Generalisable functional imaging classifiers of schizophrenia have multifunctionality as trait, state, and staging biomarkers

Authors:

Takahiko Kawashima[1*†], Ayumu Yamashita[2,3†], Yujiro Yoshihara[1], Yuko Kobayashi[1], Naohiro Okada[4,5], Kiyoto Kasai[4,5,6], Ming-Chyi Huang[7], Akira Sawa[8,9,10,11,12,13], Junichiro Yoshimoto[14,15,16], Okito Yamashita[2,17], Toshiya Murai[1], Jun Miyata[1,18], Mitsuo Kawato[2,19*], and Hidehiko Takahashi[1,20,21*]

Affiliations:

[1] Department of Psychiatry, Graduate School of Medicine, Kyoto University, Kyoto, Japan.

[2] Brain Information Communication Research Laboratory Group, Advanced Telecommunications Research Institute International, Kyoto, Japan;

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

[4] Department of Neuropsychiatry, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan.

[5] The International Research Centre for Neurointelligence (WPI-IRCN) at The University of Tokyo Institutes for Advanced Study (UTIAS), The University of Tokyo, Tokyo, Japan.

[6] UTokyo Institute for Diversity and Adaptation of Human Mind (UTIDAHM), The University of Tokyo, Tokyo, Japan

[7] Department of Psychiatry, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan.

[8] Department of Psychiatry, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

[9] Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

[10] Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore,

MD, USA.

[11] Department of Pharmacology, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

[12] Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

[13] Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA.

[14] Nara Institute of Science and Technology, Ikoma, Japan.

[15] International Centre for Brain Science (ICBS), Fujita Health University, Toyoake, Japan.

[16] Department of Biomedical Data Science, Fujita Health University School of Medicine, Toyoake, Japan.

[17] RIKEN, Centre for Advanced Intelligence Project, Tokyo, Japan.

[18] Department of Psychiatry, Aichi Medical University, Nagakute, Japan.

[19] XNef Incorporation, Kyoto, Japan.

[20] Department of Psychiatry and Behavioural Sciences, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan.

[21] Centre for Brain Integration Research, Tokyo Medical and Dental University, Tokyo, Japan.

*Corresponding authors. E-mail: tkawashima@kuhp.kyoto-u.ac.jp (T.K.), kawato@atr.jp (M.K.), hidepsyc@tmd.ac.jp (H.T.)

†These authors contributed equally to this work.

ABSTRACT (147 words)

This study introduces a novel resting-state functional connectivity (rs-FC) magnetic resonance imaging (MRI) biomarker for diagnosing schizophrenia spectrum disorder (SSD), developed using customized machine learning on an anterogradely and retrogradely harmonized dataset from multiple sites, including 617 healthy controls and 116 patients with SSD. Unlike previous rs-FC MRI biomarkers, this new biomarker demonstrated a high accuracy rate of 77.3% in an independent validation cohort, including 404 healthy controls and 198 patients with SSD from seven different sites, while overcoming across-scan variability. It specifically identifies SSD, differentiating it from other psychiatric disorders. Our analysis identified 47 important FCs significant in SSD classification, many of which are involved in the pathophysiology of SSD. These FCs could be potential trait, state, and staging markers for effectively predicting delusional tendencies, specific symptoms, and disease stages. This research underscores the potential of rs-FC as a clinically applicable neural phenotype marker for SSD.

INTRODUCTION

Schizophrenia spectrum disorder (SSD) is a chronic mental disorder affecting approximately 1% of the global population, ranking among the top ten causes of disease burden. Individuals with SSD experience symptoms such as hallucinations, delusions, and cognitive impairment[1]. SSD diagnosis is often subjective and relies on diagnostic criteria that categorise disorders based solely on signs and symptoms. Similar to the case with other psychiatric disorders, there are currently no widely applicable, objective biomarkers[2] that can distinguish individuals with SSD from their healthy counterparts with minimal overlap.

Generating a classifier of SSD using resting-state functional connectivity (rs-FC) magnetic resonance imaging (MRI) is an emerging research topic. However, the practical use of these classifiers is hindered by several key challenges[3–6]. First, there is an accuracy problem. In studies with a large sample size and external validation[7,8], classifier performance was below biomarker requirements (approximately 80%). Accuracy typically decreases with increase in sample size ($N > 200$) in whole-brain imaging biomarkers[9]. A machine learning algorithm requiring many explanatory variables (FCs) needs more data to achieve accuracy and generalisability; however, large training samples from multiple facilities might cause a large site effect, deteriorating data quality[10]. Second, there is an issue with generalisability. Most previous studies used only internal or limited external validation and lacked comprehensive external validation from independent cohorts obtained in multiple sites not involved in the initial research[11]. Third, across-session variability poses a problem. The reliability of resting-state functional (rs-f) MRI is questionable over repeated tests[12], which could affect the stability of the biomarker. Lastly, there is a lack of specificity in potential biomarkers identified in rs-FC MRI studies focusing on specific disorders. Overcoming these issues is crucial for developing a truly effective biomarker for clinical use in SSD.

While disease trait markers, including diagnostic markers, assist early intervention,

other biomarkers are also essential: state markers for estimating dynamic changes in response to treatment[13,14] and staging markers for disease prevention or personalising interventions[15]. These biomarkers have been developed separately in some studies[16,17], and other studies assumed a shared background[18–20]. In SSD research, few studies have succeeded in developing state markers using neuroimaging, and for staging markers, there is no consensus on biological staging models[21].

Patient biological data may simultaneously encompass information on traits, states, and disease stages. As machine learning exploits these features without making distinctions, diagnostic markers developed using machine learning may also function as state and/or staging markers. Previous rs-FC biomarker studies reported multiple aggregated FCs as biomarkers[22,23]; however, whether individual FCs can be used as one or more of these three types of biomarkers remains unexplored. A previous study showed that FCs originally identified as diagnostic markers dynamically changed after treatment[24]; thus, they also functioned as state markers. Nevertheless, thoroughly investigating individual FCs has not been attempted, making it difficult to determine whether any FCs could benefit early disease detection and intervention as a trait marker or treatment target selection and drug discovery as a surrogate marker.

Our research focused on two key areas (**Fig. 1**). (I) Developing Clinically Applicable SSD Classifiers. We aimed to develop a clinically viable rs-FC biomarker for SSD, addressing the challenges mentioned above. This first involved using a large, harmonized dataset from four institutions, harmonized both prospectively[25] under a national project (Strategic Research Program for Brain Science (SRPBS) & Brain/Mind Beyond) and retrospectively using the travelling subject method[10]. Second, we employed a customized sparse learning algorithm, specifically, the least absolute shrinkage and selection operator (LASSO), similar to our previous work on major depressive disorder (MDD)[26]. This algorithm classified individuals

and identified key contributing FCs. Furthermore, we attempted to enhance the performance by introducing a voting method[27]. Third, we reduced across-imaging variability and improved generalization to an independent validation cohort through spatial averaging of 100 classifiers, each analysing tens of FCs. Finally, we tested the biomarker's disorder specificity on other psychiatric disorders. In developing the biomarker, we identified the laterality effect on our biomarker, an idea originating from the findings in FC patterns. (II) Exploring Trait / State / Staging Markers of SSD. We aimed to investigate whether the diagnostic biomarker can be used as state and staging markers, assuming shared underpinnings. We demonstrated this on two levels: first, we assessed how much the important FCs (in the aggregate) had capability as trait and/or state markers by predicting symptom severity of two clinical scales; second, we investigated which individual FCs were related to any of the states or disease stages. The findings of this study could help develop new approaches for improving the diagnosis and treatment of various psychiatric disorders, not only SSD.

RESULTS

Datasets

We used two independent multi-disorder datasets: one for developing classifiers ('discovery dataset') and the other for validating the classifiers ('validation dataset'). Participant characteristics are presented in **Table 1** and **Supplementary Table S1**. Fifty-one participants whose scrubbed rs-fMRI volumes exceeded +3 standard deviations (S.D.) were excluded from all datasets. Therefore, 1,015 participants in the discovery dataset and 683 participants in the validation dataset were included in further analyses. Through the scrubbing process, 9.2 ± 17.9% (mean ± S.D.) of whole volumes per rs-fMRI scan were removed across all datasets.

Group difference of FCs and their reproducibility

First, we performed a mass univariate analysis of the difference in mean FC values between the healthy control (HC) and SSD groups. **Supplementary Fig. 1** shows the *t* values obtained from the discovery and validation datasets, where each plot corresponds to a single FC. By comparing to the data with permutated diagnosis labels (5,000 permutations), the original *t* values were significantly correlated between the two datasets ($r_{(71,631)}$ = 0.597; 95% confidence interval (CI) = [0.590, 0.605]; permutation test, P<0.05, two-sided). Hence, the distribution of the diagnostic effect across all FCs was sustained across the two mutually independent cohorts.

Performance of the LASSO classifiers

Through 10-fold cross-validation (CV) and averaging the subsample outputs, we obtained 100 classifiers. Within the discovery dataset, classification accuracy was 79.6%, with an area under the receiver operating characteristic curve (AUC) of 0.876. Sensitivity, specificity, and Matthews' correlation coefficient (MCC) were 81.5%, 79.2%, and 0.484, respectively. The density curve of HCs and patients with SSD is shown in **Fig. 3a**, where a subject with a

predicted probability of >0.5 was classified as a patient and vice versa. The curves for each site are shown in **Fig. 3b**. We found relatively high predictability for sites KUT and SWA, whereas site UTO achieved somewhat lesser scores.

To examine generalisability, we applied the classifiers to the validation dataset. The classifiers distinguished patients with SSD from HCs with 77.3% accuracy and an AUC of 0.824, similar to the results from the discovery dataset (**Fig. 3c**). Sensitivity, specificity, and MCC were 69.2%, 81.1%, and 0.490, respectively. A permutation test revealed that the AUC and MCC were highly significant (P<0.001). The classifiers showed relatively high prediction capability for sites KTT, COBRE, and TMU (**Fig. 3d**), whereas they showed lower but acceptable capability for KUP. Notably, at JHU, the classifiers distinguished HCs correctly in most cases (specificity: 83.6%) but not patients with SSD (sensitivity: 40.6%). The probability density curve of SSD was closer to that of HC (P=0.37, two-sided binomial test), although the distribution of the two curves was significantly different (P=0.013, two-sample Kolmogorov-Smirnov test), suggesting that patients with early-stage SSD at JHU fell between chronic-stage SSD and HC.

Important FCs for predicting diagnosis

We identified the important FCs that contributed more frequently to the prediction of LASSO classifiers above the chance level. We permutated class labels in the discovery dataset, created 100 quasi-classifiers, and repeated the procedure 100 times to obtain a null distribution of the maximal selection count by quasi-classifiers. At P<0.05 in the permutation test, the FCs selected by 17 or more classifiers were deemed important. Ultimately, we identified 47 important FCs, as shown in **Fig. 4** and detailed in **Supplementary Table S2**

Furthermore, we investigated the detailed patterns of FC value differences between HC and SSD groups and their reproducibility across the two datasets by plotting the mean

values for 47 FCs, facilitating a comparison between the datasets (**Fig. 6**). Notably, the relationship between the mean FC values of the HC and SSD groups was maintained in the validation dataset for 44 of 47 FCs.

Performance of the voting classifiers

Similar to the LASSO classifiers, we assessed the classification performance of the voting classifiers in two steps using the same two datasets. **Supplementary Fig. 2a, b** displays the probability density curve of the voting classifier's output through 10-fold CV in the discovery dataset. Scores were lowest at UTO among the three sites that included HC participants and those with SSD (as observed in the LASSO classifiers). However, all metrics except sensitivity indicated higher scores in voting classifiers than those in the LASSO classifiers, narrowing the performance gap between UTO and the other sites.

Independent validation for all sites in the validation dataset is presented in **Supplementary Fig. 2c**. We conducted statistical tests to objectively compare the performance of the voting classifiers with that of the LASSO classifiers. Sensitivity was significantly higher for the voting classifiers than for the LASSO classifiers ($P<0.001$, McNemar test, two-sided). However, specificity showed no significant difference between groups ($P=0.192$, McNemar test, two-sided). The AUC of the voting classifiers was 0.841, which was not significantly different from that of LASSO classifiers ($P=0.11$, DeLong's test). At the individual site level (**Supplementary Fig. 2d**), the performance (AUC, accuracy, sensitivity, specificity, and MCC) was lower at JHU and KUP among the five sites that included HCs and patients with SSD for LASSO and voting classifiers. Similar to UTO in the discovery dataset, JHU and KUP showed relatively major improvements in performance, except for specificity, when considering the minor shifts in performance indices seen at the other sites.

Supplementary analyses on classifiers' characteristics

LASSO and voting classifiers were built on subsamples in which the mean ages of the HC and SSD groups were matched to mitigate the confounding effect of age. Furthermore, we evaluated whether the classifiers' performance was influenced by confounding factors (see **Supplementary Text S1** and **Table S3**). Consequently, the performance of voting classifiers could be partly influenced by age but not by other potential confounding factors. In contrast, the LASSO classifiers were unaffected by any potential confounding variables.

We also investigated how the classification performance varied based on disease severity in the validation dataset. Following a previous report[28], we identified subgroups with different severities based on Positive and Negative Syndrome Scale (PANSS) total scores: mild (PANSS≤58, N=55), moderate (58<PANSS≤75, N=36), marked (75<PANSS≤95, N=20), severe (95<PANSS≤116, N=1), and most severe (116<PANSS, N=0). Except for the severe and most severe subgroups, each of which contained only one or no participants, the sensitivity by subgroup was highest in the marked subgroup, followed by the mild and moderate subgroups (**Supplementary Table S4**). This order of performance was consistent for the LASSO and voting classifiers, but the disparity among the subgroups decreased for the voting classifiers compared with the LASSO, suggesting that voting classifiers can predict classes with less imbalance across different disease severity levels.

Important FCs and subanalysis on laterality

Glasser's regions of interest (ROIs) included in the 47 important FCs were sorted by frequency (**Supplementary Table S5**). By considering random combinations of two ROIs from the 379 ROIs, ROIs of three times or more were deemed "significantly frequent" (P<0.05). Notably, for the frequencies of these seven ROIs, there was a laterality imbalance between the left and right hemispheres, except in the cerebellum. We assessed how this laterality might affect SSD

biomarkers in these ROIs by comparing the weighted linear sum (WLS) of the original FCs that included the concerned ROI ($WLS_{original}$) with the WLS of the mirror FCs (WLS_{mirror}) (**Supplementary Text S2**). **Supplementary Fig. S3** displays the WLS for the original and mirror FCs of the HC and SSD, respectively, for each of the seven ROIs. We found a significant group difference for $WLS_{original}$ and WLS_{mirror} in five ROIs [two sample *t*-test, Bonferroni-corrected $P<0.05/21$]. To elucidate the specific differences by diagnosis originating from laterality, we assessed the statistical significance of the difference between HC and SSD on WLS difference ($WLS_{diff} = WLS_{original} - WLS_{mirror}$), and the group difference of this index was significant in the putamen ($P=2.19\times10^{-5}$) and the superior frontal language area, a part of the supplementary motor area ($P=1.06\times10^{-4}$) [two sample *t*-test, Bonferroni corrected $P<0.05/21$].

Furthermore, we investigated whether FC laterality correlated with anatomical laterality. Previous studies have reported that volume laterality in subcortical brain regions is altered in schizophrenia[29]. Focusing on the same seven ROIs, we calculated the anatomical laterality index (LI) of each ROI (**Supplementary Text S3** for method) and conducted a correlation analysis with WLS_{diff}. We found no significant correlation between any of the ROIs (**Supplementary Fig. S4**).

Classifier specificity for SSD

We explored the generalisability of the classifiers to other mental disorders by applying them to data from participants with major depressive disorder (MDD), autism spectrum disorder (ASD), and bipolar disorder (BP). If the output probability was >0.5, the participants were assumed to have SSD-like characteristics. The discovery dataset included patients with MDD, ASD, and BP, and the validation dataset only included patients with MDD. The LASSO classifiers revealed that the patients with ASD and BP did not exhibit high- or low-SSD-like

characteristics. However, patients with MDD were less SSD-like (P=2.24×10^{-7}), resembling the HCs (**Fig. 4, Supplementary Table S7**). Voting classifiers showed a similar pattern. These results suggest that both classifiers exhibited high specificity for SSD.

In addition, we investigated whether each probability density curve was distributed differently. Both the LASSO and voting classifiers revealed that the curve of each non-SSD disorder had a significantly different distribution from those of HC and SSD in the discovery and validation datasets (**Supplementary Table S8**). Furthermore, the curves for any two non-SSD disorders showed no significant differences.

Prediction of clinical scale scores using important FCs

Another objective of this study was to determine the extendibility of the trait marker to the state or disease stage. In the first part of this investigation, we attempted to predict the scores of two clinical scales with important FCs *in the aggregate* so we could aid clinical assessment and estimate the extent to which our biomarker could function as a trait and/or state marker, using the Peters et al. Delusions Inventory (PDI) and PANSS total scores. These scales have been widely used to assess delusional thinking and psychotic syndrome[30]. PDI may reflect trait[31] as well as state[32], whereas the PANSS may reflect the overall symptom severity of psychosis at that time point (state)[28].

Using a 10-fold CV, we performed linear regression on the discovery dataset and predicted the scores of the test data for each model. Subsequently, the models were applied to the validation dataset for prediction.

Regarding PDI total scores, the results of 10-fold CV in the discovery dataset were acceptable (r=0.231, 95% CI=[0.00101, 0.438], P=0.0492, two-sided, mean absolute error [MAE]=40.0). Weak correlations between the actual and predicted scores were observed in the validation dataset (r=0.331, 95% CI=[0.0609, 0.556], R^2=0.110, P=0.0177, two-sided,

MAE=55.5) (**Fig. 5a**). A permutation test was conducted to objectively assess whether the evaluation metrics were satisfactory. The correlation coefficients were significantly high (P_{perm}=0.01) and MAE significantly low (P_{perm}=0.018).

Concerning PANSS total scores, the 10-fold CV in the discovery dataset resulted in no correlation between actual and predicted scores (r=−0.0065, 95% CI=[−0.199, 0.186], P=0.948, two-sided, MAE=17.5). The model did not predict scores in the validation dataset either (r=0.128, 95% CI=[−0.0588, 0.306], R^2=0.016, P=0.178, two-sided, MAE=16.8) (**Fig. 5b**).

In summary, we predicted the PDI total score more efficiently than the PANSS total score, suggesting that, collectively, the important FCs may function as trait and state markers.

Individual FCs that were associated with PANSS factors

In addition to the relationship between clinical scale scores and aggregated FC biomarkers, we explored the roles of individual FCs. While individual FCs are expected to be associated with certain traits, they might be linked to the state of SSD. Hence, we determined the association of each important FC with the PANSS factors using linear regression models. In total, 17 FCs were significantly associated with the PANSS factors. Using the three-factor model[33] of PANSS, we identified significant associations in FCs #45 (R.TPOJ1 & R.Thalamus, with the PANSS positive factor) and #8 (L.1 & R.3b, with the PANSS negative factor), regardless of the score conversion method (**Table 2**, **Supplementary Table S9**). With the five-factor model[34] of PANSS, we found that the significantly associated FCs, regardless of conversion method, were FCs #18 (L.FOP1 & R.Putamen), #41 (R.6a & R.PoI1), #43 (R.FOP4 & R.Putamen) with the excited factor, and FC #27 (R.RSC & R.SFL) with the depressed factor.

When the coefficient estimate was significant only for a state (not for the trait), such FCs were assumed to be “pure” state markers. We found nine FCs (**Table 2**). Among these FCs,

FC #27 (R.RSC & R.SFL) was the only one consistently identified as a "pure" state marker, regardless of conversion method.

Individual FCs that were associated with disease stage

Using a multiple regression model, we categorised important FCs into subgroups associated with trait (diagnosis), the early stage, or the chronic stage of SSD. **Fig. 5c** presents the results using the bootstrapping method (1,000 iterations). We assumed that an FC belonged to a certain category with consistent regression results for over half (>500) of the iterations. Subsequently, the important FCs were categorised most frequently as "trait and chronic stage" (61.7%), followed by "trait and early stage" (6.4%), "trait only" (2.1%), and "trait, early, and chronic stage" (2.1%). No FCs were categorised as "early stage only" or "chronic stage only," indicating that we did not identify any pure staging markers.

DISCUSSION

We developed a diagnostic rs-FC biomarker based on harmonised multicentre data and validated it using large-sample independent data. LASSO and voting classifiers achieved high performance, and we confirmed their generalisability through independent validation, demonstrating minimal machine learning-induced inflation in diagnostic biomarker performance. We also identified distinct properties of each type of classifier. By leveraging logistic regression with LASSO, we extracted the important FCs involved in SSD classification and depicted them within a constellation of large-scale networks. Our biomarker also demonstrated high specificity for SSD across broader dimensionality, encompassing several psychiatric disorders.

This study also aimed to maximise the potential of the rs-FC biomarker (originally developed as a trait marker) to function as a state marker and a staging marker. We employed two approaches, one of which involved predicting clinical scale scores based on the values of important FCs. We successfully predicted delusion proneness in patients with SSD using aggregated FCs. Multiple regression analysis focusing on individual FCs enabled us to comprehensively elucidate the associations between individual FCs and state or disease stage. This study revealed the state and staging of the biomarker characteristics of individual FCs based on generalisable diagnostic biomarkers.

The accuracy, sensitivity, and specificity of the identified diagnostic marker were approximately 80% for the discovery dataset. It demonstrated comparable performance in an external validation with seven international cohorts; this presents a sharp contrast to previous studies[11], including the report on a FC biomarker from our group[23]. This achievement was possible through bi-directional (prospective and retrospective) harmonisation, which enabled the extraction of the disease factor itself, and with optimised machine learning method. Against the concerns on session variability of rsfMRI, our previous study[35] revealed that test-retest

reliability was acceptable with the same methodology as demonstrated in this study.

The LASSO and voting classifiers exhibited distinct performance characteristics. Voting classifiers demonstrated superiority in sensitivity and a more balanced profile over LASSO classifiers, with the accuracy, sensitivity, and specificity falling within a narrow range (LASSO: 69.2–81.1%, voting: 74.7–78.8%) in independent validation.

When we categorised patients with SSD into subgroups based on disease severity, LASSO classifiers tended to identify patients with more severe symptoms correctly but not those with milder symptoms; in contrast, voting classifiers identified patients with SSD more evenly across severity subgroups. This observation suggests that voting classifiers capture the SSD trait better, whereas LASSO classifiers are more influenced by the state.

We identified important FCs that significantly contributed more frequently to SSD classification in LASSO classifiers. Notably, based on Anatomical Automatic Labelling (AAL), the putamen, insula, thalamus, and cingulum were among the top ROIs most frequently found in the important FCs (**Supplementary Table S5**). In schizophrenia research, these are consistently implicated in grey matter volume reduction[36] or FC abnormality[37,38]. Moreover, these ROIs were associated with the cortico-striatal-thalamic-cortical loop and salience network, both recognised for their pivotal roles in the psychopathology of SSD[39]. In the context of FC, these ROIs provide an intriguing discussion of SSD pathophysiology. Hypoconnectivity between the putamen and anterior cingulate cortex (FC #13, #30, #37, #38) in the SSD group aligns with a previous report[40], where greater ACC-putamen connectivity predicted better treatment response. We also observed hypoconnectivity between the putamen and insula (FC #17, #43), consistent with another report[41] indicating hypoconnectivity between the dorsal anterior insula and bilateral putamen in early-stage schizophrenia and high-risk clinical groups. Conversely, in the SSD group, the thalamus exhibited increased connectivity with various cortical ROIs (FC #11, #12, #24, #25, #35, #36, and #45), corresponding with a previous

report[42] on thalamocortical connectivity in schizophrenia, implying disrupted information filtering in SSD, consistent with the literature[43]. Sensorimotor areas (precentral and postcentral gyri, supplementary motor area) were also among the top ROIs (see below). Therefore, these important FCs aptly reflected the neural correlates of schizophrenia and were considered trait markers of SSD.

Reduced laterality in structure and function has been suggested in patients with schizophrenia[44,45]. Regarding the putamen and SFL ROIs, we observed significant group differences in WLS_{diff}, which we interpreted as the effect of FC laterality on our LASSO classifiers. Because our approach to measuring FC laterality was novel, a direct comparison of these results with previous findings was not feasible. Nevertheless, we noted a convincing agreement with a study[46] in which the authors reported meaningful laterality of target ROIs when the bilateral putamen was set as the seed. Laterality in the superior frontal language area (a part of the supplementary motor area) was decreased in the SSD group, consistent with previous studies demonstrating diminished laterality in language-related areas in patients with schizophrenia[47,48].

FC laterality and volume laterality were not correlated to any of the ROIs. A previous report showed a correlation between functional laterality and grey matter volume laterality at the whole-brain level; however, this was not specific to SSD[49]. To our knowledge, no studies have explored the relationship between FC laterality and volume laterality; thus, further investigation is needed.

We observed that the LASSO and voting classifiers exhibited high specificity for SSD, whereas other diagnoses (MDD, ASD, and BP) did not show specificity for HC or SSD. Evidence indicates that these psychiatric disorders share several characteristics with SSD in the alteration pattern of brain function[50–52]. Moreover, these disorders have similar phenotypes (e.g., cortical thickness) and genotypes[53]. Thus, our biomarkers may, to some extent, reflect

neural changes common to psychiatric disorders[54].

The second objective of this study was to dissect the biomarker into a "trait marker" and other components, such as a state marker or staging marker. Using aggregated important FCs, we achieved an acceptable level of prediction for the PDI but not for the PANSS, indicating that aggregated FCs were more strongly associated with traits. Predicting the intensity of psychotic conditions is crucial for clinical decision-making. Few groups have reported successful prediction of symptoms using FC. One study reported the prediction of the PANSS total score improvement from baseline based on striatal resting-state markers with a large sample size, but it lacked external validation[18]. Another group predicted PANSS positive and negative scores that were weak to moderate correlations with the actual scores[55], benefiting from individually defined ROIs; however, the sample size was small and lacked external validation. In this study, the predicted PDI total score was significantly correlated to the actual score in the discovery and validation datasets.

Although we successfully predicted scale scores with important FCs in the aggregate, we still lack knowledge regarding which FC was specifically related to each state. Using linear regression analysis, we identified state markers individually.

Using the three-factor model of the PANSS, we observed a significant association between positive scores and FC #45 as well as negative scores and FC #8, irrespective of the score conversion method. Evidence suggests that FC #45 significantly correlates with positive symptoms[56]. FC #8 also showed a significant association with the negative and disorganised factors of the five-factor model, which could be linked to the finding that voxel-mirrored homotopic connectivity of the precentral and postcentral gyri were negatively correlated with the PANSS total score[57].

We discovered interesting overlaps of ROIs in the potential state marker FCs. Concerning negative and depressed factors, sensorimotor areas, including the pre/postcentral

gyri and supplementary motor area, were noticeable. Specific functional alterations in these ROIs related to neurological soft signs (NSSs), a core feature of SSD, have been reported[58]. Moreover, NSSs correlate with PANSS negative scores and depression scale scores[59]. The excited factor seems to involve the putamen, insula, Rolandic operculum, and middle cingulum, regions likely to be associated with aggression, especially in studies on grey matter volume[60]. The Rolandic operculum and middle cingulum are related to aggression via disruption of the cognitive control network[61].

Although we did not discover any pure staging markers, our results provide a compelling argument for "trait and early stage" markers exemplified by FCs #16, #20, and #23. FCs #20 and #23 represented connections between regions around the superior temporal sulcus (STS) or gyrus (STG). In groups with high risk for SSD[62], the response to face recognition in STS, as well as the connectivity between bilateral STS regions, was correlated with schizotypy scores. A heightened response in carriers of a risk allele for SSD compared with non-carriers was also reported. Hence, STS activity appeared to be associated with SSD traits. However, a task-based fMRI study on working memory showed attenuated activity in the STG in patients with early-stage psychosis compared with HCs[63]. Moreover, the cortical thickness of the insula (an ROI in FC #16) and the STS region are reduced in early-stage psychosis[64]. This evidence suggests that STS/STG alterations should also be markers of early-stage disease. The fact that the probability curve for patients with early-stage psychosis at JHU lies in the middle between the HC and SSD groups may also support this argument.

This study had several limitations. First, the participants with SSD in our study were mostly medicated; hence, the applicability of our results to drug-naive patients cannot be guaranteed. Second, we only confirmed that our classifiers distinguished patients with SSD from healthy individuals not from individuals with other psychiatric disorders exhibiting psychotic symptoms. Further studies are required to test the classifiers for these disorders to

maximise the clinical applicability. Third, in the analysis on laterality, only a small number of regions were included in the important FCs, while the "lateralisation hypothesis" of psychosis covers the entire brain[65]. In future research, a broader scope is needed to verify FC laterality at the whole-brain level across different disease stages and in other disorders.

In conclusion, we developed robust and clinically usable classifiers for SSD using a combination of cutting-edge strategies. While constructing the classifiers, we identified FCs that may play key roles in the pathophysiology of SSD. We also demonstrated that these "important FCs" had multiple functions as SSD trait, state, or staging markers, which was supported by their roles in predicting clinical scale scores and regression analyses that identified individual FCs relevant to symptoms or disease stages. Our findings shed new light on the early diagnosis of SSD as well as the selection of targets for neuromodulation.

ONLINE METHODS

Participants

The discovery dataset comprised data from four sites: Kyoto University Siemens TimTrio scanner (KUT), Showa University (SWA), Centre of Innovation in Hiroshima University (COI), and University of Tokyo (UTO). The dataset included 1,045 participants consisting of 617 HCs, 116 patients with SSD (which includes schizophrenia, schizoaffective disorder, and delusional disorder), 148 patients with MDD, 125 patients with ASD, and 39 patients with BP (**Table 1** and **Supplementary Table S1**.

The validation dataset comprised international cohorts, including the Japanese cohort from Hiroshima Kajikawa Hospital (HKH), Hiroshima Rehabilitation Centre (HRC), Hiroshima University Hospital (HUH), Kyoto University Siemens Trio scanner (KTT), Kyoto University Siemens Prisma scanner (KUP). It also included the Taiwanese cohort (Taipei Medical University, TMU) and the Centre of Biomedical Research Excellence (COBRE) open

dataset (http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html), along with the Johns Hopkins University cohort (United States), which included HCs and patients with early-stage SSD. The total number of participants was 708 (405 HCs, 198 patients with SSD, and 105 patients with MDD; **Table 1** and **Supplementary Table S1**). There was no overlap of subjects between the discovery and validation datasets.

Clinical scale scores: PANSS[33] and PDI 21-item version[31], as well as information about the dosage of antipsychotics, were available from a subset of participants in both datasets (**Supplementary Table S10**).

This study was conducted in accordance with the recommendations of the review boards of institutions affiliated with the principal investigators are affiliated—namely, Hiroshima University, Kyoto University, Showa University, and University of Tokyo. The majority of the material data in this study can be downloaded from the DecNef Project Brain Data Repository[67] (https://bicr-resource.atr.jp/srpbsopen/).

Data acquisition

All data in the discovery dataset were collected using a unified protocol developed by SRPBS (https://bicr.atr.jp/decnefpro/) Project in Japan[25]. The MRI data comprised a T1-weighted structural image, rs-fMRI acquired using an echo-planar imaging technique, and field map images. The duration of rs-fMRI was 10 min. Participants were instructed to relax, stay awake, fixate on the central crosshair mark, and not concentrate on specific things. The MRI data for the validation dataset included a structural image and rs-fMRI similar to the discovery dataset; however, some of the data lacked fieldmap images. Scanning parameters for the validation dataset varied by site. The duration of rs-fMRI was approximately 4–6 min. Detailed imaging parameters for both datasets are provided in **Supplementary Table S11**.

Preprocessing

The data were preprocessed according to a previous report[26]. fMRIPrep version 20.1.11 was used for data preprocessing[66]. First, the first four volumes of the rs-fMRI scan were discarded for T1 equilibration. The preprocessing steps were as follows: slice-timing correction, realignment, coregistration, susceptibility-induced distortion correction using field maps, segmentation of a T1-weighted structural image, normalisation to Montreal Neurological Institute (MNI) space, and spatial smoothing with an isotropic Gaussian kernel of 6 mm full-width at half-maximum. For the participants without field maps in the validation dataset, we applied "fieldmap-less" distortion correction, which involves nonlinear registration of fMRI to the same-subject T1-weighted image as the undistorted target, implemented in fMRIPrep. Further details of the pipeline are available at http://fmriprep.readthedocs.io/en/latest/workflows.html.

Signal extraction

We employed two approaches for fMRI signal extraction: (1) A surface-based approach following the Human Connectome Project (HCP) pipelines. Using the ciftify toolbox version 2.3.2[67], we converted volume-based MRI data into data based on "greyordinate" (cortical grey matter surface vertices and subcortical grey matter voxels). The value of a cortical vertex was determined as the weighted mean of the voxel values in the cortical ribbon corresponding to the vertex, similar to the procedure in the *fMRIsurface* pipeline of HCP. (2) An approach based on the ROI. For reliable surface-based brain parcellation, we adopted the ROIs from Glasser et al.[68] (379 parcels in total, comprising 360 cortical parcels as surface ROIs and 19 subcortical parcels as volume ROIs). Using these approaches, we extracted BOLD signal time courses from 379 ROIs. To compare these ROIs with conventional annotations of brain areas and intrinsic brain networks, we referred to the AAL and Ji et al.[69], respectively.

Noise removal

We used component-based noise correction (CompCor)[70], a module in fMRIPrep that performs principal component analysis on the entire time series of the preprocessed rs-fMRI, to detect physiological noise. Anatomical CompCor (aCompCor) was applied to the subcortical white matter (WM) and cerebrospinal fluid (CSF) regions, and the top five principal components were estimated as physiological noise, except for one participant who had only four components. Accordingly, we regressed out these components together with six head-motion parameters and averaged the signals over the entire brain.

Temporal filtering

We employed temporal bandpass filtering ranging from 0.01–0.08 Hz[71] to the time series of rs-fMRI to extract the low-frequency brain activity characterising resting state.

Data scrubbing

We removed volumes with considerable head motion based on framewise displacement (FD)[72], which is one of the outputs of fMRIPrep. FD was calculated as the sum of the absolute displacements in translation and rotation between two consecutive volumes. We removed volumes with FD > 0.5 mm, as proposed in a previous study[72]. In addition, participants whose scrubbed volume ratio exceeded the mean +3 S.D. were excluded.

Calculation of the FC matrices

We defined FC as the temporal correlation of rs-fMRI BOLD signals between two ROIs. We calculated the Pearson's correlation coefficient of the preprocessed BOLD time series between each pair of ROIs out of Glasser's 379 ROIs. Fisher's Z-transformed values of the correlation

coefficients constituted an FC matrix for each participant, where the total number of FC was $\binom{379}{2} = 71{,}631$.

Harmonisation of the site differences

We harmonised the site effects in the discovery dataset using data from subjects who travelled between sites (travelling subjects). Site effects were separately estimated as measurement bias and sampling bias from the rs-fMRI data of these travelling subjects, each of whom underwent scans at multiple sites[10]. We subtracted the estimated measurement bias to obtain harmonised connectivity data (see **Supplementary Text S4**). For the validation dataset, we applied the ComBat harmonisation method[73–75], which estimates the site effect as one because the scanners in this dataset did not have travelling subject data. In the execution of ComBat harmonisation, we inputted the diagnosis, PANSS total score, PDI total score, age, sex, handedness, and duration of illness (for patients with SSD) as auxiliary variables to correct measurement bias.

Group difference of FCs

To delineate the cumulative effect of diagnosis on FCs, we conducted a mass univariate analysis. We calculated the difference in mean FC values (Z-transformed) between the HC and SSD groups for each FC and computed *t* statistic. To validate the reproducibility of the diagnostic effect on FCs for both datasets, we performed a permutation test (5,000 times) using *mult_comp_perm_corr* function in MATLAB (R2019a, MathWorks, USA) and assessed the significance of the correlation of the *t* values mentioned above between the two datasets.

Diagnostic classifiers for SSD

Next, we constructed classifiers to differentiate patients with SSD from HCs using machine learning based on 71,631 FC values as features. In subsequent sections, we used only the data

of HCs and patients with SSD, except for generalising the models to other disorders. Initially, we created the classifiers using logistic regression with LASSO. The sparse method in LASSO can prevent overfitting and simultaneously select important features (for the detailed methodology, refer to **Supplementary Text S5**). Although we used LASSO to identify important FCs for SSD classification, we applied a voting method to enhance the classification performance of the LASSO classifiers.

Building LASSO classifiers

As illustrated in **Fig. 6a**, we implemented a nested 10-fold CV scheme. In the outer loop, we divided the discovery dataset into ten folds. After leaving one fold as the test set, the remaining nine folds were used as the training set. To minimise the bias arising from the imbalance between the number of patients with SSD and HCs, we conducted subsampling with undersampling simultaneously. We randomly sampled the same number (N = 102) of HCs and SSDs from the training set, creating a subsample. During subsampling and undersampling, we matched the mean ages of HCs and SSDs. We fit the logistic regression model to the subsample while tuning a hyperparameter with an inner-loop 10-fold CV (for the detailed methodology, see **Supplementary Text S5**). By repeating random subsampling and fitting the model ten times, we obtained ten classifiers. We then predicted SSD probability for each participant in the test set. By applying the ten classifiers built from a training set to a test set in each CV, we obtained the probability values for the participant as the classifier outputs. When the mean probability value was >0.5, we considered the participant's class as SSD; otherwise, as HC. Finally, by repeating the above procedure ten times in the outer loop, 100 classifiers were obtained.

We evaluated the performance of the classifiers using the following indices: AUC, accuracy, sensitivity, specificity, and MCC (for MCC, see **Supplementary Text S6** for details).

Independent validation of LASSO classifiers

To assess the generalisability of the obtained classifiers, we tested them on the validation dataset. We applied the 100 classifiers to each participant in the validation dataset to compute diagnostic probability values for each participant. We classified a participant as having SSD if the average probability value was >0.5.

We also conducted a statistical analysis of the classification performance for independent validation using a permutation test. We created 100 quasi-classifiers using the same procedure as for building genuine classifiers, with the participants' classes permutated in the discovery dataset. We predicted the diagnosis in the same manner as mentioned above using the quasi-classifiers on the validation dataset for each permutation. By repeating random permutations 500 times and obtaining null distributions of the AUC and MCC, we evaluated the statistical significance of the true classifiers.

Identifying important FCs for predicting diagnosis

We investigated which FCs were utilised to predict the diagnosis in each classifier by identifying the nonzero coefficients in LASSO. We counted the number of classifiers (out of 100) that selected each FC as an explanatory variable. We performed a permutation test to identify the most informative FCs from 71,631 FCs. Each time we randomly permutated the class labels of participants in the discovery dataset, we built 100 quasi-classifiers using 10-fold CV with 10-time subsampling, following the methodology described in the previous section (see "**Building LASSO classifiers**" section). We determined the maximum counts for which each FC was selected as a predictive explanatory variable using 100 quasi-classifiers for each permutation. This procedure was repeated 100 times, resulting in a null distribution of 100 values for the maximum selection count. An FC was considered significantly informative

('important FC') when the selection count in the genuine 100 classifiers exceeded the threshold that corresponded to $P < 0.05$ in the null distribution.

Voting classifiers

In addition to logistic regression with LASSO, we further explored the diagnostic biomarker's potential by combining multiple machine-learning methods. We incorporated support vector machine (SVM)[76], random forest (RF)[77], light gradient boosting machine (LGBM)[78], and multi-layer perceptron (MLP)[79] as representative algorithms. The machine-learning packages for each algorithm were LightGBM version 3.2.1 for LGBM, Keras version 2.6.0 for MLP, and scikit-learn version 0.24.1 for SVM and RF. For each algorithm, we conducted CV, subsampling with undersampling, and training on the discovery dataset following the same procedure used to build the LASSO classifiers. Hyperparameters were tuned using Optuna version 2.9.1 in 10-fold CV within the inner loop of the nested CV. We constructed 10 classifiers by repeating the subsampling and training of the model 10 times. In each fold of the outer-loop CV, we predicted the class of each participant in a test set using these 10 classifiers (for LASSO, LGBM, and MLP, the default output was in the form of a probability value. Therefore, when the probability value was > 0.5, the predicted class was 'SSD' and vice versa). Similar to voting, we counted the number of times the participant was predicted to have an SSD by ten classifiers for each of the five machine learning algorithms (maximal count: 10 classifiers × 5 algorithms = 50). When the count exceeded 25, the participant was considered to have SSD. Using 10-time subsampling and 10-fold CV for each of the five algorithms, we obtained 500 classifiers. Next, we predicted the diagnostic class in the validation dataset using each of these classifiers, and the final prediction by voting was determined as the class predicted by the majority of the classifiers (> 250).

To statistically compare the classification capability of the voting classifiers with that

of the LASSO classifiers, we used the R programmes Compbdt and pROC. Using Compbdt[80], we compared the two classifiers based on their sensitivity and specificity using the McNemar test. We used pROC to conduct DeLong's test on the AUC. The significance level for all tests was set at $P < 0.05$.

Generalisation of the classifiers to other psychiatric disorders

Regarding the LASSO classifiers, we applied 100 classifiers to all the patients with MDD, ASD, and BP from discovery and validation datasets. If the output probability was >0.5, the participants were assumed to have SSD-like characteristics. Regarding the voting classifiers, we applied 500 classifiers to the same patients, and the participants were assumed to be SSD-like if over half (>250) of the classifiers predicted as such. The discovery dataset included patients with MDD, ASD, and BP, whereas the validation dataset only included patients with MDD. The outputs of the classifiers for HC and SSD were used for the comparison with MDD, ASD, and BP. Regarding HC and SSD, the output for test data in the 10-fold CV was used in the discovery dataset, and the output of 100 classifiers was used in the validation dataset.

To evaluate if the classifying results of patients with any of the other disorders (MDD, ASD, BP) had similarity to HC, SSD, or neither, we conducted a two-sided binomial test (the significance level: $P<0.05$). We also checked whether each probability density curve was distributed differently using the two-sample Kolmogorov-Smirnov test. We conducted this test for every combination of HC, SSD, MDD, ASD, and BP; thus, ten combinations were obtained in total. The level of significance was $P<0.05/10=0.005$ [Bonferroni-corrected].

Prediction of clinical scores using important FCs

We predicted the scores on the clinical scales (PDI and PANSS total scores) using important FCs as explanatory variables. We applied a nested 10-fold CV scheme to the discovery dataset

(only SSD participants with available target scale scores) to build prediction models (**Supplementary Fig. 5a**). In the outer loop, we divided the discovery dataset into 10 folds. After leaving one fold as the test set, we used the remaining nine folds as the training set. We fitted the linear regression model to the training set (the *LinearRegression* module in scikit-learn version 0.24.1). In the inner loop, we searched for the most suitable number of important FCs to be used as features to avoid overfitting (for the detailed methodology, see **Supplementary Text S7 & Supplementary Fig. 5b**). We predicted the score for each participant in the test set using a linear regression model with the best number of features for the fold. Using 10-fold CV, we obtained 10 models with different numbers of features. For the validation dataset, we averaged the outputs from the ten models to create the final predicted score (**Supplementary Fig. 5c**). Notably, at any step of this analysis, the predicted value was adjusted within the range of the scale score (if the prediction was lower than the lowest possible value, it was adjusted to the lowest possible value and vice versa).

We conducted a permutation test to assess the evaluation metrics statistically. Specifically, we permutated clinical labels (diagnosis and symptom scores) against FC values, created quasi-models following the same procedure as the genuine models, and repeated these steps 500 times to obtain a null distribution of the evaluation metrics. Differences were considered statistically significant if each evaluation metric for the genuine models was better than the cut-off value of P=0.05 of the null distribution.

Separately identifying FCs associated with the state

In the second part of the analysis of state and staging markers, we focused on individual FCs from the perspective of their association with various factors of state and disease stages. To identify the FCs associated with state (i.e., symptom scale scores), we conducted multiple regression analyses on each of the important FCs. We chose PANSS as the state index because

it changes dynamically with treatment and has been used as a representative scale for the state of schizophrenia. All HC participants and patients with SSD having data on PANSS scores across the discovery and validation datasets were included in the analysis. We examined two models of the PANSS: the original three-factor model[33] composed of positive, negative, and general pathological factors, and the five-factor model[34] composed of positive, negative, disorganised, excited, and depressed factors. Because the general pathological factor was not categorised as a specific symptom dimension, we used only positive and negative factors as explanatory variables for the three-factor model.

Due to the exploratory nature of this analysis, we converted the PANSS scores into explanatory variables in two ways: one by min-max normalisation and the other by binary dummy variables. The formula to fit was the same for both methods:

$$FC = \beta_0 + \beta_1 x_t + \sum_{j=1}^{n_{factor}} \left(\beta_{j+1} x_{s_j} \right) + e \qquad \text{(Eq. 1)}$$

where x_t represents the trait or diagnosis (HC: 0; SSD: 1), x_{s_j} represents the existence of symptoms for the *j*-th factor of the PANSS, and e represents a random error.

Conversion 1. Min-max normalisation of the scores.

The value x_{s_j} was determined using the following formula:

$$x_{s_j} = \frac{m_j - \min(m_j)}{\max(m_j) - \min(m_j)} \qquad \text{(Eq. 2)}$$

where m_j represents the participant's average score in the *j*-th factor of the PANSS, and $\min(m_j)$ (or $\max(m_j)$) represents the minimum (or maximum) average score of the *j*-th PANSS factor across all patients with SSD with available PANSS subscore data.

Conversion 2. Binarising the scores.

$$x_{s_j} = \begin{cases} 0 & if \ m_j \leq 2 \\ 1 & if \ m_j > 2 \end{cases} \qquad \text{(Eq. 3)}$$

The value x_{s_j} was determined as in Eq.3 since three points or more is considered pathological

in PANSS rating.

All s_j for HC participants were assumed to be zero. In the regression analysis for each FC, the FC was regarded as a potential state marker when any of the coefficient estimates for PANSS factor ($\beta_{(j+1)}, j = 1, 2, \ldots n_{\text{factor}}$) was non-zero at a significance level $P < 0.05$ with a one-sample *t*-test. Of note, we reported FCs as being significant only when the state's coefficient concerned ($\beta_{(j+1)}$) was of the same sign as w of the FC in terms of consistency with the underlying LASSO classifiers.

Identifying FCs associated with the disease stage

Lastly, we constructed another multiple regression model to identify FCs associated with the disease stage of SSD, which could be referred to as 'staging' markers. Following a pre-existing definition[81,82], we divided the SSD group into an early-stage psychosis subgroup (duration of illness (DOI) of less than 5 years) and a chronic-stage subgroup (DOI of 5 years or longer). We formulated a regression model similar to that described in the previous section.

$$FC = \beta_0 + \beta_1 x_t + \beta_2 x_e + \beta_3 x_c + \beta_4 x_a + e \qquad \text{(Eq. 4)}$$

$$such\ that \quad x_e = \begin{cases} 1 & if\ \ DOI < 5\ yrs. \\ 0 & if\ \ DOI \geq 5\ yrs. \end{cases}, \quad x_c = \begin{cases} 0 & if\ \ \text{DOI} < 5\ yrs. \\ 1 & if\ \ DOI \geq 5\ yrs. \end{cases} \qquad \text{(Eq. 5)}$$

where x_e represents the early stage of SSD, x_c represents the chronic stage, and x_t and e are defined in the same way as in Eq. 1. The value x_a, the age of the participant, was introduced as a covariate because the DOI was supposed to correlate with age. All HC participants and patients with SSD with DOI data across the discovery and validation datasets were included in the analysis. Using this model, we categorised important FCs into six groups based on which trait, early/chronic stage, was significantly associated: (1) trait only, (2) early stage only, (3) chronic stage only, (4) trait and early stage, (5) trait and chronic stage, and (6) all three. In terms of consistency with our classifiers and within the coefficients, every coefficient estimate

concerned must be of the same sign as w_i; statistical significance was determined at the level $P < 0.05$, with a one-sample *t*-test for each coefficient estimate. Considering the relative instability of the model fit due to singularity, we adopted the bootstrap method (1,000 iterations) to count the number of times a certain FC was sorted into each category.

Acknowledgements:

OY received support from the Japan Agency for Medical Research and Development (AMED) under Grant Number JP23dm0307009.

TM received support from a Grant-in-Aid for Transformative Research Areas (A) (Japan Society for The Promotion of Science, JP21H05173), Grant-in-Aid for Scientific Research (B) (Japan Society for The Promotion of Science, 21H02849), and Strategic International Brain Science Research Promotion Program (Brain/MINDS Beyond) (21dm0307102h0003) of the AMED.

JM received support from the AMED under Grant Number JP21uk1024002, and KAKENHI JP20H05064.

Funding:

This study was supported by KAKENHI JP (23H04979) from the Ministry of Education, Culture, Sports, Science and Technology of Japan, by AMED under Grant Numbers JP19dm0207069, JP18dm0307001, JP18dm0307004, JP18dm0307008, and JP18dm0307009, and by CREST (JPMJCR22P3) from the Japan Science and Technology Agency.

Author contributions:

T.K., A.Y., M.K., and H.T. designed the study. T.K., Y.Y., Y.K., N.O, K.K., M-C.H., A.S., T.M., J.M. and H.T. recruited participants for the study and collected their clinical and imaging data. T.K. and A.Y. carried out the data analysis. T.K., A.Y., and J.M. wrote the original draft of the manuscript. T.K., A.Y., Y.Y., Y.K., N.O, K.K., M-C.H., A.S., J.Y., O.Y., T.M., J.M., M.K., and H.T. 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).

REFERENCES

1. Marder, S. R. & Cannon, T. D. Schizophrenia. *N. Engl. J. Med.* **381**, 1753–1761 (2019).

2. Abi-Dargham, A. & Horga, G. The search for imaging biomarkers in psychiatric disorders. *Nat. Med.* **22**, 1248–1255 (2016).

3. Kraguljac, N. V. *et al.* Neuroimaging biomarkers in schizophrenia. *Am. J. Psychiatry* appi.ajp.2020.2 (2021) doi:10.1176/appi.ajp.2020.20030340.

4. Cai, X. *et al*. Generalizability of machine learning for classification of schizophrenia based on resting□state functional MRI data. *Hum. Brain Mapp.* **41**, 172–184 (2020).

5. Parkes, L., Satterthwaite, T. D. & Bassett, D. S. Towards precise resting-state fMRI biomarkers in psychiatry: synthesizing developments in transdiagnostic research, dimensional models of psychopathology, and normative neurodevelopment. *Curr. Opin. Neurobiol.* **65**, 120–128 (2020).

6. Steyerberg, E. W. & Harrell, F. E. Prediction models need appropriate internal, internal–external, and external validation. *J. Clin. Epidemiol.* **69**, 245–247 (2016).

7. Winterburn, J. L. *et al.* Can we accurately classify schizophrenia patients from healthy controls using magnetic resonance imaging and machine learning? A multi-method and multi-dataset study. *Schizophr. Res.* **214**, 3–10 (2019).

8. Li, C. *et al. Classification of schizophrenia spectrum disorder using machine learning and functional connectivity: Reconsidering the clinical application.* http://medrxiv.org/lookup/doi/10.1101/2020.05.30.20118026 (2020) doi:10.1101/2020.05.30.20118026.

9. Schnack, H. G. & Kahn, R. S. Detecting neuroimaging biomarkers for psychiatric disorders: Sample size matters. *Front. Psychiatry* **7**, (2016).

10. 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).

11. Porter, A. *et al.* A meta-analysis and systematic review of single vs. multimodal neuroimaging techniques in the classification of psychosis. *Mol. Psychiatry* **28**, 3278–3292 (2023).

12. Noble, S., Scheinost, D. & Constable, R. T. A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *NeuroImage* **203**, 116157 (2019).

13. Lema, Y. Y., Gamo, N. J., Yang, K. & Ishizuka, K. Trait and state biomarkers for psychiatric disorders: Importance of infrastructure to bridge the gap between basic and clinical research and industry: Trait and state biomarkers in psychiatry. *Psychiatry Clin. Neurosci.* **72**, 482–489 (2018).

14. Spellman, T. & Liston, C. Toward circuit mechanisms of pathophysiology in depression. *Am. J. Psychiatry* **177**, 381–390 (2020).

15. McGorry, P. *et al.* Biomarkers and clinical staging in psychiatry. *World Psychiatry* **13**, 211–223 (2014).

16. Taylor, J. A., Larsen, K. M. & Garrido, M. I. Multi□dimensional predictions of psychotic symptoms via machine learning. *Hum. Brain Mapp.* **41**, 5151–5163 (2020).

17. Bohaterewicz, B. *et al.* Machine learning-based identification of suicidal risk in patients with schizophrenia using multi-level resting-state fMRI features. *Front. Neurosci.* **14**, 605697 (2021).

18. Li, A. *et al.* A neuroimaging biomarker for striatal dysfunction in schizophrenia. *Nat. Med.* **26**, 558–565 (2020).

19. Yamashita, A. *et al.* Common brain networks between major depressive-disorder diagnosis and symptoms of depression that are validated for independent cohorts. *Front.*

Psychiatry **12**, 667881 (2021).

20. Miyata, J. *et al.* Functional alterations of two salience-related systems jointly and independently contribute to psychosis. *medRxiv* (2021) doi:10.1101/2021.10.02.21264326.

21. Martínez-Cao, C. *et al.* Is it possible to stage schizophrenia? A systematic review. *Transl. Psychiatry* **12**, 197 (2022).

22. Yahata, N., Morimoto, J. & Hashimoto, R. A small number of abnormal brain connections predicts adult autism spectrum disorder. *Nat. Commun.* **7**, (2016).

23. Yoshihara, Y. *et al.* Overlapping but asymmetrical relationships between schizophrenia and autism revealed by brain connectivity. *Schizophr. Bull.* **46**, 1210–1218 (2020).

24. Ichikawa, N. *et al.* Primary functional brain connections associated with melancholic major depressive disorder and modulation by antidepressants. *Sci. Rep.* **10**, 3542 (2020).

25. Tanaka, S. C. *et al.* A multi-site, multi-disorder resting-state magnetic resonance image database. *Sci. Data* **8**, 227 (2021).

26. Yamashita, A. *et al.* Generalizable brain network markers of major depressive disorder across multiple imaging sites. *PLOS Biol.* **18**, e3000966 (2020).

27. Sherazi, S. W. A., Bae, J.-W. & Lee, J. Y. A soft voting ensemble classifier for early prediction and diagnosis of occurrences of major adverse cardiovascular events for STEMI and NSTEMI during 2-year follow-up in patients with acute coronary syndrome. *PLOS ONE* **16**, e0249338 (2021).

28. Leucht, S. *et al.* What does the PANSS mean? *Schizophr. Res.* **79**, 231–238 (2005).

29. Okada, N. *et al.* Abnormal asymmetries in subcortical brain volume in schizophrenia. *Mol. Psychiatry* **21**, 1460–1466 (2016).

30. Lincoln, T. M., Ziegler, M., Lüllmann, E., Müller, M. J. & Rief, W. Can delusions be self-assessed? Concordance between self- and observer-rated delusions in schizophrenia.

Psychiatry Res. **178**, 249–254 (2010).

31. Peters, E., Joseph, S., Day, S. & Garety, P. Measuring delusional ideation: The 21-item Peters et al. Delusions Inventory (PDI). *Schizophr. Bull.* **30**, 1005–1022 (2004).

32. Balzan, R. P., Delfabbro, P. H., Galletly, C. A. & Woodward, T. S. Metacognitive training for patients with schizophrenia: Preliminary evidence for a targeted, single-module programme. *Aust. N. Z. J. Psychiatry* **48**, 1126–1136 (2014).

33. Kay, S. R., Fiszbein, A. & Opler, L. A. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophr. Bull.* **13**, 261–276 (1987).

34. Wallwork, R. S., Fortgang, R., Hashimoto, R., Weinberger, D. R. & Dickinson, D. Searching for a consensus five-factor model of the Positive and Negative Syndrome Scale for schizophrenia. *Schizophr. Res.* **137**, 246–250 (2012).

35. Okada, G. *et al.* Verification of the brain network marker of major depressive disorder: Test-retest reliability and anterograde generalization performance for newly acquired data. *J. Affect. Disord.* **326**, 262–266 (2023).

36. Chan, R. C. K., Di, X., McAlonan, G. M. & Gong, Q. -y. Brain anatomical abnormalities in high-risk individuals, first-episode, and chronic schizophrenia: An activation likelihood estimation meta-analysis of illness progression. *Schizophr. Bull.* **37**, 177–188 (2011).

37. Tu, P.-C., Hsieh, J.-C., Li, C.-T., Bai, Y.-M. & Su, T.-P. Cortico-striatal disconnection within the cingulo-opercular network in schizophrenia revealed by intrinsic functional connectivity analysis: A resting fMRI study. *NeuroImage* **59**, 238–247 (2012).

38. Li, S. *et al.* Dysconnectivity of multiple brain networks in schizophrenia: A meta-analysis of resting-state functional connectivity. *Front. Psychiatry* **10**, 482 (2019).

39. Peters, S. K., Dunlop, K. & Downar, J. Cortico-striatal-thalamic loop circuits of the salience network: A central pathway in psychiatric disease and treatment. *Front. Syst.*

Neurosci. **10**, (2016).

40. Cadena, E. J. *et al.* Cognitive control network dysconnectivity and response to antipsychotic treatment in schizophrenia. *Schizophr. Res.* **204**, 262–270 (2019).

41. Li, X.-B. *et al.* Altered resting-state functional connectivity of the insula in individuals with clinical high-risk and patients with first-episode schizophrenia. *Psychiatry Res.* **282**, 112608 (2019).

42. Anticevic, A. *et al.* Characterizing thalamo-cortical disturbances in schizophrenia and bipolar illness. *Cereb. Cortex* **24**, 3116–3130 (2014).

43. Andreasen, N. C. The role of the thalamus in schizophrenia. *Can. J. Psychiatry* **42**, 27–33 (1997).

44. Oertel, V. *et al.* Reduced laterality as a trait marker of schizophrenia—Evidence from structural and functional neuroimaging. *J. Neurosci.* **30**, 2289–2299 (2010).

45. Ribolsi, M., Daskalakis, Z. J., Siracusano, A. & Koch, G. Abnormal asymmetry of brain connectivity in schizophrenia. *Front. Hum. Neurosci.* **8**, (2014).

46. Peters, H. *et al.* Changes in extra-striatal functional connectivity in patients with schizophrenia in a psychotic episode. *Br. J. Psychiatry* **210**, 75–82 (2017).

47. Sommer, I. E. C., Ramsey, N. F. & Kahn, R. S. Language lateralization in schizophrenia, an fMRI study. *Schizophr. Res.* **52**, 57–67 (2001).

48. Alary, M. *et al.* Functional hemispheric lateralization for language in patients with schizophrenia. *Schizophr. Res.* **149**, 42–47 (2013).

49. Royer, C. *et al.* Functional and structural brain asymmetries in patients with schizophrenia and bipolar disorders. *Schizophr. Res.* **161**, 210–214 (2015).

50. Wu, X. *et al.* Functional network connectivity alterations in schizophrenia and depression. *Psychiatry Res. Neuroimaging* **263**, 113–120 (2017).

51. Koike, S. *et al.* Shared functional impairment in the prefrontal cortex affects

symptom severity across psychiatric disorders. *Psychol. Med.* **52**, 2661–70 (2020).

52. Jutla, A., Foss□Feig, J. & Veenstra□Vanderweele, J. Autism spectrum disorder and schizophrenia: An updated conceptual review. *Autism Res.* **15**, 384–412 (2021).

53. Writing Committee for the Attention-Deficit/Hyperactivity Disorder *et al.* Virtual Histology of Cortical Thickness and Shared Neurobiology in 6 Psychiatric Disorders. *JAMA Psychiatry* **78**, 47 (2021).

54. Goodkind, M. *et al.* Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry* **72**, 305 (2015).

55. Wang, D. *et al.* Individual-specific functional connectivity markers track dimensional and categorical features of psychotic illness. *Mol. Psychiatry* **25**, 2119–2129 (2020).

56. Ferri, J. *et al.* Resting-state thalamic dysconnectivity in schizophrenia and relationships with symptoms. *Psychol. Med.* **48**, 2492–2499 (2018).

57. Hoptman, M. J. *et al.* Decreased interhemispheric coordination in schizophrenia: A resting state fMRI study. *Schizophr. Res.* **141**, 1–7 (2012).

58. Schröder, J., Wenz, F., Schad, L. R., Baudendistel, K. & Knopp, M. V. Sensorimotor cortex and supplementary motor area changes in schizophrenia. A study with functional magnetic resonance imaging. *Br. J. Psychiatry* **167**, 197–201 (1995).

59. Fountoulakis, K. N., Panagiotidis, P., Gonda, X., Kimiskidis, V. & Nimatoudis, I. Neurological soft signs significantly differentiate schizophrenia patients from healthy controls. *Acta Neuropsychiatr.* **30**, 97–105 (2018).

60. Leclerc, M. P., Regenbogen, C., Hamilton, R. H. & Habel, U. Some neuroanatomical insights to impulsive aggression in schizophrenia. *Schizophr. Res.* **201**, 27–34 (2018).

61. Wong, T. Y. *et al.* Neural networks of aggression: ALE meta-analyses on trait and elicited aggression. *Brain Struct. Funct.* **224**, 133–148 (2019).

62. Yan, Z. *et al.* Hyperfunctioning of the right posterior superior temporal sulcus in response to neutral facial expressions presents an endophenotype of schizophrenia. *Neuropsychopharmacology* **45**, 1346–1352 (2020).

63. Crossley, N. A. *et al.* Superior temporal lobe dysfunction and frontotemporal dysconnectivity in subjects at risk of psychosis and in first-episode psychosis. *Hum. Brain Mapp.* **30**, 4129–4137 (2009).

64. Wen, K. *et al.* Cortical thickness abnormalities in patients with first episode psychosis: a meta-analysis of psychoradiologic studies and replication in an independent sample. *Psychoradiology* **1**, 185–198 (2021).

65. Crow, T. J. Schizophrenia as an anomaly of development of cerebral asymmetry. A postmortem study and a proposal concerning the genetic basis of the disease. *Arch. Gen. Psychiatry* **46**, 1145 (1989).

66. Esteban, O. *et al.* fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).

67. Dickie, E. W. *et al.* Ciftify: A framework for surface-based analysis of legacy MR acquisitions. *NeuroImage* **197**, 818–826 (2019).

68. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).

69. Ji, J. L. *et al.* Mapping the human brain's cortical-subcortical functional network organization. *NeuroImage* **185**, 35–57 (2019).

70. Behzadi, Y., Restom, K., Liau, J. & Liu, T. T. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage* **37**, 90–101 (2007).

71. Satterthwaite, T. D. *et al.* An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage* **64**, 240–256 (2013).

72. Power, J. D. *et al.* Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage* **84**, 320–341 (2014).

73. Fortin, J.-P. *et al.* Harmonization of multi-site diffusion tensor imaging data. *NeuroImage* **161**, 149–170 (2017).

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

75. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).

76. Boser, B. E., Guyon, I. M. & Vapnik, V. N. A training algorithm for optimal margin classifiers. in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory. Colt '92* 144–152 (ACM Press, 1992). doi:10.1145/130385.130401.

77. Tin Kam Ho. Random decision forests. in *Proceedings of 3rd International Conference on Document Analysis and Recognition* vol. 1 278–282 (IEEE Comput. Soc. Press, 1995).

78. Ke, G. *et al.* LightGBM: A Highly Efficient Gradient Boosting Decision Tree. 9.

79. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning Internal Representations by Error Propagation. in *Readings in Cognitive Science* 399–421 (Elsevier, 1988). doi:10.1016/B978-1-4832-1446-7.50035-2.

80. Roldán-Nofuentes, J. A. Compbdt: an R program to compare two binary diagnostic tests subject to a paired design. *BMC Med. Res. Methodol.* **20**, 143 (2020).

81. McGorry, P. D., Killackey, E. & Yung, A. Early intervention in psychosis: Concepts, evidence and future directions. *World Psychiatry* **7**, 148–156 (2008).

82. Newton, R. *et al.* Diverse definitions of the early course of schizophrenia—A targeted literature review. *Npj Schizophr.* **4**, 21 (2018).

1

Table 1. Demographics of the participants (HC and patients with SSD)

Site	Abbrevi-ation	HC			SSD			Total (HC & SSD)		
		Number	M/F	Age	Number	M/F	Age	Number	M/F	Age
Discovery dataset										
Kyoto University (TimTrio)	KUT	223	127/96	33.4±13.2	61	32/29	40.7±13.2	284	159/125	35.1±13.2
Showa University	SWA	101	86/15	31.4±7.9	19	15/4	42.7±8.4	120	101/19	30.7±9.5
Centre of Innovation, Hiroshima University	COI	124	46/78	51.9±13.4	-	-	-	124	46/78	51.9±13.4
University of Tokyo	UTO	169	77/92	35.7±17.5	36	24/12	23.3±10.3	205	101/104	34.9±16.5
Summary		617	336/281	36.9±16.0	116	71/45	38.5±11.4	733	407/326	37.2±15.3
Validation dataset										
Kyoto University (Trio)	KTT	72	44/28	28.7±9.4	48	23/25	37.8±9.4	120	67/53	32.4±10.4
Kyoto University (Prisma)	KUP	11	7/4	35.0±8.5	18	11/7	40.2±12.9	29	18/11	38.2±11.6
Hiroshima University Hospital	HUH	66	29/37	34.6±13.0	-	-	-	66	29/37	34.6±13.0
Hiroshima Kajikawa Hospital	HKH	29	12/17	45.4±9.5	-	-	-	29	12/17	45.4±9.5
Hiroshima Research Centre	HRC	49	13/36	41.7±11.7	-	-	-	49	13/36	41.7±11.7
COBRE	COB	73	50/23	35.7±11.6	67	57/10	37.4±13.8	140	107/33	36.6±12.7

Taipei Medical University	TMU	29	26/3	31.4±5.0		32	23/9	35.3±6.1		61	49/12	33.4±5.9
Johns Hopkins University	JHU	75	37/38	24.4±4.1		33	23/10	22.5±4.7		108	60/48	23.8±4.4
Summary		404	218/186	32.3±11.6		198	137/61	34.9±11.9		602	355/247	33.8±11.7

The distributions of age and sex were not significantly different between the HC and SSD groups in the discovery dataset ($P > 0.05$). In the validation dataset, the age distribution was not significantly different ($P > 0.05$); however, the sex distribution was significantly different ($P < 0.05$). HC: healthy control, SSD: schizophrenia spectrum disorder.

Table 2. Individual FCs significantly associated with PANSS factorial scores

Score conversion	Three factor model		Five factor model				
	Positive	Negative	Positive	Negative	Disorganised	Excited	Depressed
1. Min-max transformation	**#41: R.6a & R.Pol1** #45: R.TPOJ1 & R.Thalamus	**#1: L.3b & L.1** **#8: L.1 & R.3b**	**#5: L.5L & R.Cereb** **#42: R.i6-8 & R.PeEc** #45: R.TPOJ1 & R.Thalamus			#17: L.MI & R.Putamen #18: L.FOP1 & R.Putamen **#41: R.6a & R.Pol1** #43: R.FOP4 & R.Putamen	**#27: R.RSC & R.SFL** **#28: R.POS2 & R.SFL**
2. Binarisation	**#33: R.6ma & R.Cereb** #45: R.TPOJ1 & R.Thalamus	**#2: L.POS2 & R.SFL** #8: L.1 & R.3b **#28: R.POS2 & R.SFL**	**#3: L.PCV & L.POS1**	**#2: L.POS2 & R.SFL** **#8: L.1 & R.3b**	**#8: L.1 & R.3b**	#14: L.IFSa & R.AVI #18: L.FOP1 & R.Putamen #38: R.p32pr & R.Putamen #41: R.6a & R.Pol1 #43: R.FOP4 & R.Putamen #44: R.FOP1& R.Putamen	**#27: R.RSC & R.SFL**

Using multiple regression analyses, the FCs significantly associated with each PANSS factor were identified. When the coefficient for the factor was positive, the FC is coloured red, and when it was negative, the FC is coloured blue. "Pure state markers" (the case in which the explanatory variable of the PANSS factor was solely associated with the FC) are highlighted in bold letters. The ROIs described here are from

1 Glasser's[68] parcellation. FC: functional connectivity, PANSS: Positive and Negative Syndrome Scale, ROI: region of interest.

FIGURE LEGENDS

Figure 1. Outline of the study

This study was composed of two parts. (**I**) Constructing SSD Classifier: Using the discovery dataset, we processed rs-fMRI images into an FC matrix for each participant, which was then inputted into machine learning (LASSO) to build SSD classifiers. We obtained the classification performance through 10-fold CV and examined its external generalisability using the validation dataset. "Important FCs" were those that made the highest contribution to the classification. We conducted further analyses on the applicability of another machine learning method (voting classifiers) to other mental disorders and on FC laterality. (**II**) Exploring Trait / State / Staging markers of SSD: We investigated the different types of biomarkers inherent in important FCs. First, we attempted to predict clinical scale scores using aggregated FCs. Second, we searched for individual FCs associated with the state and/or disease stage. CV: cross-validation, SSD: schizophrenia spectrum disorder, LASSO: least absolute shrinkage and selection operator, rs-fMRI: resting state functional magnetic resonance imaging, FC: functional connectivity.

Figure 2. Probability density curves based on LASSO classifiers

The abscissa represents the predicted probability of SSD as an output of the classifiers. If a participant's probability was over 0.5 (vertical dashed line), the participant was classified as SSD, otherwise classified as HC. The ordinate indicates the density at a certain probability. (**a**) Results for all the sites combined in the discovery dataset. (**b**) Results for individual sites in the discovery dataset. COI has only an HC curve since it has no patients with SSD. (**c**) Results for all the sites combined in the validation dataset. (**d**) Results for individual sites in the validation dataset. Since three sites in Hiroshima (HKH, HRC, and HUH) did not have any patients with SSD, these sites have a curve for HCs only. HC: healthy control, SSD: schizophrenia spectrum disorder, AUC: area under the curve, MCC: Matthews' correlation coefficient, COI: Center of Innovation in Hiroshima University, KUT: Kyoto University TimTrio, SWA: Showa University, UTO: University of Tokyo, LASSO: least absolute shrinkage and selection operator, HKH: Hiroshima Kajikawa Hospital, HRC: Hiroshima Rehabilitation Centre, HUH: Hiroshima University Hospital, KTT: Kyoto University Trio, KUP: Kyoto University Prisma, TMU: Taipei Medical University, JHU: Johns Hopkins University, COBRE: Centre of Biomedical Research Excellence.

Figure 3. Important FCs ($P < 0.05$) in diagnosis prediction by LASSO classifiers

(a) Each node on the inner circle corresponds to an ROI. The line width of the FC shows how many times it was selected by the classifiers, and the line colour denotes the direction to which it contributes to the logistic regression model (red means that the higher the FC value is, the more likely the classifier's output is to be SSD; blue means a lower FC value for a higher likelihood of SSD). **(b)** The 47 important FCs were projected onto glass brains. The colours of the ROIs in relation to each intrinsic brain

network and the red/blue line colours correspond to those in (a). **(c)** The mean FC values (Z-score, on the ordinate) of the 47 important FCs for the HC and SSD groups are shown as a bar plot for the discovery and validation datasets. Error bars represent standard error. The FC numbers on the abscissa correspond to those in (a) and **Supplementary Table S2**. FC: functional connectivity, HC, healthy control; SSD: schizophrenia spectrum disorder, ROI: region of interest, LASSO: least absolute shrinkage and selection operator.

Figure 4. Probability density curves for HC, SSD, and other disorders reveal classifier specificity for SSD

(a) Results for the LASSO classifiers. **(b)** Results for the voting classifiers. HC: healthy control, SSD: schizophrenia spectrum disorder, MDD: major depressive disorder, ASD: autism spectrum disorder, BP: bipolar disorder, LASSO: least absolute shrinkage and selection operator.

Figure 5. Trait, state, and staging marker analyses.

(a, b) Prediction results of PDI total and PANSS total scores. The relationship between the predicted scores (on the ordinate) and actual scores (on the abscissa) for the validation dataset is shown in a scatter plot. The grey translucent band represents the 95% confidence interval of the regression line. **(a)** PDI total scores. **(b)** PANSS total score. (c) Individual FCs associated with disease stage. This heat map shows the results of the bootstrap method, where the number in the colour bar indicates the number of times the FC was sorted into a category out of 1,000 iterations. n.s.: not significant ($P \geq 0.05$ for all three explanatory variables or any of the coefficients' signs were inconsistent with the average weight in LASSO classifiers). MAE: mean absolute error. PDI: Peters et al. Delusion Inventory, PANSS: Positive and Negative Syndrome Scale. FC: functional connectivity, LASSO: least absolute shrinkage and selection operator.

Figure 6. Scheme of the procedures for building classifiers

(a) Workflow chart of the nested 10-fold CV applied to the discovery dataset. By repeatedly subsampling and undersampling from a training set ten times, we obtained ten subsets for each CV fold. Ultimately, 100 classifiers were obtained. **(b)** The resulting 100 classifiers were applied to the validation dataset. The average value of the classifier output was the final output. CV: cross-validation, HC: healthy control, SSD: schizophrenia spectrum disorder.

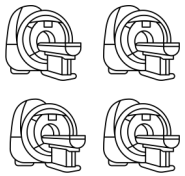
I. Developing Clinically Applicable SSD Classifiers

Concerns about rsfMRI biomarkers

1. Inadequete accuracy
2. Insufficient generalisability
3. Across-scan variability
4. Lack of disease specificity

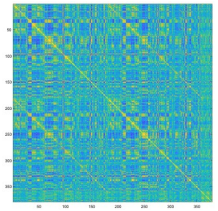
Aim: To overcome these by the methodology below

Discovery dataset
(4 sites)

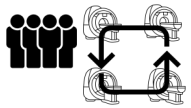


Unified imaging protocol
(prospective harmo.)

- Preprocessing
- Calculate FCs

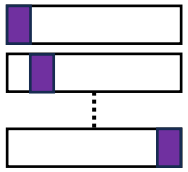


FC matrix



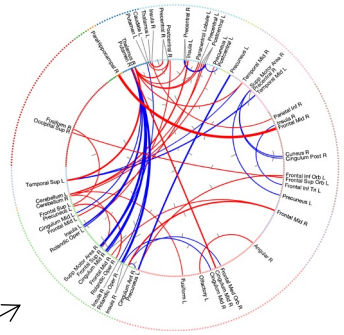
Retrospective harmo.
(travelling subjects)

Nested 10-fold CV
+ subsampling



Customised
machine learning

SSD classifiers



Important FCs

Additional analyses

- Voting classifiers
- Laterality in FCs

Validation dataset (7 sites)

Other psychiatric
disorders

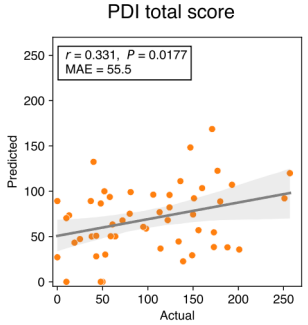
II. Exploring Trait / State / Staging markers of SSD

Trait	Diagnosis, delusional tendency etc.
State	Symptom (subject to change)
Stage	e.g. early, chronic

→ Aggregated

FC → trait, state

Scale score prediction



→ Individual

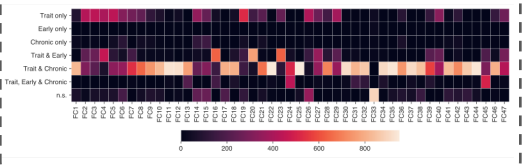
trait, state, staging → FC

Catalog of each marker type

State markers

	Three factor model	*Five factor model*
Positive	#33, #41, #45	#3, #5, #42, #45
Negative	#1, #2, #8, #28	#2, #8
Disorganised		#8
Excited		#14, #17, #18, #38, #41, #43, #44
Depressed		#27, #28

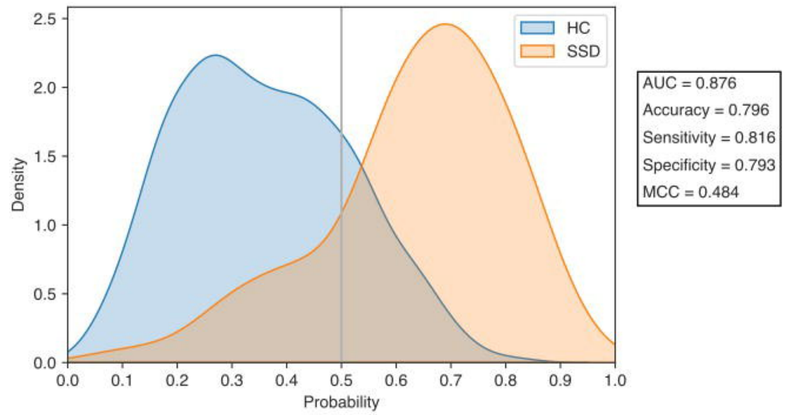
Staging markers



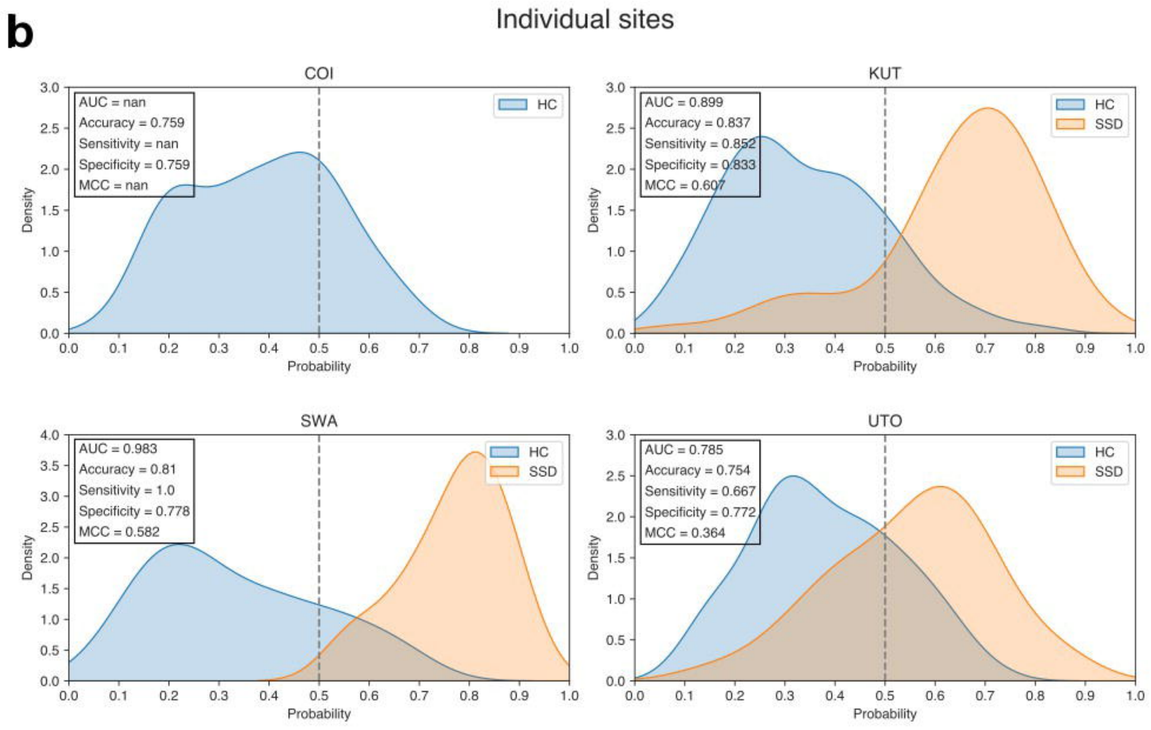
Results for LASSO Classifiers

Discovery dataset

a

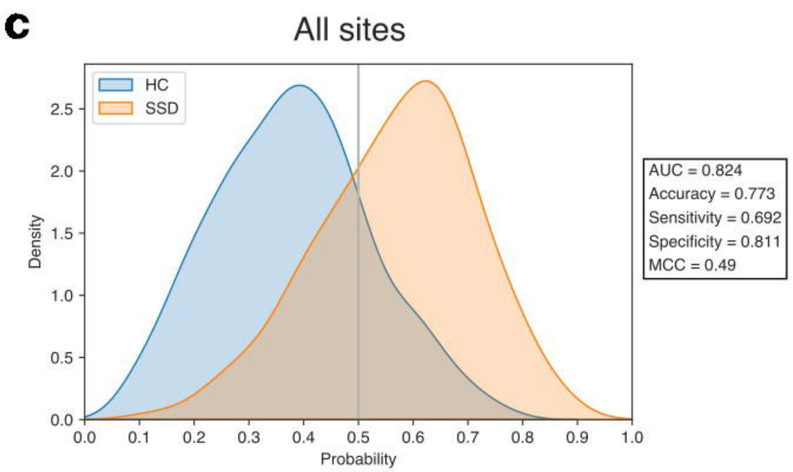


b

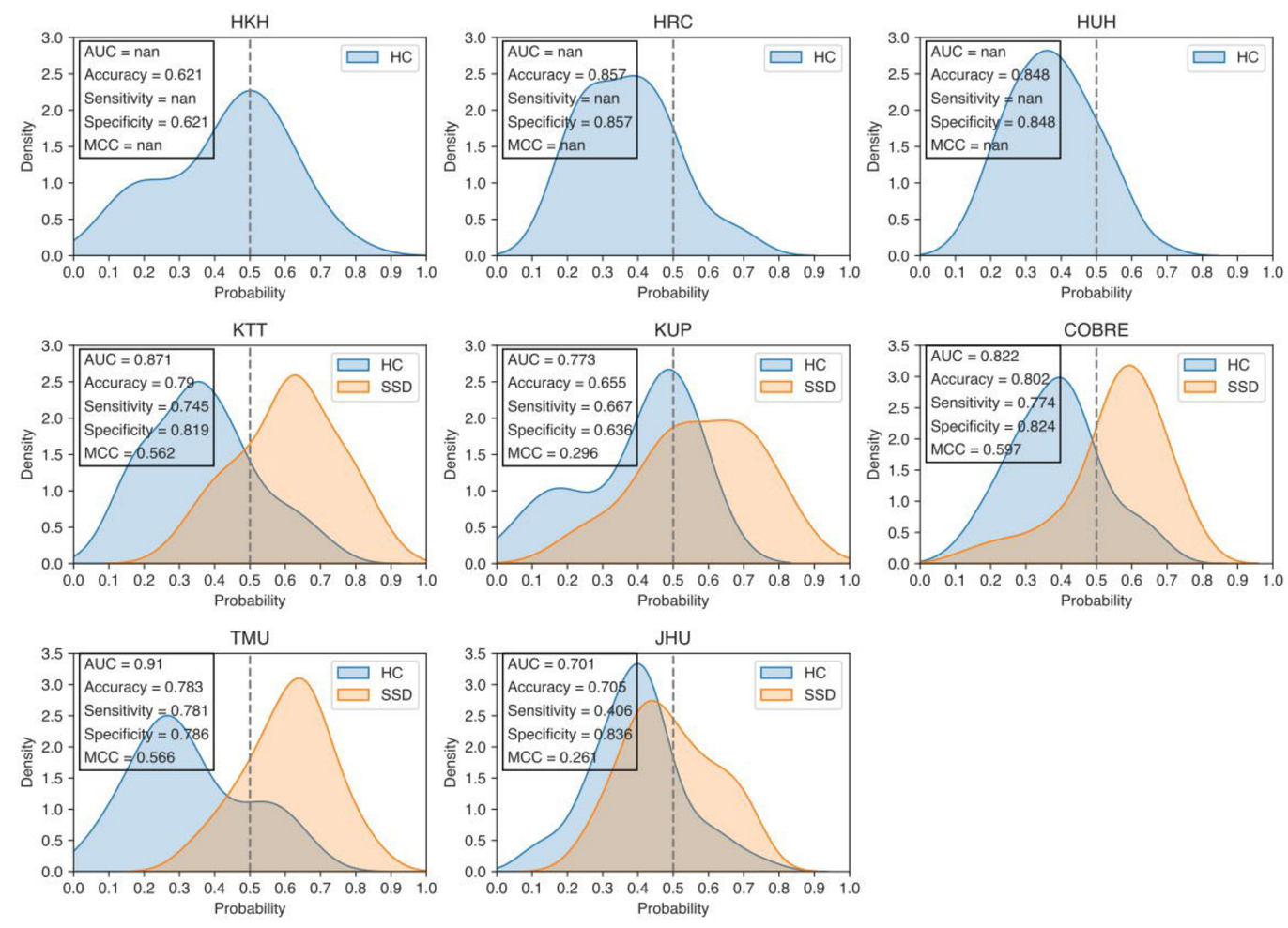


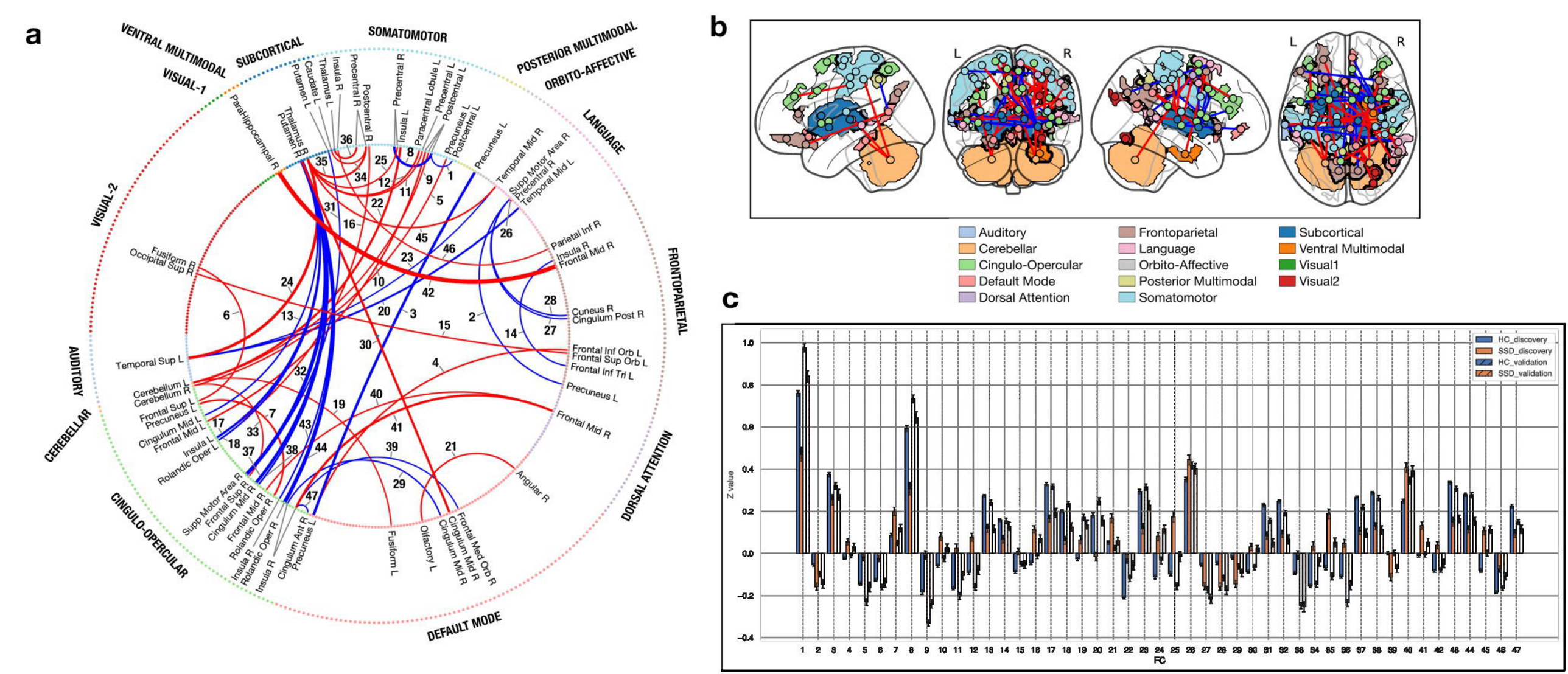
Validation dataset

c

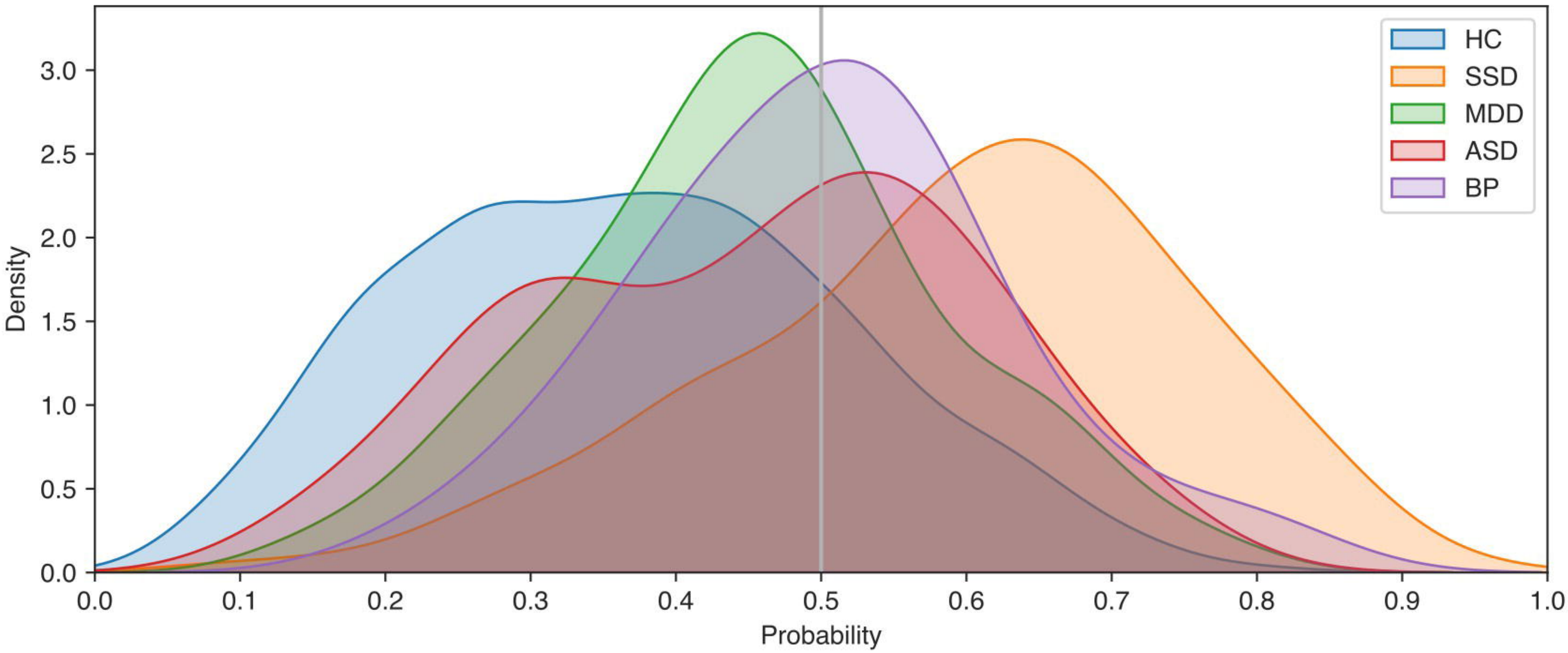


d

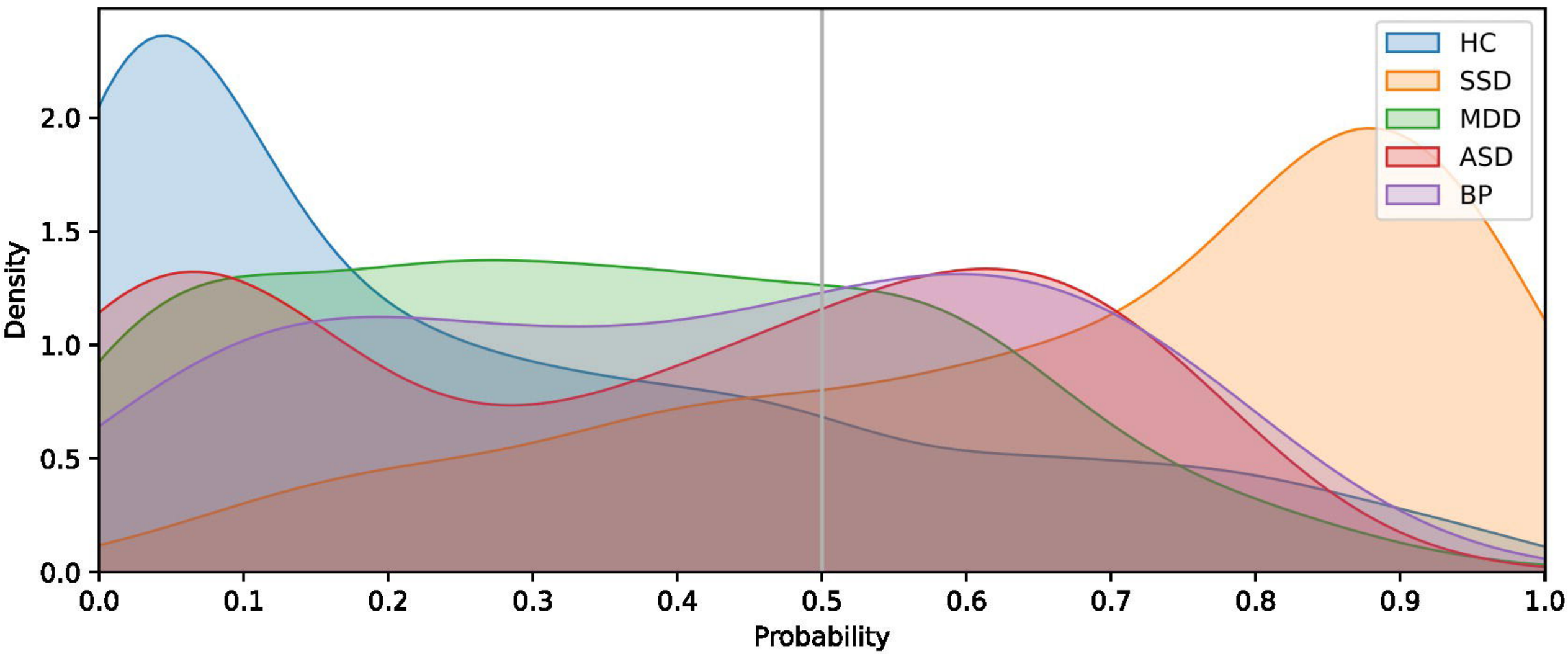



a
VENTRAL MULTIMODAL
SUBCORTICAL
SOMATOMOTOR
POSTERIOR MULTIMODAL
ORBITO-AFFECTIVE
LANGUAGE
VISUAL-1
VISUAL-2
AUDITORY
CEREBELLAR
CINGULO-OPERCULAR
DEFAULT MODE
DORSAL ATTENTION
FRONTOPARIETAL
ParaHippocampal R
Thalamus Putamen R
Putamen L
Caudate L
Thalamus L
Insula R
Precentral R
Postcentral R
Precentral R
Insula L
Paracentral Lobule L
Precentral L
Postcentral L
Precuneus L
Postcentral L
Precuneus L
Temporal Mid R
Supp Motor Area R
Precentral R
Temporal Mid L
Parietal Inf R
Insula R
Frontal Mid R
Cuneus R
Cingulum Post R
Frontal Inf Orb L
Frontal Sup Orb L
Frontal Inf Tri L
Precuneus L
Frontal Mid R
Angular R
Frontal Med Orb R
Cingulum Mid R
Olfactory L
Fusiform L
Cingulum Ant R
Precuneus L
Insula R
Rolandic Oper R
Insula R
Frontal Mid R
Cingulum Mid R
Frontal Sup R
Supp Motor Area R
Rolandic Oper L
Insula L
Frontal Mid L
Cingulum Mid L
Frontal Sup L
Precuneus L
Cerebellum L
Cerebellum R
Temporal Sup L
Occipital Sup R
Fusiform R
b
L
R
L
R
Auditory
Cerebellar
Cingulo-Opercular
Default Mode
Dorsal Attention
Frontoparietal
Language
Orbito-Affective
Posterior Multimodal
Somatomotor
Subcortical
Ventral Multimodal
Visual1
Visual2
c
HC_discovery
SSD_discovery
HC_validation
SSD_validation
Z value
FC

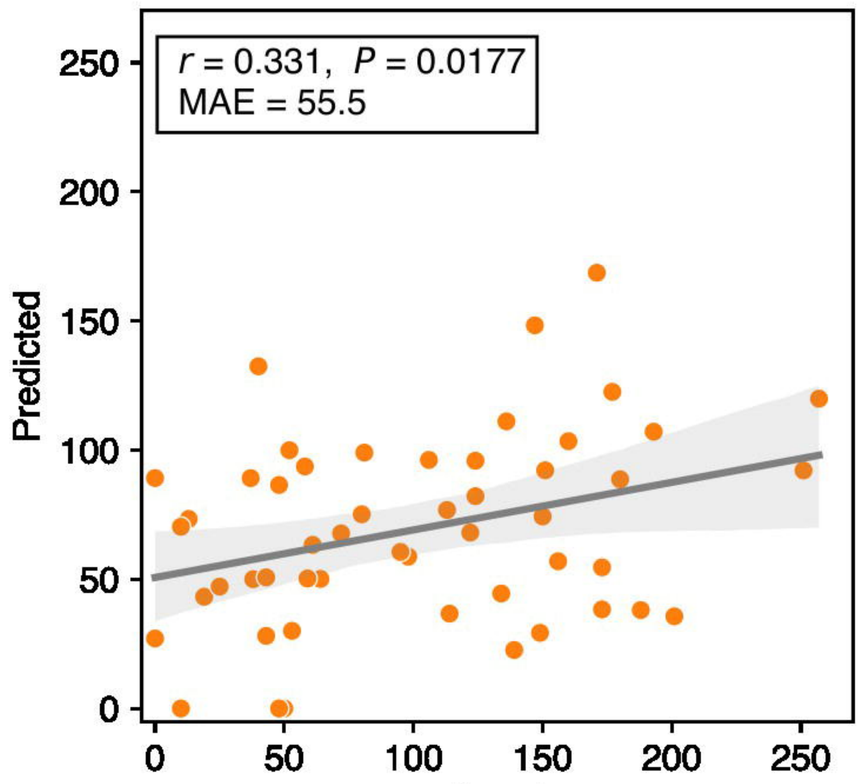
a LASSO Classifiers



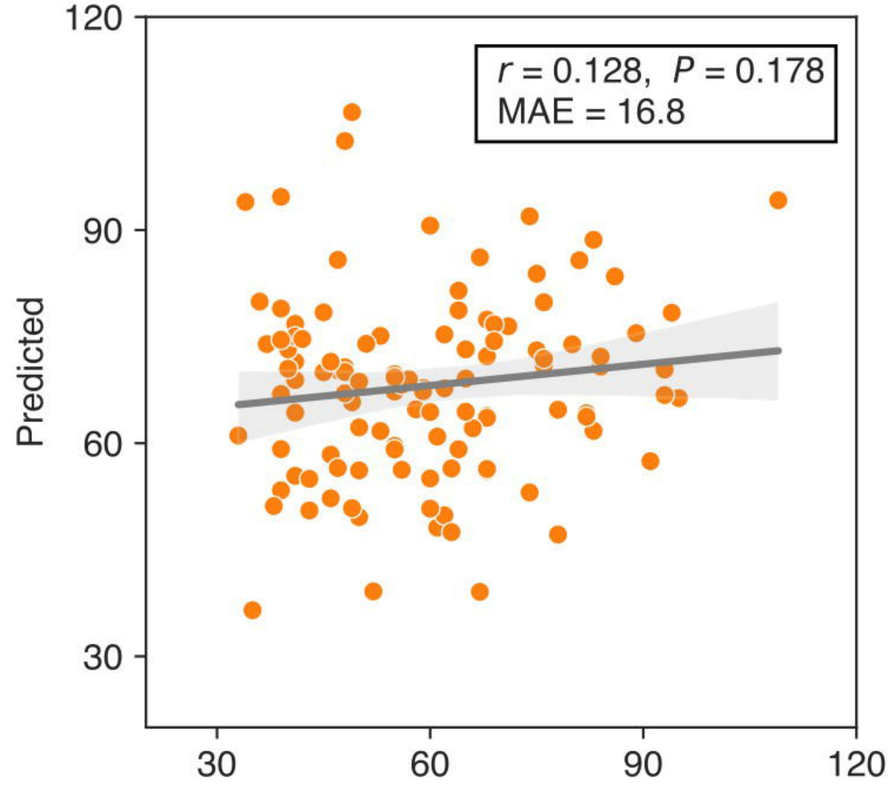
b Voting Classifiers



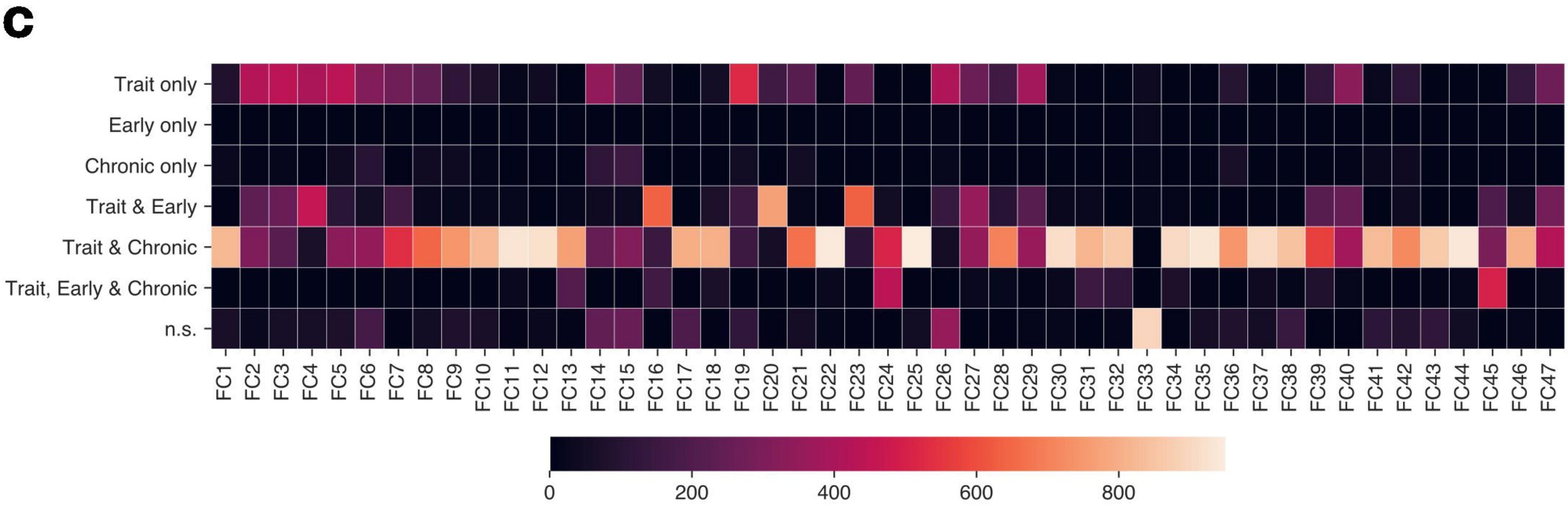
a

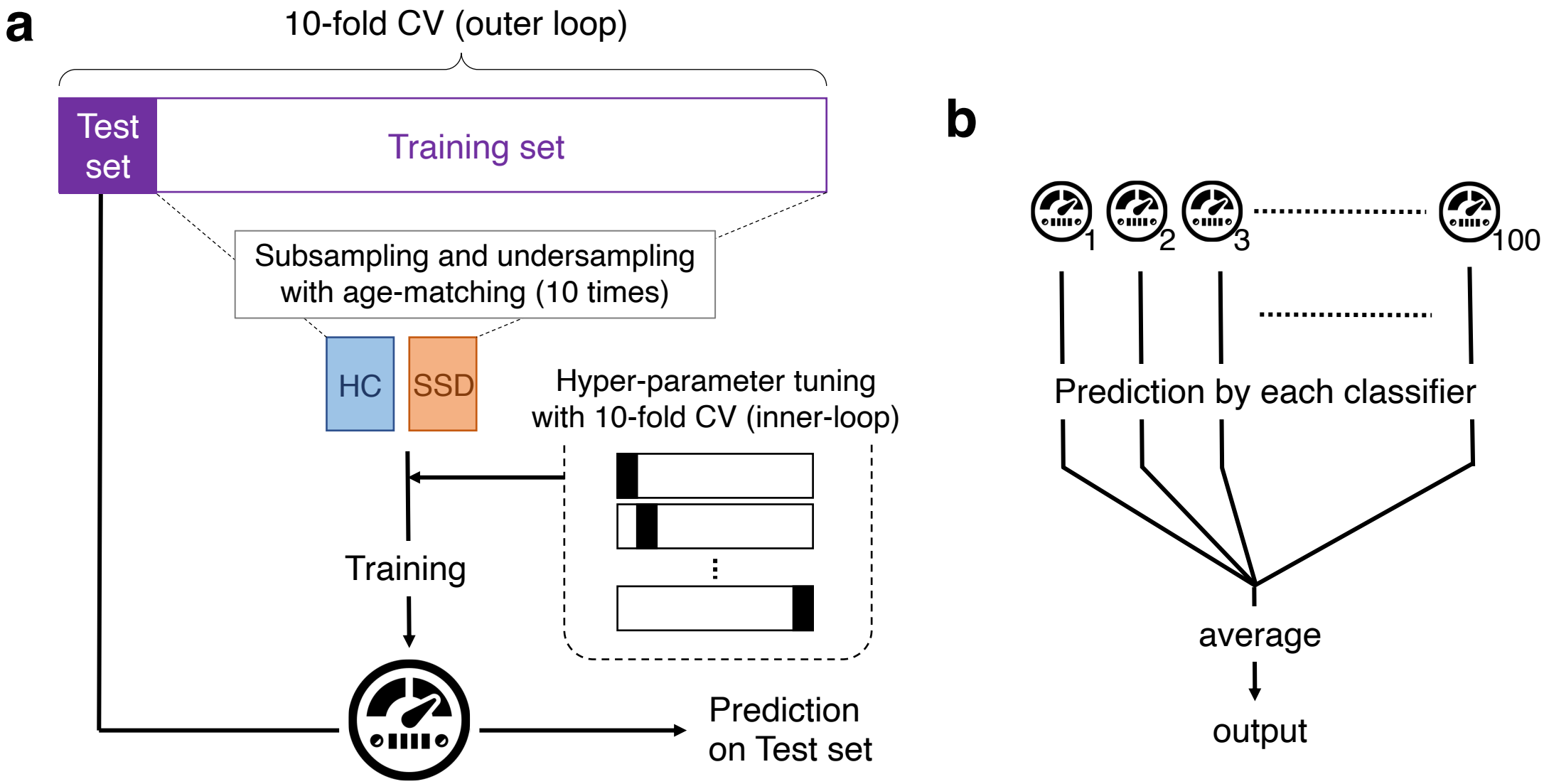

PDI total score
r = 0.331, P = 0.0177
MAE = 55.5
Predicted
Actual
0
50
100
150
200
250

b


PANSS total score
r = 0.128, P = 0.178
MAE = 16.8
Predicted
Actual
30
60
90
120

c


Trait only
Early only
Chronic only
Trait & Early
Trait & Chronic
Trait, Early & Chronic
n.s.
FC1 FC2 FC3 FC4 FC5 FC6 FC7 FC8 FC9 FC10 FC11 FC12 FC13 FC14 FC15 FC16 FC17 FC18 FC19 FC20 FC21 FC22 FC23 FC24 FC25 FC26 FC27 FC28 FC29 FC30 FC31 FC32 FC33 FC34 FC35 FC36 FC37 FC38 FC39 FC40 FC41 FC42 FC43 FC44 FC45 FC46 FC47
0
200
400
600
800



10-fold CV × 10 subsampling = 100 classifiers