Brain Regulation of Emotional Conflict Predicts Antidepressant Treatment Response for Depression

Gregory A. Fonzo, PhD[1,§], **Amit Etkin, MD PhD**[2,3,4,§,*], **Yu Zhang, PhD**[2,3,4,§], **Wei Wu, PhD**[2,3,4], **Crystal Cooper, PhD**[5], **Cherise Chin-Fatt, PhD**[5], **Manish K. Jha, MD**[5], **Joseph Trombello, PhD**[5], **Thilo Deckersbach, PhD**[6], **Phil Adams, PhD**[7], **Melvin McInnis, MD**[8], **Patrick J. McGrath**[7], **Myrna M. Weissman, PhD**[7], **Maurizio Fava, MD**[6], **Madhukar H. Trivedi, MD**[5]

[1]Department of Psychiatry, Dell Medical School, The University of Texas at Austin, Austin, TX, USA

[2]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA

[3]Stanford Neurosciences Institute, Stanford University, Stanford, CA, USA

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

[5]Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, TX, USA

[6]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA

[7]New York State Psychiatric Institute & Department of Psychiatry, College of Physicians and Surgeons of Columbia University, New York, NY, USA

[8]Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA

Abstract

The efficacy of antidepressant treatment for depression is controversial due to only modest superiority demonstrated over placebo. However, neurobiological heterogeneity within depression may limit overall antidepressant efficacy. We sought to identify a neurobiological phenotype responsive to antidepressant treatment by testing pretreatment brain activation during response to, and regulation of, emotional conflict as a moderator of the clinical benefit of the antidepressant sertraline versus placebo. Using neuroimaging data from a large randomized controlled trial, we found widespread moderation of clinical benefits by brain activity during regulation of emotional

*Corresponding author: Amit Etkin, MD PhD Department of Psychiatry and Behavioral Sciences, Stanford University, 401 Quarry Road, MC: 5797, Stanford, CA 94305-5797, amitetkin@stanford.edu, Phone: +1 650-725-5736.
§Equal contributions as first author

conflict, in which greater down-regulation of conflict-responsive regions predicted better sertraline outcomes. Treatment-predictive machine learning utilizing brain metrics outperformed a model trained on clinical and demographic variables. Our findings demonstrate antidepressant response is predicted by brain activity underlying a key self-regulatory emotional capacity. Leveraging brain-based measures in psychiatry will forge a path toward better treatment personalization, refined mechanistic insights, and improved outcomes.

Trial Registration—Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care for Depression (EMBARC),

Introduction

Major depression is a common, chronic and disabling medical condition[1], whose treatment mainstay over the past four decades has been monoaminergic antidepressant medications[2]. As a result, roughly one in eight in the US takes an antidepressant[3]. Nonetheless, large meta-analyses have found only a small overall advantage of antidepressants over placebo (Cohen's d~0.3) when used in an unselected population of depressed patients, with clinical significance only in the most severe patients[45–8] – a severity level the vast majority of depressed patients never reach[1]. The small advantage of antidepressants over placebo has led to a prevalent view that these drugs are not effective in the treatment of depression[4]. An alternative explanation, however, is that the drugs are effective for at least some individuals, but the degree of neurobiological heterogeneity inherent to the diagnosis of depression[9] diminishes overall rates of efficacy. Put differently, it may be that for some patients with a particular neurobiological phenotype, antidepressants are clearly superior to placebo, while for others there is no differential benefit[10]. Objective measures that indicate for whom clinically significant superiority of antidepressants over placebo will be observed and under what conditions are termed moderators[10]. The identification of moderators is a critical first step towards personalized medicine for mental disorders. Here, we sought to address this central question about the biology of depression and how it impacts antidepressant treatment. Specifically, we tested whether the response to, or regulation of, emotional processing – a neural function relevant for depression – could elicit phenotypic brain characteristics that differentially predict treatment outcome with an antidepressant versus with placebo (i.e. moderate the clinical effect).

The ability to respond to, and regulate, brain responses to conflicting emotional cues is a critical component of wellbeing[11–13]. Emotional conflict engages cognitive and emotion-related brain circuitry, including the lateral prefrontal cortices, anterior cingulate, insula and amygdala[14–16]. Conflict can be regulated dynamically[11,15,16], a process that engages the rostral anterior cingulate and dampens activity in the conflict-responsive regions noted above[15–17]. Patients with a range of emotional disorders (e.g. depression, generalized anxiety, panic or bipolar disorder) have perturbations in their ability to regulate conflict-related brain activity[18–22], which mirrors the broad disturbance in emotional self-regulatory capacities in these disorders. This conflict-regulatory capacity has also been hypothesized to be of direct relevance to antidepressant treatment efficacy[23]. Thus, neuroimaging this process may yield observable patterns of brain activity that discriminate likelihood of subsequent favorable antidepressant response, thereby yielding both mechanistic insight and

predictive biomarkers. Here, we sought to determine whether variation in treatment outcome amongst depressed patients reflects individual brain response patterns to emotional conflict, and/or in the individual's ability to regulate conflict-related brain activity. Addressing this question has the potential for determining: 1) whether a specific medication-responsive phenotype exists in depression, 2) what the neural mechanism driving this phenotype may be, and 3) to what degree it represents a disease abnormality or an intact capacity.

Prior antidepressant prediction studies examining cognitive or emotional functioning have suggested that better adaptive prefrontal activation and lower amygdala responses to emotional cues predict better treatment outcome[24–27]. We therefore hypothesized that response to antidepressant treatment, compared to placebo, will be moderated by an intact neural capacity for managing emotional conflict. Unfortunately, all prior studies lacked a placebo control arm, which is a critical limitation. Thus, it is unknown whether prior findings are specific to medication effects (i.e. may moderate between antidepressants and placebo) or are non-specifically predictive across medication and placebo interventions. For example, extensive prior work using electroencephalography (EEG) has found that rostral anterior cingulate theta power is predictive of antidepressant treatment outcome[28]. However, analysis of data from the Establishing Moderators and Biosignatures of Antidepressant Response in Clinic Care (EMBARC) study[29], from which we draw our neuroimaging data, found that cingulate theta power was equally and non-specifically predictive of outcome with either an antidepressant or placebo. Prior work also suggests that treatment predictive brain signals often do not differ between patients and healthy controls (i.e. do not demonstrate diagnosis-wise differences)[24,25,30]. As such, we also compared our moderators between depressed patients and controls.

To test our hypothesis, we examined functional magnetic resonance imaging (fMRI) data on a previously characterized emotional conflict task[15,16] as part of EMBARC[29]. This is the largest neuroimaging-coupled placebo-controlled randomized clinical trial (RCT) in depression to date, and the only one utilizing fMRI measures. EMBARC randomized 309 medication-free depressed outpatients to receive either the selective serotonin reuptake inhibitor (SSRI) sertraline or placebo for eight weeks (supplementary figure 1). In this unique RCT, analysis of pre-treatment neuroimaging data for moderators was done in a full intent-to-treat framework across the entire brain. Doing so minimizes potential for bias and maximizes the likelihood of replicating effects. Individual-level prediction of treatment outcome was furthermore assessed using a cross-validated machine learning analysis, also on whole-brain data, in order to test the potential for this brain metric to be developed in the future into a potentially useful prognostic tool.

Results

Overall effect of sertraline versus placebo

As expected from prior meta-analyses[6,7], a linear mixed model on our primary clinical outcome (Hamilton Rating Scale for Depression (HAMD$_{17}$)) across the entire EMBARC sample revealed a small advantage of sertraline (n=122) over placebo (n=129) (linear mixed model treatment arm $\times$ time effect: $F(1,1471)=4.6$, $p=0.032$; Cohen's d=0.27, 95%

confidence intervals = 0.02, 0.52), consistent with expectations from prior meta-analyses[45–8] (figure 1; supplementary figure 2).

Emotional conflict task behavioral effects

We next assessed the reaction times from the emotional conflict task to verify the expected patterns of effects. As detailed in prior work[15,16], incongruent trials induce a conflict effect between the facial affect (the target of detection) and the incongruent overlaid emotional word, inducing a slow down in reaction times relative to congruent trials where the facial affect and word reflect the same emotion. In this task, emotional conflict can be regulated on a trial-to-trial basis when conflict trials are preceded by other conflict trials. That is, while emotional conflict results in slowing of reaction times, this effect can be mitigated in incongruent trials that follow incongruent trials (iI trials), compared to incongruent trials that follow congruent trials (cI trials)[15,16]. This trial-to-trial adaptive regulation of emotional conflict reflects an active process by which the brain increases emotional control in response to prior trial conflict, which then benefits regulation of emotional conflict on the subsequent trial (captured by the post-incongruent incongruent (iI) minus post-congruent incongruent (cI) contrast; see supplementary methods). The task produced the expected overall behavioral effects. Conflict resulted in a reaction time slowdown (I-C difference: M 58.3ms, SD 40.8; one-sample t-test t=25.6, p<0.001, Cohen's d = 1.43, 95% confidence intervals = 1.25, 1.60), and conflict regulation resulted in mitigation of that slowdown (iI-cI difference: M=−16.6ms, SD 48.4; t=6.1, p<0.001, Cohen's d = −0.34, 95% confidence intervals = −0.47, −0.22).

Emotional conflict task main effects on brain activation

We began our investigation of the imaging data by examining task-dependent main effects across the entire randomized depressed sample to verify the task was inducing the expected patterns of brain activity. As expected from prior work[16], the main effect of the brain response to conflict (voxel level one sample t-test of incongruent minus congruent (I-C) contrast values with voxel level false discovery rate (FDR) correction of $q < 0.05$ within whole brain gray matter mask optimized to study sample (see Methods); supplementary table 3) was characterized by robust activation of the conflict response network, including the dorsomedial and dorsolateral prefrontal cortices, ventrolateral prefrontal cortex, and anterior insulae. Deactivation was observed in the ventromedial prefrontal cortex and anterior medial prefrontal cortex as well as the posterior cingulate, precuneus, hippocampus, and parahippocampal gyri.

We next assessed brain response to the regulation of emotional conflict, which in this task occurs via an implicit process when conflict trials are preceded by other conflict trials[11,13,15,16] and results in a dampening of activation in conflict-responsive regions (noted above) and, in healthy individuals, an increase in activation in the rostral cingulate and/or ventromedial prefrontal cortex[15,16]. This conflict regulation-related rostral cingulate engagement has been demonstrated to be absent in major depression[31], even in the absence of a deficit in conflict-regulatory behavior. Consistent with prior work, we observed a robust and widespread dampening of activation across the entire conflict response network (including the dorsal cingulate, dorsomedial and dorsolateral prefrontal cortex, anterior

insulae, and amygdalae; voxel level one sample t-test of contrast values for iI vs. cI trials with voxel level FDR correction of $q < 0.05$ within whole brain mask; supplementary table 4) when contrasting activation for iI vs cI (iI-cI). Notably, we found no evidence for a significant increase in activation in the rostral cingulate or ventromedial prefrontal cortex for the contrast of iI vs. cI trials, consistent with prior work noting a deficit in this response in major depression[31].

Relationships between emotional conflict regulatory behavior and brain activity

Prior work has demonstrated that individual differences in emotional conflict regulatory behavior (i.e. the difference in reaction times for iI-cI trials, in which values are typically negative due to the mitigation of the conflict effect occurring on iI trials) scale with patterns of brain activity during the iI-cI contrast[20,31], reflecting a coherent relationship between brain and behavior in this context. We therefore attempted to utilize these prior demonstrated brain-behavior relationships between individual differences in emotional conflict regulatory behavior and concomitant brain activation to better ground any observed moderating effects of brain activation on sertraline vs. placebo responses in a maximally informative context. In other words, understanding how potential moderating patterns of brain activation relate to successful (or unsuccessful) performance of this paradigm provides greater inferential power regarding their behavioral significance. Thus, we examined how individual differences in emotional conflict regulation behavior (individual subject mean reaction time differences for iI-ci trials) related to whole brain activation patterns during the iI-cI contrast across the entire randomized depressed sample, controlling for site/scanner effects. Utilizing a voxel wise robust linear model to relate individual subject iI-cI reaction time differences to individual's iI-cI brain activation contrast values, controlling for scanners, we observed that better emotional conflict regulation (more negative iI-cI mean reaction time differences) was positively associated (whole brain voxel level FDR corrected q's < 0.05 for iI-cI reaction time robust linear model coefficient) with a greater dampening (i.e. more negative iI-cI contrast values) of activation in regions of the conflict response network, including the dorsal cingulate, dorsomedial and dorsolateral prefrontal cortices, and anterior insulae (supplementary table 5). In contrast, an anterior region of the ventromedial prefrontal cortex showed the opposite effect, such that increasing activation (i.e. less negative and/or more positive iI-cI contrast values) was associated with better behavioral regulation of emotional conflict (i.e. more negative iI-cI mean reaction time differences). This finding parallels prior work demonstrating a unique role of ventral prefrontal activation in mediating successful emotional conflict regulation[15,16]. Thus, more successful emotional conflict regulation behavior in major depression is associated with a more pronounced canonical (i.e. healthy) emotional conflict regulatory neural phenotype, which is characterized by dampening of the conflict response network and engagement of the ventral prefrontal cortex[15,16].

Emotional conflict moderators of treatment outcome

We next assessed our two *a-priori* brain and behavioral measures of interest for capacity to moderate antidepressant treatment response: the response to conflict (I-C contrast; figure 2a) as well as the trial-to-trial regulation of conflict-related brain activation (iI-cI contrast). The most striking results came when examining emotional conflict regulation (iI-cI), and thus are presented first. Statistically significant moderation effects (whole brain voxel level FDR

corrected q < 0.05 for the treatment arm × time × brain activation linear mixed model interaction effect) were observed in canonical conflict-response regions including the bilateral dorsolateral prefrontal and frontopolar cortices, dorsal anterior cingulate and insula, as well as the hippocampus, visual cortex and cerebellum (figure 2b; supplementary tables 7–10). To easily convey some of the linear mixed model results, we divided the depressed sample using a median split based on iI-cI activation levels for several brain regions (i.e. as an illustrative but not a statistically meaningful split). Doing so revealed that a more favorable effect of sertraline over placebo on $HAMD_{17}$ symptom reductions was evident for individuals who better dampened activation during conflict regulation in regions such as the dorsal anterior cingulate (figure 3a), frontopolar cortex (figure 3b), and anterior insula (figure 3c) during emotional conflict regulation (see also supplementary figure 3). Many of these significant moderation effects overlapped (particularly in the dorsal cingulate, dorsolateral prefrontal cortices, anterior insulae, and frontopolar cortices) with the significant task effect for emotional conflict regulation as well as the significant brain-behavior relationships between emotional conflict regulatory behavior and task-related activation (see figure 4 for the conjunction of brain maps and supplementary figure 4 for the union of brain maps; all utilized whole brain voxel level FDR-corrected q < 0.05 effect maps). Moreover, a comparison of whole-brain moderation effects for brain activity during iI and cI conditions separately (i.e. linear mixed model F statistics for the treatment arm × time × brain activation interaction for iI and cI activation separately) revealed that the majority of iI-cI moderation effects (5,895 of 8,769 voxels) were statistically stronger (i.e. had larger F statistics for the treatment arm × time × brain activation effect) for brain activity during the iI relative to the cI condition (see Supplementary Figure 5 for a brain image comparison of statistical moderation effects), particularly in the anterior insula, dorsal cingulate/dorsomedial prefrontal cortices, and dorsolateral prefrontal cortices. Results for all clusters are reported in supplementary tables 7–10. Thus, we found strong support for our hypothesis that greater reductions in $HAMD_{17}$ with sertraline versus placebo will be moderated by better brain regulation of emotional conflict. Interestingly, the direction of the fMRI response in these sertraline-responsive patients was consistent with, and for some patients of greater magnitude, as that seen in healthy individuals (figures 3a–c), indicating these characteristics likely represent an intact capacity. Unlike brain activity, there was no significant moderation effect of iI-cI reaction time differences on $HAMD_{17}$ symptom reductions (treatment arm × time × reaction time linear mixed model effect: F=0.48, p=0.49).

Interestingly, we found no evidence that activation in the rostral anterior cingulate during conflict regulation moderated treatment outcome (whole brain voxel FDR corrected q's > 0.05). However, we found that iI-cI activation in the rostral anterior cingulate predicted outcome similarly across both arms (whole brain voxel level FDR corrected q's < 0.05 for time × brain activation effect from linear mixed model on $HAMD_{17}$ symptom reductions; supplementary table 10). In this case, lower regulation-related iI-cI activation was associated with larger $HAMD_{17}$ symptom reductions across both arms (supplementary figure 6).

In our second *a-priori* brain measure of interest, responses to emotional conflict (I-C contrast) also yielded whole-brain significant moderation effects (whole brain voxel level FDR corrected q's < 0.05 for the treatment arm × time × brain activation linear mixed model

interaction effect), albeit in vastly fewer regions than the iI-cI contrast, and not in those typically associated with response to conflict (supplementary figure 7 and supplementary table 11). I-C reaction time differences moderated outcome (treatment arm × time × reaction time interaction effect from linear mixed model: F=6.0, p=0.014). However, I-C reaction time differences did not yield a significant interaction with time on $HAMD_{17}$ symptom scores within either arm alone (time × reaction time interaction effects from linear mixed model within each treatment arm separately: F's<1.7, p's >0.19).

Individual level prediction through machine learning analysis

To assess the ability of emotional conflict regulation-related activation to predict treatment outcome at the level of individual patients, we used Relevance Vector Machines (RVM) to build a regression model for the prediction of pre- minus post-treatment change in $HAMD_{17}$ using iI-cI activation (see training and cross-validation approach in supplementary figure 8). This was done in a data-driven manner by using a parcellation of the cortex and subcortex, and 10×10-fold cross-validation for prediction performance evaluation, separately for each treatment arm. An RVM model trained on sertraline outcome yielded significant cross-validated prediction of observed treatment change scores (correlation between model-predicted $HAMD_{17}$ changes and observed $HAMD_{17}$ changes: r=0.49, p<0.001, permutation testing-verified using 1000 permutations p<0.001; figure 4a). When applied to the placebo arm, however, the sertraline-trained model did not yield a significant prediction of $HAMD_{17}$ change (correlation between model-predicted $HAMD_{17}$ changes and observed $HAMD_{17}$ changes: r=−0.06, p=0.48; figure 5b). Moreover, prediction of sertraline outcome was significantly greater than prediction of placebo outcome (Fisher's z test for difference in correlation strengths: z=4.5, p<0.001).

Interestingly, an RVM model trained on iI-cI brain activation data in the placebo arm to predict placebo outcome did not yield significant correlations between model-predict symptom changes and observed symptom changes in either the placebo or sertraline arms (r's = 0.11, p's>0.20; supplementary figure 9). Thus, the model developed on sertraline outcome reflects a sertraline-specific signal that explains a meaningful percentage of variance in treatment effects, separate of treatment effects that include (but are not limited to) placebo responses which are inherent to both treatment arms[32].

Regional weights driving the sertraline-predicting model are shown in figure 5c, and demonstrate a similar spatial pattern to that of the whole-brain moderation analysis in Figure 1b. Though positive and negative signs on feature weights cannot be easily interpreted as reflecting a per unit positive or negative relationship with treatment response, respectively, due to the combinatorial nature of the model, the absolute magnitude of the weights gives an indication of the degree of "importance" of that feature in predicting the outcome measure. Features with larger weights thus contribute more to the model predicted outcome.

The unique value of iI-cI activation in predicting sertraline outcome was likewise seen in comparison to the much less accurate prediction of outcome using I-C activation, as also expected from the mixed model results above (correlation between symptom change predicted by RVM model trained on I-C contrast values and observed symptom changes: r=0.003, p=0.98; Fisher's z test comparing correlation between I-C model-predicted

symptom changes and observed symptom changes versus iI-cI model-predicted symptom changes and observed symptom changes; z=4.0, p<0.001; Figure 5d,e). Hence, the process by which the brain regulates, but not responds to, emotional conflict is what defines the drug-sensitive phenotype within the broad clinical diagnosis of depression. This result argues, as well, against overfitting having occurred in the iI-cI analysis, as both analyses include the same number of brain activity features, yet yielded drastically different results.

Relationship of treatment-predictive emotional conflict regulation signals with clinical severity measures and patient-control differences

In order to ascertain whether our findings reflected components of disease severity, we correlated conflict regulation-related brain activity in the iI-cI contrast with several measures of baseline clinical severity, including our $HAMD_{17}$ primary measure, the Quick Inventory of Depressive Symptoms (QIDS)[33], the Spielberger State-Trait Anxiety Inventory (STAI)[34], as well as the anxious arousal, anhedonic depression and general distress subscales of the Mood and Anxiety Symptom Questionnaire (MASQ)[35]. However, no significant relationships survived Type I error correction (all whole brain voxel level FDR-corrected robust linear model coefficient q's > 0.05 relating each symptom scale score to iI-cI contrast values, controlling for scanners) for each scale, providing no evidence that our treatment-moderating brain activation findings are simply a marker of clinical severity. Moreover, the sertraline RVM model itself did not show significant relationships with clinical severity, as each individual's model-predicted $HAMD_{17}$ symptom change was not correlated with baseline $HAMD_{17}$, QIDS, STAI or MASQ subscale scores (r's<0.10, p's>0.12).

We also found no significant differences between depressed patients (as a whole) and healthy controls in conflict regulation-related brain activation (whole brain threshold free cluster enhancement with 5000 permutation tests: p's > 0.05). Likewise, when applying the sertraline predictive RVM model to iI-cI activation data from healthy controls, we did not find any significant difference in the degree to which the RVM model pattern was expressed (two-sample t-test of iI-cI RVM model predictions for depressed vs. healthy controls: t=0.08, p=0.93). Thus, the treatment-predictive emotional conflict regulation phenotype was not significantly related to either the diagnosis of depression or its clinical severity.

Finally, using brain activation to characterize the degree to which an individual manifests a sertraline-responsive phenotype may not be feasible in practice if lower-cost measures such as clinical severity scores, emotional conflict task behavior, demographic variables, or historical factors like childhood trauma exposure[36] could usefully predict outcome. This did not prove to be the case, however, as entering all of these into an RVM analysis did not yield significant prediction of $HAMD_{17}$ symptom changes with sertraline (correlation between demographic/clinical severity model-predicted $HAMD_{17}$ changes and observed $HAMD_{17}$ changes: r=0.08, p=0.39). The relationship between demographic/clinical severity model-predicted symptom changes and observed symptom changes with sertraline was also significantly weaker than the relationship between iI-cI brain activation-based RVM model-predicted symptom changes and observed symptom changes (Fisher's z test; z=3.4, p<0.001).

Discussion

We identified a medication-responsive neural phenotype within the broader clinical diagnosis of depression, which was related to the degree to which the brain could adaptively regulate emotional conflict above and beyond the brain's response to conflict itself. It was also a better predictor of treatment outcome than clinical and behavioral metrics alone. Moreover, we identified this phenotype in both linear mixed model moderation analyses and at the individual patient level using cross-validated sparse machine learning models.

Together, these findings delineate a specific neurobiological profile driving whether a patient will respond to an antidepressant differently from placebo, and they contradict assertions that antidepressants are universally ineffective. Rather, our findings argue that heterogeneity in disease mechanisms inherent in the clinical definition of depression obscures the detection of antidepressant efficacy defined by a critical treatment-relevant phenotype, and that mechanistically meaningful stratification may be possible in depression by relating biological characteristics to placebo-controlled treatment outcome. Indeed, we found no evidence that the treatment-predictive phenotype could be identified if simply comparing emotional conflict regulation in patients and controls or examining clinical severity differences between patients. This is consistent with prior non-placebo-controlled treatment predictive efforts in depression, which similarly note no group differences in treatment-predictive neural signals [24,25,30]. As such, despite no evidence for differentiating between patients and controls, our findings are of clinical importance as they illustrate how the efficacy of an intervention for depression may be strongly influenced by individual neurobiological differences between patients.

Although prior findings and theoretical considerations regarding emotional conflict regulation (and its behavioral and neural deficits) in depression have focused primarily on the rostral/ventral ACC and its lack of activation in response to emotional conflict regulation[31], the current findings indicate the most treatment-relevant aspects of the ongoing neurocircuitry processes mediating regulation of emotional conflict in major depression do not necessarily reflect this prior theoretical framework. Moreover, we detected no functional brain differences between depressed individuals and healthy comparison subjects during emotional conflict regulation, which diverges with prior findings for a rostral/ventral ACC activation failure[31]. However, the healthy comparison group utilized here was much smaller in number compared to individuals with major depression, which may have limited power to detect such differences, if present. Taken together, the lack of activation differences observed here as well as the fact that dampening of the conflict response network during conflict regulation was the most relevant to differentiating subsequent antidepressant vs. placebo responses (and not increasing activation in the rostral/ventral regions of the ACC and prefrontal cortex) suggests that this prior theoretical and neurocircuitry framework regarding emotional conflict regulation in major depression may need to be revised to better account for the current findings in a much larger sample with a rigorous randomized clinical trial methodology. We therefore recommend that the theoretical and neurobiological understanding of brain emotional conflict regulation abnormalities in depression continue to be probed utilizing large samples with roughly equivalent groups of patients and healthy comparison subjects, all acquired on the same scanner setup, and with exploration of how

clinical heterogeneity in the patient population may contribute to heterogeneity in neural phenotypes.

The impact of our conclusions is further reinforced by the size and methodological rigor of EMBARC, which is the only neuroimaging-coupled placebo-controlled antidepressant RCT designed for identification of moderators to date, as well as the whole-brain data-driven intent-to-treat framework used for all of our analyses. We took this whole-brain approach given the uniqueness of EMBARC (thus minimizing the potential of bias in reporting restricted results) and because it was unknown whether prior non-placebo-controlled treatment prediction findings would be likely moderators or nonspecific predictors across both arms. This point is illustrated, for example, by our rostral anterior cingulate findings, as noted below. Likewise, we complemented voxel-wise linear mixed model analyses with whole-brain cross-validated machine learning analyses to underscore the potential future clinical utility of this metric.

The mechanistic insights into this sertraline-responsive depression phenotype tie in more broadly to a specific aspect of self-regulatory emotional capacity. Regions such as the dorsolateral prefrontal cortex and dorsal anterior cingulate are responsive to both emotional and non-emotional conflict [14,15], and similar trial-to-trial regulation of emotional and non-emotional conflict occurs and involves reduction of activation in dorsal anterior cingulate conflict signals [37]. Hence, it may be that greater trial-to-trial regulation of any type of conflict underpins the sertraline-responsive phenotype. Future work in this area should attempt to differentiate whether the treatment-predictive conflict regulatory dampening observed here is specific to emotional conflict vs. general conflict processing, which unfortunately could not be ascertained with the current design. The specificity of this predictive signal to sertraline vs. other antidepressants (or other empirically-supported treatments such as psychotherapy or brain stimulation) is also an area of high priority for future investigations. Furthermore, the fact that the regulation of emotional conflict much more robustly moderated treatment outcome relative to overall response to the conflict itself suggests strongly that it is the specific process of adaptive regulation that matters, rather than conflict processing *per se.* As we found no evidence that emotional conflict regulation behavior (i.e. reaction times) moderates treatment outcome, this may be due to brain activation being a better assay of the relevant mechanisms relative to behavior (consistent with similar prior findings, e.g. [27]), and the fact that activity across multiple regions together contribute to single behavioral measures. These findings also delineate a particular mental capacity, and associated brain structures, as targets for development of novel treatments. These findings raise the intriguing scenario that converting a sertraline-nonresponsive patient into a sertraline-responsive patient may be possible if aspects of the brain's emotional conflict regulation capacity can be improved by new medications, brain stimulation, behavioral training, or psychotherapy.

Interestingly, not all aspects of emotional conflict regulation-related activity moderated between sertraline and placebo. Activation in the rostral anterior cingulate non-specifically predicted outcome independently of treatment, which may relate to the shared placebo component of both interventions or natural patterns of recovery over time. This is consistent with resting EEG studies, including using EMBARC data, as well as neuroimaging work

across a wide range of interventions [28]. Moreover, EEG work in non-placebo-controlled studies implicated this region in treatment prediction, but subsequent efforts found that it similarly predicted outcome with both medication and placebo [32,38]. In fact, our non-specific predictive effect partially overlapped with the cluster demonstrating non-specific prediction in the EEG analyses [32].

In future work, it will be important to replicate the current findings, which is essential to maximizing their potential clinical utility for biomarker-informed treatment prediction efforts. It is also important to determine the degree to which our findings generalize to other antidepressants or other forms of evidence-based treatment such as psychotherapy or brain stimulation, as well as to patients with multiple medications (i.e. patients here were medication-free). Importantly, as EMBARC recruited individuals without a history of antidepressant treatment failures in order to maximize power to detect treatment moderation effects, these findings may not necessarily generalize to all individuals with major depression, especially those with a prior history of failing one or more adequate antidepressant trials. Regarding methodological rigor, additional work should be done to assess and optimize test-retest reliability of emotional conflict regulation signals in depressed patients, which was outside the scope of this study. Moreover, we propose that ultimate widespread clinical application of our moderation signals will require use of EEG instead of fMRI – a transition encouraged by similar normative spatial patterns during conflict regulation in EEG as seen in fMRI [39–42]. Thus, future studies incorporating multi-modal brain assessments of emotional conflict regulation with additional design components facilitating replication of the current findings and identification of sertraline-specific predictive effects vs. those specific to other forms of treatment (or those that predict outcomes for multiple treatment modalities) will continue to advance this body of knowledge to maximally positively impact clinical care. For example, one potential clinical application could be in deciding whether to continue further medication trials after initial failure, or to switch to treatments with putatively distinct mechanisms of action (e.g. repetitive transcranial magnetic stimulation (rTMS), electroconvulsive therapy, or psychotherapy). Typically, patients undergo iterative rounds of pharmacological treatment prior to advancing to a treatment modality such as rTMS [43,44], leading to substantial morbidity and economic cost. More rapidly advancing patients with an antidepressant non-responsive brain phenotype to even costly interventions, such as rTMS, may thus make both clinical and financial sense[43,44], while additional medication trials may be better for patients with an antidepressant responsive phenotype. It is therefore noteworthy that the large-scale studies of rTMS treatment for depression have specifically enrolled medication-resistant patients, and showed benefit of real over sham rTMS [45,46]. More broadly, however, it may also be that a brain-based test that identifies an individual as having a medication-responsive brain phenotype may serve to encourage patients to seek medication treatment in the first place, and decrease the stigma associated with mental illness[47].

In sum, these findings advance a mechanistic brain profile of depression that augurs well for the possibility of an effective stratification of an otherwise heterogeneous and solely clinically defined population into segments likely or unlikely to respond to a common antidepressant treatment. While other measures will no doubt further improve this stratification and aid in its broader clinical application, our work argues for a critical

transition: from questioning whether antidepressants are effective to focusing on identifying the most robust combination of individual predictors that will shift clinical care in psychiatry to a neuroscience-informed personalized approach[48].

Methods

Participants and treatment

The study methods and participant characteristics are presented here in brief, as well as a publication detailing the rationale and design of EMBARC [29]. Written informed consent was obtained from each participant under institutional review board-approved protocols at each of the four clinical sites (University of Texas Southwestern Medical Center, Massachusetts General Hospital, Columbia University, and University of Michigan). Data reported here are based on EMBARC participants who were randomly assigned to sertraline or placebo during stage 1 of the trial (N=309 total). This sample size was based on effect size estimates for sertraline treatment from prior studies. Key eligibility for the study included being 18–65 years old, having major depression as a primary diagnosis by the Structured Clinical Interview for DSM-IV Axis I Disorders [49], at least moderate depression severity with a score 14 on the Quick Inventory of Depressive Symptomatology-Self Report (QIDS-SR) at screening and randomization, a major depressive episode beginning before age 30, either a chronic recurrent episode (duration 2 years) or recurrent MDD (at least 2 lifetime episodes), and no antidepressant failure during the current episode. Exclusion criteria included: current pregnancy, breastfeeding, no use of contraception; lifetime history of psychosis or bipolar disorder; substance dependence in the past six months or substance abuse in the past two months; unstable psychiatric or general medical conditions requiring hospitalization; study medication contraindication; clinically significant laboratory abnormalities; history of epilepsy or condition requiring an anticonvulsant; electroconvulsive therapy (ECT), vagal nerve stimulation (VNS), transcranial magnetic stimulation (TMS) or other somatic treatments in the current episode; medications (including but not limited to antipsychotics and mood stabilizers); current psychotherapy; significant suicide risk; or failure to respond to any antidepressant at adequate dose and duration in the current episode. In addition, 40 medically and psychiatrically healthy individuals were recruited across the four sites and assessed in a similar manner as the patients.

Clinical trial

EMBARC used a double-blind design, wherein participants were randomized to an 8-week course of sertraline or placebo. Randomization was stratified by site, depression symptoms severity, and chronicity using a block randomization procedure. Sertraline dosing began at 50mg using 50mg capsules and was increased as tolerated if the patient did not respond until a maximum of 200mg[29]. A similar dosing approach was used for placebo capsules. The trial was registered on ClinicalTrials.gov as: Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care for Depression (EMBARC), .

Emotional conflict task

This well-characterized paradigm assesses both emotional conflict and emotional conflict regulation [15,16]. Each trial involved presentation of an emotional face with either a fearful or

happy expression, drawn from the set of Ekman & Friesen [50], with an overlaid emotion word ("FEAR" or "HAPPY"). Participants were instructed to identify the facial emotion with a key press, while trying to ignore the emotion word. The task consisted of 148 trials, with stimuli presented for 1000 milliseconds (ms) in a fast event-related design. Interstimulus intervals were 3000–5000ms in a pseudo-randomized order counterbalanced for facial expression, gender, word, and response button. Stimuli were either congruent (e.g. fearful face with "FEAR") or incongruent (e.g. fearful face with "HAPPY"), and stimuli were furthermore balanced to achieve an equal fraction of current and prior trial congruency, while ensuring no direct stimulus repetitions. Prior to performance of the task during neuroimaging, all participants underwent a practice version to ensure task proficiency was reached (minimum 80% accuracy) and the task instructions were understood. The neuroimaging task lasted 13 minutes and 14 seconds.

Regulation in the emotional conflict task occurs via an implicit process when conflict trials are preceded by other conflict trials [11,13,15,16]. That is, while emotional conflict results in slowing of reaction times, this effect can be mitigated in incongruent trials that follow incongruent trials (iI trials), compared to incongruent trials that follow congruent trials (cI trials)[15,16]. This trial-to-trial adaptive regulation of emotional conflict reflects an active process by which the brain increases emotional control in response to prior trial conflict, which then benefits regulation of emotional conflict on the subsequent trial (captured by the iI-cI contrast). This regulation effect, captured through the same contrast, has also been extensively described for non-emotional conflict stimuli [37,51,52]. Critically, this contrast between post-incongruent incongruent and post-congruent incongruent trial compares brain responses to physically identical stimuli (i.e. incongruent trials) that differ only on the relative emotional conflict regulatory context in which they come due to prior trial congruency, and is furthermore independent of the incongruent versus congruent trial (I-C) conflict response contrast. Neuroimaging acquisition parameters are shown in Supplementary Table 2.

fMRI preprocessing and first-level modeling

FSL tools were used to preprocess imaging data [53,54]. Functional images were first realigned to structural images using an affine registration matrix and boundary-based registration based upon tissue segmentation as implemented in FSL's FLIRT, which was concatenated with a non-linear normalization of each participant's T1 image to the Montreal Neurological Institute (MNI) 152-person 1 mm^3 T1 template using FNIRT from FSL 5.0 to result in a single transformation step from individual native functional space to structurally-aligned and spatially-normalized template space. Functional images were realigned to the middle volume of the run. Nuisance signals corresponding to segmented white matter and CSF were regressed out of motion-corrected functional images. A 6 mm full-width half max (FWHM) isotropic smoothing kernel was then applied to preprocessed time series images to account for individual anatomical variability.

For individual-level analyses for each participant and timepoint, regressors modeling trials of interest were convolved with the hemodynamic response function. First-level general linear models were estimated in SPM 8 [55]. Regressors corresponded to zero-duration

markers set at the onset of stimuli, which were explicitly categorized by congruency (Incongruent or Congruent) and prior trial type (Post-incongruent or Post-congruent) in order to model conflict response and regulation effects. This resulted in 4 different trial types in total, in addition to nuisance regressors for error trials and post-error trials (when applicable) and six motion parameters.

Participant selection and quality control procedures

In order to assure the quality of imaging measures in testing moderation and prediction effects, consistent with our prior work in imaging treatment effects[56,57], we instituted cutoffs for absolute level of motion (root mean square of the absolute level of movement < 4mm across the mean of the squared maximum displacements in each of the 6 translational and rotational parameters estimated during realignment) during the baseline scans. In addition, in order to ensure brain activation measures reflect task-relevant metrics, we also instituted a minimum level of behavioral accuracy during completion of the emotional conflict task as an additional quality control metric (total accuracy 80% of trials correct). Functional runs at baseline displaying motion higher than our cutoff OR accuracy below the minimum cutoff were excluded from group-level analyses.

Sample size selection and power considerations

The EMBARC study was designed to be sufficiently powered to detect an interaction effect between treatment arm (sertraline vs. placebo) and a continuous treatment outcome-predictive marker (brain activation) based upon the following parameters: with a sample size of 160 participants in each treatment arm, there is 80% power of a two-sided significance test with $\alpha=0.05$ to detect differences between the correlation coefficients (between the marker and the outcome in each treatment group) of a magnitude 0.31. For example, if the relationship between the marker (i.e. brain activation) and the outcome ($HAMD_{17}$) were 0.31 in the sertraline group and 0 in the placebo group (or 0.41 in the sertraline group and 0.1 in the placebo group), a sample size of 160 in each treatment arm would be sufficient to detect this effect. Given that EMBARC is a hypothesis-generating study and is the largest and only placebo-controlled sertraline trial with fMRI measures, there is only indirect evidence to support that a difference in correlation coefficients between the predictive marker and the outcome in each treatment arm would be expected to be of this or higher magnitude. In fact, the magnitude of moderation effect sizes of the sertraline vs. placebo effect on a continuous outcome by a continuous measure is largely unknown. However, prior studies[24–27] examining brain-activation based predictors of treatment response in depression (but without a placebo control group, in smaller samples, and/or utilizing a dichotomous split of treatment responders and non-responders) have reported effect sizes that are moderate-to-large in magnitude, which supports the assumption that the sample sizes utilized here are at least sufficient to detect moderation effects of moderate-to-large magnitude.

Statistical analyses

All statistical tests undertaken were two-tailed. Where possible (e.g., for clinical moderation and treatment arm × time analyses), the normality and homoscedasticity of model residuals were plotted and inspected visually for adherence to model assumptions[58,59] (linear

relationship(s) between predictor(s) and outcomes, model residuals normally distributed with equivalent variances across distribution of fitted values, etc.). When this was not possible (e.g., for voxel-level moderation analyses), adherence to model assumptions was assessed utilizing within-subject extracted average cluster activation values that were exported for data visualization purposes. Visual inspection revealed no clear violations of model assumptions. Cohen's d effect size for relevant statistics was calculated according to the following methods: a) for two sample mean comparisons, d was calculated as the difference in means divided by the pooled standard deviation; b) for one sample t-tests, Cohen's d was calculated as the mean divided by the standard deviation. The confidence intervals were calculated as follows: a) for two sample mean comparisons, the 95% confidence intervals for Cohen's d were calculated using the cohen.d.ci function from the *psych* library[60] in R; b) for one sample t-tests, the 95% confidence intervals for Cohen's d were calculated according to the following formula: d +/− SE × 1.96 (critical Z value), where SE (standard error) = sqrt($(1/n)+(d^2/2n)$).

Assessing the effect of sertraline on depressive symptoms

Using the total sample after quality control procedures for poor quality imaging data, repeated measures of $HAMD_{17}$ total score at baseline, week 1, week 2, week 3, week 4, week 6, and week 8 (endpoint) were entered into a longitudinal linear mixed model in IBM SPSS version 21.0^{61} along with variables coding for participant, treatment arm, site (four centered dummy coded variables, each specifying one of the non-UT Southwestern study sites/scanners (Columbia utilized 2 separate scanners) vs. UT Southwestern) and time (scaled from 0 to 1) to assess the effect of sertraline vs. placebo on depression symptoms in an intent-to-treat framework. Treatment arm was effects coded as −0.5 for placebo and 0.5 for sertraline. Random effects were specified for the model intercept to account for within-participant variation in baseline depressive symptom severity. A fixed intercept as well as a fixed effect of time, effect of Columbia scanner 1 vs. UT Southwestern site, effect of Columbia scanner 2 vs. UT Southwestern site, effect of Massachusetts General Hospital vs. UT Southwestern site, effect of University of Michigan vs. UT Southwestern site, interaction of time with each of the aforementioned 4 site variables, treatment arm × time, and the interaction of treatment arm × time with each of the 4 aforementioned site variables were modeled, with the treatment arm × time interaction effect of interest specifying differential trajectories in symptom change over time for sertraline vs. placebo while controlling for baseline severity (random intercept), non treatment-specific symptom changes (time), potential differences in site cohorts, the interactions of site cohorts with non treatment-specific changes over time, and potential differences in medication effects by site cohort. Site variables were centered to yield an interpretable (i.e., grand average) overall treatment arm × time effect.

Assessing task effects on brain activity

The entire randomized sample was subjected to a voxelwise one-sample t-test against the null hypothesis (activation = 0) for two contrasts of interest: the effect of emotional conflict (Incongruent minus Congruent trials) and the effect of emotional conflict regulation (post-incongruent incongruent trials (iI) vs. post-congruent incongruent trials (cI); iI-cI). We constrained all imaging analyses to a search space defined by a whole-brain mask

constructed from: a) a 30% probabilistic voxelwise threshold derived from spatially normalized, segmented binary gray matter images from an independent sample of healthy comparison participants[56,57]; and b) a conjunction map of adequate signal coverage across study scanners constructed by concatenating the binarized individual-level average map of all task experimental conditions within participants acquired at each scanner, thresholding at a minimum of 50% coverage within each scanner sample, and then conjoining these maps across all scanners. Thus, the final search space was a function of both adequate signal coverage across all study scanners as well as a probabilistic assignment to the brain's gray matter. At each voxel in this search space, a one-sample t-test was conducted for each contrast of interest, and the resultant statistical t-maps were corrected for multiple comparisons on the voxel level utilizing a whole-brain (constrained within the search space) false discovery rate (FDR) correction with AFNI's program 3dFDR. Voxels displaying FDR-corrected Z-values < 1.96 (pFDR < 0.05) were then thresholded and clustered for subsequent individual-level extraction and visualization of effects in IBM SPSS version 21.0 [61]

Assessing brain-behavior relationships for emotional conflict regulation

Individual contrast maps specifying the difference in activation for iI-cI trials were subjected to a robust regression[62,63] implemented in R[64] using the fmri package[65]. The predictor of interest was the mean reaction time difference for each subject for iI minus cI trials from the emotional conflict task. Covariates of no interest included four centered dummy variables corresponding to scanner (with the scanner at UT Southwestern serving as the non-coded reference). Activation at each voxel within a whole brain mask (described above) was modeled as a function of an intercept, the grand mean-centered reaction time differences for iI minus cI trials, and the four scanner dummy variables. The corresponding model coefficients and t statistics were written out in NIFTI format, and a whole brain voxel level FDR was utilized to correct for Type I error inflation due to multiple comparisons. Voxels displaying FDR-corrected Z-values < 1.96 (pFDR < 0.05) were then thresholded and clustered for subsequent individual-level extraction and visualization of effects in IBM SPSS version 21.0 [61]

Assessing baseline brain activation moderation of the sertraline vs. placebo effect

For each contrast of interest (iI-cI and I-C), first-level, spatially-normalized activation contrast maps from each individual were loaded into R[64] using the fmri package[65] (see example R syntax available in Supplementary Software). We constrained analyses to a search space defined by a whole-brain mask constructed from: a) a 30% probabilistic voxelwise threshold derived from spatially normalized, segmented binary gray matter images from an independent sample of healthy comparison participants[56,57]; and b) a conjunction map of adequate signal coverage across study scanners constructed by concatenating the binarized individual-level average map of all task experimental conditions within participants acquired at each scanner, thresholding at a minimum of 50% coverage within each scanner sample, and then conjoining these maps across all scanners. Thus, the final search space was a function of both adequate signal coverage across all study scanners as well as a probabilistic assignment to the brain's gray matter. At each voxel within this search space, the moderating effect of baseline brain activation (mean-centered) was examined using a longitudinal linear mixed model implemented in the nlme R package.[66]

Modeled effects included a random and fixed intercept, fixed effects of time, time × treatment arm, 4 mean-centered, dummy-coded scanner variables corresponding to: a) Columbia University scanner 1 vs. UT Southwestern scanner; b) Columbia University scanner 2 vs. UT Southwestern scanner; c) Massachusetts General Hospital scanner vs. UT Southwestern scanner; and d) University of Michigan scanner vs. UT Southwestern scanner, treatment arm × time × brain activation, and all model variables in interaction with each of the 4 site variables. See supplementary table 3 for all variables in the mixed models. The effect of interest was the treatment arm × time × brain activation moderation effect from this model, controlling for all lower-order and higher-order (i.e. multiplicative) interaction terms with each of the scanner variables to fully control for potential effects due to this source of heterogeneity. At each voxel, the number of degrees of freedom was individually-defined based upon the number of individuals displaying adequate signal coverage at that voxel (which tended to be less in areas prone to EPI BOLD signal dropout, particularly for certain scanners), with individuals displaying "0" values for activation being removed from the analysis. At each voxel, the relevant F tests corresponding to effects of interest were transformed to p-values and then z-values to nullify potential statistical variation across voxels secondary to varying degrees of freedom and signal coverage. These Z value maps for effects of interest were then written out to AFNI .BRIK files using the fmri package, and voxel-level FDR-correction within the whole-brain specified search space (see above) was conducted using AFNI's program 3dFDR. Voxels displaying FDR-corrected Z-values < 1.96 ($p_{FDR} < 0.05$) were then thresholded and clustered for subsequent individual-level extraction and visualization of effects in IBM SPSS version 21.0 [61].

Machine learning analyses

Relevance Vector Machines (RVM) were used to build a regression model for the prediction of pre- minus post-treatment change in $HAMD_{17}$ (see example Matlab syntax in Supplementary Software). By exploiting an automatic relevance determination prior to penalizing unnecessary complexity in the model, RVM is able to automatically determine the feature sparsity under a Bayesian evidence framework [67]. The effectiveness of the popular L1-regularization based sparse learning algorithms [68] depends on the selection of regularization hyperparameter to a large extent. Cross-validation on training set has been usually used to determine the most appropriate hyperparameter. However, additional validation data from the training set are required by cross-validation for hyperparameter selection, which is prone to overfitting since fewer training samples can be used for model calibration [69]. Different from the L1-regularization approaches, all model parameters in RVM can be efficiently estimated based on all available training data without the need of cross-validation. RVM therefore provides a more accurate estimation of the sparse solution compared with other L1-regularization based sparse learning algorithms [70], especially when a small training set is available. In the past decades, RVM has demonstrated its strength in various fields including EEG classification for brain-computer interface [71] and bioinformatics analysis of gene expression data [72,73].

RVM learning was conducted on data from patients who had baseline $HAMD_{17}$, passed the image and task accuracy quality control criteria for the baseline scan, had imaging data at baseline, and sufficient whole-brain coverage such that data were available from almost all

ROIs for all participants (sertraline N=115, placebo N=122). For participants lacking an endpoint $HAMD_{17}$, multiple imputation by chained equations was conducted in R^{64} using the package mice[74]. The following observed variables were utilized in order to impute endpoint $HAMD_{17}$ values for missing data via Bayesian regression: baseline $HAMD_{17}$, week 1 $HAMD_{17}$, week 2 $HAMD_{17}$, week 3 $HAMD_{17}$, week 4 $HAMD_{17}$, week 6 $HAMD_{17}$, baseline Quick Inventory of Depressive Symptoms (QIDS) total score, baseline Mood and Symptom Questionnaire subscale scores for Anxious Arousal, Anhedonic Depression, and General Distress, Snaith-Hamilton Pleasure Scale (SHAPS) total score, age, years of education, gender, and Wechsler Abbreviated Scale of Intelligence (WASI) t-scores for Vocabulary and Matrix Reasoning. Extractions were conducted on cortical regions of interest (ROIs), defined based on a recently-published cortical parcellation derived from applying a combination of local gradient analysis and global signal similarity on an independent resting-state fMRI cohort [75]. Since functional parcellations typically rely on resting-state connectivity patterns, which may or may not adequately describe activity patterns in the emotional conflict task, we pooled ROIs from the 200, 400, and 600 region parcellations in order to limit parcellation-related specificity. ROIs were mapped to seven previously identified functional networks based on the spatial overlap between each ROI and each network [75]. In addition to these cortical ROIs, subcortical ROIs included striatal[76] and cerebellar[77] parcellations based on the same seven functional networks, amygdala ROIs [78], anterior and posterior hippocampal ROIs[79] and the thalamus[80]. We then regressed imaging site out of these data using multiple linear regression within the training set at each round of the RVM model, and the residualized brain signals were then used for training the model.

A regression model was then built based using RVM learning to predict each individual's pre-minus-post change in $HAMD_{17}$ scores, wherein the most important ROI features were determined under a probabilistic framework by exploiting a separate Gaussian prior with a Gaussian likelihood function. All model parameters for controlling the sparsity of regression weights were automatically estimated through Bayesian machine learning based on each training set. The prediction performance was evaluated by 10×10 cross-validation (see supplementary figure 6). Specifically, for 10 repetitions, all participants were randomly divided into 10 folds, such that each participant was left out and used as a test set once while the remaining nine folds were used as a training set for RVM model learning. Each participant was left out exactly once after running each 10-fold cross-validation. By using the estimated regression weights of the RVM model, we computed the predicted symptom change value for each of the left-out participants by the weighted sum of the all 1235 ROI features. After repeating 10-fold cross-validation 10 times, we determined each participant's predicted symptom change by taking the median of the predicted values across each of the 10 times that participant was left out. Pearson's correlation coefficient was then computed by correlating the predicted symptom changes and the actual symptom changes across all participants. The outcome of the RVM was assessed by correlating model-predicted $HAMD_{17}$ change scores with observed/imputed change scores. Significant correlations were verified using 1,000 permutations of the RVM modeling conducted by randomly shuffling observed/imputed change scores across participants. Specificity of the model prediction was tested by applying, at each round of cross-validation, the regression model (with appropriate intercepts) to the data from the other treatment arm, which was summarized for each

participant by taking the median of the 10 rounds of cross-validation. We conducted similar RVM analyses on item-level clinical, historical, demographic and behavioral data to determine whether these easier-to-get variables could perform as well as brain activation in predicting treatment outcome. Included were the Spielberger State-Trait Anxiety Inventory [34], the Quick Inventory of Depressive Symptoms [33], the Mood and Anxiety Questionnaire [35], the Childhood Trauma Questionnaire [36], age, education and I-C and iI-cI reaction time and accuracy difference scores.

Data Availability Statement

All data are publicly available through the NIMH Data Archive (https://nda.nih.gov/edit_collection.html?id=2199).

Code Availability Statement

Custom code that supports the findings of this study is available in the Supplementary Software section.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Hasin DS et al. Epidemiology of Adult DSM-5 Major Depressive Disorder and Its Specifiers in the United States. JAMA Psychiatry, doi:10.1001/jamapsychiatry.2017.4602 (2018).

2. Lopez-Munoz F & Alamo C Monoaminergic neurotransmission: the history of the discovery of antidepressants from 1950s until today. Current pharmaceutical design 15, 1563–1586 (2009). [PubMed: 19442174]

3. Moore TJ & Mattison DR Adult Utilization of Psychiatric Drugs and Differences by Sex, Age, and Race. JAMA Intern Med 177, 274–275, doi:10.1001/jamainternmed.2016.7507 (2017). [PubMed: 27942726]

4. Kirsch I The Emperor's New Drugs: Exploding the Antidepressant Myth. (Random House, 2009).

5. Khan A & Brown WA Antidepressants versus placebo in major depression: an overview. World Psychiatry 14, 294–300, doi:10.1002/wps.20241 (2015). [PubMed: 26407778]

6. Kirsch I et al. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. PLoS Med 5, e45, doi:10.1371/journal.pmed.0050045 (2008). [PubMed: 18303940]

7. Fournier JC et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. Jama 303, 47–53, doi:10.1001/jama.2009.1943 (2010). [PubMed: 20051569]

8. 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, doi:10.1016/S0140-6736(17)32802-7 (2018).

9. Drysdale AT et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nature medicine 23, 28–38, doi:10.1038/nm.4246 (2017).

10. Kraemer HC Messages for Clinicians: Moderators and Mediators of Treatment Outcome in Randomized Clinical Trials. Am J Psychiatry 173, 672–679, doi:10.1176/appi.ajp.2016.15101333 (2016). [PubMed: 26988629]

11. Etkin A, Buchel C & Gross JJ The neural bases of emotion regulation. Nat Rev Neurosci 16, 693–700, doi:10.1038/nrn4044 (2015). [PubMed: 26481098]

12. Gross JJ Handbook of Emotion Regulation. (Guilford Press, 2014).

13. Gyurak A, Gross JJ & Etkin A Explicit and implicit emotion regulation: a dual-process framework. Cogn Emot 25, 400–412, doi:10.1080/02699931.2010.544160 (2011). [PubMed: 21432682]

14. Xu M, Xu G & Yang Y Neural Systems Underlying Emotional and Non-emotional Interference Processing: An ALE Meta-Analysis of Functional Neuroimaging Studies. Front Behav Neurosci 10, 220, doi:10.3389/fnbeh.2016.00220 (2016). [PubMed: 27895564]

15. Egner T, Etkin A, Gale S & Hirsch J Dissociable neural systems resolve conflict from emotional versus nonemotional distracters. Cereb Cortex 18, 1475–1484, doi:10.1093/cercor/bhm179 (2008). [PubMed: 17940084]

16. Etkin A, Egner T, Peraza DM, Kandel ER & Hirsch J Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala. Neuron 51, 871–882, doi:10.1016/j.neuron.2006.07.029 (2006). [PubMed: 16982430]

17. Maier ME & di Pellegrino G Impaired conflict adaptation in an emotional task context following rostral anterior cingulate cortex lesions in humans. Journal of cognitive neuroscience 24, 2070–2079, doi:10.1162/jocn_a_00266 (2012). [PubMed: 22721382]

18. Chechko N et al. Brain circuitries involved in emotional interference task in major depression disorder. J Affect Disord 149, 136–145, doi:10.1016/j.jad.2013.01.013 (2013). [PubMed: 23394712]

19. Chechko N et al. Unstable prefrontal response to emotional conflict and activation of lower limbic structures and brainstem in remitted panic disorder. PLoS One 4, e5537 (2009). [PubMed: 19462002]

20. Etkin A, Prater KE, Hoeft F, Menon V & Schatzberg AF Failure of anterior cingulate activation and connectivity with the amygdala during implicit regulation of emotional processing in generalized anxiety disorder. Am J Psychiatry 167, 545–554, doi:10.1176/appi.ajp.2009.09070931 (2010). [PubMed: 20123913]

21. Etkin A & Schatzberg AF Common abnormalities and disorder-specific compensation during implicit regulation of emotional processing in generalized anxiety and major depressive disorders. Am J Psychiatry 168, 968–978, doi:10.1176/appi.ajp.2011.10091290 (2011). [PubMed: 21632648]

22. Xue S, Wang S, Kong X & Qiu J Abnormal Neural Basis of Emotional Conflict Control in Treatment-Resistant Depression. Clin EEG Neurosci 48, 103–110, doi:10.1177/1550059416631658 (2017). [PubMed: 26892803]

23. Widge AS et al. Treating refractory mental illness with closed-loop brain stimulation: Progress towards a patient-specific transdiagnostic approach. Exp Neurol 287, 461–472, doi:10.1016/j.expneurol.2016.07.021 (2017). [PubMed: 27485972]

24. Gyurak A et al. Frontoparietal Activation During Response Inhibition Predicts Remission to Antidepressants in Patients With Major Depression. Biol Psychiatry 79, 274–281, doi:10.1016/j.biopsych.2015.02.037 (2016). [PubMed: 25891220]

25. Williams LM et al. Amygdala Reactivity to Emotional Faces in the Prediction of General and Medication-Specific Responses to Antidepressant Treatment in the Randomized iSPOT-D Trial. Neuropsychopharmacology 40, 2398–2408, doi:10.1038/npp.2015.89 (2015). [PubMed: 25824424]

26. Etkin A et al. A cognitive-emotional biomarker for predicting remission with antidepressant medications: a report from the iSPOT-D trial. Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology 40, 1332–1342, doi:10.1038/npp.2014.333 (2015). [PubMed: 25547711]

27. Langenecker SA et al. Frontal and limbic activation during inhibitory control predicts treatment response in major depressive disorder. Biol Psychiatry 62, 1272–1280 (2007). [PubMed: 17585888]

28. Pizzagalli DA Frontocingulate dysfunction in depression: toward biomarkers of treatment response. Neuropsychopharmacology 36, 183–206 (2011). [PubMed: 20861828]

29. Trivedi MH et al. Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): Rationale and design. Journal of psychiatric research 78, 11–23, doi:10.1016/j.jpsychires.2016.03.001 (2016). [PubMed: 27038550]

30. Goldstein-Piekarski AN et al. Intrinsic functional connectivity predicts remission on antidepressants: a randomized controlled trial to identify clinically applicable imaging biomarkers. Transl Psychiatry 8, 57 (2018). [PubMed: 29507282]

31. Etkin A & Schatzberg AF Common abnormalities and disorder-specific compensation during implicit regulation of emotional processing in generalized anxiety and major depressive disorders. Am J Psychiatry 168, 968–978, doi:10.1176/appi.ajp.2011.10091290 (2011). [PubMed: 21632648]

32. Pizzagalli D et al. The incremental predictive validity of rostral anterior cingulate cortex activity in relation to symptom improvement in depression: A randomized clinical trial. JAMA Psychiatry (in press).

33. Rush AJ 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). [PubMed: 12946886]

34. Spielberger CD, Gorsuch RL, Lushene R, Vagg PR & Jacobs GA Manual for the State-Trait Anxiety Inventory. (Consulting Psychologists Press, 1983).

35. Wardenaar KJ et al. Development and validation of a 30-item short adaptation of the Mood and Anxiety Symptoms Questionnaire (MASQ). Psychiatry research 179, 101–106, doi:10.1016/j.psychres.2009.03.005 (2010). [PubMed: 20472297]

36. Bernstein DP et al. Initial reliability and validity of a new retrospective measure of child abuse and neglect. Am J Psychiatry 151, 1132–1136, doi:10.1176/ajp.151.8.1132 (1994). [PubMed: 8037246]

37. Kerns JG et al. Anterior cingulate conflict monitoring and adjustments in control. Science 303, 1023–1026, doi:10.1126/science.1089910 (2004). [PubMed: 14963333]

38. Korb AS, Hunter AM, Cook IA & Leuchter AF Rostral anterior cingulate cortex theta current density and response to antidepressants and placebo in major depression. Clin Neurophysiol 120, 1313–1319, doi:10.1016/j.clinph.2009.05.008 (2009). [PubMed: 19539524]

39. Clayson PE & Larson MJ Adaptation to emotional conflict: evidence from a novel face emotion paradigm. PLoS One 8, e75776 (2013). [PubMed: 24073278]

40. Tang D, Hu L, Chen A, Clayson PE & Larson MJ The neural oscillations of conflict adaptation in the human frontal region. Biol Psychol 93, 364–372, doi:10.1016/j.biopsycho.2013.03.004 (2013). [PubMed: 23570676]

41. Larson MJ, Clawson A, Clayson PE & Baldwin SA Cognitive conflict adaptation in generalized anxiety disorder. Biol Psychol 94, 408–418, doi:10.1371/journal.pone.007577610.1016/j.biopsycho.2013.08.006 (2013). [PubMed: 24005064]

42. Suzuki K & Shinoda H Transition from reactive control to proactive control across conflict adaptation: An sLORETA study. Brain Cogn 100, 7–14, doi:10.1016/j.bandc.2015.09.001 (2015). [PubMed: 26432378]

43. Voigt J, Carpenter L & Leuchter A Cost effectiveness analysis comparing repetitive transcranial magnetic stimulation to antidepressant medications after a first treatment failure for major

depressive disorder in newly diagnosed patients - A lifetime analysis. PLoS One 12, e0186950, doi:10.1371/journal.pone.0186950 (2017). [PubMed: 29073256]

44. Nguyen KH & Gordon LG Cost-Effectiveness of Repetitive Transcranial Magnetic Stimulation versus Antidepressant Therapy for Treatment-Resistant Depression. Value Health 18, 597–604, doi:10.1016/j.jval.2015.04.004 (2015). [PubMed: 26297087]

45. O'Reardon JP et al. Efficacy and safety of transcranial magnetic stimulation in the acute treatment of major depression: a multisite randomized controlled trial. Biol Psychiatry 62, 1208–1216, doi:10.1016/j.biopsych.2007.01.018 (2007). [PubMed: 17573044]

46. George MS et al. Daily left prefrontal transcranial magnetic stimulation therapy for major depressive disorder: a sham-controlled randomized trial. Arch Gen Psychiatry 67, 507–516, doi:10.1001/archgenpsychiatry.2010.46 (2010). [PubMed: 20439832]

47. Pescosolido BA et al. "A disease like any other"? A decade of change in public reactions to schizophrenia, depression, and alcohol dependence. Am J Psychiatry 167, 1321–1330, doi:10.1176/appi.ajp.2010.09121743 (2010). [PubMed: 20843872]

48. Insel TR & Cuthbert BN Medicine. Brain disorders? Precisely. Science 348, 499–500, doi:10.1126/science.aab2358 (2015). [PubMed: 25931539]

49. First M, Spitzer R, Gibbon M & William J Structured Clinical Interview fpr DSM-IV-TR Axis 1 disorders, Research Version, Patient Edition (SCID-I/P) New York, NY: New York State Psychiatric Institute, Biometric research; 2002 73 Maxwell, M. Bethesda, MD (1992).

50. Ekman P & Friesen WV Pictures of Facial Affect. (Consulting Psychologists, 1976).

51. Gratton G, Coles MG & Donchin E Optimizing the use of information: strategic control of activation of responses. J Exp Psychol Gen 121, 480–506 (1992). [PubMed: 1431740]

52. Egner T & Hirsch J Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. Nat Neurosci 8, 1784–1790, doi:10.1038/nn1594 (2005). [PubMed: 16286928]

53. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW & Smith SM FSL. Neuroimage 62, 782–790, doi:10.1016/j.neuroimage.2011.09.015 (2012). [PubMed: 21979382]

54. Smith SM et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23 Suppl 1, S208–219, doi:10.1016/j.neuroimage.2004.07.051 (2004). [PubMed: 15501092]

55. Friston KJ et al. Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp 2, 189–210 (1995).

56. Fonzo GA et al. PTSD Psychotherapy Outcome Predicted by Brain Activation During Emotional Reactivity and Regulation. Am J Psychiatry 174, 1163–1174, doi:10.1176/appi.ajp.2017.16091072 (2017). [PubMed: 28715908]

57. Fonzo GA et al. Selective Effects of Psychotherapy on Frontopolar Cortical Function in PTSD. Am J Psychiatry 174, 1175–1184, doi:10.1176/appi.ajp.2017.16091073 (2017). [PubMed: 28715907]

58. Fox J Applied regression analysis and generalized linear models, 2nd ed. (Sage Publications, Inc, 2008).

59. Loy A, Hofmann H & Cook D Model Choice and Diagnostics for Linear Mixed-Effects Models Using Statistics on Street Corners. Journal of Computational and Graphical Statistics 26, 478–492, doi:10.1080/10618600.2017.1330207 (2017).

60. psych: Procedures for Psychological, Psychometric, and Personality Research v. R package version 1.8.12 (Northwestern University, Evanston, Illinois, 2018).

61. IBM SPSS Statistics for Macintosh, Version 21.0 (IBM Corp, Armonk, NY, 2012).

62. Huber PJ Robust Regression: Asymptotics, Conjectures and Monte Carlo. The Annals of Statistics 1, 799–821 (1973).

63. Wager TD, Keller MC, Lacey SC & Jonides J Increased sensitivity in neuroimaging analyses using robust regression. Neuroimage 26, 99–113, doi:10.1016/j.neuroimage.2005.01.011 (2005). [PubMed: 15862210]

64. R: A language and environment for statistical computing. v. 3.2.3 (R Foundation for Statistical Computing, Vienna, Austria, 2015).

65. Tabelow K & Polzehl J Statistical Parametric Maps for Functional MRI Experiments in R: The Package fmri. 2011 44, 21, doi:10.18637/jss.v044.i11 (2011).

66. nlme : Linear and Nonlinear Mixed Effects Models (2015).

67. Tipping ME Sparse Bayesian Learning and the Relevance Vector Machine. Journal of Machine Learning Research 1, 211–244 (2001).

68. Tibshirani R Regression shrinkage and selection via the Lasso. J Roy Stat Soc B Met 58, 267–288 (1996).

69. Cawley GC & Talbot NL Preventing over-fitting during model selection via Bayesian regularisation of the hyper-parameters. Journal of Machine Learning Research 8, 841–861 (2007).

70. Wipf DP & Rao BD Sparse Bayesian Learning for Basis Selection. IEEE Transactions on Signal Processing 52, 2153–2164 (2004).

71. Zhang Y et al. Sparse Bayesian classification of EEG for brain-computer interface. IEEE Transactions on Neural Networks and Learning Systems 27, 2256–2267 (2016). [PubMed: 26415189]

72. Cawley GC & Talbot NL Gene selection in cancer classification using sparse logistic regression with Bayesian regularization. Bioinformatics 22, 2348–2355 (2006). [PubMed: 16844704]

73. Li Y, Campbell C & Tipping M Bayesian automatic relevance determination algorithms for classifying gene expression data. Bioinformatics 18, 1332–1339 (2002). [PubMed: 12376377]

74. van Buuren S & Groothuis-Oudshoorn K mice: Multivariate Imputation by Chained Equations in R. 2011 45, 67, doi:10.18637/jss.v045.i03 (2011).

75. Schaefer A et al. Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. Cereb Cortex, 1–20, doi:10.1093/cercor/bhx179 (2017). [PubMed: 28365777]

76. Choi EY, Yeo BT & Buckner RL The organization of the human striatum estimated by intrinsic functional connectivity. J Neurophysiol 108, 2242–2263, doi:10.1152/jn.00270.2012 (2012). [PubMed: 22832566]

77. Buckner RL, Krienen FM, Castellanos A, Diaz JC & Yeo BT The organization of the human cerebellum estimated by intrinsic functional connectivity. J Neurophysiol 106, 2322–2345, doi: 10.1152/jn.00339.2011 (2011). [PubMed: 21795627]

78. Patenaude B, Smith SM, Kennedy DN & Jenkinson M A Bayesian model of shape and appearance for subcortical brain segmentation. Neuroimage 56, 907–922, doi:10.1016/j.neuroimage. 2011.02.046 (2011). [PubMed: 21352927]

79. Chen AC & Etkin A Hippocampal network connectivity and activation differentiates post-traumatic stress disorder from generalized anxiety disorder. Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology 38, 1889–1898, doi:10.1038/npp.2013.122 (2013). [PubMed: 23673864]

80. Behrens TE et al. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nat Neurosci 6, 750–757, doi:10.1038/nn1075 (2003). [PubMed: 12808459]
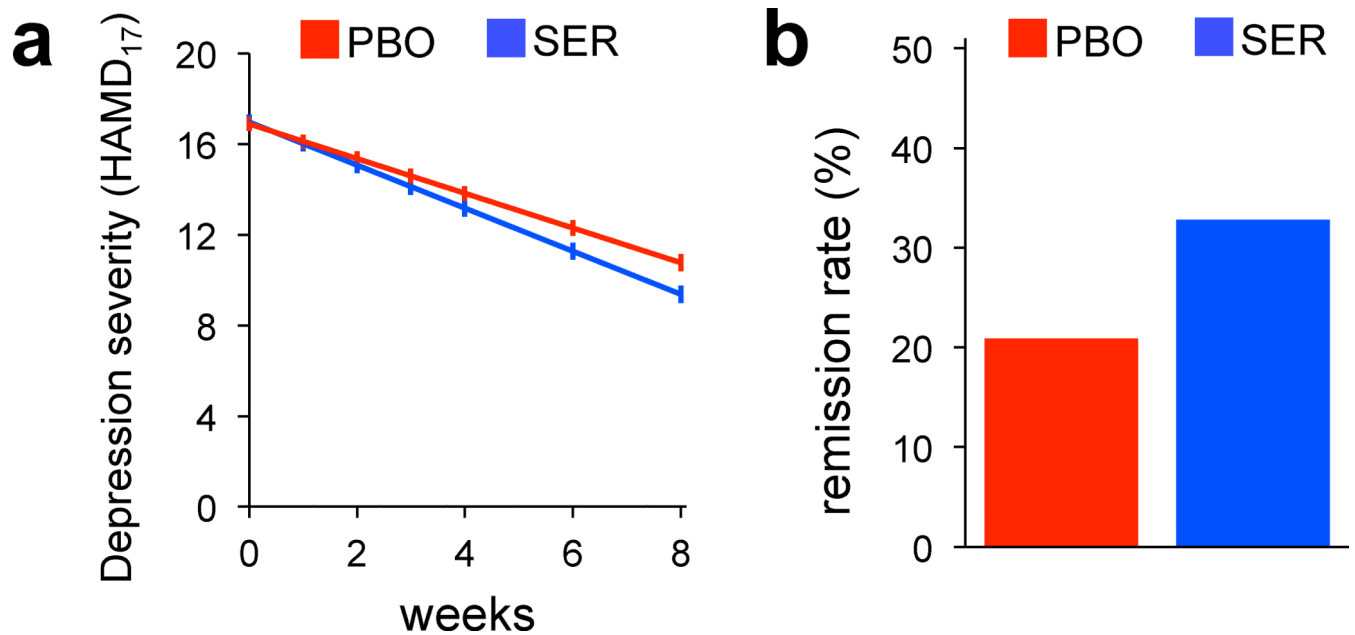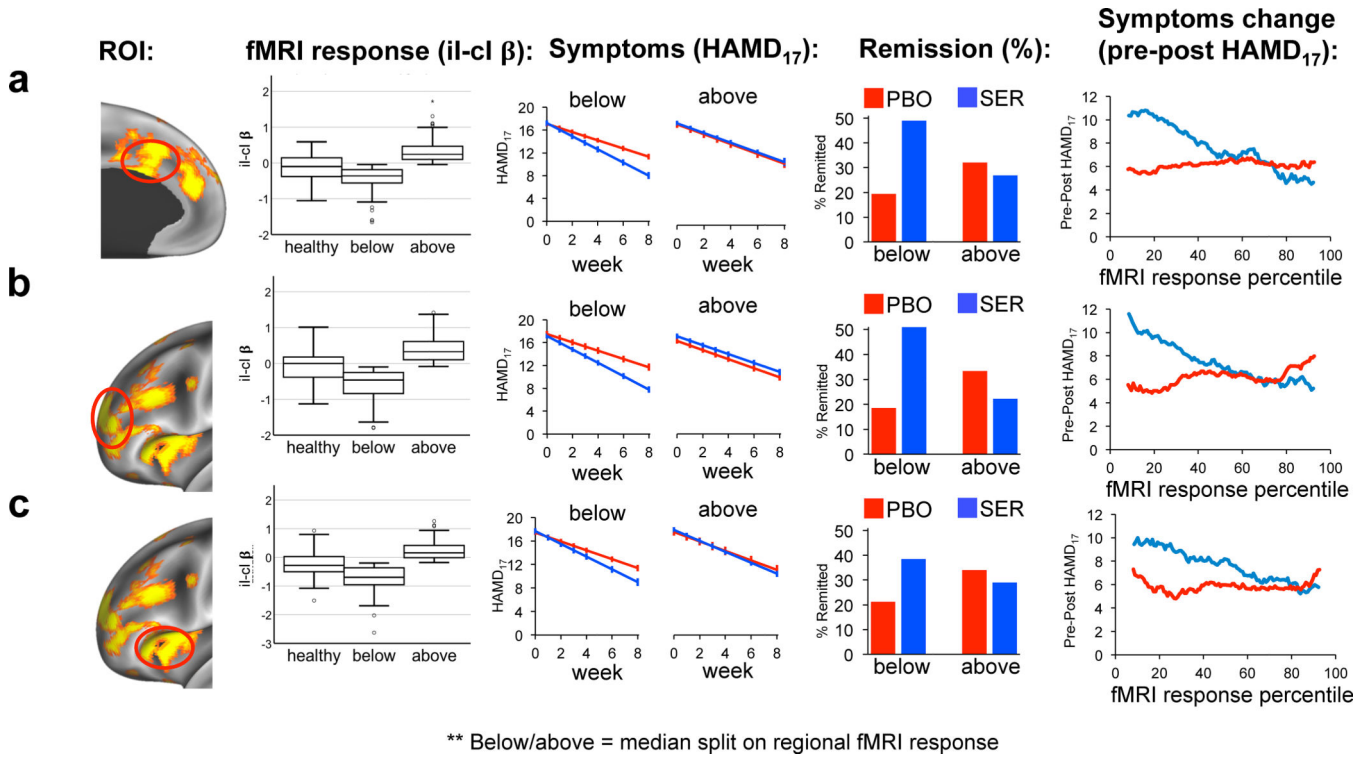
Figure 1:

Response across all patients to sertraline (SER; n=122) versus placebo (PBO; n=129) independent of emotional conflict task brain activation. Shown are: (a) predicted $HAMD_{17}$ scores from the linear mixed model (treatment arm by time interaction $F(1,1471)=4.6$, $p=0.032$; Cohen's $d=0.27$, 95% confidence intervals = 0.02, 0.52), and (b) remission rates ($HAMD_{17}$ score at study endpoint < 7) (number needed to treat (NNT)=8.4) Note no data distributions are present for the bar plots as these are simply point estimates indicating percent of sample remitted ($HAMD_{17} < 7$) at endpoint. Error bars on line graphs show +/− 1 standard error of the mean. See Supplementary Figure 2 for a depiction of these data utilizing box and whisker plots.
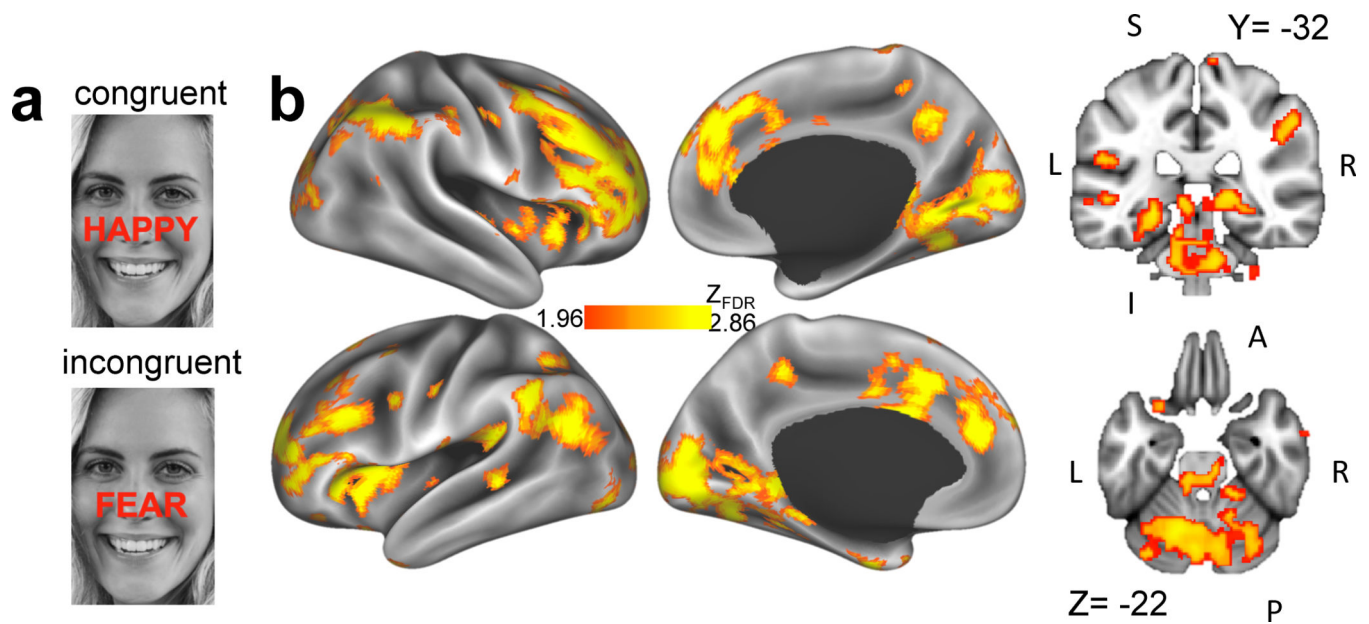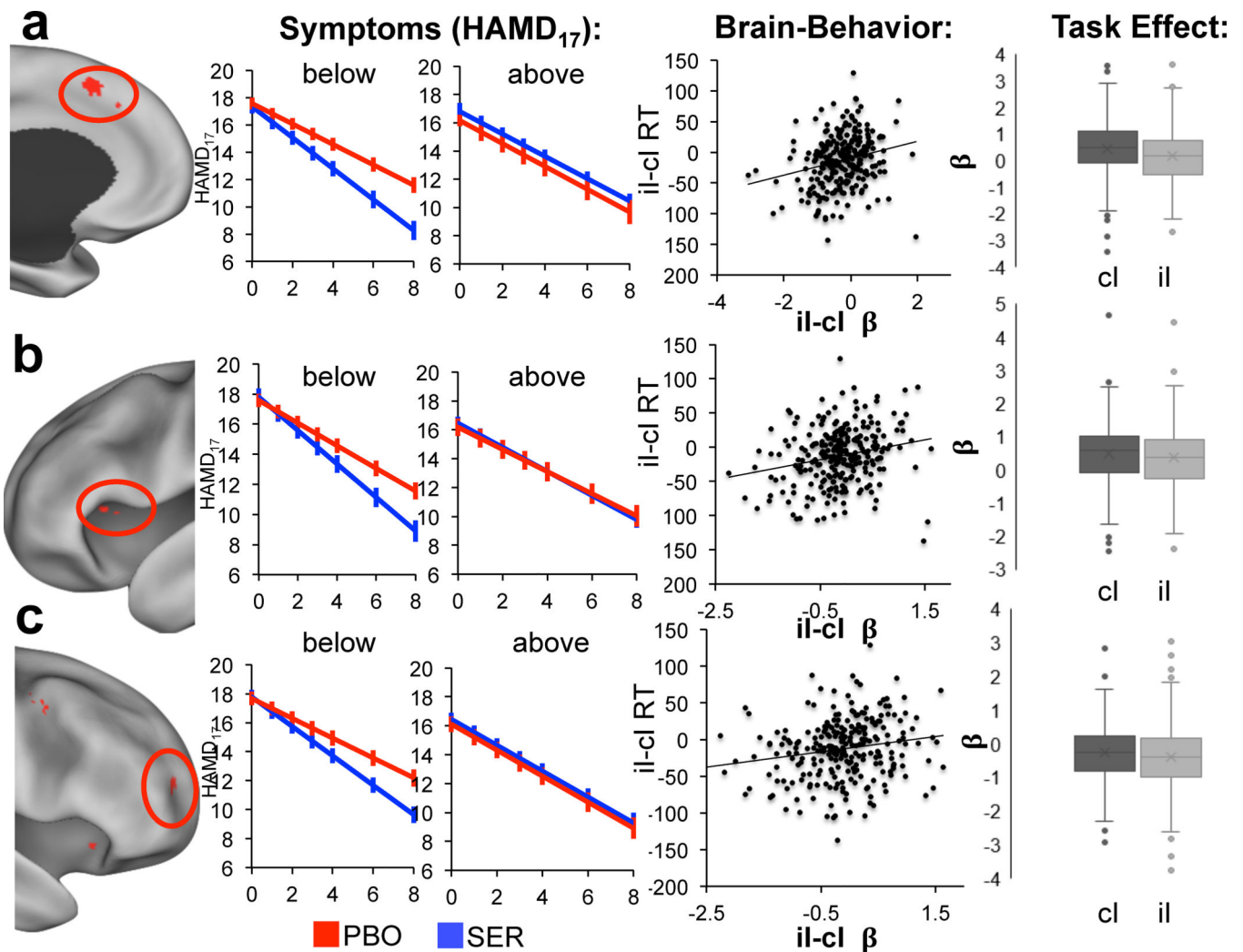
Figure 2:

Moderation by emotional conflict regulation-related brain activity between sertraline (n=122) versus placebo (n=129). (a) Example congruent and incongruent stimuli. (b) Whole-brain significant moderation results (color map shows whole brain voxel level false discovery rate corrected q < 0.05 effects for the treatment arm × time × brain activation interaction effect from the linear mixed model on a template brain surface and in cross-sectional brain slices of the MNI average anatomical, displayed as FDR z-transformed F values, which are thus all positive). Effects by treatment arm for individual peaks are reported in supplementary tables 4–6. A=anterior; L=left; I=inferior; P=posterior; R=right; S=superior.

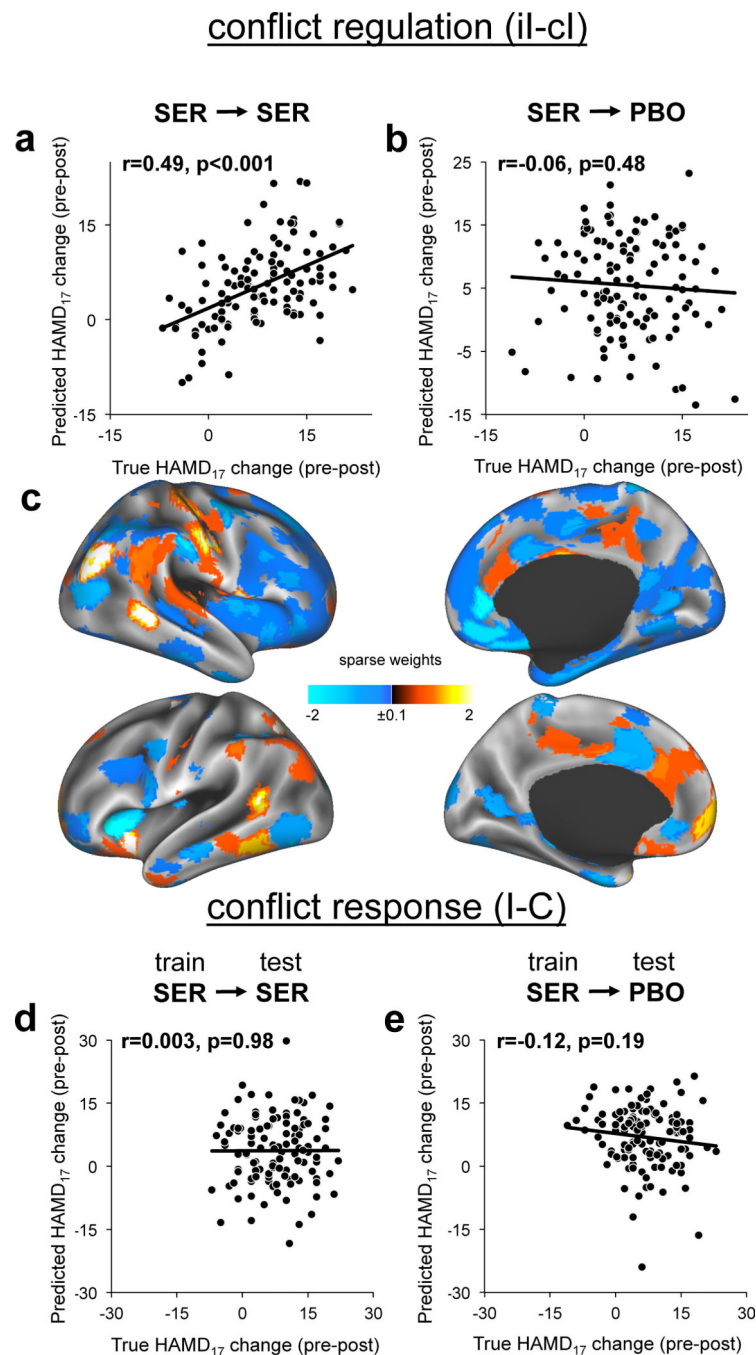** Below/above = median split on regional fMRI response

Figure 3:

Moderation by emotional conflict regulation-related brain activity between sertraline (n=122) versus placebo (n=129). (a-c) Shown are: 1) the region of interest whose effects are shown (left-hand column). 2) its fMRI signal (iI-cI contrast values) during conflict regulation comparing patients below and above the median of fMRI activity in that region, as well as healthy individuals for comparison (second column). 3) weekly depression severity (predicted $HAMD_{17}$ scores from the linear mixed models) for patients below and above the median of fMRI activity in that region. 4) predicted remission rates ($HAMD_{17} < 7$ at study endpoint) (for patients below and above the median of fMRI activity in that region, split by sertraline (SER) and placebo (PBO; third column). Note that this median split is utilized to illustrate effects detected by the linear mixed model analyses of brain activation as a continuous measure and is not statistically meaningful in itself. 5) pre-minus-post predicted symptom change in sliding window 20-patient bin averages illustrates the continuous nature of the moderation effect across different degrees of brain activation and complement the median split plots (x-axis reflects percentile for the center of the 20-patient bin with respect to fMRI signal in the regions noted). Regions shown here are: (a) dorsal anterior cingulate (below median group remission NNT=3.4, endpoint symptom difference Cohen's d=0.76), (b) left frontopolar cortex (NNT=3.1, d=0.86), and (c) left anterior insula (NNT=5.8, d=0.58). For the box and whisker plots, the center line indicates the median; the lower bound of the box indicates the median of the bottom 50% (first quartile), while the top bound of the box indicates the median of the top 50% (third quartile); the whiskers indicate the lower and upper ends of observed data values within 1.5 times the interquartile range below the first quartile (lower whisker) and above the third quartile (upper whisker); dots indicate outliers, i.e. observed data points that lie outside the bounds established by 1.5

times the interquartile range below the first quartile (dots below the bottom whisker) or above the third quartile (dots above the top whisker). For line plots, error bars show +/− 1 standard error of the mean. See Supplementary Figure 3 for depiction with box and whisker plots.

Figure 4.
Conjunction of emotional conflict regulation task effect, brain-behavior relationship effect, and treatment moderation effect in a) the dorsal cingulate/dorsomedial prefrontal cortex; b) the left anterior insula; and c) the right frontopolar cortex. Shown in columns from left to right are: 1) the region of interest whose effects are shown (left-hand column); 2) weekly depression severity (predicted $HAMD_{17}$ scores from the mixed models) for patients below and above the median of fMRI activity in that region, split by sertraline (SER; n=122) and placebo (PBO; n=129). Note that this median split is utilized to illustrate effects detected by the linear mixed model analyses of brain activation as a continuous measure and is not statistically meaningful in itself (no statistical inference utilized here as to avoid circular analyses); 3) continuous relationship between iI-cI fMRI activity (x-axis) in the depicted region and iI-cI reaction time behavior differences (y-axis), with a linear trend line, across the entire randomized depressed sample to display the positive relationship (no statistical test utilized here, as to avoid circular analyses); and 4) mean fMRI activity for iI and cI conditions in the depicted region at baseline across the entire randomized sample. For the box and whisker plots, the center line indicates the median; the lower bound of the box indicates the median of the bottom 50% (first quartile), while the top bound of the box

indicates the median of the top 50% (third quartile); the whiskers indicate the lower and upper ends of observed data values within 1.5 times the interquartile range below the first quartile (lower whisker) and above the third quartile (upper whisker); dots indicate outliers, i.e. observed data points that lie outside the bounds established by 1.5 times the interquartile range below the first quartile (dots below the bottom whisker) or above the third quartile (dots above the top whisker). For line plots, error bars show +/− 1 standard error of the mean.

conflict regulation (iI-cI)

conflict response (I-C)

Figure 5:

Machine learning selectively predicts response to sertraline (n=122) using conflict regulation-related fMRI activity. (a) A machine learing model trained (with 10-fold cross-validation) on outcome with sertraline using conflict regulation activation (iI-cI) showed a strong correlation (r=0.49, p<0.001) between predicted changes in $HAMD_{17}$ (for 10×10-fold cross-validation samples) and observed changes. (b) When applied to patients receiving placebo (n=129), the sertraline model failed to predict placebo outcome (r=−0.06, p=0.48), illustrating its sertraline-specificity. (c) Non-zero regional weights contributing to the

sertraline model. (d) Training a machine learning model on outcome with sertraline (n=122) using conflict response activation (I-C) does not yield statistically significant prediction of sertraline outcome (r=0.003, p=0.98), and (e) does not yield statistically significant prediction (r=−0.12, p=0.19) of placebo outcome (n=129).