which only explain 2.45% of risk variance even with more than 10,000 cases^{9,10}. However, there are several challenges to the development of robust neuromarkers.

One of the challenges is the requirement for a large-scale dataset that often exceeds the capacity of a single institution. Multi-site initiatives, such as the Autism Brain Imaging Data Exchange (ABIDE)^{11,12}, have enabled researchers to leverage neuroimaging and machine learning techniques to develop neuromarkers^{13,14}. The classification performance of current neuromarkers ranges from 60% to 90%, depending on the sample size, cross-validation scheme, and brain features utilized in the classification¹⁵. This multi-site, large-scale data-sharing framework has greatly facilitated the development of neuromarkers for ASD. A recent meta-analytic study, however, has reported that 93% of classification studies rely on the ABIDE data¹⁶, potentially introducing biases associated with repeatedly using the same dataset^{9,17}. It is, thus, crucial to validate these neuromarkers with completely independent datasets to ensure their generalizability and robustness.

Another challenge is the inter-site variability caused by differences in the scanning protocols and MRI scanners. The diagnostic effect is smaller than the inter-site difference inherent in brain features¹⁸, which makes it challenging to capture the reliable diagnostic effect in multi-site neuroimaging data. Various methods have been proposed to correct inter-site differences^{18,19}; prior multi-site, case-control studies have confirmed the effectiveness of these methods^{20,21}. Some classification studies have employed an inter-site cross-validation scheme to evaluate the generalizability of their classifiers to other imaging sites within a dataset²²⁻²⁴. However, this approach may not be enough as an index for generalizability, as it only repeats the analyses on the same dataset. Only a few studies have attempted to construct ASD neuromarkers and explicitly test them on completely independent datasets^{9,25-27}.

Once a generalizable neuromarker is established, it paves the way towards addressing further core questions on the neural mechanisms of ASD and psychiatric disorders more generally. One application is to examine whether persistent alterations exist in the autistic brain across different developmental stages. As suggested by developmental changes in the severity of clinical symptoms^{7,8}, the patterns of structural and functional alterations in the autistic brain may change with the developmental stages²⁸⁻³⁰. Prior classification studies have reported that classifiers trained on the adult sample cannot distinguish adolescents with ASD from typically developing adolescents and vice versa^{31,32}. On the other hand, recent large-scale, case-control studies have reported reproducible and trait-like alterations of resting-state functional connectivity (FC) in the ASD population^{21,33}. These findings suggest that if any development-independent, atypical neural substrate exists throughout the lifespan, then a neuromarker may successfully generalize across developmental stages.

Another possible application is to investigate the biological continuity between ASD and other psychiatric disorders. There has been a growing interest in recent years to clarify the biological continuum between ASD and other psychiatric disorders due to the overlaps in the symptoms and impaired cognitive functions^{34,35}. Utilizing neuromarkers as biological axes and projecting categorically-distinct disorders onto the biologically defined space allows for examining cross-disease relationships without collecting additional symptom scales²⁵. Our prior study has revealed the asymmetric relationships whereby schizophrenia (SCZ) exhibits greater proximity towards ASD than typically developing controls (TDCs) on the ASD neuromarker, while ASD exhibits greater closeness towards TDCs than SCZ on the SCZ neuromarker³⁶. Despite the known high comorbidity rate of major depressive disorder (MDD) in ASD³⁷, the relationship between ASD and MDD has yet to be examined on these biological dimensions. Replication of the relationship with SCZ and exploration of the relationship with MDD may disentangle the complicated relationships between ASD and these disorders.

In this study, we tackled the above challenges through the following procedures. First, we built an FC-based classifier using the multi-site, multi-disease dataset for the Japanese adult population, called the Strategic Research Program for the Promotion of Brain Science (SRPBS) dataset³⁸. We then applied this adult ASD classifier to the U.S., Belgium, and Japanese adult validation datasets to evaluate its generalization performance to independent imaging sites. Generalization performance was further tested across different developmental stages by applying our adult ASD classifier to child and adolescent validation datasets. These analyses identified a set of FCs associated with ASD status that are invariant across different developmental stages. Finally, we examined the relationships between ASD and two major psychiatric disorders (i.e., SCZ and MDD) in the dimensional space defined by the neuromarkers.

The strengths of our framework are multifaceted. Firstly, we have evaluated the generalization performance of our ASD classifier on different types of validation datasets representing diverse ethnic and cultural backgrounds and developmental stages. This methodology allows for the assessment of generalization performance across independent imaging sites, geographical and cultural variations and different developmental stages. Secondly, we identified vital FCs in a data-driven manner. This approach can eliminate unwanted bias in identifying the neural substrate of ASD. Indeed, this procedure has identified FCs with a consistent diagnostic effect across datasets, potentially representing the core neural substrate of ASD. Lastly, the classifier was developed based on a dataset from the adult population. The utilization of adult data, in contrast to data from children, enables us to examine the commonalities and distinctions between ASD and major psychiatric disorders that manifest post-adolescence. This is of crucial importance for a deeper understanding of the relationship between ASD and other psychiatric disorders.

Results

Japanese-adult-based neuromarker for ASD generalized to two adult validation datasets

The neuromarker built on the discovery dataset (550 TDC adults and 180 adults with ASD; see Table S1 for participant characteristics) with 71,631 FCs discriminated the ASD group from the TDC group with an accuracy of 76%, AUC of 0.84, and MCC of 0.46. The corresponding sensitivity and specificity were 76% and 75%, respectively (Fig. 1A; see Fig. S1A and Table S2 for the classification performance in each imaging site). These results demonstrate the acceptable discrimination ability of our ASD classifier on the discovery dataset.

We tested the generalizability of our neuromarker on the two independent adult validation datasets (see Table S1 for participant characteristics). For the ABIDE adult dataset, our neuromarker showed an accuracy of 62% (Fig. 1B). The corresponding AUC and MCC were 0.70 and 0.25 (all $P < 0.05$, family-wise error (FWE) corrected). The sensitivity and specificity were 67% and 58%, respectively (see Fig. S1B and Table S2 for the classification performance in each imaging site). These results indicate that our marker has significant discrimination ability in the ABIDE adult dataset. For the newly collected Japanese adult dataset, our neuromarker exhibited an accuracy of 78% (Fig. 1C and Table S2). The corresponding AUC and MCC were 0.81 and 0.52, respectively (all $P < 0.05$, FWE-corrected). The sensitivity and specificity of this marker were 64% and 87%, respectively. These results suggest that our neuromarker for the ASD diagnosis generalizes to data from independent imaging sites in the same developmental stage.

Adult ASD neuromarker is generalizable to the children and adolescents

We further tested the generalizability of our adult ASD neuromarker to children (< 12 years old) and adolescents (12 < age < 18 years old) datasets (see Table S3 for participant characteristics). Our neuromarker showed significant classification performance in the children (accuracy = 61%, AUC = 0.66 and MCC = 0.27; $P < 0.05$, FWE-corrected; Fig. 2 and Table S4) and the adolescents (accuracy = 66%, AUC = 0.71, and MCC = 0.32; $P < 0.05$, FWE-corrected). These results suggest that our neuromarker for adult ASD could generalize to other developmental stages even in independent imaging sites.

Potential factors affecting the generalization performance

We explored the impacts of head motion, harmonization, and other experimental factors (i.e., diversity in imaging sites and the choice of atlas) on the generalization performance. The details

were described in **Supplementary Materials**. We first assessed whether the head motion artificially improved the generalization performance. A significant positive correlation was not found between the AUC and mean FD ($r = -0.56$, $P = 0.002$; Fig. S2), indicating that head motion did not improve the generalizability of our neuromarker. We next assessed the impacts of the harmonization method on generalization performance. In both validation datasets, we observed degraded generalization performance without the harmonization (ABIDE adults: accuracy = 60% and AUC = 0.64; Japanese adults: accuracy = 67% and AUC = 0.73; Table S5), supporting the necessity of the harmonization method for improving the generalization performance. We then examined the impacts of imaging sites on the generalization performance because our discovery dataset comprised some imaging sites having either the imbalanced patient/control ratio (i.e., Center for Innovation in Hiroshima University [COI], Kyoto University [KUT], and University of Tokyo [UTO1]) or a different scanning protocol (i.e., UTO2). Neither the inclusion of imbalanced imaging sites nor that of a site with a different scanning protocol affects the generalizability of our classifier (Tables S6 and S7). Finally, we checked whether our generalization performance was atlas-dependent. We constructed classifiers using the Schaefer cortical atlas³⁹ that provided atlases with multiple levels of resolutions ranging from 100 to 900. We did not observe any improvements specific to the Glasser atlas in the generalization performance (Table S8).

The tests of generalizability of our ASD neuromarker to independent validation datasets depend on the assumption that altered FC patterns characteristic to ASD if any, should be reproducible between the discovery dataset and the validation datasets. To directly test this assumption, we used a mass-univariate analysis similar to previous studies⁴⁰. Briefly, we first quantified the between-group difference in each FC by calculating the t -value (the diagnosis effect) for each dataset. We then computed the Pearson correlation coefficient of t -values between the datasets.

Statistical significance was tested using a permutation test with 1,000 iterations, and the statistical threshold was set to $P < 0.05$, one-sided. The discovery dataset showed statistically significant positive correlations with other datasets (ABIDE adult: $r = 0.16$; Japanese adult: $r = 0.31$; and adolescent: $r = 0.20$, all $P < 0.05$) except for the child dataset ($r = 0.02$, $P = 0.34$; Fig. S3).

FCs associated with the ASD diagnosis

We identified a set of discriminative FCs associated with the ASD status (see **Methods** for details): 65 hyper-connections and 76 hypo-connections. We called these FCs as “discriminative FCs” hereafter. To examine the spatial distribution of these FCs, we counted the number of occurrences in which each brain region was selected as at least one of the terminations of each discriminative FC. Figs. 3A and 3B show the spatial distributions of identified FCs and their terminal regions, and Table S9 provides the connection details. In the cerebral cortex, the bilateral temporal cortices and dorsomedial and ventromedial prefrontal cortices (dmPFC and vmPFC) were notably affected. On the other hand, the right amygdala, midbrain, hippocampus, globus pallidum, and putamen were affected among the subcortical regions.

To further functionally characterize these FCs, we delineated each of the hyper- and hypo-connections into functional network anatomy⁴¹. Hyper-connections were dominantly characterized by between-network connections stemming from the subcortical and somatomotor networks to other networks, such as the default mode network (DMN), dorsal attention network (DAN), and frontoparietal network (FPN) (Fig. 3C). On the other hand, hypo-connections were notably characterized by within-network connections of the DMN and between-network connections of the DMN and visual network (Fig. 3D).

Consistency of discriminative FCs across datasets

Among 141 discriminative FCs identified by permutation tests, we further investigated which discriminative FCs showed consistent alterations across the five datasets. We computed a t -value as the effect of diagnosis in each FC of each dataset. We then counted the number of FCs showing the same sign across the five datasets. Forty-two out of 141 discriminative FCs showed the same sign of t -values across all the datasets (Hyper-connection: 52.4% and hypo-connection: 47.6%; Fig. 4A and Table S10). When considering over- and under-connectivity (i.e., the difference in the absolute of FC strength between groups), 69% of the reproducible FCs were under-connectivity (Fig. 4B). The ASD group exhibited atypical connections stemming from subcortical networks, including the midbrain, hippocampus, and amygdala, to other networks, such as the somatomotor network and FPN (Figs. 4C and 4D).

The statistical significance of the consistency among the five datasets was examined using a binomial test⁴². The number of consistent FCs within the discriminative FCs was 42, while the number of FCs showing the same sign of diagnostic effects in the whole FCs (i.e., 71,631 FCs) was 9,547. We assumed a binomial distribution, $Bi(n, p)$, where n stands for the number of discriminative FCs (i.e., $n = 141$), and p stands for the probability of being consistent across the datasets in the whole FCs (i.e., $p = 9,547/71,631$). The binomial test confirmed that discriminative FCs were reproducible across different ethnicities and cultures (i.e., U.S., Belgium, and Japan) and different developmental stages (i.e., child, adolescent, and adult) ($P < 0.05$). The binomial tests confirmed the statistical significance of the consistency between the discovery dataset and every dataset (see **Supplementary Materials** for details). These results indicate that the selection of discriminative FCs showing consistent alterations may yield the generalizability of our neuromarker across independent imaging sites and developmental stages.

Specificity of the ASD neuromarker

We applied our ASD neuromarker to SCZ and MDD to test the specificity of the neuromarker (see Table S11 for participant characteristics). Our ASD neuromarker exhibited 56% and 33% sensitivities for SCZ and MDD, respectively (Fig. 5A). As expected, the sensitivities for both disorders were greatly decreased compared with that for the ASD diagnosis. However, the sensitivity to the SCZ diagnosis was still statistically significant ($P = 0.014$), whereas the marker showed no sensitivity to the MDD diagnosis ($P = 0.638$). These results indicate that SCZ, but not MDD, was located proximate to ASD on the biological dimension defined by the ASD neuromarker.

Relationships between ASD and other psychiatric disorders on biological dimensions

To further explore the relationships of the three disorders on multiple biological dimensions, we built two additional classifiers for the SCZ and MDD (see Table S12 for the classification performance of both classifiers). We then applied each of them to the other two disorders (see **Supplementary Materials** for details). The SCZ classifier exhibited poor sensitivity to the ASD and MDD diagnoses (ASD = 34% and MDD = 44%, all $P > 0.39$; Fig. 5B). On the other hand, the MDD classifier showed poor sensitivity to the ASD diagnosis (sensitivity = 47%) but high sensitivity to the SCZ diagnosis (sensitivity = 83%; Fig. 5C). We next plotted each participant onto the two-dimensional planes defined by the neuromarkers to illustrate the relationship between ASD and other psychiatric disorders. On the two-dimensional plane characterized by the ASD and SCZ neuromarkers, ASD predominantly resided in the lower right quadrant, while SCZ was primarily situated in the upper right quadrant (Fig. 5D), suggesting an asymmetric proximity between ASD and SCZ. In contrast, on the two-dimensional space demarcated by the ASD and MDD neuromarkers, ASD was mainly

distributed in the lower right quadrant, while MDD was mainly localized in the upper left quadrant (Fig. 5E). These results suggest that, although SCZ shares characteristics with ASD, ASD and MDD appear to be distinct entities on the biological dimensions.

We further investigated why ASD and SCZ neuromarkers showed an asymmetric proximity by first checking the spatial overlap of discriminative FCs between the two neuromarkers. Permutation tests with 100 iterations identified discriminative FCs for SCZ (Fig. S4 and Table S13; see Fig S5 and Table S14 for the discriminative FCs for MDD). When comparing them with those for ASD, we observed no spatial overlap between the two neuromarkers. Next, we examined the similarities of the diagnosis effects between the two groups. We expected that the effects of diagnoses were similar in the ASD neuromarker but not in the SCZ neuromarker. We observed significant positive correlations in the ASD neuromarker (hyper-connection: $r = 0.63$, $P < 0.05$; hypo-connection: $r = 0.60$, $P < 0.05$) but not in the SCZ marker (hyper-connection: $r = 0.05$, $P > 0.86$; hypo-connection: $r = -0.44$, $P < 0.05$; Fig. S6). These results indicate that the asymmetric diagnostic effects on both neuromarkers yield such relationships between ASD and SCZ on the biological dimensions.

Discussion

In this study, we developed an FC-based neuromarker for ASD diagnosis using the Japanese adult multi-site dataset. Our ASD neuromarker was generalizable with >70% AUC value to both the ABIDE and Japanese adults from independent imaging sites. This neuromarker also exhibited acceptable generalization performance to children and adolescents, suggesting the presence of atypical neural basis that persists in ASD from childhood to adulthood. We identified 141 FCs that were pivotal in determining the ASD status. These FCs spanned multiple networks, including the DMN, somatomotor, and subcortical networks, and comprised social

brain regions, such as the vmPFC, dmPFC, STG, right amygdala, midbrain, and hippocampus. Forty-two out of 141 FCs displayed consistent effects of diagnostic status across all the five datasets. By utilizing our neuromarker as a biological axis and mapping SCZ and MDD onto the axis, we replicated previous findings suggesting proximity between ASD and SCZ and a greater distance between ASD and MDD. Our neuromarker-based analytical framework offers novel insights into the neural underpinnings of the autistic brain and provides an effective tool for elucidating transdiagnostic continuity.

We confirmed the validity of our neuromarker through stringent evaluations of its generalization performance on independent validation datasets. Most prior studies have relied on cross-validation procedures within the ABIDE dataset^{15,16}. Reliance on these methods poses the risk of overestimating the classification performance and introducing biases into the results^{9,17}. It is, therefore, critical to evaluate the generalization performance of a neuromarker using a completely independent external dataset⁴³. A handful of studies have evaluated the external validity of their neuromarkers using independent validation datasets and have reported generalization performance ranging from 67% to 75%^{9,25,26}, comparable to our results. The forte of our study is the thorough evaluation of generalization performance. Through our assessment, we verified the generalizability of our neuromarker utilizing four independent validation datasets that varied in terms of imaging sites, ethnic and cultural backgrounds, and developmental stages. Additionally, we confirmed that the generalizability of our neuromarker remained intact when utilizing another brain atlas featuring multiple resolution levels (Table S8). These thorough evaluations ensure the robustness of our neuromarker and provide a new tool for further exploring several hypotheses in the autistic brain.

Developmental-dependent changes in FCs have been a major bottleneck for transferring

neuromarkers to other developmental stages^{31,32}. Nevertheless, our adult ASD neuromarker successfully distinguished children and adolescents with ASD from TDCs. The successful generalization may be due to the LASSO method used in this study. The autistic brain likely presents a mosaic pattern of age-specific and age-unspecific atypical FCs⁴⁴⁻⁴⁶. To support this view, Alaerts and colleagues have reported that, depending on the terminal regions, the posterior superior temporal sulcus exhibits distinct age-related patterns of abnormal FCs in the autistic brain²⁹. In such a complicated situation, non-sparse machine learning algorithms such as support vector machines may have failed to select features appropriately, yielding reduced generalization performance when transferring neuromarkers to other developmental stages. The LASSO method tends to choose atypical FCs consistent within the discovery dataset during nested cross-validation. As a result, dataset-independent atypical FCs may have been selected in our adult ASD neuromarker.

Our ASD neuromarker identified 141 discriminative FCs distributed across multiple networks, most notably the DMN, somatomotor, and subcortical networks. One of the most prominent features is hypo-connectivity within the DMN, including the dmPFC, vmPFC, and temporal cortices. The hypo-connectivity within the DMN and its association with socio-communicative deficits in ASD have been repeatedly reported⁴⁷⁻⁴⁹. While, in this study, discriminative FCs stemming from temporal cortices exhibited the effect of ASD status consistently across the five datasets, those stemming from dmPFC and vmPFC did not. Given that population heterogeneity is associated with the DMN, especially dmPFC, vmPFC, and precuneus⁵⁰, the dataset-dependent effect of ASD status on these discriminative FCs across the datasets may reflect differences in the population diversity.

Another prominent feature is alterations in FCs within and between subcortical and

somatomotor networks, encompassing the amygdala, hippocampus, midbrain, and STG. In addition, we observed that 26 of 42 discriminative FCs that showed a consistent effect of ASD status across the five datasets belonged to these networks. Our findings imply that abnormalities in these networks may serve as the core neural basis for this disorder. To support this possibility, several lines of evidence suggest functional alterations within and between subcortical and somatomotor networks^{11,45,46} and their associations with the clinical symptoms, such as restricted and repetitive behaviors and atypical sensory processing^{21,41}. Furthermore, the association between atypical sensory processing and socio-communicative impairments is supported by previous literature^{51,52}. Replicability, associations with core symptoms, and consistency of abnormalities through developmental stages may distinguish these networks as a core basis for ASD.

We demonstrated the utility of our generalizable neuromarker-based analytic framework for examining the biological continuum of ASD with SCZ and MDD by mapping these disorders onto the biological dimensions defined by the neuromarkers. Previous studies have highlighted the substantial overlap between ASD and SCZ at genetic⁵³, neural⁵⁴, and behavioral levels⁵⁵. However, these findings are less replicable due to methodological differences⁵⁶. In contrast, we reproduced the asymmetric proximity between the ASD and SCZ group, whereby the SCZ group manifested stronger adjacency to the ASD group than the TDC group on the ASD neuromarker, whereas the ASD group displayed stronger proximity to the TDC group than the SCZ group on the SCZ neuromarker, by using different datasets and methods³⁶. Our replicable approach may provide a clue to investigating the spectrum structure among psychiatric disorders.

Despite the asymmetric proximity between the ASD and SCZ groups, there was no spatial

overlap in discriminative FCs between the ASD and SCZ neuromarkers. One possibility is that the SCZ neuromarker in this study was built on patients with chronic SCZ. Previous studies have reported that abnormal FCs in SCZ tend to change with the disease progression^{36,57}. Owing to this change of abnormal FCs with disease progression, patients with SCZ may show abnormalities in discriminative FCs for the ASD neuromarker, which may result in asymmetric proximity between ASD and SCZ. Indeed, we observed that ASD and SCZ status had similar diagnosis effects on discriminative FCs for the ASD neuromarker but not on those for the SCZ neuromarker (Fig. S6). The inclusion of patients with SCZ from different disease stages may provide a clue for how this asymmetry may arise.

In addition to the relationship between ASD and SCZ, we examined the relationship between ASD and MDD, but no asymmetric proximity was observed between the two disorders. Although the comorbidity rate of MDD in ASD is known to be as high as 11%³⁷, there are no case-control studies to date that examine FCs among ASD, MDD, and ASD comorbid with MDD. A few studies have examined the relationship between atypical FCs and depressive symptoms in ASD^{58,59}. For example, Kleinhans and colleagues have reported that social impairments and depressive symptoms in ASD are separately associated with abnormal amygdala circuits, and the patterns of these associations in the autistic brain differ from those in TDCs⁵⁹. Our findings may reflect that the diagnoses of ASD and MDD affect different neural circuits. The relationship between ASD and MDD on the biological dimensions could be further clarified by studies that were designed to include ASD cases comorbid with MDD.

In conclusion, we developed an ASD neuromarker generalizable to diverse validation datasets from independent imaging sites and different developmental stages. We further demonstrated the applicability of our generalizable neuromarker to examine the effects of developmental

stages on the autistic brain and the biological continuum between ASD and other disorders. Using discriminative FCs identified by our ASD neuromarker, prospective future directions will be opened, including identifying biological subtypes within ASD⁶⁰ and interventions based on those FCs as targets⁶¹.

Methods

Study design

The aim of this study was twofold: the first was to construct a generalizable neuromarker for distinguishing adults with ASD from TDCs across diverse imaging sites, and the second was to demonstrate the utility of this neuromarker by examining its generalizability to children and adolescents and investigating the transdiagnostic continuity of ASD with SCZ, and MDD. This study used three adult resting-state fMRI (R-fMRI) datasets for the analyses: one was used as the discovery dataset, and the remaining two were used as independent validation datasets. Tables [S1](#), [S15](#), and [S16](#) show the demographic information and scanning parameters for the three datasets. Data were sourced from three multi-site initiatives: the SRPBS³⁸, ABIDE^{11,12}, and Brain/MINDS Beyond (BMB) projects⁶². All the participants (if appropriate) and their respective parent/legal guardian provided written informed consent. See **Supplementary Materials** for the detailed descriptions of datasets and exclusion criteria.

Participants

The discovery dataset contained data of 550 TDCs from 5 scanners at 4 imaging sites (COI, KUT, Showa University [SWA1], and UTO1, UTO2) and 180 adults with ASD from 2 sites (SWA1 and UTO2). The first validation dataset (ABIDE adult dataset) consisted of participants from the ABIDE-I¹¹ and -II¹². The ABIDE adult dataset comprised 54 adults with ASD and 67 TDCs that were selected from the following 3 sites. The second validation dataset (i.e., the

Japanese adult dataset) consisted of 22 adults with ASD and 38 controls at Showa University (SWA2). The data were collected under the BMB project. The project details were described elsewhere⁶². See **Supplementary Materials** for more details of participants characteristics.

R-fMRI data preprocessing and network construction

We preprocessed all the R-fMRI data using fMRIPrep version 1.1.8⁶³ and the ciftify toolbox version 2.1.1⁶⁴. See **Supplementary Materials** for the details of preprocessing steps. We performed nuisance regression to remove the effects of artifactual and non-neural sources. Nuisance regressors consisted of six head-motion parameters, averaged signals from subject-specific white matter and cerebrospinal fluid masks, global signal, their temporal derivatives, and linear detrending. After nuisance regression, we applied a band-pass filter (0.008–0.1 Hz) to the residuals. We used frame-wise displacement (FD) as a measure for detecting occasional head movement, and removed volumes with $FD > 0.5$ mm, as proposed in a previous study⁶⁵. We used Glasser’s 379 surface-based brain parcellations (cortical 360 parcellations and subcortical 19 parcellations) as ROIs⁶⁶. We computed the temporal correlations of signals among all possible pairs of ROIs and applied Fisher’s r -to- z transformation, resulting in 71,631 unique FCs for each participant. Because the label of each ROI in Glasser’s atlas was not intuitive, we utilized Yeo’s resting-state network (RSN) labels⁶⁷ to assign important ROIs to the corresponding RSN label. This study added the subcortical network label to the subcortical and cerebellar regions.

Controlling for imaging site differences

We used the ComBat harmonization method⁶⁸ to control the imaging site differences inherently captured in FCs. Because of its simplicity and effectiveness, this method has been used for controlling the site differences in neuroimaging data^{69–71}. In the current study, we incorporated

disease status (0 = TDC, and 1 = ASD), age, and sex as covariates of interest into the ComBat model. Of note, we separately applied the ComBat harmonization method to both discovery and independent validation datasets.

Construction of the ASD neuromarker using the discovery dataset

Based on previous studies^{25,36,40,72–74}, we assumed that psychiatric disorder factors were associated with the limited number of FCs, rather than the whole-brain connections. We used a logistic regression analysis with a least absolute shrinkage and selection operator (LASSO) method⁷⁵ that selects an optimal subset of FCs from the whole brain connections. See **Supplementary Materials** for the construction of our ASD neuromarker. We developed the ASD neuromarker using the LASSO method with 10-fold nested cross-validation (CV) and 10 subsampling, yielding 100 trained classifiers (Fig. S7). The mean classifier output value was considered as diagnostic probability, indicating a likelihood of a participant belonging to the ASD class. We considered participants as those with ASD if their diagnostic probability values were higher than 0.5. We calculated the area under the curve (AUC) to assess the classification performance. We also computed accuracy, sensitivity, specificity, and the Matthews correlation coefficient (MCC). The MCC is suitable for the imbalanced dataset because this metric takes into account the ratio of the confusion matrix size⁷⁶. We used AUC and MCC as performance indices throughout the paper.

Generalizability of the ASD neuromarker to the ABIDE and Japanese adult datasets

We tested the generalizability of the ASD neuromarker using two independent validation datasets: the ABIDE adult and the Japanese adult datasets. We applied all the trained classifiers to the independent validation datasets and obtained 100 diagnostic probabilities for each participant. We averaged these diagnostic probabilities for each participant and considered a

participant to be a participant with ASD if the mean diagnostic probability was > 0.5 . To evaluate the generalization performance, we computed the following performance metrics: AUC, accuracy, sensitivity, specificity, and MCC.

Generalizability of the ASD neuromarker to the child and adolescent datasets

We further tested the generalizability of the adult ASD neuromarker to children (< 12 years old) and adolescents ($12 < \text{age} < 18$). We used these age ranges according to a previous study³³. We created the child and adolescent datasets using the same exclusion criteria as the ones applied to the adult dataset. The child dataset consists of 119 children with ASD and 202 those with TDC from 5 sites. The adolescent dataset consists of 121 adolescents with ASD and 132 TDCs from 8 sites. The details of demographic information and scanning parameters are in Tables [S3](#) and [S16](#). We applied the same preprocessing pipeline and harmonization method to both datasets.

Statistical analyses

Evaluation of the statistical significance of neuromarker's generalizability

The statistical significance of AUC and MCC was tested using a permutation test with 500 iterations. At each iteration, we created a permuted dataset by shuffling the diagnostic labels. We then built a classifier for the permuted dataset in the same way as for the non-permuted dataset using a 10-fold nested CV with 10 subsamples, which resulted in 100 permuted classifiers at each iteration. Each individual of the independent datasets was classified based on 100 diagnostic probability values, each of which was generated by 100 trained permuted classifiers. The threshold was set at the mean diagnostic probability value of 0.5. We calculated the AUC and MCC for a permuted classifier at each iteration. We constructed a null distribution for each performance index by aggregating those values across iterations. The statistical

threshold was set to 0.05. The Holm-Bonferroni method⁷⁷ was used to control the FWE rate across all the validation datasets.

Identifications of FCs associated with the clinical diagnosis of ASD

We identified discriminative FCs using a permutation test with 500 iterations in a similar manner to our previous study⁴⁰. First, as to each FC, we counted the number of times the LASSO selected the FC during the 10-fold CV. If a given FC was consistently important for discrimination throughout the training dataset, we expected that it would be selected significantly more times by the LASSO than the chance derived from a null distribution. We shuffled the diagnostic labels at each iteration to create a permuted dataset and constructed permuted classifiers. For each FC, we counted the number when the FC was selected by the LASSO across 10-fold cross-validation and 10 subsampling (i.e., across 100 classifiers). We used the maximum counts among all the connections at each iteration to construct a null distribution using these maxima. We considered FCs as significant contributors to the ASD status if their *P*-values were below 0.05.

Network-based characterization of discriminative FCs

To interpret the spatial distribution of discriminative FCs at the network level, we used the network anatomy⁴¹. We identified discriminative FCs using the permutation test described above. We computed the mean weights for those FCs. We considered FCs with positive mean weights as hyper-connections and those with negative mean weights as hypo-connections in the ASD brain. It turned out that we found 65 hyper-connections and 76 hypo-connections (see **Results**). Then, we computed the probability of spatial overlap between FCs within or between-RSNs defined by the Yeo atlas and hyper- or hypo-connections identified by our neuromarker.

P -values were computed using the hypergeometric cumulative density function “*hygecdf*” implemented in MATLAB as follows:

$$P = 1 - \text{hygecdf}(x, M, K, n),$$

where x stands for the number of overlapping FCs between hyper- or hypo-connections and within or between predefined RSNs. The variable, M , stands for the number of total FCs in the whole brain (i.e., $M = 71,631$), while the variables, n and K , were the total numbers of FCs in our network of interest and the canonical network of interest, respectively. We computed the P -values for hyper- and hypo-connections separately. Statistical significance was set to $P < 0.05$, adjusted with Bonferroni correction for multiple comparisons.

Sensitivity of the ASD neuromarker to other psychiatric disorders

We tested whether the developed ASD neuromarker was sensitive to SCZ and MDD. To test this, we applied the neuromarker to data from SCZ and MDD included in the SRPBS dataset³⁸. The details of their demographic information are in Table S11. We applied the same preprocessing pipeline, the exclusion criteria, and the harmonization method.

We applied 100 trained ASD classifiers to SCZ and MDD datasets and obtained 100 diagnostic probability values for each participant. We considered a participant to have the ASD label if the mean diagnostic probability was > 0.5 . Since participants with TDC were identical among the datasets, we used the diagnostic probability values during the training to compute the classification performance. We focused on sensitivity rather than other performance indices. Statistical significance was calculated based on the null distribution obtained by the permutation test with 500 iterations. The statistical threshold was set to $P < 0.05$.

References

1. Parellada, M. *et al.* In Search of Biomarkers to Guide Interventions in Autism Spectrum Disorder: A Systematic Review. *Am. J. Psychiatry* **180**, 23–40 (2023).
2. Cortese, S. *et al.* Candidate diagnostic biomarkers for neurodevelopmental disorders in children and adolescents: a systematic review. *World Psychiatry* **22**, 129–149 (2023).
3. Hallmayer, J. *et al.* Genetic heritability and shared environmental factors among twin pairs with autism. *Arch. Gen. Psychiatry* **68**, 1095–1102 (2011).
4. Colvert, E. *et al.* Heritability of Autism Spectrum Disorder in a UK Population-Based Twin Sample. *JAMA Psychiatry* **72**, 415–423 (2015).
5. Floris, D. L. *et al.* The Link Between Autism and Sex-Related Neuroanatomy, and Associated Cognition and Gene Expression. *Am. J. Psychiatry* appiajp20220194 (2022).
6. de Leeuw, A., Happé, F. & Hoekstra, R. A. A Conceptual Framework for Understanding the Cultural and Contextual Factors on Autism Across the Globe. *Autism Res.* **13**, 1029–1050 (2020).
7. Fountain, C., Winter, A. S. & Bearman, P. S. Six developmental trajectories characterize children with autism. *Pediatrics* **129**, e1112–20 (2012).
8. Waizbard-Bartov, E. *et al.* Identifying autism symptom severity trajectories across childhood. *Autism Res.* **15**, 687–701 (2022).
9. Traut, N. *et al.* Insights from an autism imaging biomarker challenge: Promises and threats to biomarker discovery. *Neuroimage* **255**, 119171 (2022).
10. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
11. Di Martino, A. *et al.* The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry* **19**, 659–667 (2014).

12. Di Martino, A. *et al.* Enhancing studies of the connectome in autism using the autism brain imaging data exchange II. *Sci Data* **4**, 170010 (2017).
13. Feng, W., Liu, G., Zeng, K., Zeng, M. & Liu, Y. A review of methods for classification and recognition of ASD using fMRI data. *J. Neurosci. Methods* **368**, 109456 (2021).
14. Horien, C. *et al.* Functional Connectome-Based Predictive Modeling in Autism. *Biol. Psychiatry* **92**, 626–642 (2022).
15. Wolfers, T. *et al.* From pattern classification to stratification: towards conceptualizing the heterogeneity of Autism Spectrum Disorder. *Neurosci. Biobehav. Rev.* **104**, 240–254 (2019).
16. Santana, C. P. *et al.* rs-fMRI and machine learning for ASD diagnosis: a systematic review and meta-analysis. *Sci. Rep.* **12**, 6030 (2022).
17. Thompson, W. H., Wright, J., Bissett, P. G. & Poldrack, R. A. Dataset decay and the problem of sequential analyses on open datasets. *Elife* **9**, e53498 (2020).
18. Yamashita, A. *et al.* Harmonization of resting-state functional MRI data across multiple imaging sites via the separation of site differences into sampling bias and measurement bias. *PLoS Biol.* **17**, e3000042 (2019).
19. Chen, A. A. *et al.* Mitigating site effects in covariance for machine learning in neuroimaging data. *Hum. Brain Mapp.* **43**, 1179–1195 (2022).
20. Xie, Y. *et al.* Alterations in Connectome Dynamics in Autism Spectrum Disorder: A Harmonized Mega- and Meta-Analysis Study Using the ABIDE Dataset. *Biol. Psychiatry* **0**, (2021).
21. Ilioska, I. *et al.* Connectome-wide mega-analysis reveals robust patterns of atypical functional connectivity in autism. *Biol. Psychiatry* **0**, (2022).
22. Spera, G. *et al.* Evaluation of Altered Functional Connections in Male Children With Autism Spectrum Disorders on Multiple-Site Data Optimized With Machine Learning.

- Front. Psychiatry* **10**, 620 (2019).
23. Duan, Y. *et al.* Identifying and Predicting Autism Spectrum Disorder Based on Multi-Site Structural MRI With Machine Learning. *Front. Hum. Neurosci.* **15**, 765517 (2021).
 24. Abraham, A. *et al.* Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example. *Neuroimage* **147**, 736–745 (2017).
 25. Yahata, N. *et al.* A small number of abnormal brain connections predicts adult autism spectrum disorder. *Nat. Commun.* **7**, 11254 (2016).
 26. Jahedi, A., Nasamran, C. A., Faires, B., Fan, J. & Müller, R.-A. Distributed Intrinsic Functional Connectivity Patterns Predict Diagnostic Status in Large Autism Cohort. *Brain Connect.* **7**, 515–525 (2017).
 27. Supekar, K. *et al.* Robust, Generalizable, and Interpretable Artificial Intelligence–Derived Brain Fingerprints of Autism and Social Communication Symptom Severity. *Biol. Psychiatry* **92**, 643–653 (2022).
 28. Uddin, L. Q., Supekar, K. & Menon, V. Reconceptualizing functional brain connectivity in autism from a developmental perspective. *Front. Hum. Neurosci.* **7**, 458 (2013).
 29. Alaerts, K. *et al.* Age-related changes in intrinsic function of the superior temporal sulcus in autism spectrum disorders. *Soc. Cogn. Affect. Neurosci.* **10**, 1413–1423 (2015).
 30. Wallace, G. L., Dankner, N., Kenworthy, L., Giedd, J. N. & Martin, A. Age-related temporal and parietal cortical thinning in autism spectrum disorders. *Brain* **133**, 3745–3754 (2010).
 31. Kazeminejad, A. & Sotero, R. C. Topological Properties of Resting-State fMRI Functional Networks Improve Machine Learning-Based Autism Classification. *Front. Neurosci.* **12**, 1018 (2018).
 32. Lanka, P. *et al.* Supervised machine learning for diagnostic classification from large-scale neuroimaging datasets. *Brain Imaging Behav.* **14**, 2378–2416 (2020).

33. Holiga, Š. *et al.* Patients with autism spectrum disorders display reproducible functional connectivity alterations. *Sci. Transl. Med.* **11**, (2019).
34. Tung, Y.-H. *et al.* Whole Brain White Matter Tract Deviation and Idiosyncrasy From Normative Development in Autism and ADHD and Unaffected Siblings Link With Dimensions of Psychopathology and Cognition. *Am. J. Psychiatry* **178**, 730–743 (2021).
35. Aoki, Y. *et al.* Association of White Matter Structure With Autism Spectrum Disorder and Attention-Deficit/Hyperactivity Disorder. *JAMA Psychiatry* **74**, 1120–1128 (2017).
36. Yoshihara, Y. *et al.* Overlapping but Asymmetrical Relationships Between Schizophrenia and Autism Revealed by Brain Connectivity. *Schizophr. Bull.* (2020) doi:10.1093/schbul/sbaa021.
37. Lai, M.-C. *et al.* Prevalence of co-occurring mental health diagnoses in the autism population: a systematic review and meta-analysis. *Lancet Psychiatry* **6**, 819–829 (2019).
38. Tanaka, S. C. *et al.* A multi-site, multi-disorder resting-state magnetic resonance image database. *Sci Data* **8**, 227 (2021).
39. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).
40. Yamashita, A. *et al.* Generalizable brain network markers of major depressive disorder across multiple imaging sites. *PLoS Biol.* **18**, e3000966 (2020).
41. Lake, E. M. R. *et al.* The Functional Brain Organization of an Individual Allows Prediction of Measures of Social Abilities Transdiagnostically in Autism and Attention-Deficit/Hyperactivity Disorder. *Biol. Psychiatry* **86**, 315–326 (2019).
42. Yamashita, M., Kawato, M. & Imamizu, H. Predicting learning plateau of working memory from whole-brain intrinsic network connectivity patterns. *Sci. Rep.* **5**, 7622 (2015).

43. Scheinost, D. *et al.* Ten simple rules for predictive modeling of individual differences in neuroimaging. *Neuroimage* **193**, 35–45 (2019).
44. Padmanabhan, A., Lynn, A., Foran, W., Luna, B. & O’Hearn, K. Age related changes in striatal resting state functional connectivity in autism. *Front. Hum. Neurosci.* **7**, 814 (2013).
45. Cerliani, L. *et al.* Increased Functional Connectivity Between Subcortical and Cortical Resting-State Networks in Autism Spectrum Disorder. *JAMA Psychiatry* **72**, 767–777 (2015).
46. Park, S. *et al.* Atypical integration of sensory-to-transmodal functional systems mediates symptom severity in autism. *Front. Psychiatry* **12**, (2021).
47. Hong, S.-J. *et al.* Atypical functional connectome hierarchy in autism. *Nat. Commun.* **10**, 1022 (2019).
48. Yerys, B. E. *et al.* Default mode network segregation and social deficits in autism spectrum disorder: Evidence from non-medicated children. *Neuroimage Clin* **9**, 223–232 (2015).
49. Padmanabhan, A., Lynch, C. J., Schaer, M. & Menon, V. The Default Mode Network in Autism. *Biol Psychiatry Cogn Neurosci Neuroimaging* **2**, 476–486 (2017).
50. Benkarim, O. *et al.* Population heterogeneity in clinical cohorts affects the predictive accuracy of brain imaging. *PLoS Biol.* **20**, e3001627 (2022).
51. Kaneko, A. *et al.* Sensory and social subtypes of Japanese individuals with autism spectrum disorders. (2021) doi:10.31234/osf.io/umvcn.
52. Thye, M. D., Bednarz, H. M., Herringshaw, A. J., Sartin, E. B. & Kana, R. K. The impact of atypical sensory processing on social impairments in autism spectrum disorder. *Dev. Cogn. Neurosci.* **29**, 151–167 (2018).
53. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk

- loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371–1379 (2013).
54. Chen, H. *et al.* Shared atypical default mode and salience network functional connectivity between autism and schizophrenia. *Autism Res.* **10**, 1776–1786 (2017).
 55. Oliver, L. D. *et al.* Social Cognitive Performance in Schizophrenia Spectrum Disorders Compared With Autism Spectrum Disorder: A Systematic Review, Meta-analysis, and Meta-regression. *JAMA Psychiatry* **78**, 281–292 (2021).
 56. Jutla, A., Foss-Feig, J. & Veenstra-VanderWeele, J. Autism spectrum disorder and schizophrenia: An updated conceptual review. *Autism Res.* **15**, 384–412 (2022).
 57. Zhao, W. *et al.* Common abnormal connectivity in first-episode and chronic schizophrenia in pre- and post-central regions: Implications for neuromodulation targeting. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **117**, 110556 (2022).
 58. Hogeveen, J., Krug, M. K., Elliott, M. V. & Solomon, M. Insula-Retrosplenial Cortex Overconnectivity Increases Internalizing via Reduced Insight in Autism. *Biol. Psychiatry* **84**, 287–294 (2018).
 59. Kleinhans, N. M. *et al.* Subregional differences in intrinsic amygdala hyperconnectivity and hypoconnectivity in autism spectrum disorder. *Autism Res.* **9**, 760–772 (2016).
 60. Urchs, S. G. W. *et al.* Functional connectivity subtypes associate robustly with ASD diagnosis. *Elife* **11**, (2022).
 61. Yamada, T. *et al.* Resting-State Functional Connectivity-Based Biomarkers and Functional MRI-Based Neurofeedback for Psychiatric Disorders: A Challenge for Developing Theranostic Biomarkers. *Int. J. Neuropsychopharmacol.* **20**, 769–781 (2017).
 62. Koike, S. *et al.* Brain/MINDS beyond human brain MRI project: A protocol for multi-level harmonization across brain disorders throughout the lifespan. *Neuroimage Clin* **30**,

- 102600 (2021).
63. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).
 64. Dickie, E. W. *et al.* Ciftify: A framework for surface-based analysis of legacy MR acquisitions. *Neuroimage* **197**, 818–826 (2019).
 65. Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L. & Petersen, S. E. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* **59**, 2142–2154 (2012).
 66. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
 67. Yeo, B. T. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).
 68. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
 69. Fortin, J.-P. *et al.* Harmonization of cortical thickness measurements across scanners and sites. *Neuroimage* **167**, 104–120 (2018).
 70. Fortin, J.-P. *et al.* Harmonization of multi-site diffusion tensor imaging data. *Neuroimage* **161**, 149–170 (2017).
 71. Yu, M. *et al.* Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data. *Hum. Brain Mapp.* **39**, 4213–4227 (2018).
 72. Ichikawa, N. *et al.* Primary functional brain connections associated with melancholic major depressive disorder and modulation by antidepressants. *Sci. Rep.* **10**, 3542 (2020).
 73. Almuqhim, F. & Saeed, F. ASD-SAENet: A Sparse Autoencoder, and Deep-Neural Network Model for Detecting Autism Spectrum Disorder (ASD) Using fMRI Data. *Front. Comput. Neurosci.* **15**, 654315 (2021).

74. Kwon, H. *et al.* Sparse Hierarchical Representation Learning on Functional Brain Networks for Prediction of Autism Severity Levels. *Front. Neurosci.* **16**, 935431 (2022).
75. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc.* **58**, 267–288 (1996).
76. Chicco, D. Ten quick tips for machine learning in computational biology. *BioData Min.* **10**, 35 (2017).
77. Holm, S. A Simple Sequentially Rejective Multiple Test Procedure. *Scand. Stat. Theory Appl.* **6**, 65–70 (1979).

Acknowledgments:

We are very grateful to the participants in the SRPBS, ABIDE-I, ABIDE-II, and BMB projects.

Funding:

This study was supported by JSPS KAKENHI Grant Numbers JP21H05171 and JP21H05174, AMED under Grant Numbers JP19dm0207069, JP18dm0307001, JP18dm0307004, JP18dm0307008, and JP18dm0307009, Moonshot R&D Grant Number JPMJMS2021, and by UTokyo Institute for Diversity and Adaptation of Human Mind (UTIDAHM), and the International Research Center for Neurointelligence (WPI-IRCN) at The University of Tokyo Institutes for Advanced Study (UTIAS). ABIDE-I was supported by NIMH (K23MH087770), NIMH (R03MH096321), Leon Levy Foundation, Joseph P. Healy, and the Stavros Niarchos Foundation. ABIDE-II was supported by NIMH (5R21MH107045), NIMH (5R21MH107045), Nathan S. Kline Institute of Psychiatric Research, Joseph P. Healey, Phyllis Green, and Randolph Cowen.

Author contributions:

T.I., A.Y., M.K., and R.H. designed the study. T.I., N.Y., J.F., Y.Y., M.N., R.A., Y.Y.A., H.O., M.T., N.I., G.O., N.O., K.K., Y.O. N.K. and H.T. recruited participants of this study and collected their clinical and imaging data. T.I. and A.Y. carried out the analysis of the data. T.I., A.Y., and R.H. wrote the original draft of manuscript. T.I., A.Y., Y.T., N.Y., Y.Y.A., J.F., Y.Y., M.N., R.A., H.O., Y.S., M.T., N.I., G.O., N.O., K.K., S.C.T., H.I., N.K., Y.O., H.T., M.K., O.Y. and R.H. reviewed and revised the manuscript.

Competing interests:

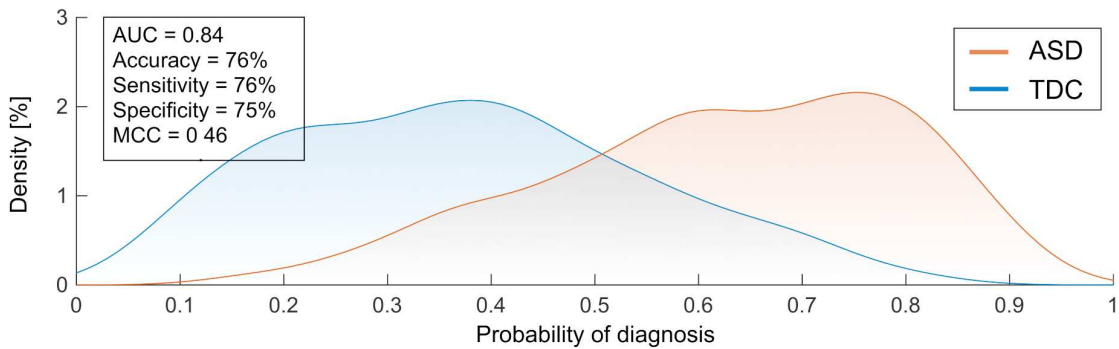
MK is an inventor of patents owned by the Advanced Telecommunications Research Institute International related to the present work (PCT/JP2014/061544 [WO2014178323] and JP2015-228970/6195329). AY and MK are inventors of a patent application submitted by the Advanced Telecommunications Research Institute International related to the present work (JP2018-192842). YT is an employee of SHIONOGI & CO., LTD.

Data and materials availability

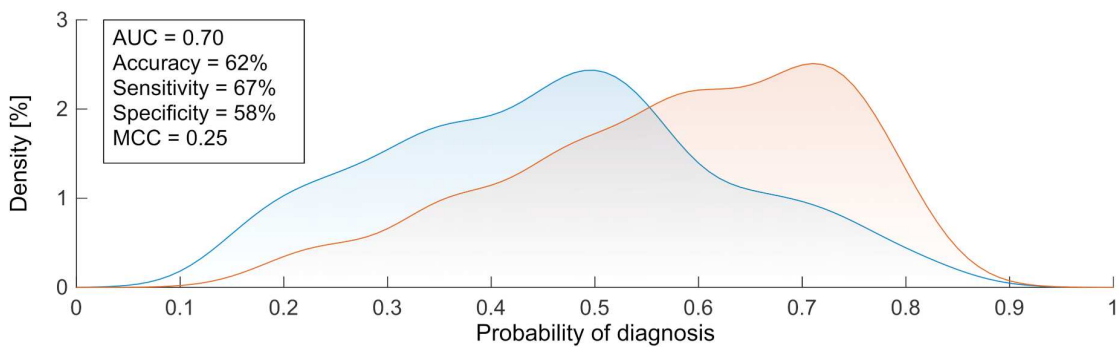
All the discovery data, except for a part of the discovery dataset (i.e., SWA1 and UTO2), were available from the Decoded Neurofeedback (DecNef) Project Brain Data Repository (<https://bicr-resource.atr.jp/decnefpro/>). ABIDE-I and ABIDE-II were available from https://fcon_1000.projects.nitrc.org/indi/abide/. The BMB data will be available from its project website (<https://hbm.brainminds-beyond.jp/>).

Figures

A. Probability distribution of the ASD diagnosis in the discovery dataset



B. Probability distribution of the ASD diagnosis in the ABIDE adult dataset



C. Probability distribution of the ASD diagnosis in the Japanese adult dataset

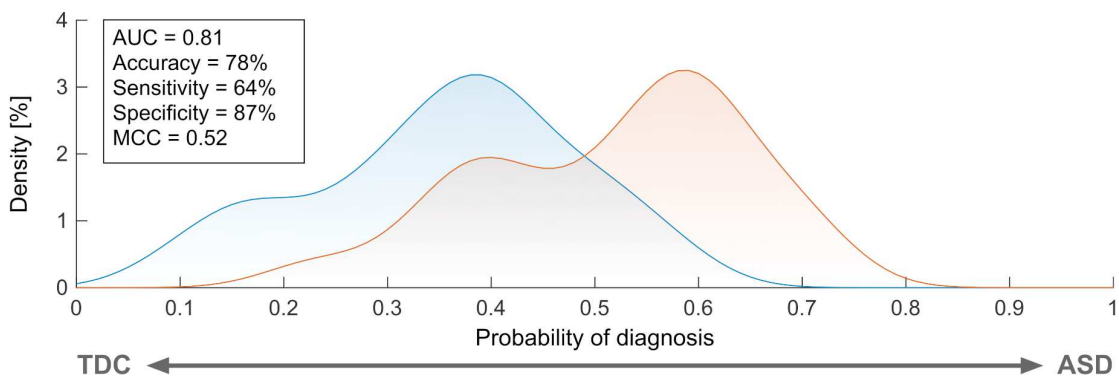
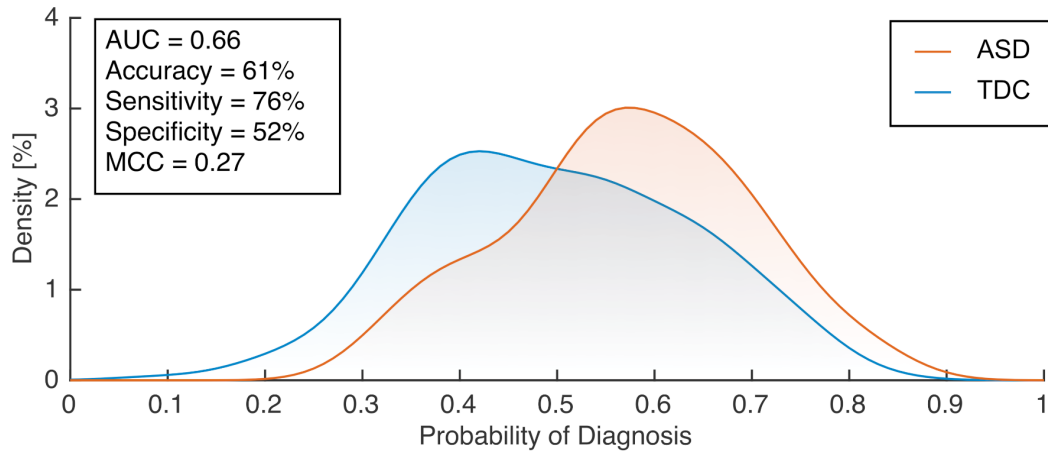


Fig. 1. The classification performance of the ASD classifier in the discovery and adult validation datasets. (A) The probability of the ASD diagnosis in the discovery dataset. (B) The probability of the ASD diagnosis in the ABIDE adult validation dataset. (C) The probability of the ASD diagnosis in the Japanese adult validation dataset. **Abbreviations:** AUC: area under the curve, ASD: autism spectrum disorder, MCC: Matthews correlation coefficient, and TDC: typically developing control.

A. Probability distribution of the ASD diagnosis in the child dataset



B. Probability distribution of the ASD diagnosis in the adolescent dataset

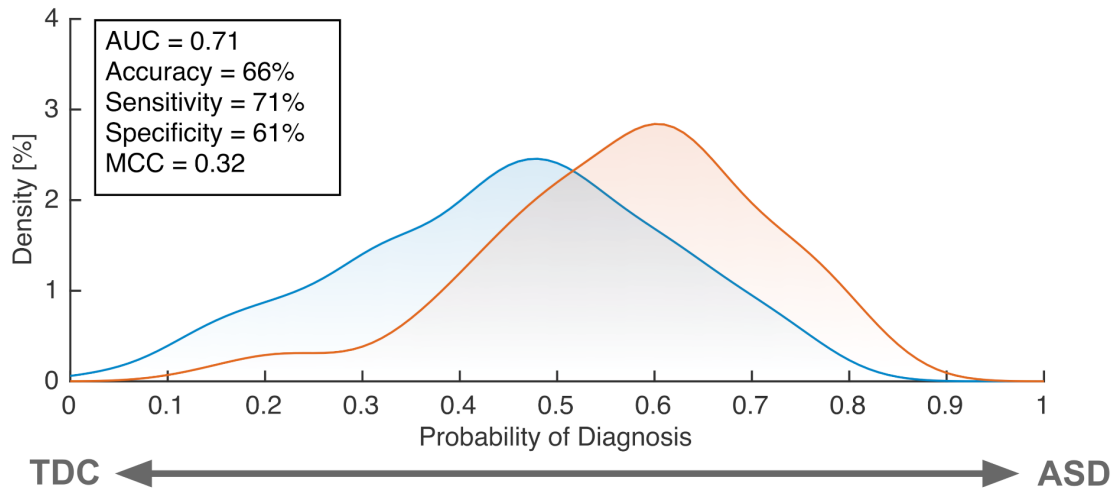


Fig. 2. The classification performance of the ASD classifier in the child and adolescent validation datasets. (A) The probability of ASD diagnosis in children. (B) The probability of ASD diagnosis in adolescents. Abbreviations: AUC: area under the curve, ASD: autism spectrum disorder, MCC: Matthews correlation coefficient, and TDC: typically developing control.

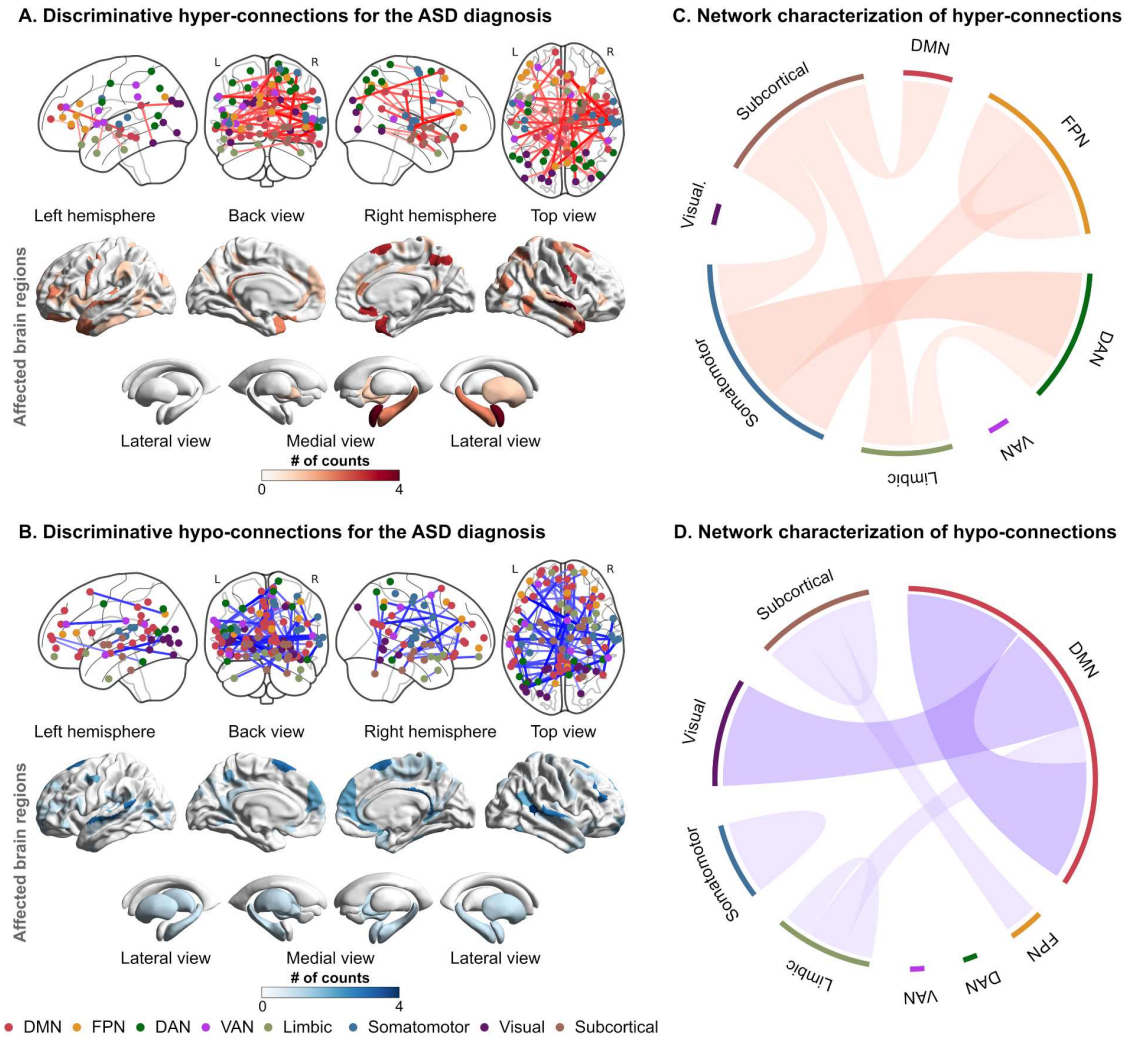


Fig. 3. The spatial distribution and network-based characterization of discriminative functional connections for the ASD diagnosis. (A) The spatial distribution of discriminative hyper-connections and affected brain regions for the ASD diagnosis. (B) The spatial distribution of discriminative hypo-connections and affected brain regions for the ASD diagnosis. The node color represents the corresponding resting-state network. (C) Network-based characterization of hyper-connections. (D) Network-based characterization of hypo-connections. **Abbreviations:** ASD: autism spectrum disorder, DAN: dorsal attention network, DMN: default mode network, FPN: frontoparietal network, TDC: typically developing control, and VAN: ventral attention network.

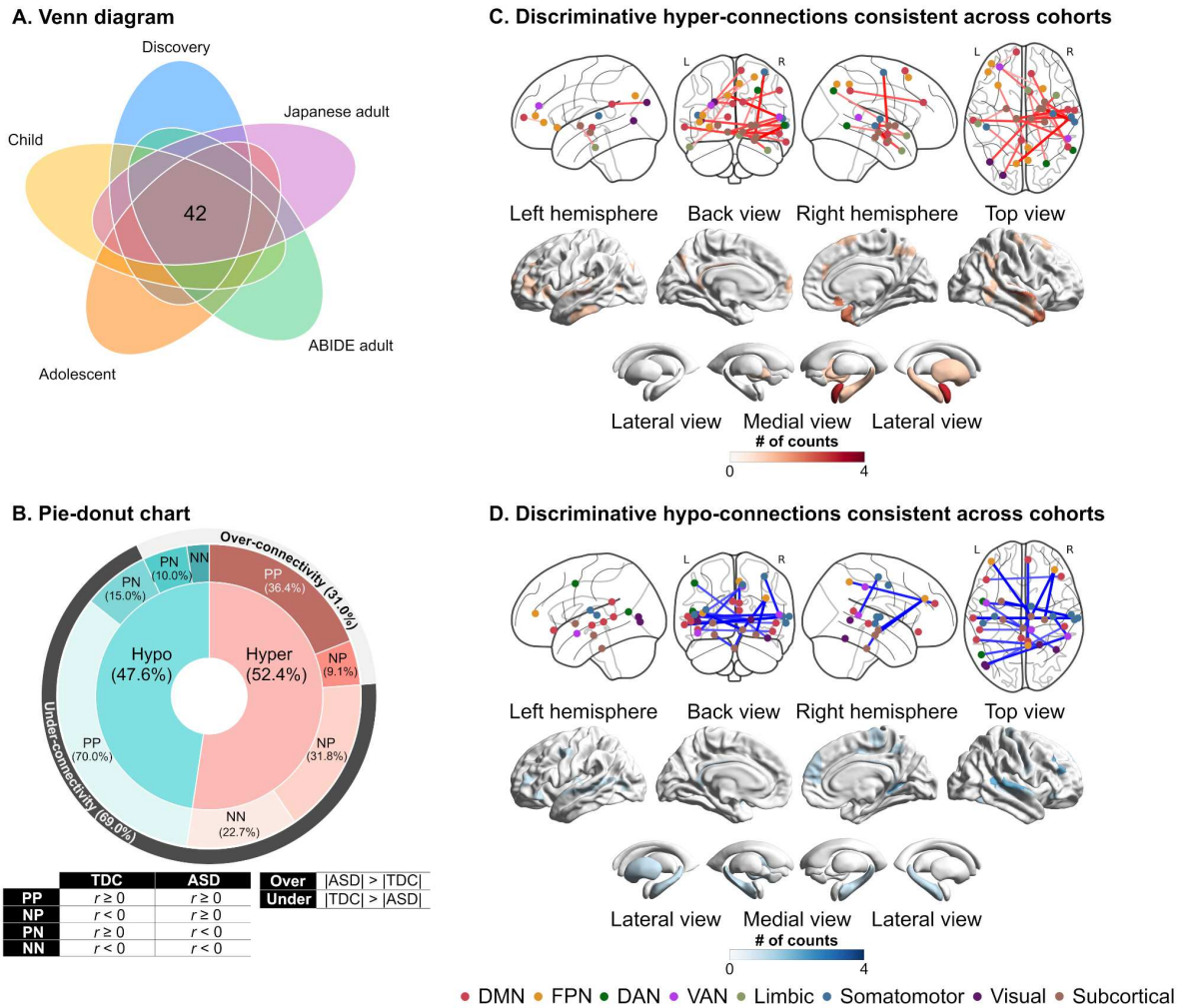


Fig. 4. The distribution of atypical functional connections that are consistent across the datasets. (A) The spatial distribution of discriminative hyper-connections consistent across cohorts. (B) The spatial distribution of discriminative hypo-connections consistent across cohorts. (C) Pie-donut chart showing the details of reproducible hypo- and hyper-connections. **Abbreviations:** ABIDE: autism brain imaging data exchange, DAN: dorsal attention network, DMN: default mode network, FPN: frontoparietal network, L: left, N: Negative, P: positive, R: right, and VAN: ventral attention network.

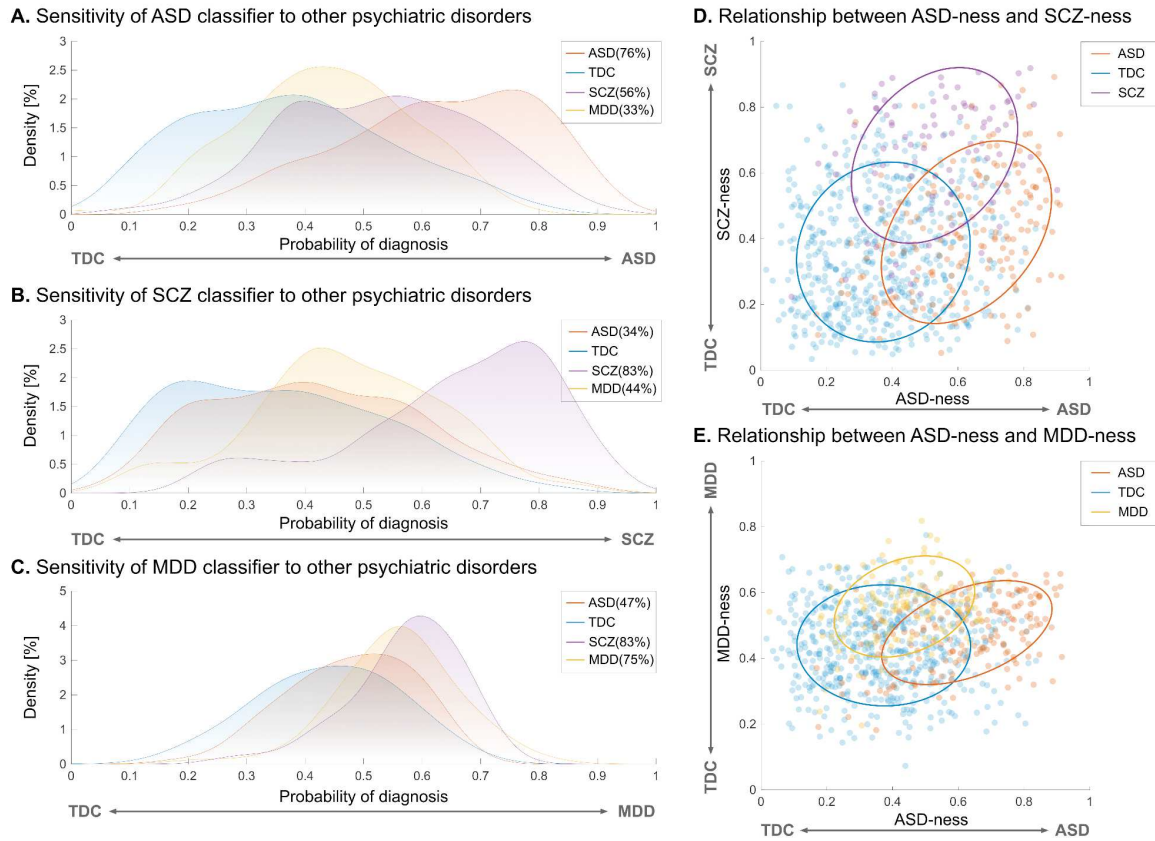


Fig. 5. Sensitivity of each neuromarker to other psychiatric disorders and dimensional relationships between ASD and other psychiatric disorders. (A) Sensitivity of the neuromarker for autism spectrum disorder (ASD) to schizophrenia (SCZ) and major depressive disorder (MDD). (B) Sensitivity of the SCZ neuromarker to ASD and MDD. (C) Sensitivity of the MDD neuromarker to ASD and SCZ. (D) Dimensional relationship between the ASD-ness and SCZ-ness defined by the diagnostic probability in ASD and SCZ classifiers, respectively. (E) Dimensional relationship between the ASD-ness and MDD-ness defined by the diagnostic probability in ASD and MDD neuromarkers, respectively.

Supplementary Files

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

- [SI.pdf](#)