

Contrastive Functional Connectivity Defines Neurophysiology-informed Symptom Dimensions in Major Depression

Hao Zhu¹, Xiaoyu Tong¹, Nancy B. Carlisle², Hua Xie³, Corey J. Keller^{4,5,6}, Desmond J. Oathes^{7,8}, Charles B. Nemeroff⁹, Gregory A. Fonzo⁹, Yu Zhang^{1,10,*}

1 Department of Bioengineering, Lehigh University, Bethlehem, PA, USA

2 Department of Psychology, Lehigh University, Bethlehem, PA, USA

3 Center for Neuroscience Research, Children's National Hospital, Washington, DC, USA

4 Wu Tsai Neuroscience Institute, Stanford University, Stanford, CA, USA

5 Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, USA

6 Veterans Affairs Palo Alto Healthcare System, and the Sierra Pacific Mental Illness, Research, Education, and Clinical Center (MIRECC), Palo Alto, CA, 94394, USA

7 Center for Brain Imaging and Stimulation, Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

8 Center for Neuromodulation in Depression and Stress, Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA

9 Center for Psychedelic Research and Therapy, Department of Psychiatry and Behavioral Sciences, Dell Medical School, The University of Texas at Austin, Austin, TX, USA

10 Department of Electrical and Computer Engineering, Lehigh University, Bethlehem, PA, USA

* Corresponding author: Yu Zhang
Assistant Professor of Bioengineering
Assistant Professor of Electrical and Computer Engineering
Lehigh University
Bethlehem, PA 18015, USA
Email: yuzi20@lehigh.edu

Trial Registration: Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care for Depression (EMBARC), NCT#01407094

Abstract

Background: Major depressive disorder (MDD) is a prevalent psychiatric disorder characterized by substantial clinical and neurobiological heterogeneity. Conventional studies that solely focus on clinical symptoms or neuroimaging metrics often fail to capture the intricate relationship between these modalities, limiting their ability to disentangle the complexity in MDD. Moreover, patient neuroimaging data typically contains normal sources of variance shared with healthy controls, which can obscure disorder-specific variance and complicate the delineation of disease heterogeneity.

Methods: We employed contrastive principal component analysis to extract disorder-specific variations in fMRI-based resting-state functional connectivity (RSFC) by contrasting MDD patients (N=233) with age-matched healthy controls (N=285). We then applied sparse canonical correlation analysis to identify latent dimensions in the disorder variations by linking the extracted contrastive connectivity features to clinical symptoms in MDD patients.

Results: Two significant and generalizable dimensions linking distinct brain circuits and clinical profiles were discovered. The first dimension, associated with an apparent “internalizing-externalizing” symptom dimension, was characterized by self-connections within the visual network and also associated with choice reaction times of cognitive tasks. The second dimension, associated with personality facets such as extraversion and conscientiousness typically inversely associated with depression symptoms, is primarily driven by self-connections within the dorsal attention network. This “depression-protective personality” dimension is also associated with multiple cognitive task performances related to psychomotor slowing and cognitive control.

Conclusions: Our contrastive RSFC-based dimensional approach offers a new avenue to dissect clinical heterogeneity underlying MDD. By identifying two stable, neurophysiology-informed symptom dimensions in MDD patients, our findings may enhance disease mechanism insights and facilitate precision phenotyping, thus advancing the development of targeted therapeutics for precision mental health.

Introduction

Major depressive disorder (MDD) is a highly prevalent mental disorder with a lifetime prevalence of over 20% in women and 11% in men, impacting a substantial number of individuals globally. MDD is diagnosed based on the Diagnostic and Statistical Manual of Mental Disorders 5th edition¹. Unfortunately, such symptom-based diagnoses include patients with a bewildering mixture of symptoms and have resulted in substantial clinical and neurobiological heterogeneity, obscuring the underlying mechanisms of the cognitive and behavioral dysfunctions in MDD patients. This oversight may contribute to suboptimal treatment efficacy^{2,3} and hinder the development of more effective therapeutics, necessitating the dissection of MDD heterogeneity.

Prior studies have attempted to examine the clinical heterogeneity within MDD utilizing symptom-based dimensional approaches or subtyping analyses⁴⁻⁶. For instance, efforts have been made to identify symptom dimensions shared among patients and their associations with personality⁷, treatment outcomes⁸, or comorbid mental health conditions⁹. While current clinical measures are useful for inexpensively characterizing a patient symptom profile, they fall short in capturing underlying neurophysiological variations, limiting the discovery of neurobiological basis underlying the clinical heterogeneity. A recent systematic review⁶ reported that there is wide diversity of identified symptom dimensions and subtypes among previous studies, indicating the failure of addressing disease heterogeneity and the necessity of a more reliable and objective delineation of MDD pathological dimensions.

Neuroimaging techniques, such as functional magnetic resonance imaging (fMRI), have demonstrated promise in probing neurobiology of various psychiatric disorders. However, many previous neuroimaging studies have followed a case-control design that focuses on the group difference in aberrant brain circuits between MDD patients and healthy individuals, either during the resting state¹⁰⁻¹² or in response to specific tasks¹³. Such an approach can only extract population-level information and has a limited ability to characterize the heterogeneity in MDD¹⁴. Recent research efforts have shifted to neuroimaging-based dimensional analyses to examine variations in brain structure and function beyond the conventional group-level comparison^{15,16}. In particular, leveraging functional connectivity that measures neural coupling between brain regions, recent studies have revealed novel subtypes among mood and anxiety disorders^{17,18}, which are not identifiable via a conventional case-control approach or clinically-defined MDD categories. To further enhance the clinical relevance of subtype or dimension discovery, increasing research efforts have delved deeper into not only dimensional patterns of brain circuits but also their behavioral or cognitive profiles^{17,19-21}. Typically employing machine learning techniques such as canonical correlation analysis (CCA) or partial least squares, these studies have identified latent dimensions by jointly examining functional connectivity and symptom/behavioral measures in a data-driven manner¹⁹⁻²¹. These approaches revealed patterns of association between neurophysiological characteristics and specific symptom combinations, offering insights into the underlying mechanisms behind these disorders.

Neuroimaging data from patients contains both disorder-specific variation and disorder-irrelevant variation shared with healthy controls²². Given the high variability observed across individual-level brain metrics, one challenge in characterizing heterogeneity among clinical populations lies in the fact that it is usually muddied by overall normative variability. This has substantially decreased the signal-to-noise ratio and leads to suboptimal identification of meaningful patterns. In addressing this challenge, contrastive learning emerges as a powerful tool for training models to selectively retain group-specific features while filtering out

confounding information inherent in the broader population. This process enables a focused analysis by isolating and emphasizing the distinctive characteristics pertinent to the target group. Encouragingly, some recent works have integrated the contrastive learning technique into the exploration of brain biomarkers in psychiatric disorders such as autism^{22,23}, demonstrating its unique advantages in uncovering the intricate nuances of brain morphology²² and functional connectivity²³ associated with psychiatric conditions. Contrastive learning holds great potential in capturing the inadequately explored pathology-relevant variations for defining symptom dimensions more precisely within MDD patients. Hence, we aim to shed light on the underlying heterogeneity specific to MDD through the identification of linked dimensions between clinical symptoms and MDD-specific neurophysiology components obtained by contrastive learning.

In this study, we applied a contrastive learning-based brain-symptom analytical framework²³ to identify latent dimensions that link resting-state functional connectivity (RSFC) and clinical symptoms in MDD patients. This framework comprises two main steps. First, we conducted the contrastive principal component analysis (cPCA) on the fMRI-based RSFC data of MDD patients (N=233), with age-matched healthy controls (N=285) as background data, i.e. normative sources of variance to be removed from consideration in subtyping analyses. This process allowed us to extract contrastive RSFC features that disentangled disorder-specific variations from those shared with the healthy population. We then applied sparse CCA between these contrastive RSFC features and representative symptoms to reveal informative latent dimensions that link MDD-specific brain functional variations and clinical symptoms. Our study successfully identified two robust latent dimensions showing distinct neural circuit patterns and clinical profiles, primarily involving internalizing-externalizing symptoms and depression-protective personality. Rigorous stability analysis and cross-validation further confirmed their robustness and generalizability. We also examined associations of the identified dimensions with performance on various cognitive tasks, offering valuable insights into MDD-related cognitive dysfunction. Collectively, our contrastive RSFC-based dimensional approach provides a new avenue to examine neurophysiology-informed symptom dimensions for an improved understanding of mechanisms underlying MDD symptom heterogeneity. This may lead to a more objective classification of psychiatric conditions, thereby advancing targeted therapeutics for precision mental health.

Materials and Methods

Participants

Patient Population: The patient population used in our study include MDD patients from the *Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care for Depression* (EMBARC) dataset²⁴. EMBARC is a large, randomized placebo-controlled clinical trial for examining biomarkers for MDD and antidepressant treatment response. Written informed consent was obtained from each participant under the Institutional Review Board (IRB) approved protocols at each of the four study sites, including University of Texas Southwestern Medical Center, Columbia University/Stony Brook, Massachusetts General Hospital, University of Michigan, University of Pittsburgh, and McLean Hospital. Subjects were required to meet the SCID criterion for an MDD episode and have a Quick Inventory of Depressive Symptomatology score²⁵ of ≥ 14 at both screening and randomization visits. This dataset includes 296 MDD

patients who were randomly prescribed sertraline or placebo for eight weeks, with the primary outcome of treatment response measured by the 17-item Hamilton Depression Rating Scale (HAMD₁₇)²⁶. For our brain-symptom linked dimension analysis, we included 25 baseline clinical scales and subscales: symptoms of depression (the 17-item HAMD; Quick Inventory of Depressive Symptomatology); Childhood Trauma Questionnaire subscales (emotional abuse, emotional neglect, physical abuse, physical neglect, sexual abuse)²⁷; suicide-related symptoms (Concise Associated Symptoms Tracking²⁸; Concise Health Risk Tracking-Self Report: propensity and risk scores²⁹); Mood and Anxiety Symptoms Questionnaire subscales (anxious arousal, anhedonic depression, general distress)³⁰; Mood Disorder Questionnaire³¹; personalities (NEO-Five Factor Inventory subscales, including neuroticism, extraversion, openness, agreeableness, and conscientiousness³²; Standardised Assessment of Personality – Abbreviated Scale); Self-Administered Comorbidity Questionnaire³³; Snaith-Hamilton Pleasure Score³⁴; and Social Adjustment Scale overall mean score. Due to incomplete clinical measures, 63 subjects were excluded. The remaining 233 subjects (aged 18-65) were used for our analyses. Demographics of these subjects are listed in Supplementary Table S1.

Healthy Population: We compiled and harmonized 285 healthy controls (aged 18-77.5) as a background group for cPCA. Demographic information of these subjects is summarized in Supplementary Table S2. These subjects were selected by matching the age of the healthy population with the patient population (Kolmogorov–Smirnov test $p = 0.052$, Figure S1). Among them, 27 are from the EMBARC dataset and others are from the following three datasets. Specifically, we used 83 healthy controls from the University of California Los Angeles Consortium for Neuropsychiatric Phenomics (UCLA-CNP)³⁵, which was approved by the IRB at University of California, Los Angeles and the Los Angeles County Department of Mental Health. 72 subjects were selected from the Amsterdam Open MRI Collection - Population Imaging of Psychology (AOMIC-PIOP1)³⁶ dataset, which was approved by the faculty's ethical committee at the University of Amsterdam. The remaining 103 subjects were from the Leipzig Study for Mind-Body-Emotion Interactions (LEMON)³⁷ dataset, which was approved by the ethics committee at the medical faculty of the University of Leipzig.

MRI Acquisition and Preprocessing

EMBARC. MRI data were acquired using 3T MRI systems at four different sites. At each site, resting-state fMRI data were scanned via T2* weighted images using a single-shot gradient echo-planar pulse sequence lasting for six minutes, with parameter settings: repetition time 2000 ms, echo time 28 ms, flip angle 90°, matrix size 64 × 64, voxel size 3.2 × 3.2 × 3.1 mm³, and 39 axial slices. Each subject underwent one or two runs of fMRI scans within a single day.

UCLA-CNP. MRI data were acquired on two 3T Siemens Trio scanners at UCLA. Resting-state MRI data were scanned using a T2*-weighted echoplanar imaging sequence lasting for 304 seconds, with the following parameters: repetition time 2000 ms, echo time 30 ms, flip angle 90°, matrix size 64 × 64, voxel size 3 × 3 × 4 mm³, and 34 slices.

AOMIC-PIOP1. MRI data were acquired on a Philips 3T scanner. During the resting state scans, participants were instructed to keep their gaze fixated on a fixation cross in the middle of the screen with a gray background and to let their thoughts run freely. Resting-state fMRI scanning lasted six minutes (i.e., 480 volumes with a repetition time 750 ms).

LEMON. MRI data were scanned using a 3T Siemens Verio scanner. Resting-state fMRI data were scanned using a T2*-weighted gradient echo planar imaging sequence lasting for 15 minutes and 30 seconds. The sequence parameters were specified as follows: repetition time 1400 ms, echo time 30 ms, flip angle 69°, matrix size 88 × 88, voxel size 2.3 × 2.3 × 2.3 mm³, and 64 slices.

All the resting-state fMRI data were preprocessed using the fMRIPrep pipeline³⁸ and aggregated into 100 regions-of-interest (ROIs) level time series according to the Schaefer parcellation³⁹. RSFC features were then calculated as Pearson's correlation coefficient in fMRI time series between every pair of ROIs. For MDD patients with multiple fMRI runs, we took the average of the RSFC across runs to generate the RSFC feature data used for the subsequent analysis. Given that the fMRI data from healthy controls were derived from different studies, we applied the well-established ComBat harmonization technique⁴⁰ to their RSFC data to mitigate site effects. During the removal of site effects, age and gender were designated as biological covariates to be preserved, with EMBARC as the reference batch.

Disentangling MDD-specific RSFC Variations via Contrastive PCA

In our study, we employed cPCA⁴¹ on the RSFC data of MDD patients as the target data and the RSFC data of healthy controls as the background data, aiming to pinpoint the disorder-specific foreground components by contrasting the healthy population. cPCA operates under the assumption that the target data comprises both domain-specific (foreground) information and domain-unrelated (background) variance. Utilizing the covariance matrix C_p derived from the target data (MDD patients in our case) and the covariance matrix C_{bg} derived from the background data (healthy controls), cPCA identifies the linear components most closely associated with the foreground data by subtracting the background covariance matrix from the target covariance matrix to obtain the foreground covariance matrix: $C_{fg} = C_p - \alpha C_{bg}$, where α represents a hyperparameter quantifying the degree of contrast. This hyperparameter was determined based on the cross-validation results through grid search. The top 200 contrastive principal components, which explain more than 70% of the data variance, were used in subsequent analyses. This selection was made because the dimensionality of informative components is constrained by the sample size of patients.

Identifying Neurophysiology-informed Symptom Dimensions using sCCA

We first applied PCA to the clinical variables and retained 80% of the total variance to reduce dimensionality. We then applied sparse canonical correlation analysis (sCCA)⁴² to identify latent dimensions linking contrastive connectivity features with clinical symptoms. sCCA optimizes the correlation between two data matrices, thus yielding symptom dimensions maximally associated with neuroimaging features. Specifically, we utilized sCCA to analyze the disorder-specific components extracted from cPCA and corresponding clinical measures, aiming to identify neurophysiology-informed symptom dimensions. Sparsity constraint is imposed on both connectivity features and clinical measures to improve interpretability and alleviate overfitting. The overall minimization objective is formulated as $-Cov(Au, Bv) + \lambda_1 \|u\|_1 + \lambda_2 \|v\|_1$, where A and B are data matrices of connectivity features and clinical measures respectively, u , v are the FC

and symptom dimension loadings, and λ_1, λ_2 are sparsity hyperparameters on the loadings. We performed grid search to find the hyperparameters with best performance.

Cross-Validation and Statistical Significance Verification

We performed 10 repetitions of 10-fold cross-validation to evaluate the generalizability and stability of identified latent dimensions. In each fold, 90% of the patients and all healthy subjects were compiled as the training set of cPCA. As the order of identified CCA dimensions might vary across folds due to training data variance, we employed the dimensions identified from the whole dataset as reference dimensions to align the dimensions acquired from different folds. Pearson's correlation coefficient was then computed to assess the generalizability of the identified dimensions on the validation set. Intra-class correlation coefficient was employed to confirm the stability of dimension loadings across cross-validation folds.

Furthermore, to evaluate the significance of the identified dimensions, we conducted permutation tests 1000 times. Specifically, we randomly permuted the clinical symptoms for each patient to others and followed the same procedures to generate the Pearson correlation R values on cross validation test set as described above. Test R values from dimensions with the same ranking across different folds were aggregated to generate the null distribution.

Network-level Connectivity Importance

Network-level connectivity provides a broader view of which neural systems and their interactions contribute most to the identified latent dimensions. To measure this, we first retained only the top 10% loadings with the highest absolute values, setting the rest to zero to exclude the less important and stable connections. Then, we averaged all absolute loadings within and between each pair of networks to obtain the network-level connectivity importance.

Post-hoc Examination of Associations between Dimensions and Cognitive Task Performance

We investigated the correlation between the FC scores of identified dimensions and the cognitive task performance of MDD patients. This analysis may provide insights into how different cognitive abilities, such as psychomotor slowing⁴³, cognitive control⁴⁴, working memory⁴⁵, reward learning⁴⁶, and resolution and adjustment behavior in response to emotional conflict⁴⁷, relate to these brain-symptom latent dimensions. These cognitive and emotional measurements are selected because they are potential predictors for MDD treatment response²⁴, thus having the probability to be correlated with the identified MDD-specific dimensions. The specific task item we included can be found in Table S3.

Results

Contrastive RSFC Defines Two Generalizable Brain-Symptom Linked Dimensions in MDD

We applied the contrastive learning-based brain-symptom association identification framework to the 233 MDD patients. Two dimensions with generalizable association between neurophysiology and symptom profile have been identified through 10 rounds of 10-fold cross-validation and permutation test. The composition and essence of the first identified dimension was interpreted, based upon loadings, to indicate an axis of internalizing-externalizing symptoms ($R_{\text{train}} = 0.551$, $R_{\text{cv}} = 0.287$, $p_{\text{permutation}} < 0.001$, Figures 2, S4). A higher internalizing-externalizing symptoms dimension score indicates a greater tendency towards internalizing behavior such as lack of pleasure or energy, and a lower tendency towards externalizing behavior such as mania and anger attack. The second identified dimension incorporates the concept of depression-protective personality ($R_{\text{train}} = 0.437$, $R_{\text{cv}} = 0.207$, $p_{\text{permutation}} = 0.001$, Figure 2, S4). Individuals with higher scores on this personality dimension score exhibited higher scores on personality facets typically negatively associated with depression symptoms, such as extraversion and conscientiousness, while a low dimension score indicates higher scores on personality facets typically associated with greater depression symptoms, such as neuroticism⁴⁸. Both dimensions significantly outperformed the performance yielded from its non-contrastive counterpart integrating standard PCA and sCCA (paired t-test: internalizing-externalizing symptoms dimension: $p = 5.2 \times 10^{-4}$; depression-protective personality dimension: $p = 0.008$), demonstrating the unique advantage of contrastive learning for improving the identification of brain-symptom associations. Afterward, we calculated the intra-class correlation coefficient (ICC) of dimension composition in each cross-validation fold to assess the stability of the dimensions. These two dimensions both showed high stability. For internalizing-externalizing symptoms, the ICC value is 0.87 (95% CI: [0.87, 0.88]) for FC loadings, 0.92 (95% CI: [0.88, 0.96]) for symptom loadings. For depression-protective personality, the ICC value is 0.94 (95% CI: [0.94, 0.94]) for FC loadings, 0.94 (95% CI: [0.90, 0.97]) for symptom loadings (Figure S4).

RSFC Signatures Associated with MDD Symptoms

Next, we investigated the most important ROI-level and network-level connectivity within each dimension. (Figure 3). In the internalizing-externalizing symptoms dimension, the correlation between the left and right cuneus and other regions, particularly the left fusiform gyrus, the left inferior temporal gyrus, and the right fusiform gyrus, contributes the most. In terms of anti-correlation, the connection between the left and right posterior cingulate cortex and other regions, mainly the left superior occipital gyrus and the right middle occipital gyrus, contributes the most. On the network level, the dimension is characterized by the self-connection of visual network, the connection between limbic network and sensorimotor network, and the connection between limbic network and dorsal attention network. In depression-protective personality dimension, the correlation between the right middle cingulate cortex and the left superior parietal cortex contributes the most. In terms of anti-correlation, the key connections are between the left fusiform gyrus, the left inferior temporal gyrus, and the left superior parietal cortex, as well as with the left inferior frontal gyrus and the left middle frontal gyrus. On network level, this dimension is highly related to self-connections within the dorsal attention network.

Dimensions Are Associated with Cognitive Task Performance

Internalizing-externalizing symptoms score showed a significant correlation with reaction time of choice reaction time task ($r=0.20$, $p_{\text{fdr}}=0.0084$) and reaction time difference in a flanker task ($r=0.18$, $p_{\text{fdr}}=0.024$). This indicates that a higher score on this dimension is associated with longer reaction time, suggesting lower performance. Reaction time difference in the flanker task measures selective attention and inhibitory function; longer reaction times indicate a greater interference effect from incongruent stimuli. Depression-protective personality dimension score had significant correlations with multiple cognitive tasks, including the accuracy of A-not-B task ($r=-0.24$, $p_{\text{fdr}}=0.0016$), reaction time of choice reaction time task ($r=0.31$, $p_{\text{fdr}}=2.8 \times 10^{-5}$), and both the accuracy difference ($r=-0.15$, $p_{\text{fdr}}=0.046$) and reaction time difference ($r=0.20$, $p_{\text{fdr}}=0.0084$) in the flanker task. This suggests that this dimension has a wider range of correlations with different cognitive tasks measuring working memory, interference adjustments, and cognitive control. No significant correlations were observed between these two dimensions and the word fluency task, emotion conflict task, or probabilistic reward task (Figure 4).

We also examined the capabilities of these dimensions in predicting antidepressant treatment response using chi-square test. However, no significant performance was observed for either dimension in differentiating remission versus non-remission based on whether the FC dimension scores were above or below the median (internalizing-externalizing symptoms dimension: $p=0.27$ for sertraline arm, $p=1.00$ for placebo arm; depression-protective personality dimension: $p=0.40$ for sertraline arm, $p=0.49$ for placebo arm) (Supplementary Table S4).

Discussion

In this study, we implemented a data-driven framework that integrated contrastive machine learning with multivariate correlation analysis to uncover MDD-specific brain functional variations and their associations with clinical symptoms. We identified two robust and generalizable latent dimensions linking neurophysiological and clinical profiles, offering novel and objective biomarkers for dissecting heterogeneity in MDD.

While neuroimaging data conveys important information about circuit dysfunction in MDD patients, it also contains undesired variance shared with healthy individuals. Conventional dimensional methods obtain and identify behavior or symptom-related biomarkers from clinical populations, which may result in suboptimal findings without removing the variance shared by healthy populations. Alternatively, contrastive learning-based frameworks can better extract disorder-specific neurophysiology components through data distillation. In our results, the dimensions derived from contrastive learning show a reduced correlation with age compared to those without contrastive learning (Figures S2, S3). This suggests that contrastive learning successfully removed some disorder-unrelated components, such as age. Other than the undesired variance, MDD itself also exhibits substantial neurobiological and clinical heterogeneity. Previous studies based on case-control analysis or dimensional approaches considering only clinical symptoms or neuroimaging may therefore inadequately address the heterogeneity^{6,49}. Our framework links both modalities, addresses the underlying heterogeneity, and provides interpretability.

For the two neurophysiology-informed symptom dimensions we have identified, internalizing-externalizing symptoms dimension is associated with key brain regions such as the cuneus and fusiform gyrus, consistent with findings from studies on neurodevelopmental trajectories of internalizing-externalizing symptoms⁵⁰. Our results suggest that these associations likely persist into adulthood. Our network-level analysis on this dimension aligns with the literature, which identifies the visual network and dorsal attention network as strong predictors of internalizing-externalizing traits⁵¹. Our analysis of the behavioral task is also consistent with findings in the literature that associate cuneus activation with facilitating vigilance⁵².

The depression-protective personality dimension is marked by personalities such as extraversion, conscientiousness and neuroticism. Studies have identified a strong association between key brain regions, such as the fusiform gyrus, and personality disorders^{53,54}, which is consistent with our findings. A recent study also identified the significant correlations between dorsal attention network and personalities including neuroticism and agreeableness⁵⁵, and borderline personality disorder⁵⁶. Additionally, dorsal attention network has also shown great importance in multiple cognitive control tasks^{57,58}. All these results align with our findings to the depression-protective personality dimension.

Despite our contributions, this study has several limitations. First, while cPCA is effective for extracting disorder-specific components, it requires fine-tuning of hyperparameters to control the degree of contrast with the background data (i.e., healthy population). Future research could focus on developing automatic frameworks for extracting contrastive connectivity features. Additionally, our study harmonized data from different sources to enlarge the sample size, enhancing the robustness of the identified latent dimensions, followed by verifying their generalizability through rigorous cross-validation. However, future work should further involve independent datasets with comparable clinical measurements to replicate our findings. Lastly, given the recognized importance of subcortical regions and cerebellum in depression^{59,60}, extending the analysis to these regions may provide a more comprehensive understanding of brain dysfunctions associated with MDD psychopathology.

In summary, this study employed a combination of contrastive learning and sparse canonical component analysis on a depression patient dataset, unveiling two generalizable disorder-specific dimensions linking neurophysiology and symptom profiles. Our findings hold potential for advancing the understanding of disease mechanisms, facilitating precise diagnosis, formulating individualized treatment plans, and fostering the development of innovative treatment methods.

Acknowledgements

This work was supported by NIH grant nos. R01MH129694, R21MH130956, R21AG080425, Alzheimer's Association Grant (AARG-22-972541), and Lehigh University FIG (FIGAWD35), CORE, and Accelerator grants. Portions of this research were conducted on Lehigh University's Research Computing infrastructure partially supported by NSF Award 2019035. G.A.F. was also supported by philanthropic funding and NIH grant nos. R01MH132784 and R01MH125886, and grants from the One Mind – Baszucki Brain Research Fund, the SEAL Future Foundation, and the Brain and Behavior Research Foundation.

Financial Disclosures

G.A.F. received monetary compensation for consulting work for SynapseBio AI and owns equity in Alto Neuroscience. C.J.K. reports equity from Alto Neuroscience. The remaining authors declare no competing interests. C.N. is a consultant for ANeuroTech (division Anima BV), Janssen Research and Development, BioXcel Therapeutics, Engrail Therapeutics, Clexio Biosciences LTD, EmbarkNeuro, Galen Mental Health LLC, Goodcap Pharmaceuticals, ITI Inc, LUCY Scientific Discovery, Relmada Therapeutics, Sage Therapeutics, Senseye Inc, Precisement Health, Autobahn Therapeutics Inc, EMA Wellness, Skyland Trails, Denovo Biopharma, and the Brain & Behavior Research Foundation. C.N. owns the following patents: Method and devices for transdermal delivery of lithium (US 6,375,990B1), Method of assessing antidepressant drug therapy via transport inhibition of monoamine neurotransmitters by ex vivo assay (US 7,148,027B2), Compounds, Compositions, Methods of Synthesis, and Methods of Treatment (CRF Receptor Binding Ligand) (US 8,551, 996 B2). C.N. owns stock in Corcept Therapeutics Company, EMA Wellness, Precisement Health, Relmada Therapeutics, Signant Health, Galen Mental Health LLC, and Senseye Inc. The remaining authors have no conflicts of interest to declare.

References

1. Edition, F. Diagnostic and statistical manual of mental disorders. *Am Psychiatric Assoc* **21**, 591-643 (2013).
2. Cipriani, A., *et al.* Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet* **391**, 1357-1366 (2018).
3. Nierenberg, A.A., *et al.* A brief review of antidepressant efficacy, effectiveness, indications, and usage for major depressive disorder. *J Occup Environ Med* **50**, 428-436 (2008).
4. Andrews, G., *et al.* Dimensionality and the category of major depressive episode. *Int J Methods Psychiatr Res* **16 Suppl 1**, S41-51 (2007).
5. Enns, M.W. & Cox, B.J. Personality dimensions and depression: review and commentary. *The Canadian Journal of Psychiatry* **42**, 274-284 (1997).
6. Van Loo, H.M., De Jonge, P., Romeijn, J.-W., Kessler, R.C. & Schoevers, R.A. Data-driven subtypes of major depressive disorder: a systematic review. *BMC medicine* **10**, 1-12 (2012).
7. Hur, J.W. & Kim, Y.K. Comparison of Clinical Features and Personality Dimensions between Patients with Major Depressive Disorder and Normal Control. *Psychiatry Investig* **6**, 150-155 (2009).
8. Vrieze, E., *et al.* Dimensions in major depressive disorder and their relevance for treatment outcome. *J Affect Disord* **155**, 35-41 (2014).
9. Biehn, T.L., *et al.* Underlying Dimensions of DSM-5 Posttraumatic Stress Disorder and Major Depressive Disorder Symptoms. *Psychological Injury and Law* **6**, 290-298 (2013).
10. Greicius, M.D., *et al.* Resting-state functional connectivity in major depression: abnormally increased contributions from subgenual cingulate cortex and thalamus. *Biol Psychiatry* **62**, 429-437 (2007).
11. Sen, B., Mueller, B., Klimes-Dougan, B., Cullen, K. & Parhi, K.K. Classification of major depressive disorder from resting-state fMRI. in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 3511-3514 (IEEE, 2019).
12. Shen, T., *et al.* Increased cognition connectivity network in major depression disorder: a FMRI study. *Psychiatry Investig* **12**, 227-234 (2015).
13. Davey, C.G., Yucel, M., Allen, N.B. & Harrison, B.J. Task-related deactivation and functional connectivity of the subgenual cingulate cortex in major depressive disorder. *Front Psychiatry* **3**, 14 (2012).
14. Brakowski, J., *et al.* Resting state brain network function in major depression - Depression symptomatology, antidepressant treatment effects, future research. *J Psychiatr Res* **92**, 147-159 (2017).
15. Fu, C.H.Y., *et al.* AI-based dimensional neuroimaging system for characterizing heterogeneity in brain structure and function in major depressive disorder: COORDINATE-MDD consortium design and rationale. *BMC Psychiatry* **23**, 59 (2023).
16. Fu, C.H.Y., *et al.* Neuroanatomical dimensions in medication-free individuals with major depressive disorder and treatment response to SSRI antidepressant medications or placebo. *Nature Mental Health* (2024).
17. Drysdale, A.T., *et al.* Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat Med* **23**, 28-38 (2017).
18. Zhang, Y., *et al.* Identification of psychiatric disorder subtypes from functional connectivity patterns in resting-state electroencephalography. *Nature biomedical engineering* **5**, 309-323 (2021).
19. Dunlop, K., *et al.* Dimensional and Categorical Solutions to Parsing Depression Heterogeneity in a Large Single-Site Sample. *Biol Psychiatry* (2024).
20. Kebets, V., *et al.* Somatosensory-Motor Dysconnectivity Spans Multiple Transdiagnostic Dimensions of Psychopathology. *Biol Psychiatry* **86**, 779-791 (2019).
21. Xia, C.H., *et al.* Linked dimensions of psychopathology and connectivity in functional brain networks. *Nat Commun* **9**, 3003 (2018).
22. Aglinskias, A., Hartshorne, J.K. & Anzellotti, S. Contrastive machine learning reveals the structure of neuroanatomical variation within autism. *Science* **376**, 1070-1074 (2022).

23. Tong, X., *et al.* Symptom dimensions of resting-state electroencephalographic functional connectivity in autism. *Nature Mental Health* (2024).
24. Trivedi, M.H., *et al.* Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): Rationale and design. *J Psychiatr Res* **78**, 11-23 (2016).
25. Rush, A.J., *et al.* The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry* **54**, 573-583 (2003).
26. Hamilton, M. A rating scale for depression. *Journal of neurology, neurosurgery, and psychiatry* **23**, 56 (1960).
27. Bernstein, D.P., *et al.* Initial reliability and validity of a new retrospective measure of child abuse and neglect. *The American journal of psychiatry* **151**, 1132-1136 (1994).
28. Trivedi, M.H., *et al.* Concise Associated Symptoms Tracking scale: a brief self-report and clinician rating of symptoms associated with suicidality. *The Journal of clinical psychiatry* **72**, 3587 (2011).
29. Trivedi, M.H., *et al.* Concise Health Risk Tracking scale: a brief self-report and clinician rating of suicidal risk. *J Clin Psychiatry* **72**, 757-764 (2011).
30. Watson, D. & Clark, L.A. Mood and anxiety symptom questionnaire. *Journal of Behavior Therapy and Experimental Psychiatry* (1991).
31. Hirschfeld, R.M., *et al.* Development and validation of a screening instrument for bipolar spectrum disorder: the Mood Disorder Questionnaire. *American journal of psychiatry* **157**, 1873-1875 (2000).
32. McCrae, R.R. & Costa, P.T. The structure of interpersonal traits: Wiggins's circumplex and the five-factor model. *Journal of personality and social psychology* **56**, 586 (1989).
33. Sangha, O., Stucki, G., Liang, M.H., Fossel, A.H. & Katz, J.N. The Self-Administered Comorbidity Questionnaire: a new method to assess comorbidity for clinical and health services research. *Arthritis Care & Research: Official Journal of the American College of Rheumatology* **49**, 156-163 (2003).
34. Snaith, R.P., *et al.* A scale for the assessment of hedonic tone the Snaith-Hamilton Pleasure Scale. *The British Journal of Psychiatry* **167**, 99-103 (1995).
35. Poldrack, R.A., *et al.* A phenome-wide examination of neural and cognitive function. *Sci Data* **3**, 160110 (2016).
36. Snoek, L., *et al.* The Amsterdam Open MRI Collection, a set of multimodal MRI datasets for individual difference analyses. *Sci Data* **8**, 85 (2021).
37. Babayan, A., *et al.* A mind-brain-body dataset of MRI, EEG, cognition, emotion, and peripheral physiology in young and old adults. *Sci Data* **6**, 180308 (2019).
38. Esteban, O., *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat Methods* **16**, 111-116 (2019).
39. Schaefer, A., *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb Cortex* **28**, 3095-3114 (2018).
40. Fortin, J.P., *et al.* Harmonization of cortical thickness measurements across scanners and sites. *Neuroimage* **167**, 104-120 (2018).
41. Abid, A., Zhang, M.J., Bagaria, V.K. & Zou, J. Exploring patterns enriched in a dataset with contrastive principal component analysis. *Nat Commun* **9**, 2134 (2018).
42. Witten, D.M., Tibshirani, R. & Hastie, T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* **10**, 515-534 (2009).
43. Taylor, B.P., *et al.* Psychomotor slowing as a predictor of fluoxetine nonresponse in depressed outpatients. *American Journal of Psychiatry* **163**, 73-78 (2006).
44. Julian, L.J. & Mohr, D.C. Cognitive predictors of response to treatment for depression in multiple sclerosis. *The Journal of neuropsychiatry and clinical neurosciences* **18**, 356-363 (2006).
45. Gorlyn, M., *et al.* Neuropsychological characteristics as predictors of SSRI treatment response in depressed subjects. *J Neural Transm (Vienna)* **115**, 1213-1219 (2008).
46. Pizzagalli, D.A., *et al.* Single dose of a dopamine agonist impairs reinforcement learning in humans: behavioral evidence from a laboratory-based measure of reward responsiveness. *Psychopharmacology (Berl)* **196**, 221-232 (2008).
47. Etkin, A., Egner, T., Peraza, D.M., Kandel, E.R. & Hirsch, J. Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron* **51**, 871-882 (2006).

48. Rosellini, A.J. & Brown, T.A. The NEO Five-Factor Inventory: Latent structure and relationships with dimensions of anxiety and depressive disorders in a large clinical sample. *Assessment* **18**, 27-38 (2011).
49. Dichter, G.S., Gibbs, D. & Smoski, M.J. A systematic review of relations between resting-state functional-MRI and treatment response in major depressive disorder. *J Affect Disord* **172**, 8-17 (2015).
50. Blok, E., Geenjaar, E.P., Geenjaar, E.A., Calhoun, V.D. & White, T. Neurodevelopmental trajectories in children with internalizing, externalizing and emotion dysregulation symptoms. *Frontiers in Psychiatry* **13**, 846201 (2022).
51. Qu, Y.L., *et al.* Distinct brain network features predict internalizing and externalizing traits in children and adults. *bioRxiv* (2023).
52. Chen, R., *et al.* Acute Tai Chi Chuan exercise enhances sustained attention and elicits increased cuneus/precuneus activation in young adults. *Cerebral Cortex* **33**, 2969-2981 (2023).
53. Dickey, C.C., *et al.* A MRI study of fusiform gyrus in schizotypal personality disorder. *Schizophrenia research* **64**, 35-39 (2003).
54. Kruschwitz, J., *et al.* 5-HTTLPR/rs25531 polymorphism and neuroticism are linked by resting state functional connectivity of amygdala and fusiform gyrus. *Brain Structure and Function* **220**, 2373-2385 (2015).
55. Simon, S.S., Varangis, E. & Stern, Y. Associations between personality and whole-brain functional connectivity at rest: Evidence across the adult lifespan. *Brain and behavior* **10**, e01515 (2020).
56. Hall, N.T. & Hallquist, M.N. Right anterior insula effective connectivity impairs intrinsic BOLD fluctuations in dorsal attention network in adolescents and young adults with borderline personality symptoms. *bioRxiv*, 2022.2008. 2008.503183 (2022).
57. Majerus, S., Péters, F., Bouffier, M., Cowan, N. & Phillips, C. The dorsal attention network reflects both encoding load and top-down control during working memory. *Journal of Cognitive Neuroscience* **30**, 144-159 (2018).
58. Machner, B., *et al.* Resting-state functional connectivity in the dorsal attention network relates to behavioral performance in spatial attention tasks and may show task-related adaptation. *Frontiers in human neuroscience* **15**, 757128 (2022).
59. Depping, M.S., Schmitgen, M.M., Kubera, K.M. & Wolf, R.C. Cerebellar contributions to major depression. *Frontiers in psychiatry* **9**, 634 (2018).
60. Tong, X., *et al.* Optimizing Antidepressant Efficacy: Multimodal Neuroimaging Biomarkers for Prediction of Treatment Response. *medRxiv* (2024).

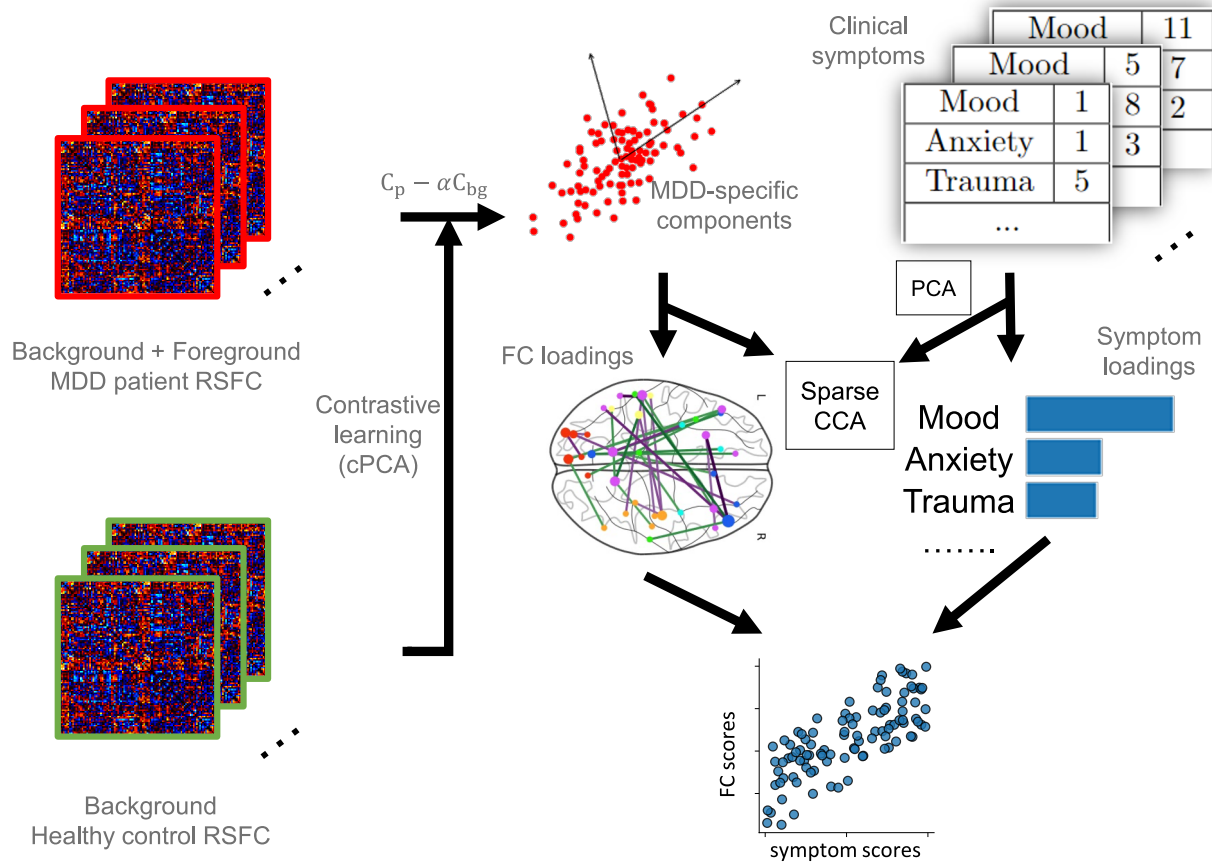


Figure 1 Illustration of the proposed framework. MDD patient RSFC data is considered to include both disorder-specific information (foreground) and shared noise with healthy control data (background). Contrastive learning is used for removing the background noise from patient data to generate MDD-specific components. Specifically, the covariance matrix of patient RSFC (C_p) and healthy control RSFC (C_{bg}) are computed to represent the group variances. MDD-specific components are derived from the updated covariance matrix $C_p - \alpha C_{bg}$, which represents the foreground variance, and α represents a hyperparameter quantifying the degree of contrast. These components are used as MDD-specific RSFC features in the next step. Subsequently, sparse CCA is performed to identify the linked dimensions between neurophysiology and PCA-transformed symptoms, represented by FC and symptom loadings. These loadings serve as linear combinations which project the RSFC features and symptoms into a common space with high correlations.

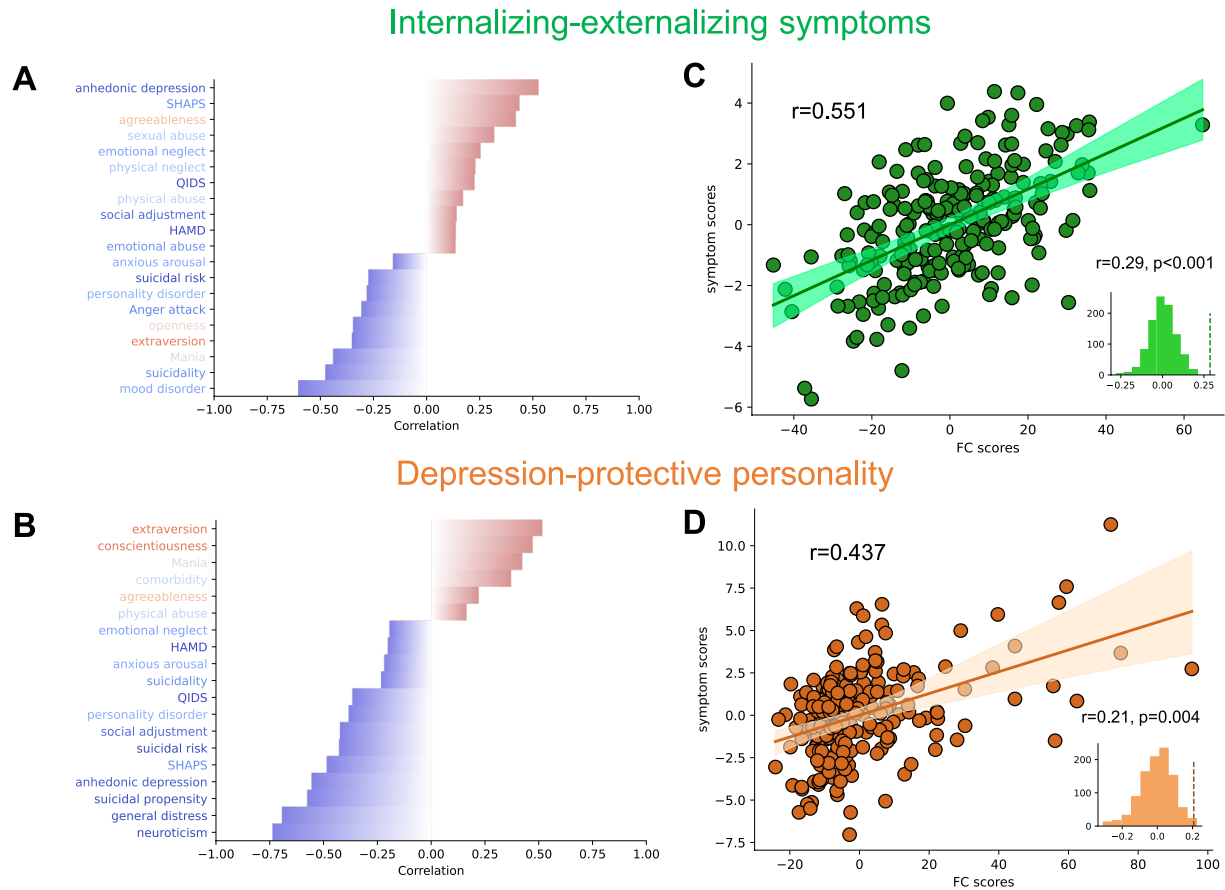
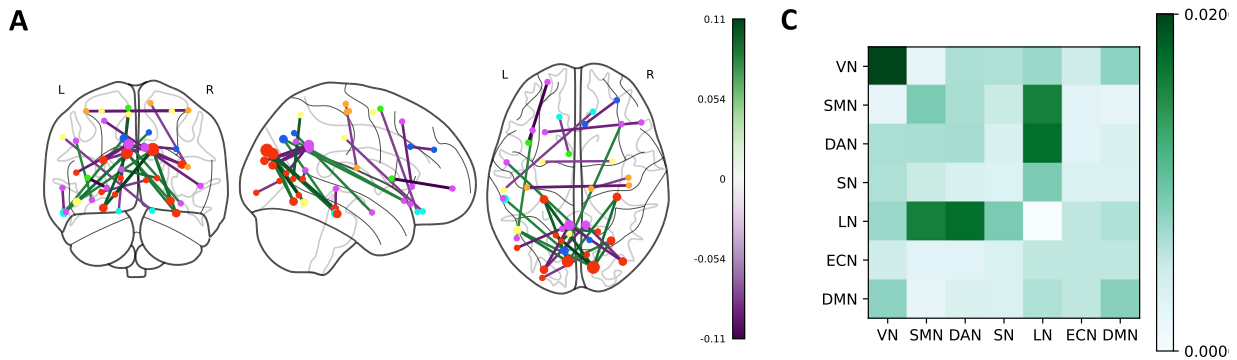


Figure 2 Identified linked dimensions between fMRI RSFC features and clinical measurements. (A) Top correlations between the symptom score of internalizing-externalizing symptoms dimension and 25 clinical measures. Leading negative correlations include Mood Disorder Questionnaire, Concise Associated Symptoms Tracking, and Altman Self-Rating Mania Scale. Anhedonic depression of Mood and Anxiety Symptoms Questionnaire is the leading positive correlation. The text color of the clinical measures represents the direction of the differences between healthy controls and MDD patients. Red indicates that healthy controls have higher average scores, while blue indicates that patients have higher average scores. (B) Top correlations of depression-protective personality dimension. NEO-Five Factor Inventory subscales (extraversion, conscientiousness, neuroticism) contribute the most. (C, D) Correlation between FC and symptom scores of the two linked dimensions on all patients. Cross-validation permutation results of 1000 repeats are shown in the small panels.

Internalizing-externalizing symptoms



Depression-protective personality

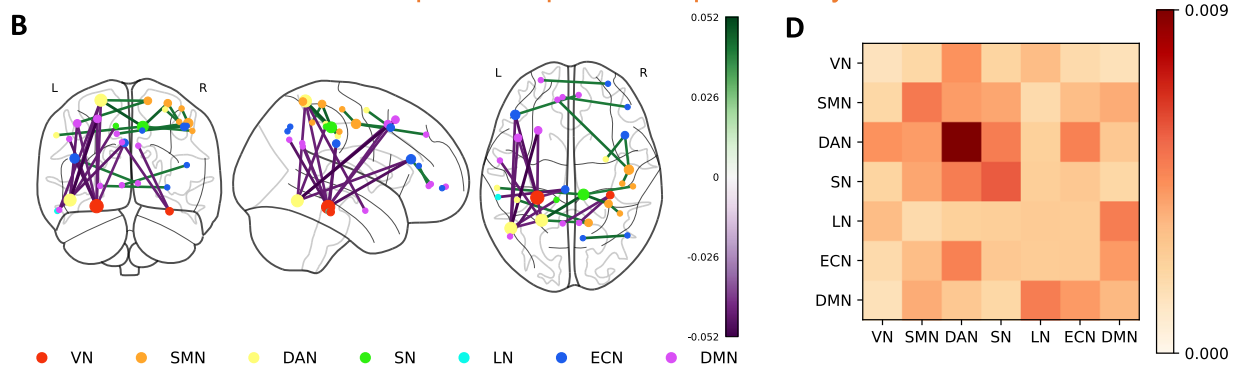
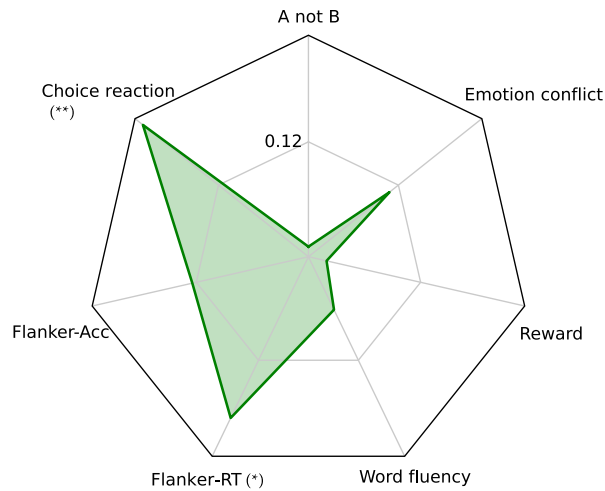


Figure 3 RSFC loadings and network-level importance of internalizing-externalizing symptoms dimension (A,C) and depression-protective personality dimension (B,D). (A, B) Top 15 positive and negative loadings of FC dimensions. Node size indicates the involvement of ROI in top loadings. (C, D) Network-level importance within two FC dimensions among 7 brain networks: visual network, sensorimotor network, dorsal attention network, salient network, limbic network, executive control network and default mode network. The values of network-level importance are calculated by first keeping only the top 10% loadings with highest absolute values and setting the rest to 0, and averaging all absolute loadings between every pair of networks.

Internalizing-externalizing symptoms



Depression-protective personality

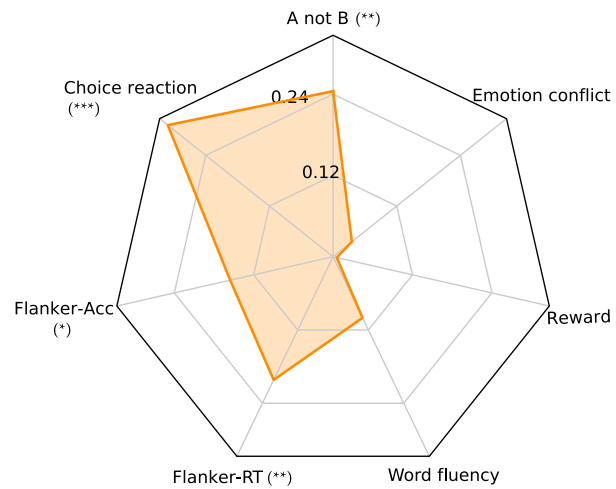


Figure 4 Pearson correlation between two FC dimensions and cognitive task performances. Left: internalizing-externalizing symptoms dimension; Right: depression-protective personality dimension. Cognitive task items include: 1) the overall accuracy of the A not B task, 2) the overall reaction time of the choice reaction time task, 3) the accuracy difference between incongruent and congruent trials in the flanker task, 4) the reaction time difference between incongruent and congruent trials in the flanker task, 5) the total valid number of words in the word fluency task, 6) the overall response bias in the probabilistic reward task, and 7) the reaction time difference between incongruent-following and congruent-following incongruent trials in the emotion conflict task. *** $p_{\text{fdr}} < 0.001$, ** $p_{\text{fdr}} < 0.01$, * $p_{\text{fdr}} < 0.05$.

Supplementary Materials

Table S1. Distribution of demographic information and clinical measurements of EMBARC MDD patients included in our study (N=233).

Item	Sub-scales	Mean±Std	Range
Demographics	Age	36.3±18.5	18-65
	Gender	67.4% Female	
HAMD (Hamilton Depression Rating Scale)	17-item total score	18.4±4.2	7-32
ASRM (Altman Self-Rating Mania Scale)	/	1.5±1.9	0-11
AAQ (Anger Attack Questionnaire)	/	0.4±0.5	0-1
CTQ (Childhood Trauma Questionnaire)	emotional abuse	13.1±5.6	5-25
	emotional neglect	13.5±4.9	5-24
	physical abuse	8.3±4.3	5-25
	physical neglect	8.3±3.5	5-20
	sexual abuse	8.3±5.6	5-25
CAST (Concise Associated Symptoms Tracking)	/	30.3±9.2	7-57
CH RTP (Concise Health Risk Tracking-Self Report)	propensity	26.4±8.1	3-49
	risk	5.2±2.4	0-10
MASQ (Mood and Anxiety Symptoms Questionnaire)	anxious arousal	17.4±5.4	10-35
	anhedonic depression	43.7±5.3	16-50
	general distress	32.1±7.9	13-50
MDQ (Mood Disorder Questionnaire)	/	4.1±3.0	0-13
NEO-FFI (NEO-Five Factor Inventory)	agreeableness	32.0±7.0	7-47
	conscientiousness	23.6±8.3	3-44
	extraversion	19.4±7.6	4-43
	neuroticism	35.2±6.3	15-48
	openness	31.8±7.5	14-48
QIDS (Quick Inventory of Depressive Symptomatology)	/	16.8±2.8	10-26
SCQ (Self-Administered Comorbidity Questionnaire)	/	1.9±2.7	0-17
SHAPS (Snaith–Hamilton Pleasure Scale)	continuous score	33.5±5.4	18-51
SAS (Social Adjustment Scale)	overall mean score	2.6±0.6	1.1-4
SAPAS (Standardised Assessment of Personality – Abbreviated Scale)	/	3.9±1.4	1-8

Table S2. Demographic information of healthy controls included in our study.

	UCLA CNP (N=83)	LEMON (N=103)	AOMIC (N=72)	EMBARC (N=27)	Total (N=285)
Age	35.31±8.26	45.22±20.72	21.96±1.86	42.75±12.84	36.73±16.48
Female %	43.3	33.0	56.9	60.8	46

Table S3. Cognitive and emotional tasks included in our study and their measurements.

Task	Item	Measurements
A not B task	Overall accuracy	Working memory
Choice reaction time task	Overall reaction time	Interference and post-error adjustments
Flanker task	Accuracy difference between incongruent and congruent trials	Cognitive control, conflict monitoring
	Reaction time difference between incongruent and congruent trials	Cognitive flexibility, conflict processing speed
Word fluency task	Total valid number of words	Psychomotor slowing
Probabilistic reward task	Overall response bias	Reward learning
Emotion conflict task	Reaction time difference between incongruent-following and congruent-following incongruent trials	Resolution and adjustment behavior in response to emotional conflict

Table S4. Chi-square test between the population distribution in high/low FC dimension score groups and remission rates within different treatment groups (sertraline and placebo). Remission is defined as HAM-D score ≤ 7 at week 8. For each dimension, high/low groups are divided by the median FC dimension score from all population. All p-values > 0.05 , indicating that there is no significant relation between high/low dimension scores and remission.

Internalizing-externalizing symptoms dimension						
Sertraline	Observed		Expected		Statistics	
	Remission	Non-remission	Remission	Non-remission	χ^2	1.20
High group	14	42	17.19	38.81	p	0.27
Low group	21	37	17.81	40.19	samples	114
Placebo	Observed		Expected		Statistics	
	Remission	Non-remission	Remission	Non-remission	χ^2	0.00
High group	16	45	15.89	45.11	p	1.00
Low group	15	43	15.11	42.89	samples	119
Depression-protective personality dimension						
Sertraline	Observed		Expected		Statistics	
	Remission	Non-remission	Remission	Non-remission	χ^2	0.71
High group	21	39	18.42	41.58	p	0.40
Low group	14	40	16.58	37.42	samples	114
Placebo	Observed		Expected		Statistics	
	Remission	Non-remission	Remission	Non-remission	χ^2	0.48
High group	17	40	14.85	42.15	p	0.49
Low group	14	48	16.15	45.85	samples	119

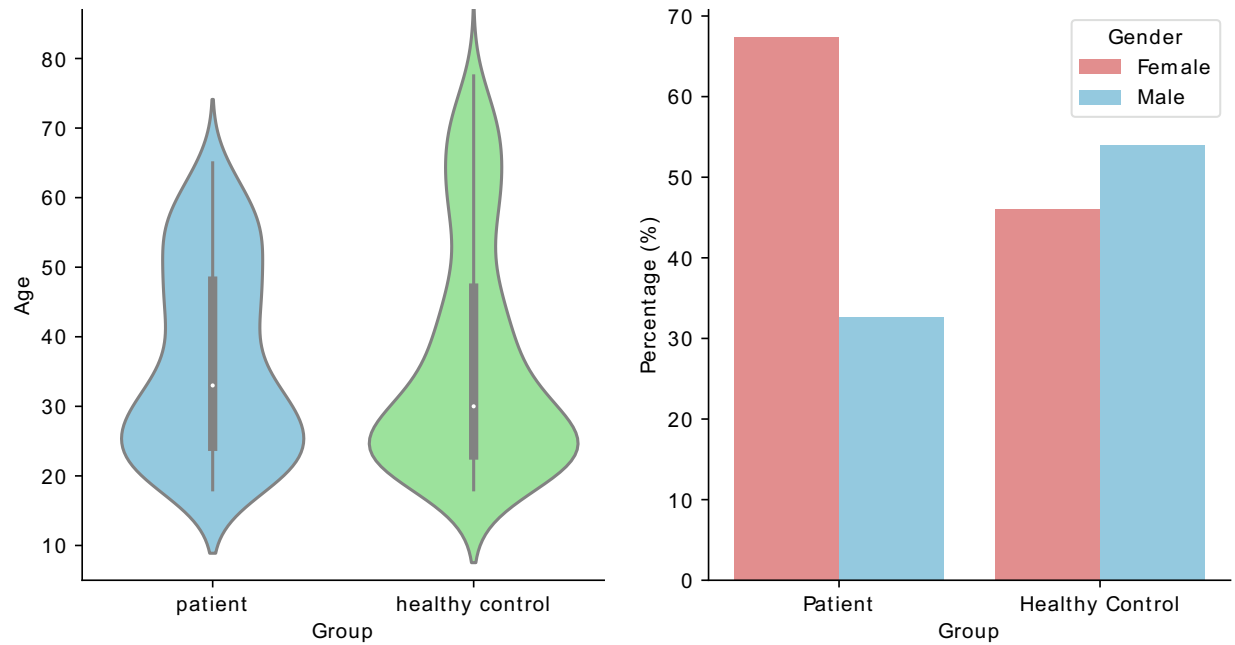


Figure S1. Demographic information comparison between patient (N=233) and healthy control groups (N=285). The age distribution of the patient population and the healthy population was matched, with a Kolmogorov-Smirnov test yielding a p-value of 0.052.

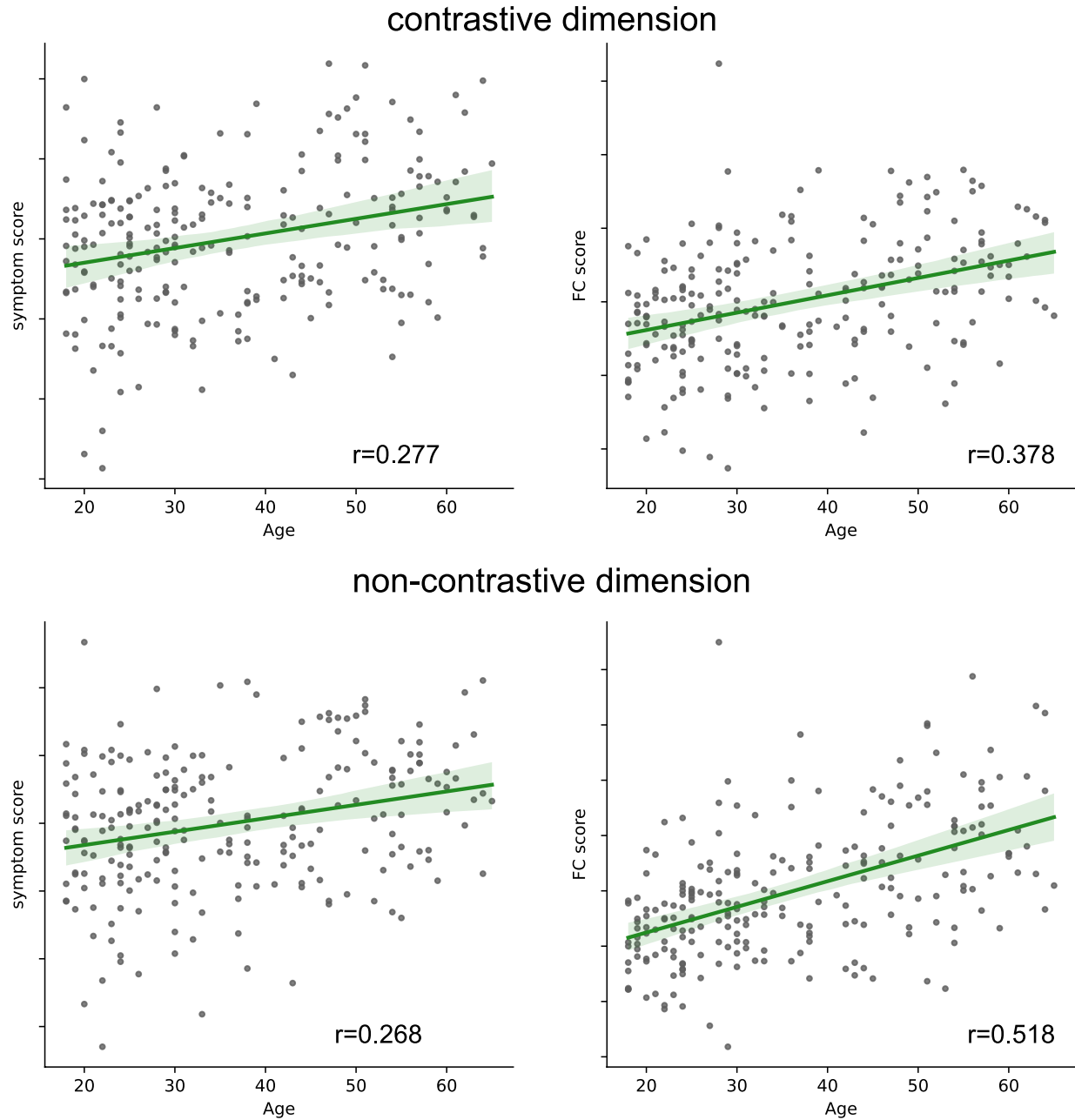


Figure S2. Comparison of the age-correlation of the best symptom/FC dimensions derived from cPCA and regular PCA components. Contrastive learning marginally reduced the age-related component in its FC dimension, with a Fisher's z test showing near-significant results ($p = 0.059$).

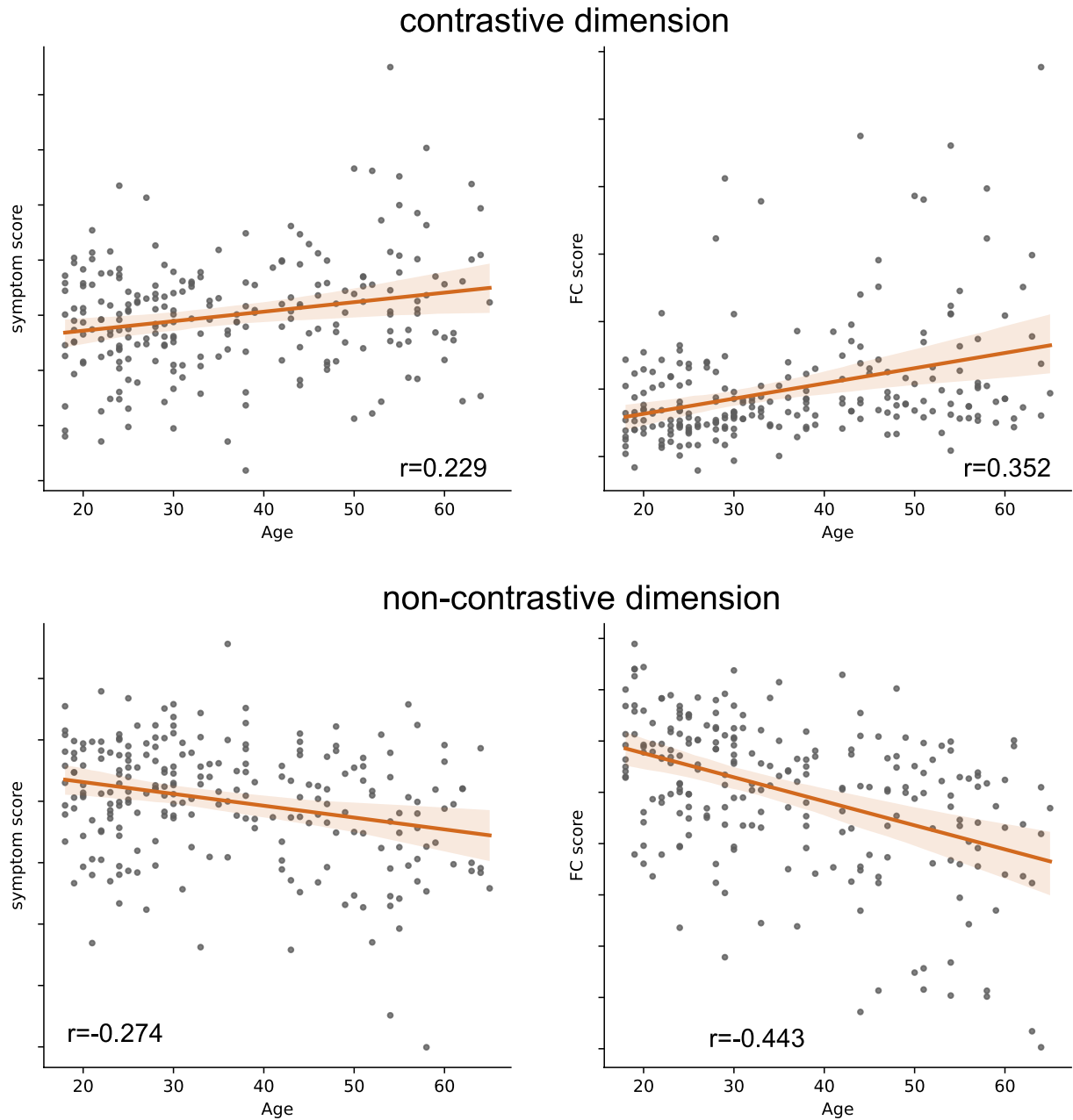


Figure S3. Comparison of the age correlation in the second-best symptom/FC dimensions derived from cPCA and regular PCA components. Contrastive learning reduced the age-related component in both symptom and FC dimensions, but the reduction was not statistically significant ($p > 0.1$).

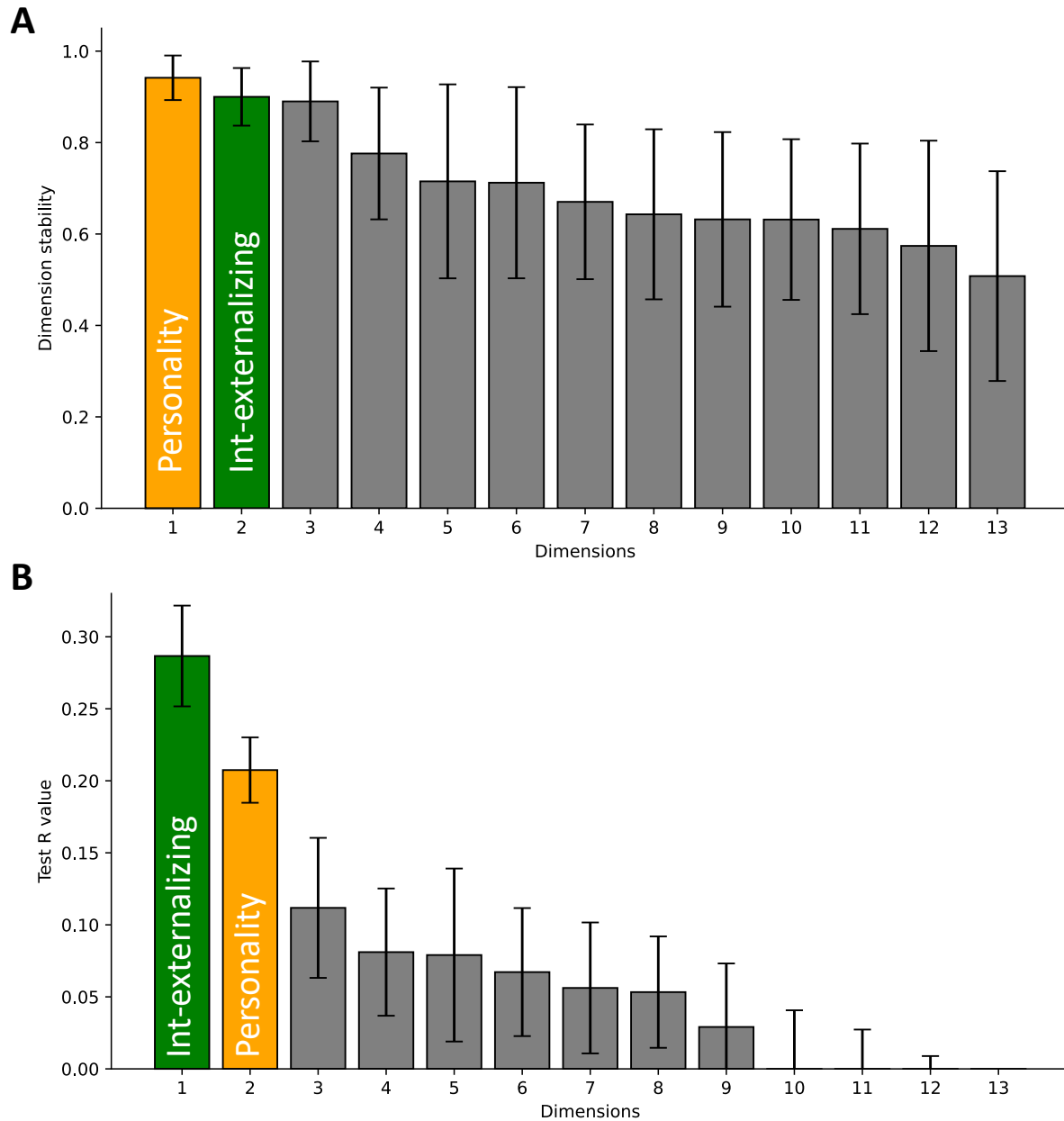


Figure S4. (A) Overall dimension stability for all 13 dimensions given by sCCA, measured by the average of the cosine similarity between each pair of FC dimensions and each pair of symptom dimensions from 10×10 folds. All dimensions have an overall stability over 0.5. (B) For the 13 dimensions, only two dimensions survived FDR correction in permutation test, shown in colors. The raw p-values are: $p < 0.001$ for internalizing-externalizing symptoms dimension, and $p = 0.001$ for depression-protective personality dimension.