

Title: A Neural Marker of Obsessive-Compulsive Disorder from Whole-Brain Functional Connectivity

Authors and author addresses:

Yu Takagi^{1,2†}, Yuki Sakai^{1,3†}, Lisi Giuseppe¹, Noriaki Yahata^{4,5}, Yoshinari Abe³, Seiji Nishida³, Takashi Nakamae³, Jun Morimoto^{1,2}, Mitsuo Kawato^{1,2}, Jin Narumoto³, Saori C Tanaka^{1*}

¹ATR Brain Information Communication Research Laboratory Group, Kyoto 619-0288, Japan

²Graduate School of Information Science, Nara Institute of Science and Technology, Nara 630-0192, Japan

³Department of Psychiatry, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, Kyoto 602-8566, Japan

⁴Department of Youth Mental Health, Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan

⁵Department of Molecular Imaging and Theranostics, National Institute of Radiological Sciences, National Institutes for Quantum and Radiological Science and Technology, Chiba 263-8555, Japan

† Contributed equally to this work

*** Corresponding author:** xsaori@atr.jp

Abstract

Obsessive-compulsive disorder (OCD) is a common psychiatric disorder with a lifetime prevalence of 2–3%. Recently, brain activity in the resting state is gathering attention as a new means of exploring altered functional connectivity in psychiatric disorders. Although previous resting-state functional magnetic resonance imaging studies investigated neurobiological abnormalities of patients with OCD, there are concerns that should be addressed. One concern is the validity of the hypothesis employed. Most studies used seed-based analysis of the fronto-striatal circuit, despite the potential for abnormalities in other regions. A hypothesis-free study is a promising approach in such a case, while it requires researchers to handle a dataset with large dimensions. Another concern is the reliability of biomarkers derived from a single dataset, which may be influenced by cohort-specific features. Here, by employing a recently developed machine-learning algorithm to avoid these concerns, we identified the first OCD biomarker that is generalized to an external dataset. We also demonstrated that the functional connectivities that contributed to the classification were widely distributed rather than locally constrained. Our

generalizable classifier has the potential not only to deepen our understanding of the abnormal neural substrates of OCD but also to find use in clinical applications.

Introduction

Obsessive-compulsive disorder (OCD) is a common psychiatric disorder with a lifetime prevalence of 2–3%¹ that is characterized by obsessions (recurrent intrusive thoughts with excessive anxiety) and compulsions (excessive repetitive actions for reducing obsession-induced anxiety). Previous neuroimaging studies using structural and task-based functional magnetic resonance imaging (fMRI) have revealed neurobiological dysfunctions in OCD, most notably in the fronto-striatal circuit^{2–6}. A meta-analysis of task-based fMRI studies using the symptom provocation paradigm revealed consistent increased activation within fronto-striatal regions⁷. Likewise, a multicenter voxel-based morphometric study revealed altered front-striatal gray and white matter volumes in patients with OCD⁸. Structural- or functional-MRI OCD classifiers constructed based on these findings have been reported^{9–12}. Furthermore, methods for modulating the neural activity of the brain regions within the fronto-striatal circuit, such as deep brain stimulation (DBS), have been applied as clinical therapy for OCD¹³.

Besides structural and task-based fMRI studies, resting-state fMRI (rs-fMRI) is gathering attention as a new means of exploring altered functional connectivity in OCD¹⁴. Several studies have reported that rs-fMRI can detect differences in functional connectivity (FC) between healthy controls (HCs) and patients with OCD¹⁵⁻¹⁷, or find correlations with treatment response to medication¹⁸ and behavioral therapy^{19,20}. Furthermore, DBS reduced excessive FC within the fronto-striatal circuit, and the DBS-induced changes in FC and changes in symptom severity were correlated²¹.

There are two types of rs-fMRI studies in OCD: hypothesis-driven, seed-based analyses and hypothesis-free, data-driven analyses. Many of the OCD rs-fMRI studies have used seed-based FC analyses with a focus on the hypothesis of local abnormalities, especially within the fronto-striatal circuit^{15,16}. Recently, other data-driven studies have revealed more global abnormalities, involving a more complex combination of activity throughout the brain^{22,23}. The latter approach does not require an *a priori* hypothesis; therefore, it has the potential to quantitatively evaluate the contribution of the fronto-striatal circuit relative to other brain regions².

Although previous rs-fMRI studies revealed neurobiological abnormalities in patients with OCD, the generalizability of these findings is still elusive. In fact, even for the most promising fronto-striatal circuit hypothesis, the findings were inconsistent^{15,16,23–25}; that is, the fronto-striatal circuit in OCD is hypoconnected in some studies^{24,25} and hyperconnected in others^{15,16,23}. Furthermore, other studies have suggested that there were abnormalities in addition to the fronto-striatal circuit, including the fronto-parietal and default mode networks^{5,26,27}. No study has quantitatively evaluated the relative importance of the fronto-striatal circuit relative to the whole brain. Thus, it is necessary to construct a reliable biomarker using a fully data-driven approach, which consists of the most essential FCs to discriminate patients with OCD and HCs, and to evaluate the previous findings comprehensively. However, to our knowledge, no data-driven study of OCD has investigated the generalizability of its findings using an external dataset. Although only one rs-fMRI study²² has predicted the diagnosis of OCD in a data-driven and cross-validated manner, the generalizability of the observations was not verified with an external dataset. Indeed, it is very difficult to construct a classifier with sufficient generalizability because of

two major difficulties²⁸. First, the number of subjects in an rs-fMRI dataset is usually small relative to the high dimensionality of FCs. It is a well-known problem that applying a naïve machine-learning classifier to such a dataset leads to over-fitting²⁹. Second, findings obtained from a single dataset are heavily influenced by cohort-specific features, that is, nuisance variables (NVs), which may lead to catastrophic over-fitting. Gruner et al. also mentioned that they were concerned about the effect of medicine as an NV²²; therefore, we should confirm whether we can predict patients regardless of the presence/absence of medication.

Here, we aimed to construct a reliable whole-brain rs-fMRI biomarker using a data-driven approach with recently developed machine-learning algorithms²⁹ that can overcome the above issues. We employed a cascade of two algorithms: L1-Sparse Canonical Correlation Analysis (SCCA) and Sparse Logistic Regression (SLR)^{30,31}. By adopting a cascade of the sparse estimation method, our procedure leads to sparse parameters with higher generalizability, while at the same time excluding features correlating with NVs. We hypothesized that our method could be used to distinguish patients with OCD from HCs,

even in an external dataset. Furthermore, its predictions were unaffected by NVs such as the usage of medicine. Finally, we quantitatively evaluated the contribution of the fronto-striatal circuit relative to other brain regions for classification.

Results

Constructing an rs-FC-based classifier

All rs-fMRI data (N = 108) were collected at Kyoto Prefectural University of Medicine (KPUM), Kyoto, Japan. Fifty-two patients with OCD and 56 HC participants were included.

Table 1 summarizes the demographic data of the participants. There were 16 participants using medication (16 participants using antidepressants, 4 using antipsychotics, and 2 using anxiolytics). All patients were surveyed for obsessive symptoms using the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS)³².

Figure 1 shows the overview of the analysis. Pairwise, interregional FC was evaluated for each participant after standard preprocessing among 140 regions of interests (ROIs) covering the entire brain. The time courses of the voxels in each ROI were averaged to extract its time course. Then, for each participant, a matrix of FC between all ROIs was calculated by evaluating pairwise temporal Pearson correlations of the time course of blood oxygenation level-dependent signals. Further, to avoid multicollinearity between the input features, we used principal component analysis (PCA) and kept all obtained principal

components (PCs). This procedure enabled us to reduce the dimensionality of the input feature space from nearly 10,000 to the number of participants, thereby allowing the classifier to learn more stably. It should be noted that PCA was conducted using the whole training dataset; that is, an external dataset was not used to obtain the transformation matrix.

To avoid problems of over-fitting due to small sample size or irrelevant NVs, we applied the method developed in our previous study (see Methods)²⁹. We constructed the classifier by combining two machine-learning algorithms: L1-SCCA and SLR^{30,31}. Our method can avoid the problem of over-fitting by adopting a cascade of the sparse estimation method, a well-known approach for handling small sample sizes. Furthermore, it also avoids extracting cohort-specific OCD irrelevant features or NVs.

Reliable classifier for OCD in the training set

Leave-one-out cross-validation (LOOCV) was used to assess classification accuracy (see Methods). Participants with OCD could be separated from HCs with 73% accuracy and an area under the curve (AUC) of 0.81 (1,000-repetition permutation test, $P < 0.001$). Thus,

the discriminatory ability of the classifier was high. The weighted linear summation (WLS or linear discriminant function) of the identified PC values of the classifier predicted the diagnostic label of each participant. Participants with a positive WLS were classified as OCD patients and those with a negative WLS as HC. Figure 2a shows that the WLS distributions of the OCD and HC participants were separated to the right (OCD) and left (HC).

Generalization of the classifier for the external dataset

The generalizability of the classifier was tested by using an external dataset collected on a different MRI scanner from that used to collect the training dataset (see Methods). We used the same dataset as Sakai et al.¹⁶ The patients were recruited at KPUM. None of the participants had been taking any kind of psychotropic medication for at least 8 weeks. Fifteen participants were entered into both experiments. In such a case, we used them in the training dataset and excluded them from the external dataset. Finally, 28 participants, including 18 patients with OCD, were used as the external dataset. Thus, there was no overlap between the training and external datasets. For this external dataset, the present

classifier, trained with a different MRI scanner, performed well, with an AUC of 0.70 (1,000 repetitions permutation test, $P = 0.049$) (Fig. 2b). Notably, this performance was achieved when we applied the classifier to the external dataset for the first time without any adjustment of the machine-learning procedure. Thus, this external dataset was the final validation dataset.

Effects of NVs

Next, we investigated the effects of medication on classification accuracy. For the training set, the accuracy of LOOCV was 75% (12 of 16) for patients with medication, and 67.5% for patients without medication (27 of 40). The classification accuracies were not significantly different between the two populations (chi-squared test, $P = 0.581$). None of the patients in the external cohort were medicated.

Contribution to the WLSs of each FC

To understand how each FC contributed individually to the WLSs, for each FC, its contribution to the WLSs through the selected PCs was calculated. As both PCA and the classifier are linear methods, the contribution of each FC could be calculated by examining

the transformation matrix of PCA and the weight of the classifier. We considered 200 FCs that contributed the most to the WLSs. Figure 3a shows the spatial distribution of these 200 FCs that were identified from the dataset for the reliable classification of OCD and HC participants.

Next, to interpret their contributions in macroscale regions, all ROIs were grouped into 18 macroscale brain regions that were defined functionally in a previous study³³ (e.g., the default mode network³⁴) and examined the number of FCs between each pair of regions in each network. Figure 3b shows the matrices for the 200 FCs in the macroscale regions.

Diagonal and non-diagonal elements show within- and between-network FCs, respectively.

Figure 3c shows a circle plot of the 200 FCs in the macroscale regions. The number of FCs in each of the two macroscale regions is presented as the thickness of the connection lines.

Some trends were observed, for example, the right-lateral fronto-parietal network contributed strongly relative to the other regions. However, the FCs were distributed widely rather than locally constrained. As for the FCs between the bilateral basal ganglia-thalamus and orbitofrontal cortex, only 2 FCs between the thalamus and orbitofrontal cortex were

included among the 200 most contributing FCs (highlighted by the blue box in Fig. 3b). It is noteworthy that no FC between the orbitofrontal cortex and striatum was included in the 200 most contributing FCs.

Discussion

A reliable neuroimaging-based classifier for OCD was developed in this study by investigating whole-brain FC patterns using rs-fMRI data. This classifier incorporated the PCs of FCs distributed across the brain, and achieved a high AUC of 0.81 with an accuracy of 73%. Further, the classifier was generalized to an external dataset (AUC of 0.70). To our knowledge, no neuroimaging-based classifier for OCD has been shown to be generalizable using an external dataset. Furthermore, by interpreting the classifier, we first evaluated the relative contribution of the fronto-striatal circuit to the classification of OCD quantitatively.

We found that the FCs contributing to the classification were distributed widely rather than locally constrained. Specifically, many of them were involved in the fronto-parietal or default mode network. It is noteworthy that there are not as many studies investigating the fronto-parietal and default mode networks compared to those targeting the fronto-striatal circuit, both seed-based^{17,26,27} and data-driven studies^{20,22,35} have reported abnormalities of these networks besides the fronto-striatal circuit. Although a previous study found DBS-induced changes in the fronto-striatal circuit, our result suggests that we may also find

broad changes in whole brain FC patterns induced by DBS.

Although successful construction of a structural- or functional-MRI OCD classifier has been reported previously⁹⁻¹², the present study is the first to successfully classify OCD across internal and external datasets. This was achieved because our analysis pipeline was fully data-driven and cross-validated, instead of using the seed-based analysis employed in most of the previous studies. Furthermore, we employed a cascade of sparse estimation methods by using L1-SCCA and SLR²⁹. We were able to avoid the over-fitting problem by extracting optimal PCs that were relevant only to the core OCD characteristics. At the same time, we could eliminate the effects of NVs such as age, sex, and medication by feature selection. Specifically, we did not observe a clear difference in classification accuracy between patients with and without medication. Medication reportedly significantly affects rs-FC patterns³⁶, and a naïve algorithm might over-fit the difference induced by the use of medicine, which leads to a reduction of generalization accuracy for non-medicated OCD patients in the validated data.

The output of the OCD classifier might provide a reliable measure of an individual's

‘OCD-ness’ along one of the biological dimensions in psychiatric disorders, because our OCD classifier was successfully generalized to an external dataset for the first time. In the field of psychiatry, we have been unable to find any neuroscientific evidence to support the breakdown of complex psychiatric disorders into separate categories. Therefore, the hypothesis of a multiple psychiatric disorder spectrum is gaining attention³⁷. According to this view, psychiatric disorders are the product of shared risk factors, or dimensions, that lead to abnormalities. Although the findings from brain imaging^{29,38} and genetic studies³⁹ support this idea, this hypothesis is still premature because of the scarcity of reliable dimensions. We believe our OCD classifier may provide an objective, reliable dimension for the spectrum. Further studies evaluating the relationship between the classifiers of multiple psychiatric disorders are needed for a deeper understanding of psychiatric disorders and for clinical application.

A limitation of the present study is that we cannot directly compare our finding with previous studies investigating local brain regions^{15,16}. This is because we employed the PCs of FCs, and they represent a linear combination of whole brain FCs. This is the

conventional approach in the field of machine learning to avoid the over-fitting problem when using a dataset with a small sample size and high dimensionality. In addition, unlike our previous study that employed a multi-site dataset²⁹, all participants in the training dataset were scanned in the same site. It might also lead to difficulties with generalization at another site without PCA because of the presence of uncontrolled site-specific NVs. A future study with a much larger sample size assessed at multiple sites will investigate the contribution of each FC independently.

In summary, we have developed the first generalizable rs-fMRI-based classifier for OCD. It distinguished participants with OCD from HCs even in an external validation dataset. We believe that our whole-brain biomarker will shed light on the abnormality of whole-brain FCs as a neural substrate of OCD.

Methods

Training dataset used for the construction of the OCD classifier

Participants: All resting state fMRI data (N = 108) were collected at KPUM; 69 of these participants were also included in the study of Abe et al.⁴⁰ The demographic data for all experiments are shown in Table 1. Patients with OCD were recruited at KPUM. Trained, experienced clinical psychiatrists and psychologists assessed all participants. All patients were primarily diagnosed using the Structured Clinical Interview for DSM-IV Axis I Disorders-Patient Edition (SCID)⁴¹. Exclusion criteria were 1) cardiac pacemaker or other metallic implants or artifacts; 2) significant disease, including neurological diseases, disorders of the pulmonary, cardiac, renal, hepatic, or endocrine systems, or metabolic disorders; 3) prior psychosurgery; 4) DSM-IV diagnosis of mental retardation and pervasive developmental disorders based on a clinical interview and psychosocial history; and 4) pregnancy. We excluded patients with current DSM-IV Axis I diagnosis of any significant psychiatric illness except OCD as much as possible and only 4 patients with trichotillomania, 1 patients with tic disorder, 1 patients with tic disorder and specific

phobia, and 1 patients with bulimia nervosa were included as patients with comorbidity. There was no history of psychiatric illness in the control group as determined by the SCID-Non-Patient Edition⁴². In addition, they reported no history of psychiatric treatment in any of their first-degree relatives. Handedness was classified based on a modified 25-item version of the Edinburgh Inventory. The Medical Committee on Human Studies at KPUM approved all procedures in this study. All participants gave written, informed consent after receiving a complete description of the study. All methods were carried out in accordance with the approved guidelines and regulations.

Image acquisition: A whole-body 3-T MR system (Achieva 3.0T Quasar Gyroscan Intera; Philips Medical Systems, Best, The Netherlands) with an 8-channel phased-array head coil at the Kajiicho Medical Imaging Center was used to generate magnetic resonance images. Functional data were collected using gradient Echo Planar Imaging (EPI) sequences (echo time/repetition time, 30/2000 ms; flip angle, 80°; field of view, 192 mm²; imaging matrix, 64 × 64, 39 slices; slice thickness, 3.0 mm, no gaps). High-resolution (1.0 × 1.0 × 1.0 mm) T1-weighted magnetization-prepared rapid gradient echo images were also acquired before

scanning the functional data. The first 6 (additional) images were discarded to allow magnetization to reach equilibrium. All participants underwent an approximately 6 min and 40 s resting-state scan, resulting in a total of 200 volumes. They were instructed simply to keep their eyes closed, not to think of anything, and not to fall asleep.

External validation dataset

Participants: We used the same dataset as Sakai et al.¹⁶ Fifteen participants were also included in the training dataset; therefore, they were excluded from the validation dataset. Finally, 28 participants were used as external validation data. Thus, there was no overlap between the training and external validation datasets. Patients with current DSM-IV Axis I diagnosis of any significant psychiatric illness except OCD were excluded. The other settings were the same as for the training dataset. The Medical Committee on Human Studies at KPUM approved all procedures in the study. All participants gave written, informed consent after receiving a complete description of the study. All methods were carried out in accordance with the approved guidelines and regulations.

Image acquisition: A whole-body 1.5-T MR system (Gyrosan Intera; Philips Medical Systems, Best, The Netherlands) with a 6-channel phased-array head coil was used to generate MR images. Foam pads were used to reduce head motion and scanner noise. Functional data were collected using gradient EPI sequences (echo time/repetition time, 40/2411 ms; flip angle, 80°; field of view, 192 mm²; imaging matrix, 64 × 64, 35 slices; slice thickness, 3.6 mm, no gaps). High-resolution (1 × 1 × 1.5 mm) T1-weighted magnetization-prepared rapid gradient echo images were acquired before each resting image. All participants underwent an approximately 8 min resting-state scan, resulting in a total of 200 volumes. The experimental settings for the resting-state scan were the same as for the training dataset.

Preprocessing

We used a preprocessing method similar to that of Yahata et al.²⁹ for both the training and external datasets. We used Statistical Parametric Mapping 8 (Wellcome Trust Centre for Neuroimaging, London, UK; <http://www.fil.ion.ucl.ac.uk/spm/software/>) in MATLAB (The MathWorks, Inc., Natick, MA) for preprocessing and statistical analyses. First, head motion

was compensated for by collecting raw functional images for slice-timing and realigning them to the mean image of that sequence. Second, the structural images were co-registered to the mean functional image and segmented into 3 tissue classes in the Montreal Neurological Institute (MNI) space. Using associated parameters, we normalized the functional images and resampled them in a $2 \times 2 \times 2$ mm grid. Third, the images were smoothed by a Gaussian function with a full width at half-maximum of 6 mm. To avoid the effects of motion artifacts, the pre-processed sequence of functional images was examined as follows. First, the mean relative displacement in each of the 6 motion parameters (translation along and rotation with respect to the x, y, and z axes) was evaluated by calculating the mean of the absolute frame-to-frame relative changes in each parameter through a given time series (namely, the mean of $|\Delta_p(i)| = |p_{i+1} - p_i|$ across the time series, where p is one of the 6 motion parameters and i specifies the time point). In both the training and external datasets, no statistically significant difference between the groups were noted in this measure for the 6 motion parameters (two-sample t-test, $P > 0.05$ for all parameters in both datasets). Next, the frame displacement (FD) was calculated for each

participant at each time point by summing all 6 parameters. Using this FD, we used the “scrubbing” procedure to identify and exclude any frame affected by excessive head motion⁴³. Specifically, a frame was flagged and removed, along with the previous and two subsequent frames, from correlation analysis, if the associated FD exceeded 0.5 mm. For both datasets, there was no difference in the number of frames that passed this procedure between the HC and OCD populations (two-sample t-test, $P > 0.05$).

Interregional correlation analysis

A pairwise, interregional FC was evaluated for each participant among 140 ROIs covering the entire brain. Each region’s spatial extent was defined anatomically according to the digital atlas of the BrainVISA Sulci Atlas (BSA)⁴⁴. As this atlas does not include the cerebellum, the 3 subregions of the cerebellum were appended to it based on the anatomical automatic labeling (AAL) package⁴⁵. This BSA-AAL composite atlas was resampled in the $2 \times 2 \times 2$ mm grid MNI space. The time course of the voxels in each region was averaged to extract its representative time course. Further, we excluded the ROIs with zero-variance in at least 1 participant. The time course sets were band-pass filtered (0.008–0.1 Hz) prior

to the following regression procedure. The filtered time courses were linearly regressed by the temporal fluctuations of the white matter, cerebrospinal fluid, and entire brain as well as the 6 head motion parameters. The fluctuation in each tissue class was determined from the average time course of the voxels within a mask created by the segmentation procedure of the T1 image. The mask for the white matter was eroded by 1 voxel to consider a partial volume effect. These extracted time courses were band-pass filtered (0.008–0.1 Hz) before linear regression, as was performed for the regional time courses. Then, for each participant, a matrix of FCs between all ROIs was calculated while discarding flagged frames, if any, in the previous procedure (scrubbing). The scrubbing procedure removed any frames exhibiting abrupt head movements that could be the source of high-frequency fluctuations in the filtered time course⁴⁶. The FC matrices are symmetric, so values on only one side of the diagonal were kept, resulting in the number of samples \times number of FC matrices. Further, to reduce the dimensionality of the matrix from nearly 10,000 to the number of participants, we used PCA and kept all obtained PCs for the following classification analyses, resulting in the number of samples \times number of PC matrices. This procedure

allowed the classifier to avoid multicollinearity between the input features and to learn in a stable manner. PCA was conducted using the whole training dataset.

Selecting FCs as the OCD classifier

To avoid the problems of over-fitting because of the small sample size or irrelevant NVs, we applied the method developed by Yahata et al.²⁹ The procedure for selecting relevant PCs, training the predictive model, and assessing its generalization ability was performed as a sequential process of nested feature-selection and LOOCV. In each LOOCV fold, all-but-one participant was used to train the SLR classifier, while the remaining participants were used for evaluation. SLR can train a logistic regression model while objectively pruning PCs that are not useful for classifying OCD. Before training SLR, the input dimension must be reduced to some extent and, at the same time, the effects of NVs that may lead to over-fitting must be reduced. Therefore, before LOOCV, nested feature selection was performed using L1-SCCA. L1-SCCA identifies the latent relationships between PCs and various attributes of each participant, including the diagnostic label and available demographic information. By selecting PCs connected to a canonical variable

related to only the “Diagnosis” label and not to NVs, we aimed to reduce the inferential effects of NVs. Here, we defined age, sex, handedness, and drug use (anxiolytics, antidepressants, or antipsychotics) as NVs.

Prediction of the diagnostic label

Logistic regression analysis was used as the classifier to diagnostically label from the identified PCs. A logistic function is used to define the probability of a participant belonging to the OCD class:

$$P(y = 1 | \hat{\mathbf{z}}; \mathbf{w}) = \frac{1}{1 + \exp(-\mathbf{w}^T \hat{\mathbf{z}})}$$

Here y is the diagnosis class label (OCD, $y = 1$; HC, $y = 0$). $\hat{\mathbf{z}} = [\mathbf{z}^T, 1]^T \in \mathbb{R}^{m+1}$ is a feature vector with an augmented input, where the feature vector \mathbf{z} is the PCs of a participant’s rs-fMRI sample. Using the augmented input “1” is a standard approach to introduce constant (bias) input for the classifier. $\mathbf{w} \in \mathbb{R}^{m+1}$ is the weight vector of the logistic function.

To decrease the dimension of the feature vector further, which was already reduced by L1-SCCA according to the equation, we used an SLR method. SLR automatically selects OCD-classification-related features as input for the logistic function. In SLR, the

probability distribution of the parameter vector is estimated by the hierarchical Bayesian estimation approach, in which the prior distribution of each element of the parameter vector is represented as a Gaussian distribution. Due to the automatic relevance determination property of the hierarchical Bayesian estimation method, some of the Gaussian distributions have a sharp peak at zero so that irrelevant features are not used in the classification.

References

1. Ruscio, A., Stein, D., Chiu, W. & Kessler, R. The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. *Mol. Psychiatry* **15**, 53–63 (2008).
2. Menzies, L. *et al.* Integrating evidence from neuroimaging and neuropsychological studies of obsessive-compulsive disorder: the orbitofronto-striatal model revisited. *Neurosci. Biobehav. Rev.* **32**, 525–49 (2008).
3. Admon, R. *et al.* Functional and structural neural indices of risk aversion in obsessive-compulsive disorder (OCD). *Psychiatry Res. - Neuroimaging* **203**, 207–213 (2012).
4. Nakamae, T. *et al.* Altered fronto-striatal fiber topography and connectivity in obsessive-compulsive disorder. *PLoS One* **9**, (2014).
5. Hou, J. *et al.* Morphologic and functional connectivity alterations of corticostriatal and default mode network in treatment-naïve patients with obsessive-compulsive disorder. *PLoS One* **8**, (2013).

6. Jung, W. H. *et al.* Abnormal corticostriatal-limbic functional connectivity in obsessive-compulsive disorder during reward processing and resting-state. *NeuroImage Clin.* **3**, 27–38 (2013).
7. Rotge, J.-Y. *et al.* Provocation of obsessive-compulsive symptoms: a quantitative voxel-based meta-analysis of functional neuroimaging studies. *J. Psychiatry Neurosci.* **33**, 405–12 (2008).
8. De Wit, S. J. *et al.* Multicenter voxel-based morphometry mega-analysis of structural brain scans in obsessive-compulsive disorder. *Am. J. Psychiatry* **171**, 340–349 (2014).
9. Soriano-Mas, C. *et al.* Identifying patients with obsessive-compulsive disorder using whole-brain anatomy. *Neuroimage* **35**, 1028–1037 (2007).
10. Weygandt, M. *et al.* fMRI pattern recognition in obsessive-compulsive disorder. *Neuroimage* **60**, 1186–1193 (2012).
11. Li, F. *et al.* Multivariate pattern analysis of DTI reveals differential white matter in individuals with obsessive-compulsive disorder. *Hum. Brain Mapp.* **35**, 2643–2651 (2014).

12. Hu, X. *et al.* Multivariate pattern analysis of obsessive-compulsive disorder using structural neuroanatomy. *Eur. Neuropsychopharmacol.* **26**, 246–254 (2016).
13. Denys, D. & Mantione, M. *Deep brain stimulation in obsessive-compulsive disorder. Progress in Brain Research* **175**, (2009).
14. Fox, M. D. & Greicius, M. Clinical applications of resting state functional connectivity. *Front. Syst. Neurosci.* **4**, 19 (2010).
15. Harrison, B. J. *et al.* Altered corticostriatal functional connectivity in obsessive-compulsive disorder. *Arch. Gen. Psychiatry* **66**, 1189–1200 (2009).
16. Sakai, Y. *et al.* Corticostriatal functional connectivity in non-medicated patients with obsessive-compulsive disorder. *Eur. Psychiatry* **26**, 463–469 (2011).
17. Jang, J. H. *et al.* Functional connectivity in fronto-subcortical circuitry during the resting state in obsessive-compulsive disorder. *Neurosci. Lett.* **474**, 158–162 (2010).
18. Shin, D.-J. *et al.* The effects of pharmacological treatment on functional brain connectome in obsessive-compulsive disorder. *Biol. Psychiatry* **75**, 606–14 (2014).
19. Göttlich, M., Krämer, U. M., Kordon, A., Hohagen, F. & Zurowski, B. Resting-state

- connectivity of the amygdala predicts response to cognitive behavioral therapy in obsessive compulsive disorder. *Biol. Psychol.* **111**, 100–109 (2015).
20. Yang, X.-Y. *et al.* Regional homogeneity of spontaneous brain activity in adult patients with obsessive-compulsive disorder before and after cognitive behavioural therapy. *J. Affect. Disord.* **188**, 243–251 (2015).
 21. Figeo, M. *et al.* Deep brain stimulation restores frontostriatal network activity in obsessive-compulsive disorder. *Nat. Neurosci.* **16**, 386–7 (2013).
 22. Gruner, P. *et al.* Independent component analysis of resting state activity in pediatric obsessive-compulsive disorder. *Hum. Brain Mapp.* **0**, (2014).
 23. Beucke, J. C. *et al.* Abnormally high degree connectivity of the orbitofrontal cortex in obsessive-compulsive disorder. *JAMA psychiatry* **70**, 619–29 (2013).
 24. Posner, J. *et al.* Reduced functional connectivity within the limbic cortico-striato-thalamo-cortical loop in unmedicated adults with obsessive-compulsive disorder. *Hum. Brain Mapp.* **35**, 2852–2860 (2014).
 25. Bernstein, G. A. *et al.* Abnormal striatal resting-state functional connectivity in

- adolescents with obsessive-compulsive disorder. *Psychiatry Res. - Neuroimaging* **247**, 49–56 (2016).
26. Stern, E. R., Fitzgerald, K. D., Welsh, R. C., Abelson, J. L. & Taylor, S. F. Resting-state functional connectivity between fronto-parietal and default mode networks in obsessive-compulsive disorder. *PLoS One* **7**, (2012).
27. Beucke, J. C. *et al.* Default mode network subsystem alterations in obsessive-compulsive disorder. *Br. J. Psychiatry* **205**, 376–382 (2014).
28. Whelan, R. & Garavan, H. When optimism hurts: Inflated predictions in psychiatric neuroimaging. *Biol. Psychiatry* **75**, 746–748 (2014).
29. Yahata, N. *et al.* A small number of abnormal brain connections predicts adult autism spectrum disorder. *Nat. Commun.* **7**, (2016).
30. Witten, D. M., Tibshirani, R. & Hastie, T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* **10**, 515–34 (2009).
31. Yamashita, O., Sato, M., Yoshioka, T., Tong, F. & Kamitani, Y. Sparse estimation

automatically selects voxels relevant for the decoding of fMRI activity patterns.

Neuroimage **42**, 1414–29 (2008).

32. Goodman, W. K. *et al.* The Yale-Brown obsessive compulsive scale: I. Development, use, and reliability. *Arch. Gen. Psychiatry* **46**, 1006–1011 (1989).
33. Laird, A. R. *et al.* Behavioral interpretations of intrinsic connectivity networks. *J. Cogn. Neurosci.* **23**, 4022–37 (2011).
34. Fox, M. D. *et al.* The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 9673–8 (2005).
35. Tian, L. *et al.* Abnormal functional connectivity of brain network hubs associated with symptom severity in treatment-naive patients with obsessive-compulsive disorder: A resting-state functional MRI study. *Prog. Neuro-Psychopharmacology Biol. Psychiatry* **66**, 104–111 (2016).
36. Schaefer, A. *et al.* Serotonergic modulation of intrinsic functional connectivity. *Curr. Biol.* **24**, 2314–2318 (2014).
37. Adam, D. On the spectrum. *Nature* **496**, 6–8 (2013).

38. Dichter, G. S., Damiano, C. A. & Allen, J. A. Reward circuitry dysfunction in psychiatric and neurodevelopmental disorders and genetic syndromes: animal models and clinical findings. *J. Neurodev. Disord.* **4**, 19 (2012).
39. 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).
40. Abe, Y. *et al.* Hyper-influence of the orbitofrontal cortex over the ventral striatum in obsessive-compulsive disorder. *Eur. Neuropsychopharmacol.* **25**, 1898–1905 (2015).
41. First, M. B., Spitzer, R. L., Miriam, G. & Williams, J. B. W. *Structured Clinical Interview for Axis I DSM-IV Disorders-Patient Edition (SCID-I/P)*. New York: Biometrics Research Department. (New York State Psychiatric Institute, 1994).
42. First, M. B., Spitzer, R. L., Gibbon, M. & Williams, J. B. W. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP)*. New York: Biometrics Research Department. (New York State Psychiatric Institute, 2002).

43. 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).
44. Perrot, M., Rivière, D. & Mangin, J. F. Cortical sulci recognition and spatial normalization. *Med. Image Anal.* **15**, 529–550 (2011).
45. Tzourio-Mazoyer, N. *et al.* Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* **15**, 273–89 (2002).
46. Griffanti, L. *et al.* ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. *Neuroimage* **95**, 232–247 (2014).

Author contributions:

Y.T., Y.S., M.K., J.N., and S.C.T. conceived and designed the study; Y.S., Y.A., S.N., and T.N. recruited the participants of the study and collected their clinical and imaging data; Y.T., Y.S., L.G., N.Y., and J.M. analyzed the data; Y.T., Y.S., and S.C.T. wrote the manuscript. All authors reviewed the manuscript.

Acknowledgements:

This study is the result of ‘Brain Machine Interface Development’, carried out under the Japan Agency for Medical Research and Development. We thank T. Okada, H. Ito, and technical engineers for their assistance in MRI data acquisition. We thank K. Tamura, S. Kimura, and K. Inoue for their assistance in the assessment of the patients. We thank the Joint Usage/Research Center at ISER, Osaka University for financial support.

Competing financial interests:

The authors declare no competing financial interests.

Figure /Table Legends

Figure 1: Schematic diagram of the procedure for selecting FCs as an OCD biomarker and assessing their predictive power.

A sketch map of prediction analysis. Resting state functional connectivity (rs-FC) matrices were processed through the cascading feature selection procedure. Left-out participants and all participants in the external validation dataset were classified based on the classifier derived from the rs-FC matrix from other participants.

Figure 2: Distribution of weighted linear summations (WLS) of functional connections used for the classification of the OCD and HC populations.

(a) The number of HC (white) and OCD (black) participants in the internal dataset in a specific WLS interval of width 5 is shown as a histogram. (b) WLS for the validation dataset in a specific WLS interval of width 2 is shown as a histogram.

Figure 3: Functional connections used in the classification of the OCD and HC populations.

(a) The 200 most contributing FCs from the left (left top), posterior (left bottom), and top (right) to the WLSs are visualized. (b) Matrices for the most contributing 200 FCs in 18 macroscale regions that were functionally defined in a previous study³³. Diagonal and non-diagonal elements show within- and between-network FCs, respectively. The blue box highlights the corresponding area in the matrix discussed in the main text, i.e., FC between the orbitofrontal and basal ganglia-thalamus networks. (c) Circle plot of the 200 most contributing 200 FCs in 18 macroscale regions. The number of FCs in each of the 2 macroscale regions are presented as the thickness of the connection lines (edges).

Table 1: Demographic information of the participants used to construct the classifier for the OCD and HC populations (mean \pm standard deviation). All demographic distributions are matched between the OCD and HC populations in the internal and external datasets ($P > 0.05$).

NR, not recorded.

NA, not applicable.

Table 1

	Internal		External	
	OCD	HC	OCD	HC
Male/Female	23/33	26/26	4/6	6/12
Age (years)	32.64 ± 9.63	29.40 ± 7.46	31.50 ± 10.31	29.89 ± 8.69
Handedness (R/L)	51/5	50/2	9/1	15/3
Y-BOCS	21.26 ± 6.65	NR	23.8 ± 5.77	NR
Medication (with medication/without medication)	16/40	NA	0/10	NA

Figure 1

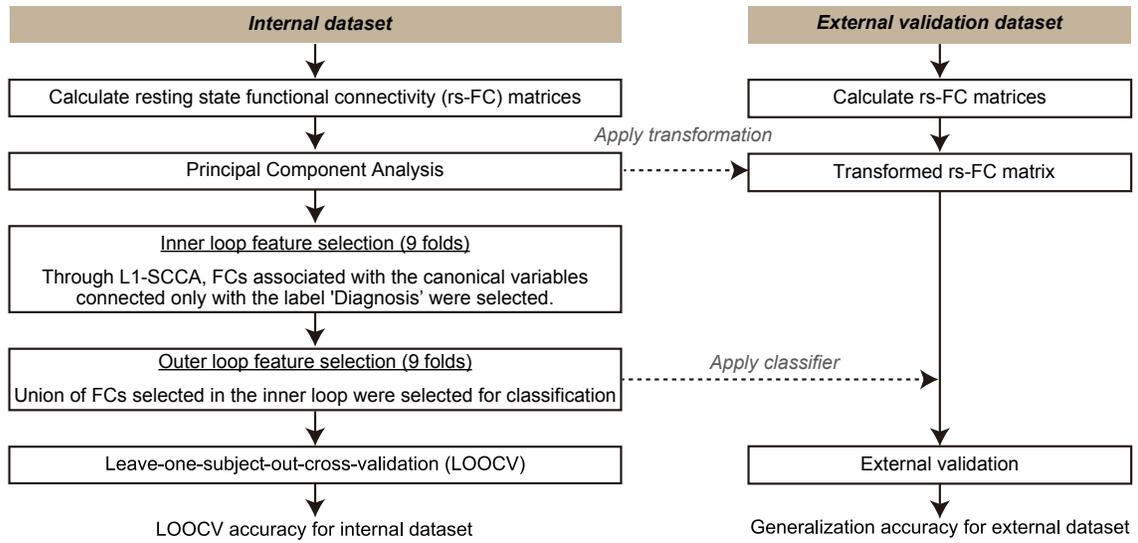


Figure 2

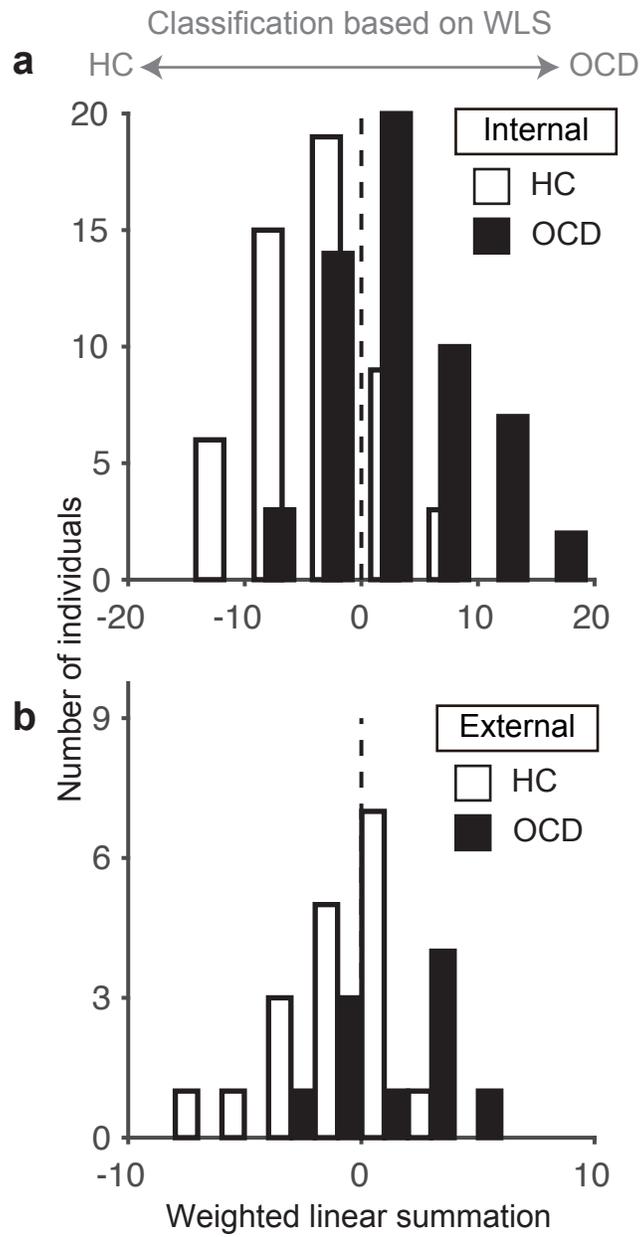


Figure 3

