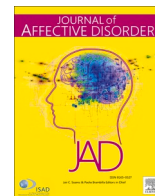




Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Short Communication

Verification of the brain network marker of major depressive disorder: Test-retest reliability and anterograde generalization performance for newly acquired data



Go Okada^{a,*}, Toshinori Yoshioka^{b,c}, Ayumu Yamashita^{b,d}, Eri Itai^a, Satoshi Yokoyama^a, Toshiharu Kamishikiryo^a, Hotaka Shinzato^a, Yoshikazu Masuda^a, Yuki Mitsuyama^a, Shigeyuki Kan^a, Akiko Kurata^a, Masahiro Takamura^e, Atsuo Yoshino^f, Akio Mantani^g, Osamu Yamamoto^h, Norio Yokotaⁱ, Tatsuji Tamura^j, Hiroaki Jitsuiki^k, Mitsuo Kawato^{b,c}, Okito Yamashita^{b,l}, Yuki Sakai^{b,c}, Yasumasa Okamoto^a

^a Department of Psychiatry and Neurosciences, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan

^b Brain Information Communication Research Laboratory Group, Advanced Telecommunications Research Institute International, Kyoto, Japan

^c XNef, Inc., Kyoto, Japan

^d Department of Psychiatry, Boston University School of Medicine, MA, USA

^e Department of Neurology, Faculty of Medicine, Shimane University, Izumo, Japan

^f Health Service Center, Hiroshima University, Higashihiroshima, Japan

^g Mantani Mental Clinic, Hiroshima, Japan

^h Ujina Mental Clinic, Hiroshima, Japan

ⁱ Yokota Mental Clinic, Hiroshima, Japan

^j Tamura Mental Clinic, Hiroshima, Japan

^k Jitsuiki Clinic, Hiroshima, Japan

^l Center for Advanced Intelligence Project, RIKEN, Tokyo, Japan

ARTICLE INFO

Keywords:

Major depressive disorder
Resting-state functional magnetic resonance imaging
Brain network marker
Reliability
Generalizability

ABSTRACT

Background: Recently, we developed a generalizable brain network marker for the diagnosis of major depressive disorder (MDD) across multiple imaging sites using resting-state functional magnetic resonance imaging. Here, we applied this brain network marker to newly acquired data to verify its test-retest reliability and anterograde generalization performance for new patients.

Methods: We tested the sensitivity and specificity of our brain network marker of MDD using data acquired from 43 new patients with MDD as well as new data from 33 healthy controls (HCs) who participated in our previous study. To examine the test-retest reliability of our brain network marker, we evaluated the intraclass correlation coefficients (ICCs) between the brain network marker-based classifier's output (probability of MDD) in two sets of HC data obtained at an interval of approximately 1 year.

Results: Test-retest correlation between the two sets of the classifier's output (probability of MDD) from HCs exhibited moderate reliability with an ICC of 0.45 (95 % confidence interval, 0.13–0.68). The classifier distinguished patients with MDD and HCs with an accuracy of 69.7 % (sensitivity, 72.1 %; specificity, 66.7 %).

Limitations: The data of patients with MDD in this study were cross-sectional, and the clinical significance of the marker, such as whether it is a state or trait marker of MDD and its association with treatment responsiveness, remains unclear.

Conclusions: The results of this study reaffirmed the test-retest reliability and generalization performance of our brain network marker for the diagnosis of MDD.

* Corresponding author at: Department of Psychiatry and Neurosciences, Hiroshima University, 1-2-3 Kasumi, Minami-Ku, Hiroshima 734-8553, Japan.

E-mail address: gookada@hiroshima-u.ac.jp (G. Okada).

<https://doi.org/10.1016/j.jad.2023.01.087>

Received 8 February 2022; Received in revised form 23 December 2022; Accepted 4 January 2023

Available online 27 January 2023

0165-0327/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Psychiatric disorders including major depressive disorder (MDD) are thought to be the result of brain circuit dysfunction (Insel and Cuthbert, 2015). However, in current medical practice, MDD is diagnosed by evaluating various symptoms, such as depressed mood and loss of interest, by detailed interviews, and there is no objective biological diagnostic method.

In recent years, resting-state functional magnetic resonance imaging (rs-fMRI) and machine learning techniques have been applied in an effort to create a classifier that can be used for an objective auxiliary diagnosis of MDD that reflects brain function. A meta-analysis of studies attempting to classify MDD and healthy individuals using such techniques was also performed, and it was reported that the results of studies using rs-fMRI had an average sensitivity of 85 % and specificity of 83 % (Kambeitz et al., 2017). However, few studies have verified the practical generalization of classifiers to data from other facilities. A recent study using multimodal MRI data from two centers reported being able to discriminate patients from healthy controls (HCs) with an area under the curve of 0.916 and accuracy of 84.8 % in a validation cohort (Sun et al., 2022), indicating the importance of multicenter joint research projects.

In the Strategic Research Program for Brain Sciences, which was conducted from November 2013 to March 2018, we were engaged in multi-center data analysis and the standardization of imaging protocols to verify the reproducibility of results as well as technological developments to overcome the differences between MRI scanners (Yamashita et al., 2019). By utilizing this technology, we developed a diagnostic brain network marker that can distinguish patients with MDD and HCs based on resting-state functional connectivity (FC) patterns with a probability of 66 %, even with a completely independent large dataset (Yamashita et al., 2020). However, all previous studies on the generalization of MRI biomarkers for psychiatric disorders were retrograde. That is, independent validation cohorts were acquired before or in parallel to the acquisition of the discovery cohorts and construction of the MRI biomarkers. A more stringent generalization test should utilize independent validation cohorts that are acquired after the biomarker is constructed, as in prospective clinical trials. This was one of the two main objectives of the current research.

In rs-fMRI, in addition to inter-facility differences, variability between measurements in the same person (Noble et al., 2019) is a barrier to the development of reliable biomarkers. Although we reported that our brain network marker achieved high accuracy in approximately 50 scans of nine healthy participants (mean accuracy of all participants = 84.5, 1SD = 12.8, across participants) (Yamashita et al., 2020), this result was for data that were measured at a short interval from subjects who were particularly accustomed to scanning. In order to verify that the reproducibility of our marker is sufficient for clinical use, it is necessary to check its reproducibility at longer intervals with a sufficient number of participants.

In this study, we applied this brain network marker for MDD diagnosis (Yamashita et al., 2020) to data collected after the end of the Strategic Research Program for Brain Sciences project, in order to examine if we could achieve the same result if we measure the same person twice at an interval of approximately 1 year and that sufficient sensitivity can be reproduced using data from new patients.

2. Materials and methods

2.1. Participants

A total of 47 patients with MDD (22 males and 25 females; average age, 44.62 ± 11.78 years; 43 right-handed and 4 left-handed) were recruited from Hiroshima University Hospital and local clinics between April 2018 and December 2018. The patients had been diagnosed by an expert clinician using the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; the Japanese version of the Mini-International

Neuropsychiatric Interview (M.I.N.I.) (Otsubo et al., 2005; Sheehan et al., 1998) was administered at the time of participation to confirm the diagnosis. Exclusion criteria for the patient group included current or past manic episodes, psychotic episodes, alcohol dependence or/and abuse, substance dependence or/and abuse, and antisocial personality disorder based on M.I.N.I.

Thirty-nine HCs (21 males and 18 females; average age, 60.26 ± 12.83 years; 36 right-handed and 3 left-handed) were recruited between April 2018 and December 2018 from those who participated in our previous study (Yamashita et al., 2020) in 2017, and in the present study, we examined test-retest reliability between measurements taken at an interval of approximately 1 year. HCs were also screened with the M.I.N.I. to confirm they had no psychiatric disorder. The depressive symptoms of each participant on the day of the MRI scan were assessed using the Japanese version of Beck Depression Inventory-II (BDI-II) (Beck et al., 1996; Kojima et al., 2002).

The study was conducted in compliance with the relevant guidelines and regulations and the latest version of the World Medical Association's Declaration of Helsinki. The current study protocol was approved by the Ethics Committee of Hiroshima University. Before the administration of any experimental procedure, written informed consent was obtained from all participants.

2.2. MRI acquisition

The scanner and imaging protocol were the same as one of the eight used in our previous study (Yamashita et al., 2020). Rs-fMRI scans were acquired for 10 min from each participant using a 3.0 Tesla MRI system (Siemens MAGNETOM Verio; Siemens, Erlangen, Germany) at Hiroshima University. The following scanning parameters were used: slice number, 40; matrix size, 64 × 64; FOV, 212 mm; voxel size, 3.3 × 3.3 × 3.2 mm (slice gap, 0.8 mm); TR, 2500 ms; TE, 30 ms; flip angle, 80°. T1-weighted structural images were also acquired using a 3D magnetization-prepared rapid gradient echo sequence: 178 1-mm thick sagittal slices with a 256 × 256 acquisition matrix; FOV, 256 mm; TI/TR/TE = 900/2300/2.98 ms; flip angle, 9°. The acquisition time for rs-fMRI scans and T1-weighted structural images was 617 s and 312 s, respectively.

2.3. Preprocessing and calculation of the resting-state FC matrix

The preprocessing and calculation of the resting-state FC matrix used have been described in detail elsewhere (Yamashita et al., 2020). We preprocessed the rs-fMRI data using FMRIprep version 1.3.2 (Esteban et al., 2019). The first 10 s of the data were discarded to allow for T1 equilibration. Preprocessing steps included slice-timing correction, realignment, coregistration, distortion correction using a field map, segmentation of T1-weighted structural images, normalization to the Montreal Neurological Institute space, and spatial smoothing with an isotropic Gaussian kernel of 6 mm full width at half maximum. To analyze the data, we used ciftify toolbox version 2.0.X (Dickie et al., 2019). We used Glasser's 379 surface-based parcellations as regions of interest (ROIs) (Glasser et al., 2016); the blood oxygen level-dependent (BOLD) signal time courses were extracted from these 379 ROIs. A temporal bandpass filter was applied to the time series using a first-order Butterworth filter with a pass band between 0.01 and 0.08 Hz. Framewise displacement, which represents head motion between two consecutive volumes as a scalar quantity, i.e., the summation of absolute displacements in translation and rotation, was calculated for each functional session, and we removed volumes with a framewise displacement > 0.5 mm (Power et al., 2014). If the ratio of the excluded volumes after scrubbing exceeded 47 %, the participants were excluded from the analysis (Yamashita et al., 2020). As a result, the rs-fMRI data from four patients with MDD and six HCs were removed. Thus, we included 43 patients with MDD (20 males and 23 females; average age, 45.33 ± 11.78 years; 39 right-handed and 4 left-handed) and 33 HCs (17

males and 16 females; average age, 59.39 ± 12.38 years; 30 right-handed and 3 left-handed) for further analysis. FC was calculated as the temporal correlation of rs-fMRI BOLD signals across 379 ROIs for each participant. Fisher's z -transformed Pearson's correlation coefficients were calculated between the preprocessed BOLD signal time courses of each possible pair of ROIs and used to construct 379×379 symmetrical connectivity matrices. We used 71,631 FC values ($[379 \times 378] / 2$) of the lower triangular matrix of the connectivity matrix for further analysis. In our previous study (Yamashita et al., 2020), in order to construct the classifiers, logistic regression analyses using the least absolute shrinkage and selection operator method were used to select the optimal subset of FCs. In this study, the performance of the classifiers created in our previous study (Yamashita et al., 2020) was evaluated with newly acquired data.

2.4. Test-retest reliability and generalization performance of the classifier developed by Yamashita et al. (2020)

As 100 classifiers of MDD (10-fold cross-validation \times 10 subsamples) were created in the original study (Yamashita et al., 2020), we applied all of these classifiers to the new data set. Next, we averaged the 100 outputs (diagnostic probability) for each participant. Importantly data newly acquired at Hiroshima University were sent to Advanced Telecommunications Research Institute International without a diagnostic label, and the diagnostic probability calculated by Advanced Telecommunications Research Institute International was compared with the actual diagnostic label stored at Hiroshima University, so the persons responsible for the diagnostic probability computation were completely blinded to the diagnostic label. As mentioned earlier, the recruited HCs also participated in our previous study, allowing for longitudinal comparisons. We examined the test-retest reliability of the brain network marker by calculating intraclass correlation coefficients (ICCs) from the classifier's outputs in two sets of HC data from the same individuals acquired in 2018 and 2017. We then assessed the correlation between the averaged diagnostic probability value (probability of MDD) and the BDI-II score, and tested the sensitivity and specificity of the network marker by using the data newly acquired in 2018 from HCs and new patients with MDD. We considered a participant to be a patient with MDD if the probability of MDD was $>50\%$.

3. Results

3.1. Test-retest reliability at a 1-year interval

The test-retest correlation between the two sets of classifier's output of HCs is shown in Fig. 1. The classifier's output exhibited moderate reliability with an ICC of 0.45 (95 % confidence interval = 0.13–0.68; $P = 0.004$).

3.2. Generalization performance for new MDD patients

We found significant correlations between the classifier's output (probability of MDD) and depressive symptoms (BDI-II score) ($r = 0.26$, $P = 0.024$), and the classifier distinguished patients with MDD and HCs with an accuracy of 69.7 % (sensitivity, 72.1 %; specificity, 66.7 %) (Fig. 2). As this study included people aged over 60 years who may already have functional and structural changes, we examined the effect of age on the classifier's output. As a result, there was no relationship between the classifier's output and age ($r = 0.093$, $P = 0.423$) (Fig. S2), and the correlation between the classifier's output and BDI-II score demonstrated stronger significance ($r = 0.31$, $P = 0.006$) in partial correlation analysis with the effect of age as a covariate.

4. Discussion

In this study, we confirmed that similar results can be obtained by

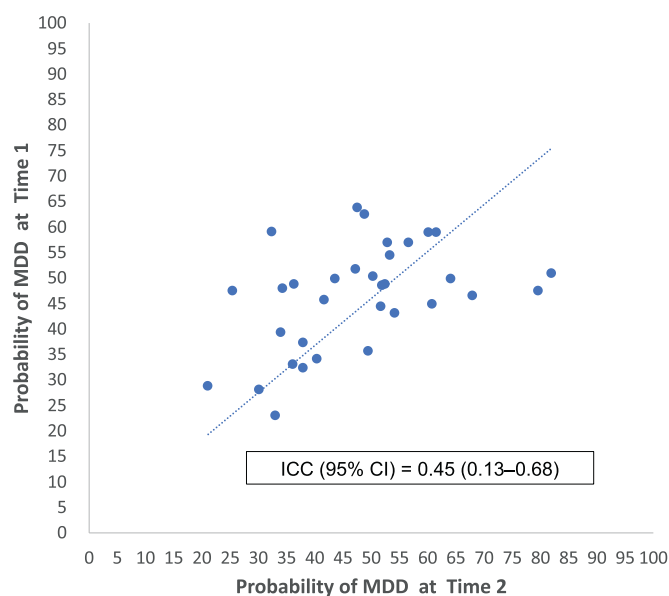


Fig. 1. Test-retest correlation between classifier outputs of two sets of data obtained from the same healthy controls at an interval of approximately 1 year. Scatter plot of the classifier's output (probability of major depressive disorder [MDD]) at Time 1 and Time 2. The blue dotted line represents the line of identity. Each data point represents one participant. Intraclass correlation coefficient (ICC) and 95 % confidence interval (CI) are shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

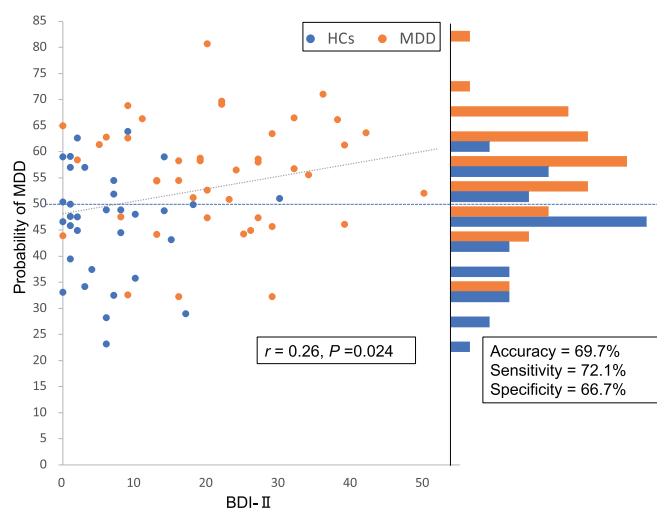


Fig. 2. Generalization performance of the brain network marker for major depressive disorder (MDD). Scatter plot of the Beck Depression Inventory-II (BDI-II) score and the classifier's output (probability of MDD). The black dotted line indicates the linear regression of the BDI score from the probability of MDD. The correlation coefficient (r) and P -value are shown. Each data point represents individual healthy controls (HCs) and patients with MDD. The histogram displays the distribution of the probability of MDD for HCs and patients with MDD.

measuring the same person twice with the brain network marker for MDD developed in our previous study (Yamashita et al., 2020), and that sufficient sensitivity can be reproduced for new patients with MDD.

Although we have reported the generalizability of the network marker to completely independent data obtained at different facilities in our previous study (Yamashita et al., 2020), repeated confirmation of the reproducibility of network markers is important when considering

their clinical significance. An important advance in the present study was that whereas our previous study divided the existing data into discovery and validation cohorts to evaluate the performance of the network marker, we evaluated its performance on data acquired after it was developed. In addition, regarding the reliability of the network marker, we had verified its reliability using only a small number of healthy subjects at a short interval, but in the present study, we increased the number of subjects and verified its reliability at a relatively long interval. By utilizing a longitudinal rs-fMRI dataset of HCs obtained at an interval of approximately 1 year, we determined that the classifier's output (probability of MDD) was acceptably stable. Compared to a recent review of the test-retest reliability of FC indicating that individual edges show a weak ICC of 0.29 on average (Noble et al., 2019), our brain network marker exhibited moderate reliability with an ICC of 0.45. We assessed the magnitude of the variation of individual FCs in repeated measurements in our rs-fMRI by using the data from our previous study (Yamashita et al., 2019). As a result, we found that the variation between measurements was approximately 2.5 times the subject effect (slightly reduced by harmonization), and even in analysis limited to the same machine used in this study, it was slightly less than twice the subject effect (Fig. S1). Since the variability of individual FCs was very large in our own data, the result for test-retest reliability suggests that our brain network marker consists of a small number of connections that are robust against not only facility-to-facility differences but also inter-measurement variations.

By applying our brain network marker for MDD diagnosis to newly acquired data, the classifier distinguished patients with MDD and HCs with an accuracy of 69.7 % (sensitivity, 72.1 %; specificity, 66.7 %). This result was comparable to or even better than the 66 % accuracy for the validation dataset and 66 % accuracy for the discovery dataset in our previous study (Yamashita et al., 2020). Generalization performance was confirmed again with the same or even better accuracy, indicating that our brain network marker for MDD diagnosis is highly stable. In the future, we will clarify the clinical significance of the brain network marker by conducting prospective studies and examining the relationship between longitudinal changes and treatment responsiveness.

5. Limitations

The data of patients with MDD in this study were cross-sectional, and the clinical significance of the marker, such as whether it is a state or trait marker of MDD and its association with treatment responsiveness, remains unclear. Repeated measurements of rs-fMRI were performed only on HCs in this study. Given that damage to brain structure or function may be more rapid and severe in patients with MDD compared to HCs, the test-retest reliability of the marker in patients with MDD needs further investigation.

6. Conclusions

The results of this study reaffirmed the test-retest reliability in HCs and generalization performance of our brain network marker for the diagnosis of MDD (Yamashita et al., 2020) using newly acquired and longitudinal data.

CRedit authorship contribution statement

GO, MK, and YO designed the study. GO, EI, SY, TK, YM, MT, AY, AM, OY, NY, TT, HJ, and YO collected the data. GO, TY, and AY performed analysis under the supervision of MK and YO. GO drafted the manuscript. GO, MK, and YO discussed the results and conclusions for editing the manuscript. All authors contributed to the article and approved the submitted version.

Conflict of interest

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).

Acknowledgements

This research was supported by AMED under grant numbers JP18dm0307002, JP18dm0307008, and JP20pc0101061. The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript. A draft of this manuscript was edited by NAI, Inc. (Yokohama, Japan).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jad.2023.01.087>.

References

- Beck, A., Steer, R., Brown, G., 1996. Manual for the Beck Depression Inventory-II. Psychological Corporation, San Antonio, TX.
- Dickie, E.W., Anticevic, A., Smith, D.E., Coalson, T.S., Manogaran, M., Calarco, N., Viviano, J.D., Glasser, M.F., Van Essen, D.C., Voineskos, A.N., 2019. Ciftify: a framework for surface-based analysis of legacy MR acquisitions. *NeuroImage* 197, 818–826.
- Esteban, O., Markiewicz, C.J., Blair, R.W., Moodie, C.A., Isik, A.I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S.S., Wright, J., Durnez, J., Poldrack, R.A., Gorgolewski, K.J., 2019. fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* 16, 111–116.
- Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., Smith, S.M., Van Essen, D. C., 2016. A multi-modal parcellation of human cerebral cortex. *Nature* 536, 171–178.
- Insel, T.R., Cuthbert, B.N., 2015. Medicine. Brain disorders? Precisely. *Science* 348, 499–500.
- Kambeitz, J., Cabral, C., Sacchet, M.D., Gotlib, I.H., Zahn, R., Serpa, M.H., Walter, M., Falkai, P., Koutsouleris, N., 2017. Detecting neuroimaging biomarkers for depression: a meta-analysis of multivariate pattern recognition studies. *Biol Psychiatry* 82, 330–338.
- Kojima, M., Furukawa, T.A., Takahashi, H., Kawai, M., Nagaya, T., Tokudome, S., 2002. Cross-cultural validation of the Beck depression inventory-II in Japan. *Psychiatry Res.* 110, 291–299.
- Noble, S., Scheinost, D., Constable, R.T., 2019. A decade of test-retest reliability of functional connectivity: a systematic review and meta-analysis. *NeuroImage* 203, 116157.
- Otsubo, T., Tanaka, K., Koda, R., Shinoda, J., Sano, N., Tanaka, S., Aoyama, H., Mimura, M., Kamijima, K., 2005. Reliability and validity of Japanese version of the mini-international neuropsychiatric interview. *Psychiatry Clin. Neurosci.* 59, 517–526.
- Power, J.D., Mitra, A., Laumann, T.O., Snyder, A.Z., Schlaggar, B.L., Petersen, S.E., 2014. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage* 84, 320–341.
- Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The mini-international neuropsychiatric interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* 59 (Suppl. 20), 22–33 quiz 34–57.
- Sun, K., Liu, Z., Chen, G., Zhou, Z., Zhong, S., Tang, Z., Wang, S., Zhou, G., Zhou, X., Shao, L., Ye, X., Zhang, Y., Jia, Y., Pan, J., Huang, L., Liu, X., Liu, J., Tian, J., Wang, Y., 2022. A two-center radiomic analysis for differentiating major depressive disorder using multi-modality MRI data under different parcellation methods. *J. Affect. Disord.* 300, 1–9.
- Yamashita, A., Sakai, Y., Yamada, T., Yahata, N., Kunimatsu, A., Okada, N., Itahashi, T., Hashimoto, R., Mizuta, H., Ichikawa, N., Takamura, M., Okada, G., Yamagata, H., Harada, K., Matsuo, K., Tanaka, S.C., Kawato, M., Kasai, K., Kato, N., Takahashi, H., Okamoto, Y., Yamashita, O., Imamizu, H., 2020. Generalizable brain network markers of major depressive disorder across multiple imaging sites. *PLoS Biol.* 18, e3000966.
- Yamashita, A., Yahata, N., Itahashi, T., Lisi, G., Yamada, T., Ichikawa, N., Takamura, M., Yoshihara, Y., Kunimatsu, A., Okada, N., Yamagata, H., Matsuo, K., Hashimoto, R., Okada, G., Sakai, Y., Morimoto, J., Narumoto, J., Shimada, Y., Kasai, K., Kato, N., Takahashi, H., Okamoto, Y., Tanaka, S.C., Kawato, M., Yamashita, O., Imamizu, H., 2019. 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.