Check for updates

ARTICLE   OPEN

Clinical response to neurofeedback in major depression relates to subtypes of whole-brain activation patterns during training

Masaya Misaki [1,2 ✉], Kymberly D. Young [3], Aki Tsuchiyagaito [1,2], Jonathan Savitz [1,2] and Salvador M. Guinjoan [1,2,4]

Major Depressive Disorder (MDD) poses a significant public health challenge due to its high prevalence and the substantial burden it places on individuals and healthcare systems. Real-time functional magnetic resonance imaging neurofeedback (rtfMRI-NF) shows promise as a treatment for this disorder, although its mechanisms of action remain unclear. This study investigated whole-brain response patterns during rtfMRI-NF training to explain interindividual variability in clinical efficacy in MDD. We analyzed data from 95 participants (67 active, 28 control) with MDD from previous rtfMRI-NF studies designed to increase left amygdala activation through positive autobiographical memory recall. Significant symptom reduction was observed in the active group ($t = -4.404$, $d = -0.704$, $p < 0.001$) but not in the control group ($t = -1.609$, $d = -0.430$, $p = 0.111$). However, left amygdala activation did not account for the variability in clinical efficacy. To elucidate the brain training process underlying the clinical effect, we examined whole-brain activation patterns during two critical phases of the neurofeedback procedure: activation during the self-regulation period, and transient responses to feedback signal presentations. Using a systematic process involving feature selection, manifold extraction, and clustering with cross-validation, we identified two subtypes of regulation activation and three subtypes of brain responses to feedback signals. These subtypes were significantly associated with the clinical effect (regulation subtype: $F = 8.735$, $p = 0.005$; feedback response subtype: $F = 5.326$, $p = 0.008$; subtypes' interaction: $F = 3.471$, $p = 0.039$). Subtypes associated with significant symptom reduction were characterized by selective increases in control regions, including lateral prefrontal areas, and decreases in regions associated with self-referential thinking, such as default mode areas. These findings suggest that large-scale brain activity during training is more critical for clinical efficacy than the level of activation in the neurofeedback target region itself. Tailoring neurofeedback training to incorporate these patterns could significantly enhance its therapeutic efficacy.

INTRODUCTION

Major Depressive Disorder (MDD) presents a significant public health challenge, with approximately one-third of diagnosed patients not responding to first-line treatments such as antidepressants and psychotherapy. This results in substantial disability and economic losses due to treatment costs and lost productivity [1, 2]. Real-time functional magnetic resonance imaging neurofeedback (rtfMRI-NF) has emerged as a promising alternative, demonstrating large to medium effect sizes in treating depressive symptoms [3–5]. This noninvasive brain modulation technique involves the real-time analysis and visualization of brain activation signals, thereby enabling participants to self-regulate their brain activity. Its efficacy in training participants to modulate their brain activation is well-supported by many studies, including several meta-analyses [3–11]. However, the direct impact of rtfMRI-NF on symptom relief is not yet fully understood due to incomplete knowledge of the neural mechanisms by which this training alleviates symptoms through the regulation of specific brain activations.

Previous studies on the mechanisms of neurofeedback (NF) training [12–15], including investigations into brain responses to feedback signals [15–19], have identified a broad spectrum of brain activities associated with NF-mediated self-regulation training. This training process is considered to include aspects of reinforcement learning, and two types of brain activation - evaluation of feedback values and modulation of brain activation - are common components of the reinforcement learning process [20]. Thus, to elucidate the learning mechanisms of NF-mediated brain regulation, investigating these two epochs is crucial. Active regions during these epochs typically include the prefrontal cortex, salience network, and reward processing areas [12–15]. However, while many studies have focused on the success of regulating target brain activities, the relationship between these activities and subsequent symptom relief remains elusive. Furthermore, interindividual variability in the clinical efficacy of NF necessitates further investigation to identify brain response subtypes associated with therapeutic outcomes.

This study aimed to characterize whole-brain activation patterns during rtfMRI-NF training in individuals with MDD, with the goal of identifying brain activation subtypes associated with interindividual variability in therapeutic efficacy. To this end, we analyzed a large dataset from rtfMRI-NF studies where participants with MDD
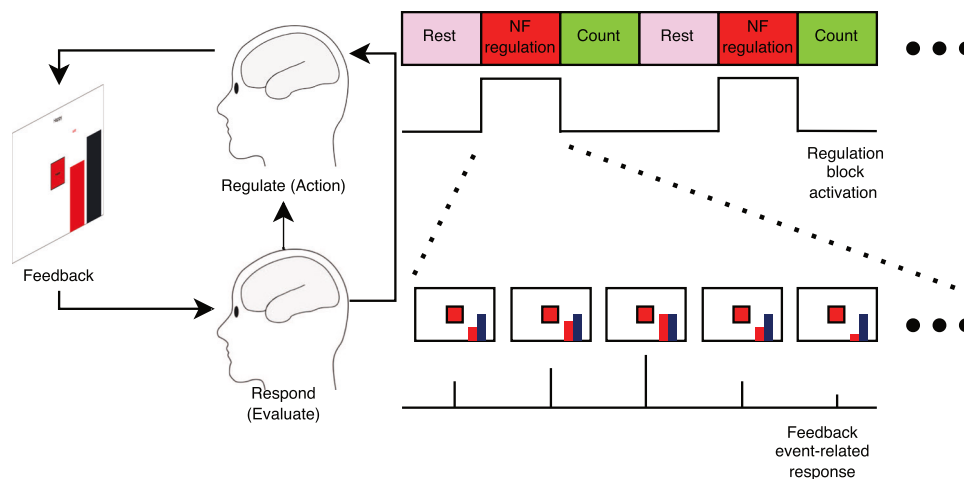
Fig. 1 Schematic diagram of the neurofeedback-based brain regulation training loop (left panel) and the models used to estimate brain activation at each epoch (right panel). The reinforcement learning process is typically characterized by the evaluation of a feedback signal followed by the adjustment of action (mental regulation) to increase the reward (feedback) signal. Brain activations corresponding to these two components were analyzed using a block-wise response model during the neurofeedback regulation blocks and an event-related response model for each neurofeedback presentation at every TR, modulated by feedback amplitude.

were trained to regulate left amygdala activity through NF [21–24]. These studies consistently observed significant reductions in depressive symptoms on average post-training, albeit with variations in therapeutic outcomes among participants.

We hypothesized that the observed variability in treatment efficacy could be explained by variations in whole-brain activation patterns during self-regulation training, extending beyond the NF target region (amygdala). The involvement of large-scale networks in NF training has been demonstrated in a meta-analysis of NF studies [12], and burgeoning evidence suggests that the effects of NF training may extend beyond the targeted brain region [25–27]. Specifically, we focused on two types of brain activation critical for NF training: regulatory task activity throughout the task block and instantaneous responses to NF signal presentations (Fig. 1). These two types of activation are thought to correspond to two critical components of the reinforcement learning process [20] and should characterize the training process for each individual.

METHODS
Participants
The present study was a secondary analysis of data from our previously published studies [21, 22, 24] and a preliminary study utilizing the same rtfMRI-NF protocols in individuals diagnosed with MDD. The University of Oklahoma Institutional Review Board (IRB) or the Western IRB reviewed and approved the study protocols, ensuring adherence to the ethical principles of the Declaration of Helsinki. Participants provided written informed consent prior to participation and were financially compensated. Participants met DSM-IV-TR [28] criteria for MDD based on the Structural Clinical Interview for DSM-IV disorders [29] or DSM-5 criteria for MDD based on the Mini-International Neuropsychiatric Interview (MINI) [30]. Previous articles [21, 22, 24] detailed each study's inclusion and exclusion criteria. Common inclusion criteria across studies are ages 18–65, current diagnosis of MDD, and common exclusion criteria are current diagnosis of PTSD, substance use disorder, bipolar disorder, active suicidal ideation or behavior within a year, a history of psychosis, pregnancy, and MRI contraindicators.

The present analysis included data from 95 participants with MDD (68 females; mean age ± SD = 33.6 ± 10.4 years), consisting of 67 in the active NF group (46 females) who received left amygdala neurofeedback and 28 in the control group (22 females) who received neurofeedback from a brain region not associated with emotional processing.

Real-time fMRI neurofeedback paradigm
The NF training was designed to enhance the activation signal in the left amygdala while recalling happy autobiographical memories [31]. The task

sequence utilized a blocked design consisting of alternating periods of rest, self-regulation with NF, and number counting, each lasting 40 s. This sequence was repeated four times in a single training run, with participants undergoing three such runs per session. Because the number of sessions differed among the studies, the present analyses were confined to data from the three runs of the initial session. Depressive symptom severity was estimated with the Montgomery-Åsberg Depression Rating Scale (MADRS) [32] immediately before the NF session, and one-week post-training. The raters of the MADRS were not informed of the experimental conditions of the participants.

MRI imaging in all experiments was conducted using the same 3 T Discovery MR750 scanner (GE Healthcare). Blood Oxygenation Level-Dependent (BOLD) fMRI data acquisition employed a T2*-weighted gradient-echo planar imaging (EPI) sequence, with parameters set as follows: TR/TE = 2000/30 ms, acquisition matrix = 96 × 96, field of view (FOV)/slice thickness = 240 mm/2.9 mm, flip angle = 90°, and 34 axial slices, using a SENSE acceleration factor of 2. EPI images were then reconstructed to a 128 × 128 matrix, resulting in an fMRI voxel size of 1.875 × 1.875 × 2.9 mm³. Anatomical reference was obtained using a T1-weighted magnetization-prepared rapid gradient-echo (MPRAGE) sequence.

A detailed description of the rtfMRI-NF procedure can be found in our previous publications [21, 22, 24]. Briefly, a custom in-house rtfMRI system was utilized for the experiments [31]. For the active group, the NF signal was extracted from the left amygdala, defined by a 7-mm diameter sphere centered at Talairach coordinates (x, y, z = −21, −5, −16 mm), and then mapped to each participant's brain space. For the control group, the NF signal was sourced from the horizontal segment of the intraparietal sulcus at Talairach coordinates (−42, −48, 48 mm), a region suggested to be unrelated to emotion regulation [33]. During the happy memory recall block, the NF signal was quantified as the percent signal change from the mean signal in the preceding rest block. The signal was updated every 2 s and visually presented to participants as a red bar.

MRI data processing
The Analysis of Functional NeuroImages (AFNI; http://afni.nimh.nih.gov) software suite was utilized for image processing. After discarding the first three volumes to achieve signal equilibrium, preprocessing steps were conducted, including despike, RETROICOR [34] along with respiratory volume per time (RVT) regression [35] for physiological noise correction, slice-timing and motion corrections, nonlinear warping to the MNI template brain with resampling to 2 mm³ voxels using the Advanced Normalization Tools (ANTs; http://stnava.github.io/ANTs/), spatial smoothing with a 6 mm full-width at half maximum (FWHM) Gaussian kernel, and scaling of signals to percent change relative to the voxel-wise mean.

Activation during the self-regulation was assessed using a general linear model (GLM) analysis. We extracted the two types of brain response in this process (Fig. 1). The GLM regressors included response models for the
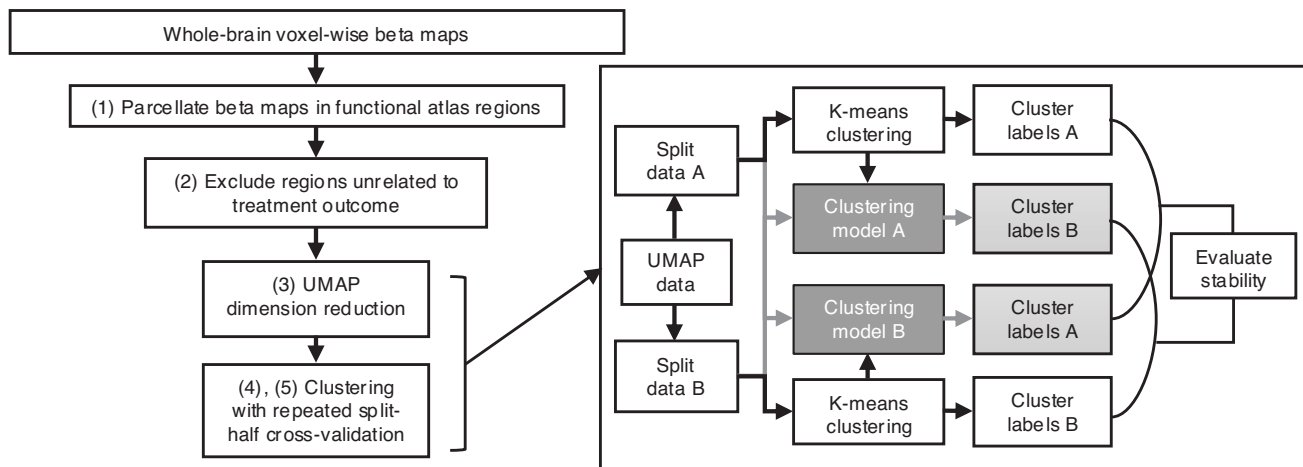
Fig. 2 Procedures of subtyping whole-brain beta maps. The flowchart illustrates the steps taken to identify subtypes of brain activation associated with treatment outcomes.

regulation and counting blocks, each modeled with a boxcar function convolved with the canonical hemodynamic response function (HRF), with the rest period serving as the baseline. The regressor for the regulation block was used for estimating the first type of brain activation, regulation block activation (Fig. 1). Another type of brain response, the feedback event-related response (Fig. 1), was evaluated using additional event-related regressors for each feedback presentation. The response was modeled as a delta function, modulated by the feedback amplitude normalized in each run, and convolved with the hemodynamic response function (HRF). Since the feedback signal was presented during the regulation block and this event regressor could be collinear with the regulation block regressor, we orthogonalized the feedback-response-event regressor with respect to the regulation block regressor. The GLM regressors also included twelve motion parameters (three rotations, three translations, and their temporal derivatives), three principal components of ventricle signals, local white matter signals (ANATICOR) [36], and an event-related regressor for the onset of any condition block (modeled as a delta function convolved with HRF) as nuisance covariates. Volumes with frame-wise displacement greater than 0.3 and their preceding volume were censored in the GLM. The beta coefficients from the general linear model (GLM) analysis were used as response estimates.

These response estimates were assessed for each block independently using the least squares - separate (LS-S) approach [37], in which separate regressors for the target block and all other blocks were included to estimate the response in a block, and repeated for each block in separate GLM analyses. The use of this block-wise response facilitates subsequent group analysis using linear mixed-effect model and provides the estimates with better test-retest reliability [38, 39].

Clustering analysis

Clustering analysis was conducted on the whole-brain beta maps solely for the active group participants to categorize their brain activation patterns. This approach was chosen because including the control group could introduce brain activation patterns with various unspecific effects. Such inclusion might increase the dimensionality of latent subtypes and complicate the extraction of subtypes relevant to the active NF training process.

A significant challenge posed by this analysis is the high dimensionality of the whole-brain beta maps; the problem often referred to as the "curse of dimensionality" [40]. This phenomenon refers to the issues arising in high-dimensional spaces, where distances between points become uniformly large, data points are sparsely distributed, and there is a high risk of overfitting, leading to clustering solutions that are difficult to reproduce.

To address this challenge, we implemented a multi-step strategy to reduce dimensionality and identify an informative space associated with treatment outcomes. Additionally, we employed cross-validation to ensure the clustering solution's stability and reproducibility. Figure 2 illustrates the flowchart of the analysis performed to delineate subtypes of brain activation. Our approach involved (1) aggregating voxel-wise responses into regional averages using a functional brain atlas, (2) removing the regions irrelevant to treatment outcomes, (3) applying Uniform Manifold Approximation and Projection (UMAP) [41] to extract a low-dimensional representation, (4) applying k-means clustering in the UMAP space, and (5) employing repeated cross-validation to assess the robustness of the clustering solution [42].

Initially, voxel-wise beta values were averaged within each functional region defined by the Shen 268 atlas [43, 44]. This atlas was chosen for its demonstrated performance in various predictive modeling studies [43, 45–49]. The data for each participant were averaged either across the three runs or for each individual run. We tested both averaging approaches (across runs and run-wise) as a hyperparameter to determine which yielded the most stable clustering solution. Subsequently, we identified regions correlating with changes in the MADRS score (change relative to the baseline score expressed as a ratio) after adjusting for the effects of age and sex. A threshold of $p < 0.1$ was used for this selection, prioritizing dimensionality reduction rather than finding significantly related regions. After excluding the regions irrelevant to the treatment outcome, we utilized UMAP for dimensionality reduction, followed by k-means clustering.

The UMAP parameters were set to optimize the clustered distribution within the reduced-dimensional space, with 'n_neighbors' (the number of neighboring points used in the local manifold approximation) set to 50, and 'min_dist' (the minimum distance apart that points are allowed to be in the low-dimensional representation) set to 0. A larger 'n_neighbors' value allows UMAP to focus on global structure rather than local structure, preventing the expansion of within-cluster local distributions and maintaining a clear cluster structure. Meanwhile, a low 'min_dist' allows sample points to agglomerate in the extracted space, further encouraging the formation of distinct clusters. Subsequently, k-means clustering was executed within the UMAP-defined space. UMAP's dimensionality reduction and clustering were treated as a unified process to search for a low-dimensional space that produced the most stable clusters. The search was conducted across the hyperparameter space of response summary (average across runs or a sequence of three runs), UMAP random seed (50 values), UMAP dimension (ranging from 2 to 50), and the number of clusters (ranging from 2 to 5).

The stability of the clusters was rigorously evaluated through cross-validation, facilitated by the 'reval' Python package [42]. This involved dividing the dataset into two halves, applying clustering to one half, and then applying the derived clustering rule to the opposite half, and vice versa, to assess the consistency of the cluster labels between the two rules. This procedure was repeated 10 times, each with a unique data split, to calculate the average stability score. The stability measure is the mean normalized Hamming distance between the two solutions, ranging from 0 to 1. A smaller value indicates a more stable and reproducible solution. Since this measure is larger with a larger number of clusters, it was scaled by the stability of random labeling of the same number of clusters [50].

In summary, these procedures treated UMAP dimensionality reduction and clustering as a single entity to search for the most reproducible clustering solution, ensuring that the results were not driven by random processes.
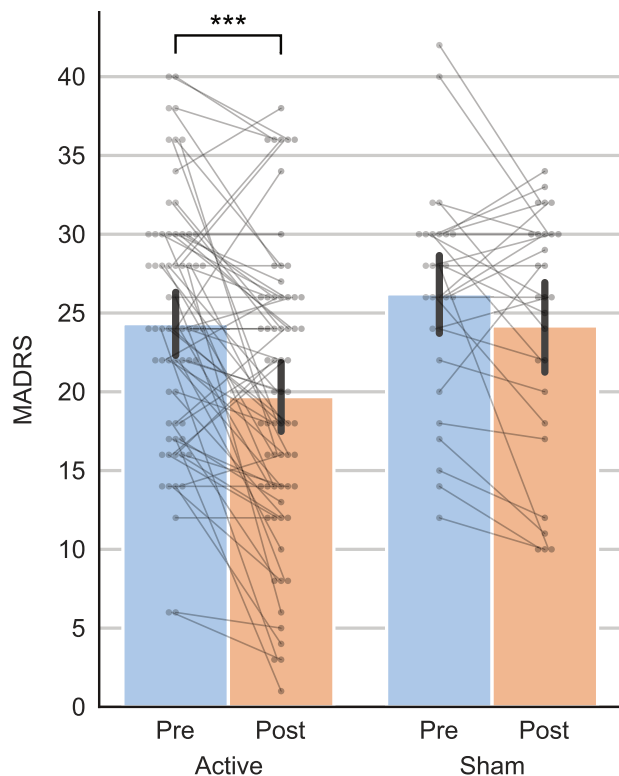
Fig. 3 MADRS scores before and after NF training in the active and control groups. The bars show the mean MADRS scores for each group, with the error bars representing their standard error. Individual participant scores are shown as dots, with lines connecting pre- and post-training scores to highlight changes for each individual. A statistically significant decrease in post-training MADRS scores was observed in the active group (***$p < 0.001$), but not in the control group ($p = 0.111$). However, the time by group interaction was not statistically significant ($p = 0.090$).

Mapping the brain responses in each subtype
After identifying the subtypes in the clustering analysis, we evaluated the response patterns for each subtype using a linear mixed-effect (LME) model analysis performed voxel-wise with the 'lme4' package [51] in R language and environment for statistical computing. This LME analysis was applied to the beta values, incorporating fixed effects for the run, the identified subtype from the clustering analysis, age, and sex, along with a random effect for participants at the intercept level. The mean response for each subtype was calculated using the 'emmeans' package [52] in R. The mean maps of the subtypes were thresholded at a voxel-wise $p < 0.001$, corrected for cluster size with a $p < 0.05$ using AFNI 3dClustSim.

Statistical testing
LME analysis was performed to investigate changes in MADRS scores one week after a NF session for both active and control groups. The LME model included fixed effects for time (pre-/post-NF), group (active/control), age, and sex, as well as a random effect for participants at the intercept. We also examined whether treatment efficacy was correlated with NF training performance measures, including the mean NF amplitude, mean left amygdala response during the regulation block, and the change in left amygdala response between the first and the last training runs using linear model analysis.

We then examined whether the identified subtypes of brain activation during NF training were associated with demographic and training performance variables, as well as the depressive symptom score (MADRS) at the baseline and its change post-NF. Each variable was examined using linear model analysis, with subtypes serving as independent variables. The post-hoc analysis for the mean response for each subtype and the difference between the subtypes was performed using the 'emmeans' package [52], and p-values for the post-hoc comparisons were corrected by multivariate t adjustment.

In addition, the subtyping rules established in the active group were applied to the control group to assess if similar patterns in MADRS score changes could be observed across the subtypes. The presence of consistent subtype associations with symptom reduction in both the active and control groups would suggest the operation of a common treatment mechanism. This could indicate that the efficacy of NF training is not specific to the neurofeedback signal from an emotion-related region.

RESULTS
Data selection
Post-training MADRS scores were not available for two participants in the active group. fMRI data from five participants were excluded from the analysis due to problems with physiological signal acquisition (two from the active group and one from the control group) or excessive head motion (two from the active group) with more than 30% of time points censored (Framewise Displacement [FD] >0.3) in all training runs. Additionally, four participants in the active group who exhibited excessive head motion in any one of the training runs were excluded only when analyzing data with a series of three training runs. Analyses were performed using the largest sample size available for each test measure, regardless of missing data on other variables (see Supplementary Section 1). No significant differences in age, sex, and baseline MADRS scores were observed between the active and control groups in any of these selected datasets.

Reduction of depressive symptoms following NF training
Figure 3 shows the MADRS scores before and after NF training for each group, with average scores represented by the height of bars and individual participant scores depicted as points connected by lines. LME analysis identified a significant main effect of time. Post-hoc analysis revealed a significant decrease in post-session scores ($t = -4.404$, $d = -0.704$, $p < 0.001$). Supplementary Table S2a provides the ANOVA tables for the LME analysis. While the time by group interaction was not statistically significant ($p = 0.090$), the active group showed a significant symptom reduction ($t = -5.576$, $d = -0.978$, $p < 0.001$) but not the control group ($t = -1.609$, $d = -0.430$, $p = 0.111$).

Left amygdala activation increased during the regulation block relative to rest in the active group ($t = 2.842$, corrected $p = 0.011$), but not in the control group ($t = -0.032$, corrected $p = 0.999$) (Supplementary Fig. S2). However, no significant main effect of group ($F = 2.870$, $p = 0.094$) or interaction between group and run ($F = 0.597$, $p = 0.551$) was observed. Furthermore, within the active group, no significant associations were found between changes in MADRS scores and measures of NF training performance, including mean NF amplitude, average left amygdala response during regulation blocks, or changes in left amygdala response from the first to the last training runs (see Supplementary Section 2 for comprehensive details).

Clustering results
In clustering of activation maps related to the regulation block, the optimal stability score of 0.217 was obtained using the following hyperparameters: averaging responses across runs for each participant, extracting a 40-dimensional UMAP space, and clustering into two subtypes, which included 32 and 31 participants respectively. Similarly, in the clustering of activation maps associated with the feedback event, an optimal stability score of 0.229 was reached by analyzing a series of responses across three runs, extracting a 14-dimensional UMAP space, and forming three subtypes consisting of 23, 15, and 22 participants each. No significant association was found between the subtypes identified in the regulation block and those in the feedback event ($\chi^2 = 3.143$, $p = 0.208$).

Figure 4 shows t-maps of the mean beta values for the subtypes of regulation block activations (REG subtypes), with detailed peak
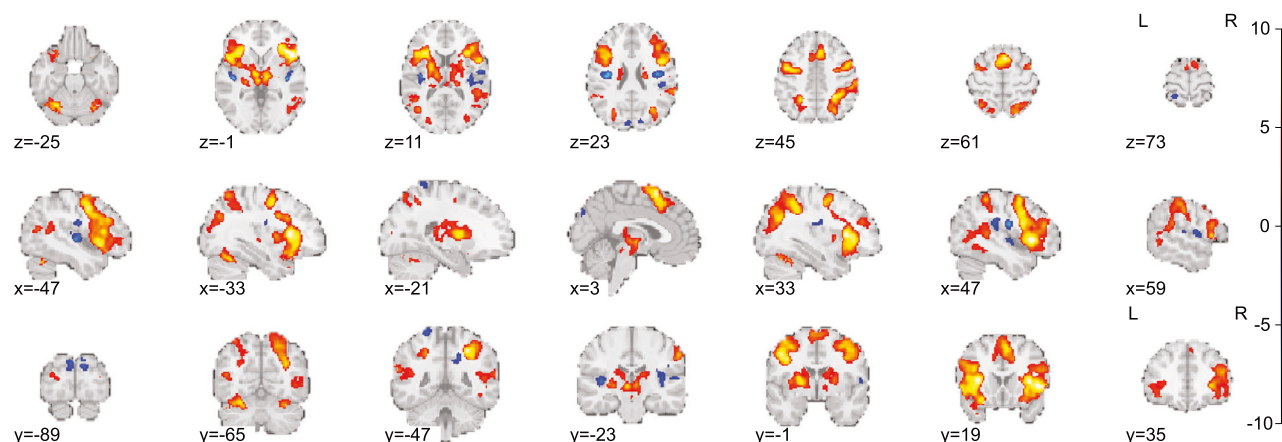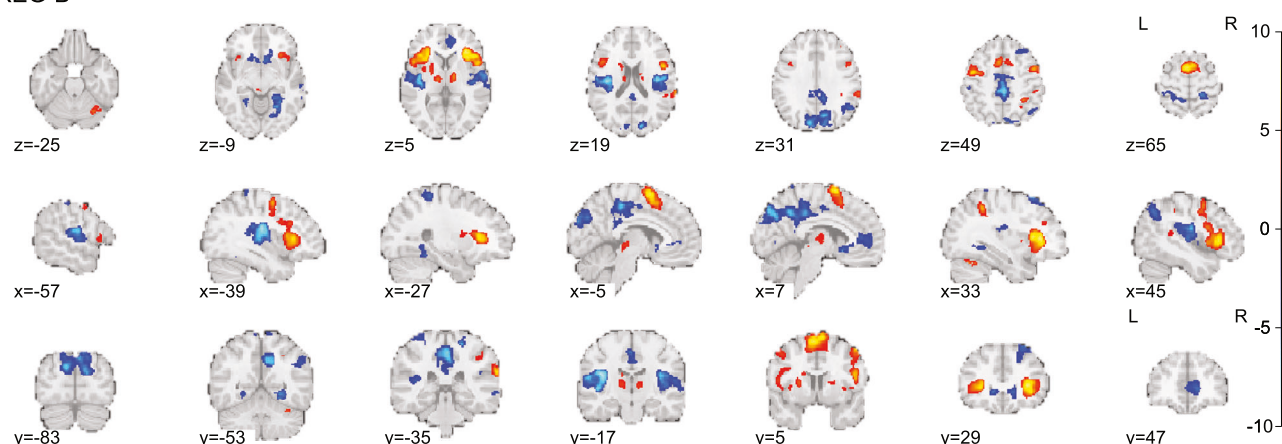
Fig. 4 t maps of mean activation in the regulation block for the REG subtypes. The maps are thresholded at a voxel-wise $p < 0.001$, with cluster-size correction at $p < 0.05$.

coordinates available in Supplementary Section 3, Supplementary Table S3. The first subtype, REG-A, exhibited increased BOLD signals in regions commonly identified in NF studies [12], including the lateral prefrontal cortex, superior and inferior parietal lobules, supplementary motor area (SMA), anterior insula, thalamus, and cerebellum. In contrast, decreased BOLD signals were observed in Heschl's gyrus, posterior insula, middle cingulate cortex, and cuneus. The second subtype, REG-B, demonstrated increased BOLD signals in the lateral prefrontal cortex, SMA, and anterior insula, with decreased activity in default mode network (DMN) areas such as the precuneus, posterior cingulate cortex, mediodorsal prefrontal cortex, Rolandic operculum to posterior insula, and fusiform gyrus.

Figure 5 presents t-maps of the mean beta values for subtypes of the brain responses to the feedback event (FBR subtypes), with peak coordinates detailed in Supplementary Section 4, Supplementary Table S4. The first subtype, labeled FBR0, demonstrated a limited response, particularly in the Rolandic operculum and right supramarginal gyrus. The second subtype, labeled FBR−, was characterized by a negative association of BOLD signal changes in response to NF signals, affecting regions including the SMA, middle cingulate cortex, Heschl's gyrus, Rolandic operculum, and visual cortex. In contrast, the third subtype, labeled FBR+, displayed a positive association of BOLD signal changes in areas such as the precuneus, middle cingulate cortex, lateral prefrontal cortex, nucleus accumbens, cerebellum, and inferior occipital regions.

Subtype associations with demographics, NF-training characteristics, and the depressive symptom in the active group

Linear model analyses examining associations of the REG and FBR subtypes with age, mean NF signal amplitude, mean left amygdala activation across runs, and baseline MADRS scores revealed no significant differences between the subtypes. Similarly, χ^2 tests showed no significant associations of these subtypes with sex or study participation (refer to Supplementary Section 5 for further detailed statistics).

However, changes in MADRS scores, indicative of clinical efficacy, were significantly associated with these subtypes. Figure 6 shows the MADRS scores before and after NF training for each REG and FBR subtype. We tested multiple linear models, both with and without the REG and FBR factors, on MADRS change. The model that included both REG and FBR factors, along with their interaction term, explained the variability in MADRS change significantly better than the reduced models (refer to SI section 6). In this model, the main effects of both REG and FBR subtypes, as well as their interaction effect on changes in MADRS scores, were statistically significant (REG, $F = 8.735$, $p = 0.005$; FBR, $F = 5.326$, $p = 0.008$; interaction $F = 3.471$, $p = 0.039$). Post-hoc analysis revealed a significant decrease in MADRS scores within the REG-B subtype ($t = -6.077$, $d = -1.702$, $p < 0.001$), in contrast to REG-A ($t = -2.170$, $d = -0.608$, $p = 0.068$). For the FBR subtypes, significant decreases in MADRS scores were observed in both FBR− ($t = -4.995$, $d = -1.399$, $p < 0.001$) and FBR+ ($t = -3.818$,
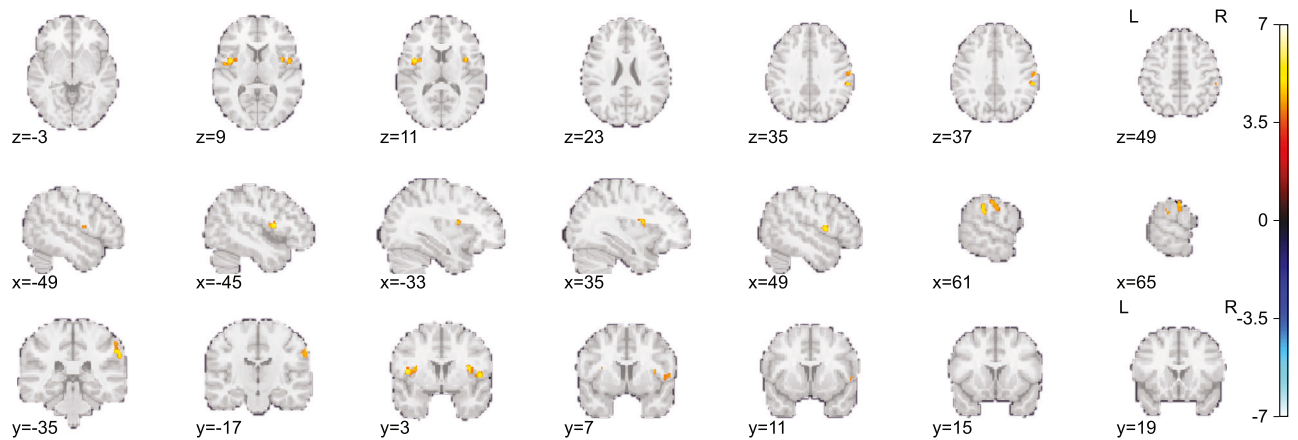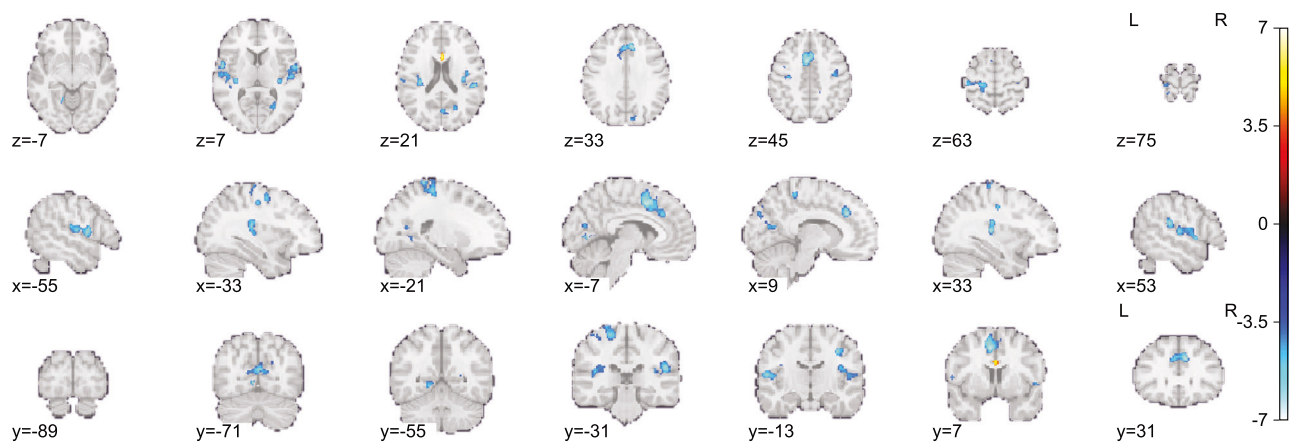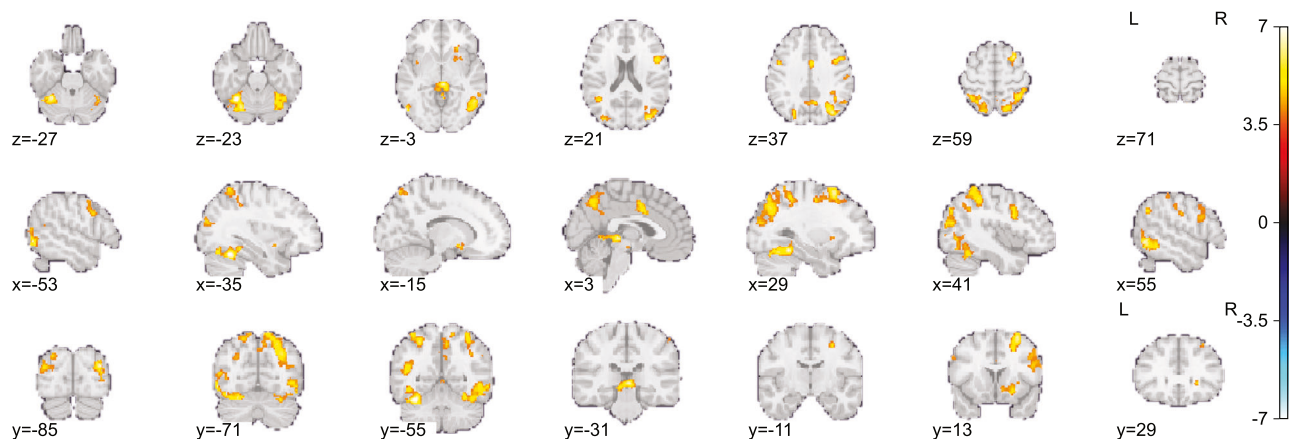
Fig. 5 t maps of mean activation in response to NF signals for the FBR subtypes. The maps are thresholded at a voxel-wise $p < 0.001$, with cluster-size correction at $p < 0.05$.

$d = -1.069$, $p = 0.001$), whereas the FBR0 subtype exhibited no significant change ($t = -0.839$, $d = -0.235$, $p = 0.786$).

Moreover, changes in MADRS scores varied by subtype combination, as illustrated in Fig. 6. A significant reduction in scores was noted when REG-A was paired with FBR- ($t = -3.327$, $d = -0.931$, $p = 0.003$). Within the REG-B group, no significant reduction was observed with the FBR0 subtype ($t = -1.221$,

$d = -0.342$, $p = 0.401$). However, significant reductions were seen when REG-B was paired with FBR- ($t = -3.789$, $d = -1.061$, $p < 0.001$) and FBR+ ($t = -6.018$, $d = -1.685$, $p < 0.001$).

In summary, participants in the active group classified as REG-A with FBR− ($d = -0.931$), REG-B with FBR− ($d = -1.061$), and REG-B with FBR+ ($d = -1.685$) experienced significant reductions in MADRS scores, with large effect sizes, one week after NF training.
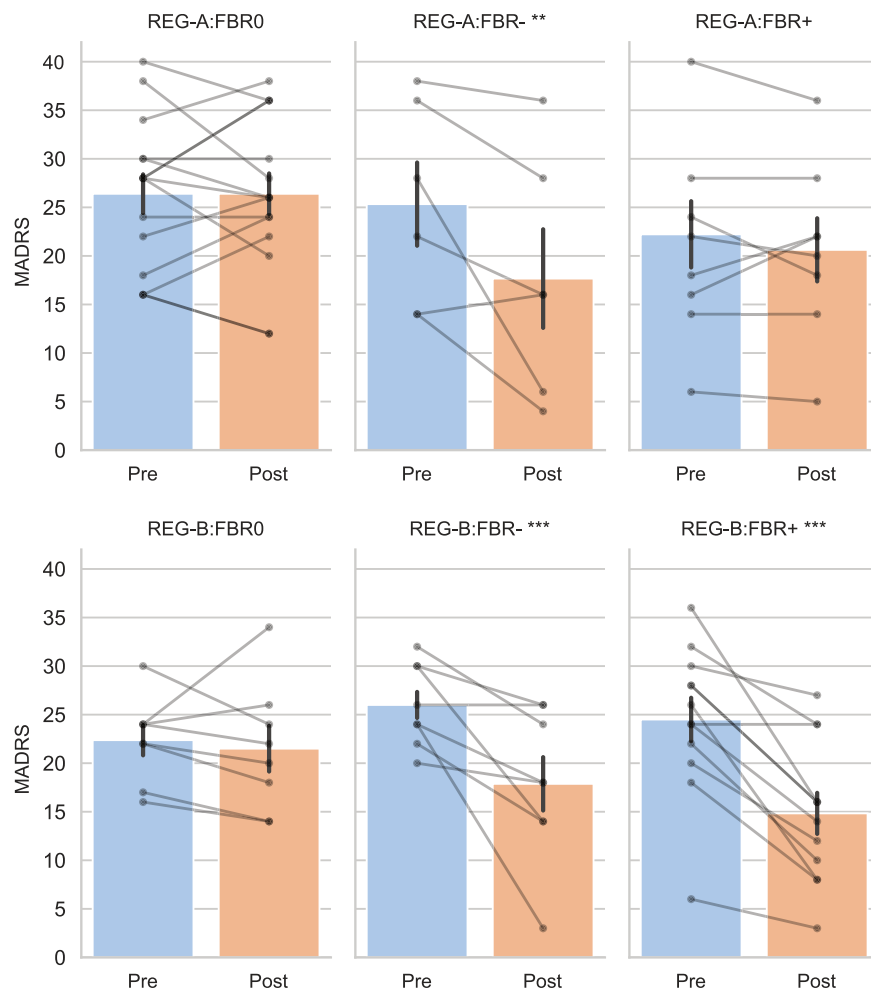
Fig. 6 MADRS scores before and after NF training for each REG and FBR subtype. Significant reductions were observed in REG-A:FBR− (**$p < 0.005$), REG-B:FBR− (***$p < 0.001$), and REG-A:FBR+ (***$p < 0.001$).

Application of clustering rules to the control group

When the clustering rules derived from the active group were applied to the control group, participants were classified as follows: 13 into REG-A and 14 into REG-B for the REG subtypes; and 11 into FBR0, 5 into FBR−, and 11 into FBR+ for the FBR subtypes. No significant association was observed between the REG and FBR subtypes within the control group ($\chi^2 = 0.345$, $p = 0.842$). Additionally, the distribution of participants across subtypes did not significantly differ from that observed in the active group for both REG ($\chi^2 = 0$, $p = 1.000$) and FBR ($\chi^2 = 0.526$, $p = 0.768$) subtypes.

In the control group, no significant associations were found between the subtypes and variables such as age, sex, study participation, mean NF signal amplitude, changes in left amygdala activation, or baseline MADRS scores, as detailed in SI Section 7. Contrary to the active group, there were no significant differences in MADRS score changes across subtypes within the control group.

DISCUSSION

The primary objective of this study was to identify subtypes of brain activation patterns during NF training that could explain interindividual differences in clinical response. Our analysis revealed that training characteristics within the target brain region, such as mean NF signal amplitude, mean left amygdala activation, and signal changes in the left amygdala, were not associated with changes in depressive symptoms. However, we

found significant associations between subtypes of whole-brain activation patterns during NF training and changes in MADRS scores. In contrast, the control group showed no significant associations between the subtypes and changes in MADRS scores, underscoring that the effects of NF treatment are distinct from placebo effects.

Significant symptom reduction was observed across multiple subtype combinations, indicating the existence of multiple brain functional pathways to successful treatment. Notably, individuals classified within the FBR0 subtype, characterized by their non-responsiveness to feedback signals, did not experience significant symptom reduction, highlighting the critical role of brain response to feedback signals in effective NF training. Significant symptom reduction was evident in individuals exhibiting brain activation correlated with feedback signals, either in a positive (FBR+) or negative (FBR-) manner. However, the efficacy of this response pattern was dependent on their regulation subtype. Individuals exhibiting increased activation across broad brain areas during self-regulation (REG-A), which overlap with areas commonly reported in NF studies [12], required a negative response to NF signals (FBR-) to achieve significant symptom reduction. Conversely, individuals in the REG-B subtype, who showed increased activation in executive control areas such as the lateral prefrontal cortex and SMA and salience network regions such as the anterior insula (areas that overlap with those involved in general skill learning and emotion regulation [53, 54]), alongside suppressive activation in DMN regions during self-regulation, achieved

symptom reduction both in a positive (FBR+) or negative (FBR-) responses to feedback signals. Common activation regions between REG-A and REG-B, including areas within the salience network, have been identified as essential for the emotion regulation process [55]. Additionally, common deactivation, including the posterior insula, may be associated with high task engagement, as task-related deactivation can extend to the posterior insula with increasing task demands [56]. These findings suggest that both groups made a concerted effort to regulate the left amygdala signal. However, the differences in therapeutic outcomes indicate that merely increasing emotion regulation effort is not sufficient for treatment efficacy. Instead, selective modulation of brain responses is crucial for the success of NF therapy.

Increased activation of the DMN, often associated with self-referential thoughts [57, 58], has been frequently observed in individuals with MDD compared to healthy controls, and is linked to repetitive negative thinking [59–67]. Particularly relevant to the pathophysiology of MDD and its treatment mechanisms could be the significant role of deactivating the posterior DMN, especially the posterior cingulate cortex (PCC). Current evidence suggests that the PCC acts as a major cortical hub for various self-referential phenomena, ranging from spatial body localization to self-talk and autobiographical information processing [68–71]. In this context, the characteristic deactivation of the PCC observed in brain activation subtypes associated with clinical response in our study supports the hypothesis that this region may underlie the persistent and intensified negative self-referential mentation that characterizes depression.

Among the subtype combinations with significant treatment effects, REG-B with FBR+ demonstrated the largest treatment effect ($d = -1.685$). Viewing NF training as a reinforcement learning process, maintaining a favorable brain state (REG-B) while strongly responding to positive rewards associated with it (FBR+) maximizes the therapeutic effect, consistent with reinforcement learning mechanisms. The involvement of reward-responsive areas, such as the nucleus accumbens, in the FBR+ response further supports this process. Additionally, a strong response to positive rewards could enhance the antidepressant effect on its own, which may have contributed to the large treatment effect observed in this subtype. Indeed, a study on the placebo antidepressant effect using NF demonstrated that the placebo effect is amplified by expectations of efficacy and positive (sham) feedback signals [72]. While this suggests that a strong reward response can enhance treatment, the present findings indicate that this response alone was insufficient for significant symptom reduction, as the REG-A with FBR+ group did not show a significant reduction in MADRS scores (Fig. 6).

The absence of suppressive activation during the regulation task can be compensated for by negative responses to feedback signals in the SMA, middle cingulate cortex (MCC), auditory cortex (including Heschl's gyrus and Rolandic operculum), and visual cortex. Although the functional mechanism of these negative responses to NF signals remains unclear, they may counteract maladaptive brain activation patterns associated with MDD. Notably, alterations in SMA activity, frequently reported in MDD [73–76], and changes in middle cingulate activation during reward anticipation have also been highlighted in a meta-analysis [77]. The association between less MCC activation during NF training and greater symptom reduction was also reported [26]. Furthermore, increased connectivity in the auditory cortex was associated with repetitive negative thinking [78], and heightened activity or connectivity in the visual cortex have also been reported in individuals with MDD [48, 79].

Selective increases and decreases in brain activation associated with successful NF training have also been documented in a meta-analysis of amygdala NF studies [4]. This study reported that successful amygdala modulation was linked to deactivation in the posterior insula and DMN regions, including the ventromedial prefrontal cortex and PCC, during the regulation period. Additionally, negative activations in successful modulators were noted in the left parahippocampal gyrus and cuneus, aligning with the present results observed in the FBR- subtype. Notably, this meta-analysis [4] defined training success as the ability to down-regulate amygdala activation, in contrast to our study, which focused on up-regulating amygdala activation and evaluated training success by changes in depressive symptoms. This suggests a common underlying mechanism in emotion regulation training that leads to depressive symptom reduction, despite varying approaches to NF signal definition and operational definitions of success.

In light of our findings, the therapeutic effects of NF targeting the left amygdala cannot be solely attributed to changes in its activity alone. This observation aligns with previous research showing that activations in brain regions beyond the intended target during NF training can mediate symptom reduction [26]. Such evidence supports the idea that brain self-regulation through NF involves a large-scale network, particularly the activation of externally directed systems and the deactivation of internally directed systems, extending beyond a single target region.

This raises two critical questions regarding the efficacy of the NF signal. First, did the feedback from the left amygdala specifically contribute to the therapeutic effect? The results suggest that it did, as the control group, undergoing the same procedure but with a feedback signal unrelated to emotional state, showed a smaller treatment effect than the experimental group. This indicates that NF from the left amygdala contributed to the enhanced therapeutic outcome. However, the interaction between groups and pre- and post-training outcomes ($p = 0.090$, Supplementary Table S2a), as well as the group difference ($p = 0.089$, Supplementary Table S2a), on the MADRS were marginal. This is likely due to the moderate ($d = -0.430$), yet not statistically significant ($p = 0.111$), effect seen in the control group, suggesting that some therapeutic benefit arises from the experimental procedure and mental strategies used.

Nevertheless, the addition of NF, which provided real-time feedback on emotional arousal states via left amygdala activity, further enhanced the effect. Neurofeedback is unique in offering online, targeted feedback, allowing participants to refine their cognitive control in ways not typically achievable through other methods. This, we believe, is what makes NF such a powerful tool for promoting clinical improvement.

Another critical question is whether feedback from regions other than the left amygdala could be more effective in treating depressive symptoms. The efficacy of the treatment was not solely determined by the success of left amygdala regulation, but rather by how the regulation was achieved. We believe that amygdala activity is a reliable marker of emotional states and can indicate successful regulatory efforts, as demonstrated in many NF protocols [11]. Therefore, its feedback signal was valuable for participants to evaluate and refine their cognitive control strategies based on this feedback. However, our findings also highlight the importance of the regulation process itself in achieving treatment success and in explaining the variability in outcomes. Incorporating feedback that targets both regulation-related activities and responses to feedback within large-scale brain networks could potentially enhance treatment efficacy.

As NF methodologies continue to evolve, focusing on the process of regulating affective states rather than merely activating specific brain regions might yield more effective treatments. In this context, pattern-based NF approaches such as DecNef [80], connectome-based neurofeedback [81], and semantic neurofeedback [82, 83], which emphasize extracting feedback signals from multivariate brain activation patterns, offer a promising direction for refining training protocols. This shift toward whole-brain activity and network-based feedback could lead to more robust and individualized therapeutic outcomes.

This study's exploration of interindividual differences in brain activation during NF training and their association with therapeutic effects on depressive symptoms represents a significant step toward understanding the mechanisms of NF as a treatment for psychiatric disorders. The findings highlight the importance of whole-brain regulation processes over the simple modulation of amygdala activity, which could inform future improvements in NF protocols. However, it is important to recognize that these mechanisms may not be consistent across all NF training, as they likely vary depending on the target cognitive functions. For instance, in the context of unconscious perceptual enhancement through implicit training without a specific mental strategy, modulating neural activity in the target region via NF might directly influence performance [80]. Focusing on therapeutic mechanisms specific to individual NF strategies, rather than seeking universal rules or competing theories, may be more beneficial for advancing NF as a practical treatment tool.

The limitations of this study warrant careful consideration. The left amygdala signal in these studies was extracted using a 7 mm-diameter spherical ROI, which likely reduced the region specificity of the NF signal. This noisiness may have contributed to the training effects not being confined to the amygdala. While the REG and FBR responses were extracted separately, their regressors were naturally collinear because feedback only happened during the regulation block. To solve this, we made the FBR event regressors orthogonal to the REG block regressor. This means that some parts of the FBR response might have been included in the REG regressor, but only the parts that were not related to REG were isolated for the FBR activity pattern. Importantly, even after removing the influence of REG, the FBR still contributed significantly to explaining the variability in the treatment effect. This highlights the importance of considering the feedback response style in understanding how NF treatment works.

Although the clustering analysis was designed to classify individuals into distinct subtypes, it is possible that these subtypes do not represent distinct groups per se. Rather, the subtypes may be part of a continuous distribution with no clear demarcation between groups. This is suggested by the fact that the obtained optimal stability scores, which indicate the discrepancy between cross-validated solutions, were greater than 0.2, indicating potential ambiguity in cluster definition. In scenarios where the samples have a continuous distribution, splitting into an equal number of participants could emerge as a more stable approach. This phenomenon may explain why our analysis resulted in nearly equal numbers of participants being categorized into each subtype, highlighting that these identified subtypes may not delineate distinct groups but rather capture different facets of the continuous response spectrum. It is also noteworthy that subtype classification was not related to age, sex, or baseline severity of depressive symptoms. These suggest that the observed subtypes are not an inherent trait but may be a modifiable target to improve the therapeutic effect of the NF treatment. However, the relationship between the subtypes and other demographic factors or neurobiological indicators before and after treatment remains unexplored. Further investigation with more systematic longitudinal study, may reveal subtypes of treatment response based on specific pretreatment characteristics.

Another limitation of this study is the focus on data from the first training session, which may not capture the full dynamics of NF treatment over multiple sessions. While this approach allowed for a larger sample size and more consistent data across participants, it did not account for potential changes in subtypes or treatment responses that may occur with additional sessions. However, the relationship between the number of training sessions and the therapeutic effect is likely complex. A recent study on NF treatment for treatment-resistant MDD found no improvement in regulating amygdala activation after a single session [84], and symptom relief has been reported to increase with time after training [85]. Given these factors, distinguishing between the effects of the number of training sessions, the delayed therapeutic effect, and the subtypes during NF training is challenging. To clarify these relationships, future studies will require careful designs with multiple comparison groups addressing each factor.

In conclusion, the present study demonstrates that the therapeutic outcomes of NF training are significantly influenced by whole-brain activation patterns both during the process of affect regulation and in response to feedback signals. Our findings reveal multiple patterns of brain activity associated with significant therapeutic effects, suggesting a variety of potential pathways to recovery through NF training. In future studies, the optimization of NF training may involve real-time monitoring of brain activity during regulation efforts and in response to feedback signals, as well as tailoring feedback to the individual's current state of training progress. Although real-time fMRI NF requires expensive equipment, training effects have been observed after relatively short sessions [84], and its therapeutic benefits persist over time [85]. This suggests that the overall cost may not be significantly higher than long-term cognitive therapy or pharmacotherapy. Moreover, real-time fMRI's ability to provide personalized feedback based on each individual's unique brain activity holds promise for improving treatment precision and cost-effectiveness by enhancing intervention effectiveness beyond traditional approaches that rely on normative group patterns. Such an approach could increase the effectiveness of NF training by adjusting it to match each participant's unique neural response patterns.

DATA AVAILABILITY
The data that support the findings of this study are available from the corresponding author upon reasonable request. Access to the data will require a data sharing agreement between the requesting institution and the Laureate Institute for Brain Research, in accordance with institutional policies and ethical guidelines.

REFERENCES
1. Institute for Health Metrics and Evaluation (IHME). Global Burden of Disease Study 2019. IHME, https://www.healthdata.org/research-analysis/gbd2020.
2. Greenberg PE, Fournier A-A, Sisitsky T, Simes M, Berman R, Koenigsberg SH, et al. The Economic Burden of Adults with Major Depressive Disorder in the United States (2010 and 2018). PharmacoEconomics. 2021;39:653–65.
3. Pindi P, Houenou J, Piguet C, Favre P. Real-time fMRI neurofeedback as a new treatment for psychiatric disorders: A meta-analysis. Prog Neuro Psychopharmacol Biol Psychiatry. 2022;119:110605.
4. Goldway N, Jalon I, Keynan JN, Hellrung L, Horstmann A, Paret C, et al. Feasibility and utility of amygdala neurofeedback. Neurosci Biobehav Rev. 2022;138:104694.
5. Fernandez-Alvarez J, Grassi M, Colombo D, Botella C, Cipresso P, Perna G, et al. Efficacy of bio- and neurofeedback for depression: a meta-analysis. Psychological Med. 2022;52:201–16.
6. Taylor SF, Martz ME. Real-time fMRI neurofeedback: the promising potential of brain-training technology to advance clinical neuroscience. Neuropsychopharmacology. 2023;48:238–9.
7. Girges C, Vijiaratnam N, Zrinzo L, Ekanayake J, Foltynie T. Volitional Control of Brain Motor Activity and Its Therapeutic Potential. Neuromodulation. 2022;25:1187–96.
8. Dudek E, Dodell-Feder D. The efficacy of real-time functional magnetic resonance imaging neurofeedback for psychiatric illness: A meta-analysis of brain and behavioral outcomes. Neurosci Biobehav Rev. 2021;121:291–306.
9. Trambaiolli LR, Kohl SH, Linden DEJ, Mehler DMA. Neurofeedback training in major depressive disorder: A systematic review of clinical efficacy, study quality and reporting practices. Neurosci Biobehav Rev. 2021;125:33–56.
10. Martz ME, Hart T, Heitzeg MM, Peltier SJ. Neuromodulation of brain activation associated with addiction: A review of real-time fMRI neurofeedback studies. NeuroImage Clin. 2020;27:102350.
11. Barreiros AR, Almeida I, Baia BC, Castelo-Branco M. Amygdala Modulation During Emotion Regulation Training With fMRI-Based Neurofeedback. Front Hum Neurosci. 2019;13:89.
12. Emmert K, Kopel R, Sulzer J, Bruhl AB, Berman BD, Linden DEJ, et al. Meta-analysis of real-time fMRI neurofeedback studies using individual participant data: How is brain regulation mediated? Neuroimage. 2016;124:806–12.

13. Zotev V, Phillips R, Misaki M, Wong CK, Wurfel BE, Krueger F, et al. Real-time fMRI neurofeedback training of the amygdala activity with simultaneous EEG in veterans with combat-related PTSD. NeuroImage Clin. 2018;19:106–21.

14. Herwig U, Lutz J, Scherpiet S, Scheerer H, Kohlberg J, Opialla S, et al. Training emotion regulation through real-time fMRI neurofeedback of amygdala activity. Neuroimage. 2019;184:687–96.

15. Paret C, Zahringer J, Ruf M, Gerchen MF, Mall S, Hendler T, et al. Monitoring and control of amygdala neurofeedback involves distributed information processing in the human brain. Hum Brain Mapp. 2018;39:3018–31.

16. Lubianiker N, Paret C, Dayan P, Hendler T. Neurofeedback through the lens of reinforcement learning. Trends Neurosci. 2022;45:579–93.

17. Sitaram R, Ros T, Stoeckel L, Haller S, Scharnowski F, Lewis-Peacock J, et al. Closed-loop brain training: the science of neurofeedback. Nat Rev Neurosci. 2017;18:86–100.

18. Lawrence EJ, Su L, Barker GJ, Medford N, Dalton J, Williams SC, et al. Self-regulation of the anterior insula: Reinforcement learning using real-time fMRI neurofeedback. Neuroimage. 2014;88:113–24.

19. Skottnik L, Sorger B, Kamp T, Linden D, Goebel R. Success and failure of controlling the real-time functional magnetic resonance imaging neurofeedback signal are reflected in the striatum. Brain Behav. 2019;9:e01240.

20. Sutton RS, Barto AG. Reinforcement learning: An introduction. 2nd edn. Cambridge, MA: MIT Press; 2018.

21. Young KD, Zotev V, Phillips R, Misaki M, Yuan H, Drevets WC, et al. Real-time FMRI neurofeedback training of amygdala activity in patients with major depressive disorder. PLoS One. 2014;9:e88785.

22. Young KD, Siegle GJ, Bodurka J, Drevets WC. Amygdala Activity During Autobiographical Memory Recall in Depressed and Vulnerable Individuals: Association With Symptom Severity and Autobiographical Overgenerality. Am J Psychiatry. 2016;173:78–89.

23. Yuan H, Young KD, Phillips R, Zotev V, Misaki M, Bodurka J. Resting-state functional connectivity modulation and sustained changes after real-time functional magnetic resonance imaging neurofeedback training in depression. Brain Connectivity. 2014;4:690–701.

24. Tsuchiyagaito A, Smith JL, El-Sabbagh N, Zotev V, Misaki M, Al Zoubi O, et al. Real-time fMRI neurofeedback amygdala training may influence kynurenine pathway metabolism in major depressive disorder. NeuroImage Clin. 2021;29:102559.

25. Misaki M, Mulyana B, Zotev V, Wurfel BE, Krueger F, Feldner M, et al. Hippocampal volume recovery with real-time functional MRI amygdala neurofeedback emotional training for posttraumatic stress disorder. J Affect Disord. 2021;283:229–35.

26. Misaki M, Phillips R, Zotev V, Wong CK, Wurfel BE, Krueger F, et al. Brain activity mediators of PTSD symptom reduction during real-time fMRI amygdala neurofeedback emotional training. NeuroImage Clin. 2019;24:102047.

27. Misaki M, Phillips R, Zotev V, Wong CK, Wurfel BE, Krueger F, et al. Real-time fMRI amygdala neurofeedback positive emotional training normalized resting-state functional connectivity in combat veterans with and without PTSD: a connectome-wide investigation. NeuroImage Clin. 2018;20:543–55.

28. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR Fourth Edition (Text Revision). Washington, DC: American Psychiatric Publishing; 2000.

29. First MB. Structured clinical interview for DSM-IV-TR axis I disorders, research version, patient edition (SCID-I/P). New York, NY: Biometrics research; 2002.

30. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. 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. 1998;59:34–57.

31. Zotev V, Krueger F, Phillips R, Alvarez RP, Simmons WK, Bellgowan P, et al. Self-regulation of amygdala activation using real-time FMRI neurofeedback. PLoS One. 2011;6:e24522.

32. Montgomery SA, Åsberg M. A new depression scale designed to be sensitive to change. Br J Psychiatry J Ment Sci. 1979;134:382–9.

33. Fias W, Lammertyn J, Caessens B, Orban GA. Processing of Abstract Ordinal Knowledge in the Horizontal Segment of the Intraparietal Sulcus. J Neurosci. 2007;27:8952–6.

34. Glover GH, Li TQ, Ress D. Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. Magn Reson Med. 2000;44:162–7.

35. Birn RM, Smith MA, Jones TB, Bandettini PA. The respiration response function: the temporal dynamics of fMRI signal fluctuations related to changes in respiration. Neuroimage. 2008;40:644–54.

36. Jo HJ, Saad ZS, Simmons WK, Milbury LA, Cox RW. Mapping sources of correlation in resting state FMRI, with artifact detection and removal. Neuroimage. 2010;52:571–82.

37. Mumford JA, Turner BO, Ashby FG, Poldrack RA. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage. 2012;59:2636–43.

38. Chen G, Taylor PA, Stoddard J, Cox RW, Bandettini PA, Pessoa L. Sources of Information Waste in Neuroimaging: Mishandling Structures, Thinking Dichotomously, and Over-Reducing Data. Aperture Neuro. 2022;2:1–22.

39. Chen G, Padmala S, Chen Y, Taylor PA, Cox RW, Pessoa L. To pool or not to pool: Can we ignore cross-trial variability in FMRI? Neuroimage. 2021;225:117496.

40. Altman N, Krzywinski M. The curse(s) of dimensionality. Nat Methods. 2018;15:399–400.

41. McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv. 2018. https://arxiv.org/abs/1802.03426.

42. Landi I, Mandelli V, Lombardo MV. reval: A Python package to determine best clustering solutions with stability-based relative clustering validation. Patterns. 2021;2:100228.

43. Shen X, Finn ES, Scheinost D, Rosenberg MD, Chun MM, Papademetris X, et al. Using connectome-based predictive modeling to predict individual behavior from brain connectivity. Nat Protoc. 2017;12:506–18.

44. Shen X, Tokoglu F, Papademetris X, Constable RT. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. Neuroimage. 2013;82:403–15.

45. Greene AS, Gao S, Scheinost D, Constable RT. Task-induced brain state manipulation improves prediction of individual traits. Nat Commun. 2018;9:2807.

46. Greene AS, Gao S, Noble S, Scheinost D, Constable RT. How Tasks Change Whole-Brain Functional Organization to Reveal Brain-Phenotype Relationships. Cell Rep. 2020;32:108066.

47. Ju S, Horien C, Shen X, Abuwarda H, Trainer A, Constable RT, et al. Connectome-based predictive modeling shows sex differences in brain-based predictors of memory performance. Front Dement. 2023;2:1126016.

48. Misaki M, Tsuchiyagaito A, Guinjoan SM, Rohan ML, Paulus MP. Trait repetitive negative thinking in depression is associated with functional connectivity in negative thinking state rather than resting state. J Affect Disord. 2023;340:843–54.

49. Yoo K, Rosenberg MD, Hsu WT, Zhang S, Li CR, Scheinost D, et al. Connectome-based predictive modeling of attention: Comparing different functional connectivity features and prediction methods across datasets. Neuroimage. 2018;167:11–22.

50. Lange T, Roth V, Braun ML, Buhmann JM. Stability-based validation of clustering solutions. Neural Comput. 2004;16:1299–323.

51. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Usinglme4. J Stat Softw. 2015;67:1–48.

52. emmeans: Estimated Marginal Means, aka Least-Squares Means. 2022. https://CRAN.R-project.org/package=emmeans. Accessed Date 2022.

53. Linhartova P, Latalova A, Kosa B, Kasparek T, Schmahl C, Paret C. fMRI neurofeedback in emotion regulation: A literature review. Neuroimage. 2019;193:75–92.

54. Etkin A, Buchel C, Gross JJ. The neural bases of emotion regulation. Nat Rev Neurosci. 2015;16:693–700.

55. Picó-Pérez M, Alemany-Navarro M, Dunsmoor JE, Radua J, Albajes-Eizagirre A, Vervliet B, et al. Common and distinct neural correlates of fear extinction and cognitive reappraisal: A meta-analysis of fMRI studies. Neurosci Biobehav Rev. 2019;104:102–15.

56. Harrison BJ, Pujol J, Contreras-Rodríguez O, Soriano-Mas C, López-Solà M, Deus J, et al. Task-Induced Deactivation from Rest Extends beyond the Default Mode Brain Network. PLOS ONE. 2011;6:e22964.

57. Axelrod V, Rees G, Bar M. The default network and the combination of cognitive processes that mediate self-generated thought. Nat Hum Behav. 2017;1:896–910.

58. Andrews-Hanna JR, Smallwood J, Spreng RN. The default network and self-generated thought: component processes, dynamic control, and clinical relevance. Ann N Y Acad Sci. 2014;1316:29–52.

59. Hamilton JP, Farmer M, Fogelman P, Gotlib IH. Depressive Rumination, the Default-Mode Network, and the Dark Matter of Clinical Neuroscience. Biol Psychiatry. 2015;78:224–30.

60. Wise T, Marwood L, Perkins AM, Herane-Vives A, Joules R, Lythgoe DJ, et al. Instability of default mode network connectivity in major depression: a two-sample confirmation study. Transl Psychiatry. 2017;7:e1105.

61. Zhu X, Zhu Q, Shen H, Liao W, Yuan F. Rumination and Default Mode Network Subsystems Connectivity in First-episode, Drug-Naive Young Patients with Major Depressive Disorder. Sci Rep. 2017;7:43105.

62. Bessette KL, Jenkins LM, Skerrett KA, Gowins JR, DelDonno SR, Zubieta JK, et al. Reliability, Convergent Validity and Time Invariance of Default Mode Network Deviations in Early Adult Major Depressive Disorder. Front Psychiatry. 2018;9:244.

63. Jacob Y, Morris LS, Huang KH, Schneider M, Rutter S, Verma G, et al. Neural correlates of rumination in major depressive disorder: A brain network analysis. NeuroImage Clin. 2020;25:102142.

64. Makovac E, Fagioli S, Rae CL, Critchley HD, Ottaviani C. Can't get it off my brain: Meta-analysis of neuroimaging studies on perseverative cognition. Psychiatry Res Neuroimaging. 2020;295:111020.

65. Misaki M, Tsuchiyagaito A, Al Zoubi O, Paulus M, Bodurka J, Tulsa I. Connectome-wide search for functional connectivity locus associated with pathological

rumination as a target for real-time fMRI neurofeedback intervention. Neuro-Image Clin. 2020;26:102244.

66. Stern ER, Eng GK, De Nadai AS, Iosifescu DV, Tobe RH, Collins KA. Imbalance between default mode and sensorimotor connectivity is associated with perse-verative thinking in obsessive-compulsive disorder. Transl Psychiatry. 2022;12:19.

67. Yang MH, Guo ZP, Lv XY, Zhang ZQ, Wang WD, Wang J, et al. BMRMI Reduces Depressive Rumination Possibly through Improving Abnormal FC of Dorsal ACC. Neural Plast. 2022;2022:8068988.

68. Agathos J, Steward T, Davey CG, Felmingham KL, Ince S, Moffat BA et al. Dif-ferential engagement of the posterior cingulate cortex during cognitive restructuring of negative self- and social beliefs. Social Cognit Affect Neurosci. 2023;18:nsad024.

69. Natu VS, Lin JJ, Burks A, Arora A, Rugg MD, Lega B. Stimulation of the Posterior Cingulate Cortex Impairs Episodic Memory Encoding. J Neurosci. 2019;39:7173–82.

70. Guterstam A, Bjornsdotter M, Gentile G, Ehrsson HH. Posterior cingulate cortex integrates the senses of self-location and body ownership. Curr Biol. 2015;25:1416–25.

71. Foster BL, Koslov SR, Aponik-Gremillion L, Monko ME, Hayden BY, Heilbronner SR. A tripartite view of the posterior cingulate cortex. Nat Rev Neurosci. 2023;24:173–89.

72. Pecina M, Chen J, Karp JF, Dombrovski AY. Dynamic Feedback Between Anti-depressant Placebo Expectancies and Mood. JAMA Psychiatry. 2023;80:389–98.

73. Andreescu C, Butters M, Lenze EJ, Venkatraman VK, Nable M, Reynolds IIICF, et al. fMRI activation in late-life anxious depression: a potential biomarker. Int J Geriatr Psychiatry. 2009;24:820–8.

74. Ionescu DF, Niciu MJ, Mathews DC, Richards EM, Zarate JrCA. Neurobiology Of Anxious Depression: A Review. Depression Anxiety. 2013;30:374–85.

75. Sarkheil P, Odysseos P, Bee I, Zvyagintsev M, Neuner I, Mathiak K. Functional connectivity of supplementary motor area during finger-tapping in major depression. Compr Psychiatry. 2020;99:152166.

76. Northoff G, Hirjak D, Wolf RC, Magioncalda P, Martino M. All roads lead to the motor cortex: psychomotor mechanisms and their biochemical modulation in psychiatric disorders. Mol Psychiatry. 2021;26:92–102.

77. Jauhar S, Fortea L, Solanes A, Albajes-Eizagirre A, McKenna PJ, Radua J. Brain acti-vations associated with anticipation and delivery of monetary reward: A systematic review and meta-analysis of fMRI studies. PLOS ONE. 2021;16:e0255292.

78. Tsuchiyagaito A, Sanchez SM, Misaki M, Kuplicki R, Park H, Paulus MP, et al. Intensity of repetitive negative thinking in depression is associated with greater functional connectivity between semantic processing and emotion regulation areas. Psychological Med. 2023;53:5488–99.

79. Wu F, Lu Q, Kong Y, Zhang Z. A Comprehensive Overview of the Role of Visual Cortex Malfunction in Depressive Disorders: Opportunities and Challenges. Neurosci Bull. 2023;39:1426–38.

80. Shibata K, Lisi G, Cortese A, Watanabe T, Sasaki Y, Kawato M. Toward a com-prehensive understanding of the neural mechanisms of decoded neurofeedback. Neuroimage. 2019;188:539–56.

81. Scheinost D, Hsu TW, Avery EW, Hampson M, Constable RT, Chun MM, et al. Connectome-based neurofeedback: A pilot study to improve sustained attention. Neuroimage. 2020;212:116684.

82. Ciarlo A, Russo AG, Ponticorvo S, di Salle F, Lührs M, Goebel R, et al. Semantic fMRI neurofeedback: a multi-subject study at 3 tesla. J Neural Eng. 2022;19:036020.

83. Russo AG, Luhrs M, Di Salle F, Esposito F, Goebel R. Towards semantic fMRI neurofeedback: navigating among mental states using real-time representational similarity analysis. J Neural Eng. 2021;18:046015.

84. Compere L, Siegle GJ, Lazzaro S, Riley E, Strege M, Canovali G, et al. Amygdala real-time fMRI neurofeedback upregulation in treatment resistant depression: Proof of concept and dose determination. Behav Res Ther. 2024;176:104523.

85. Rance M, Walsh C, Sukhodolsky DG, Pittman B, Qiu M, Kichuk SA, et al. Time course of clinical change following neurofeedback. Neuroimage. 2018;181:807–13.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

Conceptualization, MM, KDY, and AT; Methodology, MM; Investigation, MM, KDY and AT; Writing – Original Draft, MM; Writing – Review & Editing, MM, KDY, AT, JS, and SMG.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. The study protocol was reviewed and approved by the University of Oklahoma Institutional Review Board (IRB) or the Western IRB reviewed the WCG IRB (https://www.wcgirb.com). All participants provided written informed consent prior to their inclusion in the study.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41380-024-02880-3.

Correspondence and requests for materials should be addressed to Masaya Misaki.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.