nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-019-0732-1

Brain regulation of emotional conflict predicts antidepressant treatment response for depression

Gregory A. Fonzo[1,9], Amit Etkin[2,3,4,9*], Yu Zhang[2,3,4,9], Wei Wu[2,3,4], Crystal Cooper[5], Cherise Chin-Fatt[5], Manish K. Jha[5], Joseph Trombello[5], Thilo Deckersbach[6], Phil Adams[7], Melvin McInnis[8], Patrick J. McGrath[7], Myrna M. Weissman[7], Maurizio Fava[6] and Madhukar H. Trivedi[5]

The efficacy of antidepressant treatment for depression is controversial due to the 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 conflict, in which greater downregulation of conflict-responsive regions predicted better sertraline outcomes. Treatment-predictive machine learning using brain metrics outperformed a model trained on clinical and demographic variables. Our findings demonstrate that 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.

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 United States takes an antidepressant[3]. Nonetheless, large meta-analyses have found only a small overall advantage of antidepressants over placebo (Cohen's $d \approx 0.3$) when used in an unselected population of patients with depression, with clinical significance only in the most severe patients[4–8], a severity level the vast majority of patients with depression 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 affects antidepressant treatment. Specifically, we tested whether the response to or regulation of emotional conflict—a neural function relevant for depression—could elicit phenotypic brain characteristics that differentially predict treatment outcome with an antidepressant versus with placebo (that is, 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 (for example, 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 favourable antidepressant response, thereby yielding both mechanistic insight and predictive biomarkers. Here, we sought to determine whether variation in treatment outcome among patients with depression 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.

[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]Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA, USA. [4]Sierra Pacific Mental Illness Research, Education and Clinical Center in the Veterans Affairs Palo Alto Healthcare System, 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 and 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. [9]These authors contributed equally: Gregory A. Fonzo, Amit Etkin, Yu Zhang. *e-mail: amitetkin@stanford.edu

Previous 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 previous studies lacked a placebo control arm, which is a critical limitation. Thus, it is unknown whether previous findings are specific to medication effects (that is, they may moderate between antidepressants and placebo) or are non-specifically predictive across medication and placebo interventions. For example, extensive previous 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. Previous work also suggests that treatment-predictive brain signals often do not differ between patients and healthy controls (that is, they do not demonstrate diagnosis-wise differences)[24,25,30]. As such, we also compared our moderators between patients with depression 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) of depression to date and the only one using fMRI measures. EMBARC randomized 309 medication-free outpatients with depression to receive either the selective serotonin reuptake inhibitor sertraline or placebo for 8 weeks (Supplementary Fig. 1). In this unique RCT, analysis of pretreatment 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 with a cross-validated machine-learning analysis, also on whole-brain data, 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 previous 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 × time effect: $F(1,1471)=4.6$, $P=0.032$; Cohen's $d=0.27$; 95% confidence interval,0.02–0.52), consistent with expectations from previous meta-analyses[4–8] (Fig. 1 and Supplementary Fig. 2).

Emotional conflict task behavioural effects. We next assessed the reaction times from the emotional conflict task to verify the expected patterns of effects. As detailed in previous 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 slowdown in reaction times relative to congruent trials in which 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 previous trial conflict, which then benefits regulation of emotional conflict on the subsequent trial (captured by the postincongruent incongruent (iI) minus postcongruent incongruent (cI) contrast; see Supplementary Methods). The task produced the expected overall behavioural effects. Conflict resulted in a reaction time slowdown (incongruent minus congruent (I–C) difference: $M=58.3$ ms, s.d. 40.8; one-sample t-test, $t=25.6$; $P<0.001$, Cohen's $d=1.43$; 95% confidence interval, 1.25–1.60), and conflict regulation resulted in mitigation of that slowdown (iI–cI difference: $M=-16.6$ ms, s.d. 48.4; $t=6.1$; $P<0.001$; Cohen's $d=-0.34$; 95% confidence interval,−0.47 to −0.22).

Emotional conflict task-related brain activation. We began our investigation of the imaging data by examining task-dependent main effects across the entire randomized depressed sample to verify that the task was inducing the expected patterns of brain activity. As expected from previous work[16], the main effect of the brain response to conflict (voxel-level one-sample t-test of I–C contrast values with voxel-level false discovery rate (FDR) correction of $q<0.05$ within whole-brain grey 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[21], even in the absence of a deficit in conflict regulatory behaviour. Consistent with previous 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 versus cI (il–cl) trials with voxel-level FDR correction of $q<0.05$ within whole-brain mask; Supplementary Table 4) when contrasting activation for 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–cI trials, consistent with previous work noting a deficit in this response in major depression[21].

Relationships between emotional conflict regulatory behaviour and brain activity. Previous work has demonstrated that individual differences in emotional conflict regulatory behaviour (that is, 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[20,21], reflecting a coherent relationship between brain and behaviour in this context. We therefore attempted to use these demonstrated brain-behaviour relationships between individual differences in emotional conflict regulatory behaviour and concomitant brain activation to better ground any observed moderating effects of brain activation on sertraline versus 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 behavioural significance. Thus, we examined how individual differences in emotional conflict regulation behaviour (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,

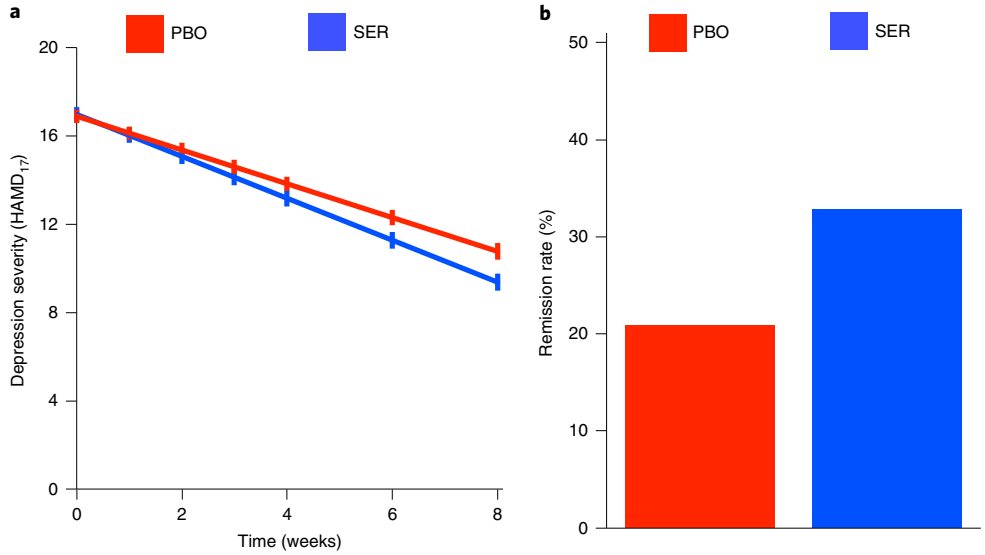

Fig. 1 | Response across all patients to sertraline ($n = 122$) versus placebo ($n = 129$) independent of emotional conflict task brain activation. a, Predicted $HAMD_{17}$ scores from the linear mixed model (treatment arm by time interaction $F(1, 1{,}471) = 4.6$, $P = 0.032$; Cohen's $d = 0.27$; 95% confidence interval, 0.02–0.52). **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 percentage of sample remitted ($HAMD_{17} < 7$) at endpoint. Error bars on line graphs show ±1 s.e.m. See Supplementary Fig. 2 for a depiction of these data with box and whisker plots. SER, sertraline; PBO, placebo.

controlling for site/scanner effects. Using a voxel-wise robust linear model to relate an individual subject's 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 $q < 0.05$ for iI–cI reaction time robust linear model coefficient) with a greater dampening (that is, 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 (that is, less negative and/or more positive iI–cI contrast values) was associated with better behavioural regulation of emotional conflict (that is, more negative iI–cI mean reaction time differences). This finding parallels previous work demonstrating a unique role of ventral prefrontal activation in mediating successful emotional conflict regulation[15,16]. Thus, more successful emotional conflict regulation behaviour in major depression is associated with a more pronounced canonical (that is, 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 behavioural measures of interest for capacity to moderate antidepressant treatment response: the response to conflict (I–C contrast; Fig. 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 (Fig. 2b and Supplementary Tables 7–10). To easily convey some of the linear mixed model results, we divided the depressed sample with a median split based on iI–cI activation levels for several brain regions (that is, as an illustrative but not a statistically meaningful split). Doing so revealed that a more favourable 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 (Fig. 3a), frontopolar cortex (Fig. 3b) and anterior insula (Fig. 3c) (see also Supplementary Fig. 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-behaviour relationships between emotional conflict regulatory behaviour and task-related activation (see Fig. 4 for the conjunction of brain maps and Supplementary Fig. 4 for the union of brain maps; all used 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 (that is, 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 (that is, had larger F statistics for the treatment arm × time × brain activation effect) for brain activity during the iI than the the cI condition (see Supplementary Fig. 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 patients responsive to sertraline was consistent with, and for some patients of greater magnitude, that seen in healthy individuals (Fig. 3a–c), indicating that these characteristics probably 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: $F = 0.48$, $P = 0.49$).

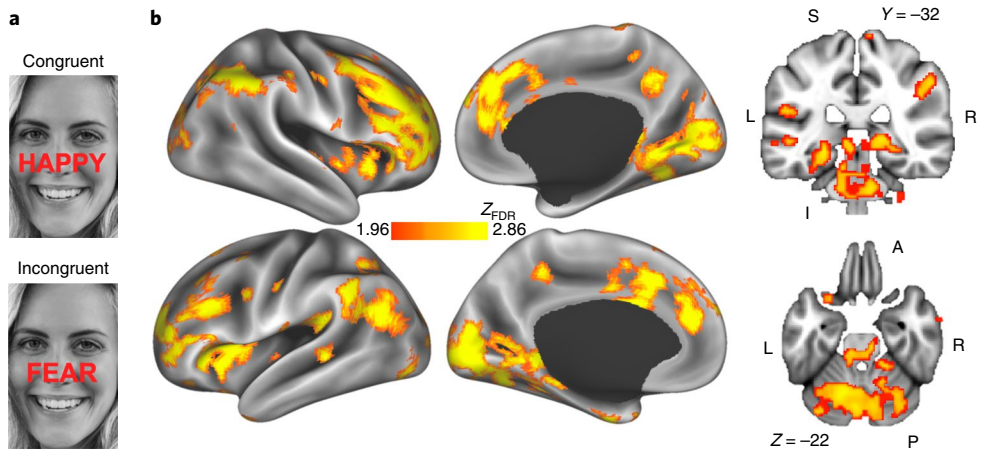

Fig. 2 | Moderation by emotional-conflict-regulation-related brain activity of the antidepressant effect of sertraline ($n = 122$) vs placebo ($n = 129$). a, Example congruent and incongruent stimuli. **b**, Whole-brain significant moderation results (colour map shows whole-brain voxel-level FDR-corrected $q < 0.05$ 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 Montreal Neurological Institute (MNI) standard-space T1-weighted average anatomical map, 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. Credit: **a**, Photo by Nick Huizinga on Unsplash

Interestingly, we found no evidence that activation in the rostral anterior cingulate during conflict regulation moderated treatment outcome (whole-brain voxel FDR-corrected $q > 0.05$). However, we found that iI–cI activation in the rostral anterior cingulate predicted outcome similarly across both arms (whole-brain voxel-level FDR-corrected $q < 0.05$ 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 Fig. 6).

In our second a priori brain measure of interest, responses to emotional conflict (I–C contrast) also yielded significant whole-brain moderation effects (whole-brain voxel-level FDR-corrected $q < 0.05$ 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 Fig. 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: $F < 1.7$, $P > 0.19$).

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 (RVMs) 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 Fig. 8). This assessment was done in a data-driven manner by use of a parcellation of the cortex and subcortex and a (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 with 1,000 permutations $P < 0.001$; Fig. 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: $r = -0.06$, $P = 0.48$; Fig. 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-predicted symptom changes and observed symptom changes in either the placebo or sertraline arms ($r = 0.11$, $P > 0.20$; Supplementary Fig. 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[31].

Regional weights driving the sertraline-predicting model are shown in Fig. 5c; they demonstrate a similar spatial pattern to that of the whole-brain moderation analysis in Fig. 1b. Although positive and negative signs on feature weights cannot be easily interpreted as reflecting a per unit positive or negative relationship with treatment response, respectively, because of 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; $z = 4.0$, $P < 0.001$; Fig. 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 but yielded drastically different results.

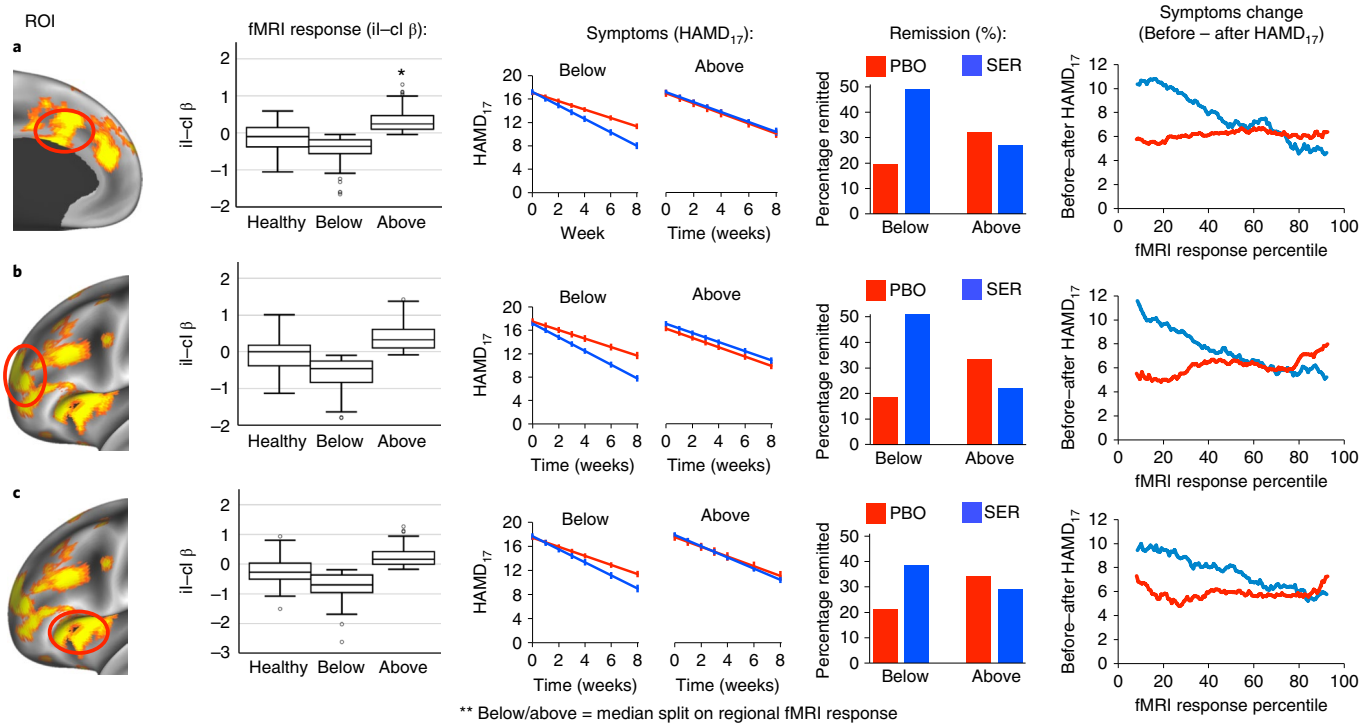

Fig. 3 | Decomposing emotional-conflict-regulation-related brain activity moderation of the antidepressant effect of sertraline ($n = 122$) vs placebo ($n = 129$). **a**–**c**, Shown in columns from left to right: the ROI whose effects are shown (left-hand column). Second column: its fMRI signal (iI–cI contrast values, i.e., beta coefficient) during conflict regulation comparing patients below and above the median of fMRI activity in that region, as well as healthy individuals for comparison. Third column: 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. Fourth column: 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)). Note that this median split is used to illustrate effects detected by the linear mixed model analyses of brain activation as a continuous measure and is not statistically meaningful in itself. Last column: before minus after 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 complements the median split plots (*x* axis reflects percentile for the centre of the 20-patient bin with respect to fMRI signal in the regions noted). **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, $d = 0.86$). **c**, Left anterior insula (NNT = 5.8, $d = 0.58$). For the box and whisker plots, the centre 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, that is, 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 s.e.m.. See Supplementary Fig. 3 for depiction with box and whisker plots.

Relationship of treatment-predictive emotional conflict regulation signals with clinical severity measures and patient-control differences. 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)[32], the Spielberger State-Trait Anxiety Inventory[33], as well as the anxious arousal, anhedonic depression and general distress subscales of the Mood and Anxiety Symptom Questionnaire[34]. However, no significant relationships survived Type I error correction (all whole-brain voxel-level FDR-corrected robust linear model coefficient $q > 0.05$ 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, Spielberger State-Trait Anxiety Inventory or Mood and Anxiety Symptom Questionnaire subscale scores ($r < 0.10$, $P > 0.12$).

We also found no significant differences between patients with depression (as a whole) and healthy controls in conflict regulation-related brain activation (whole-brain threshold free cluster enhancement with 5,000 permutation tests: $P > 0.05$). 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 versus 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 behaviour, demographic variables or historical factors like childhood trauma exposure[35] 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-pre-

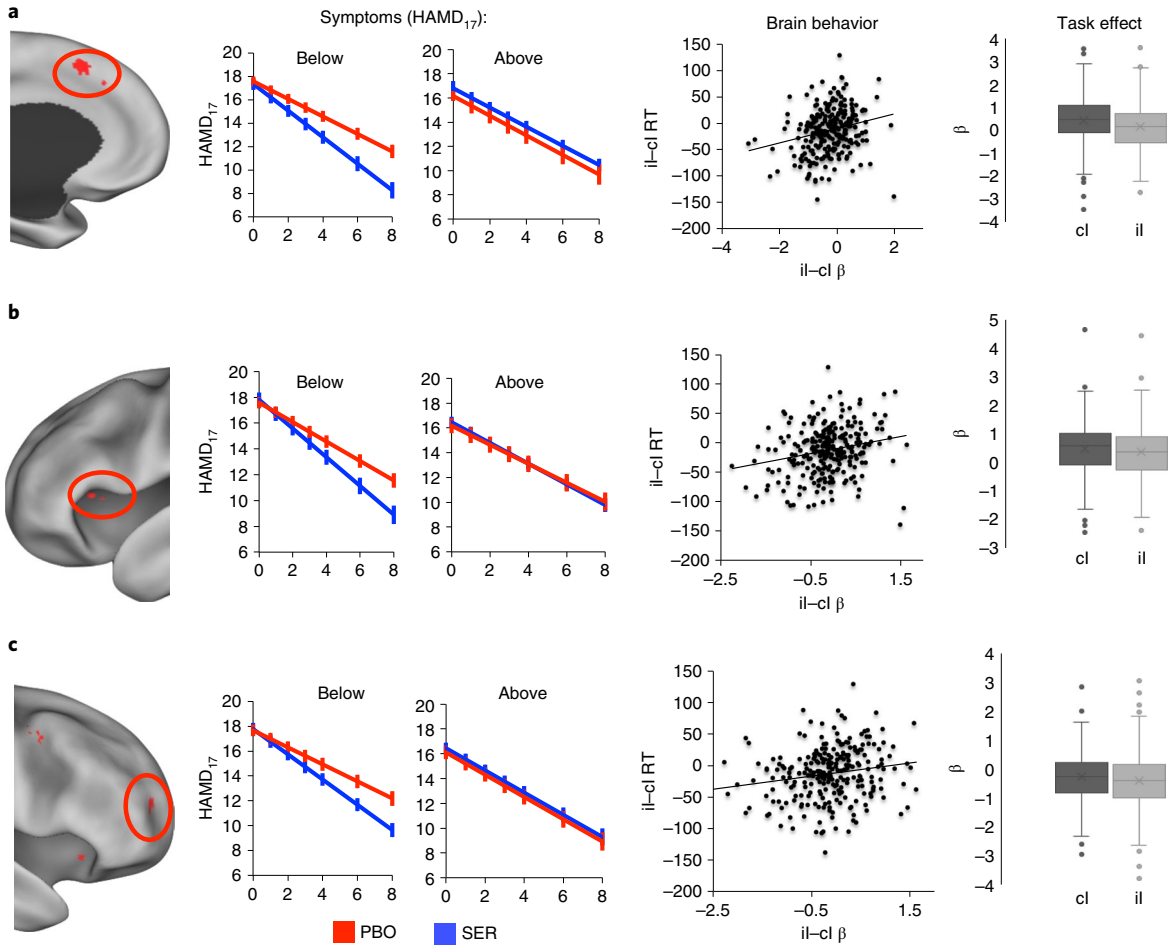

Fig. 4 | Overlap of treatment-moderating conflict-regulation-related brain activation, task-dependent brain activation and brain–behaviour relationships. a–c, Conjunction of emotional conflict regulation task effect, brain-behaviour relationship effect and treatment moderation effect in the dorsal cingulate–dorsomedial prefrontal cortex (**a**); the left anterior insula (**b**); and the right frontopolar cortex (**c**). Shown in columns from left to right are (leftmost) the ROI whose effects are shown; (second column) 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 used 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 used here so as to avoid circular analyses); (third column) continuous relationship between il–cl fMRI activity (*x* axis) in the depicted region and il–cl reaction time (RT) behaviour differences (*y* axis), with a linear trend line, across the entire randomized depressed sample to display the positive relationship (no statistical test used here, so as to avoid circular analyses); and (rightmost column) mean fMRI activity for il and cl conditions in the depicted region at baseline across the entire randomized sample. For the box and whisker plots, the centre 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, that is, 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 s.e.m.

dicted $HAMD_{17}$ 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 behavioural 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, which can be defined by a specific treatment-relevant neural phenotype. Moreover, mechanistically meaningful stratification may be possible in depression by relating biological characteristics to placebo-controlled treatment outcome. Indeed, we found no evi-

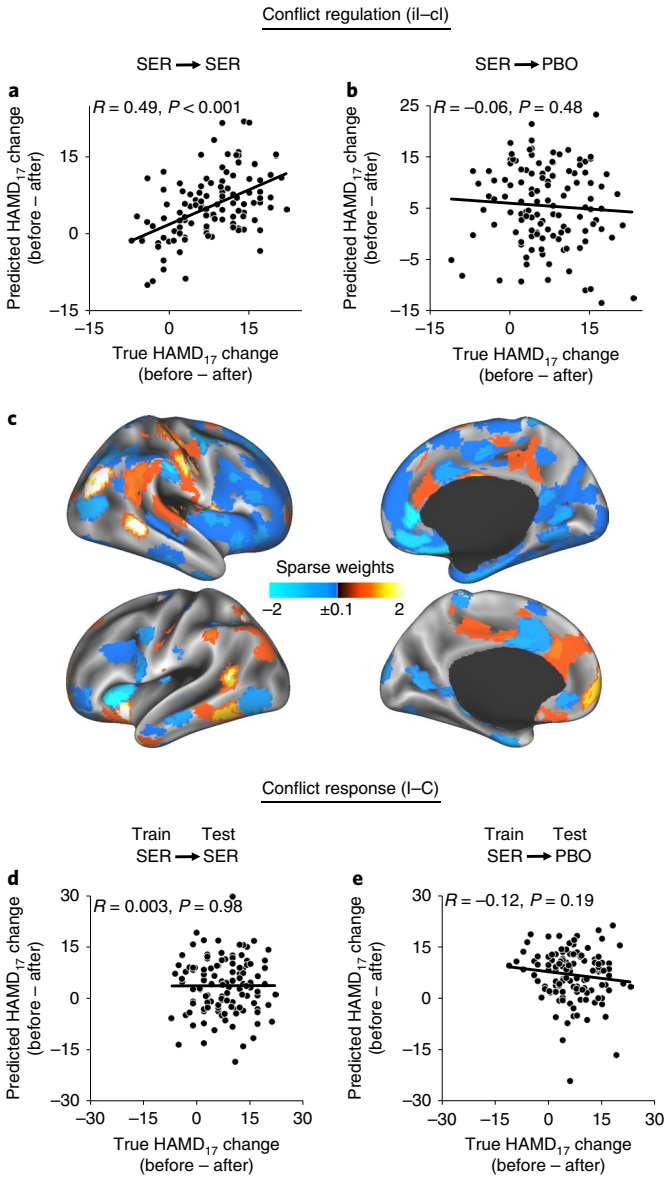

Fig. 5 | Machine learning selectively predicts response to sertraline ($n = 122$) with conflict-regulation-related fMRI activity. a, A machine-learning model trained (with 10-fold cross-validation) on outcome with sertraline, using conflict-regulation activation (il–cl), showed a strong correlation ($r = 0.49$, $P < 0.001$) 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**,**e**, Training a machine-learning model on outcome with sertraline ($n = 122$) with conflict-response activation (I–C) does not yield statistically significant prediction of sertraline outcome ($r = 0.003$, $P = 0.98$) (**d**) and does not yield statistically significant prediction ($r = -0.12$, $P = 0.19$) of placebo outcome ($n = 129$) (**e**).

dence 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 finding is consistent with previous 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 previous findings and theoretical considerations regarding emotional conflict regulation (and its behavioural and neural deficits) in depression have focused primarily on the rostral/ventral ACC and its lack of activation in response to emotional conflict regulation[21], the current findings indicate that the most treatment-relevant aspects of the ongoing neurocircuitry processes mediating regulation of emotional conflict in major depression do not necessarily reflect this previous theoretical framework. Moreover, we detected no functional brain differences between depressed individuals and healthy comparison subjects during emotional conflict regulation, which diverges with previous findings for a rostral/ventral ACC activation failure[21]. However, the healthy comparison group used here was much smaller than the group of 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 versus placebo responses (and not increasing activation in the rostral–ventral regions of the anterior cingulate cortex (ACC) and prefrontal cortex) suggests that this previous 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 RCT methodology. We therefore recommend that the theoretical and neurobiological understanding of brain emotional conflict regulation abnormalities in depression continue to be probed with large samples and 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 rigour 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 previous non–placebo-controlled treatment prediction findings would be likely moderators or non-specific 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[36]. 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 versus general conflict processing, which unfortunately could not be ascertained with the current design. The specificity of this predictive signal to sertraline versus 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 than 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 behaviour (that is, reaction times) moderates treatment outcome, this result may be due to brain activation being a better assay of the relevant mechanisms relative to behaviour (consistent with similar previous findings, for example ref. [27]) and the fact that activity across multiple regions together contribute to single behavioural 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 patient non-responsive to sertraline into a patient responsive to sertraline may be possible if aspects of the brain's emotional conflict regulation capacity can be improved by new medications, brain stimulation, behavioural 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 finding is consistent with resting EEG studies, including those 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[31,37]. In fact, our non-specific predictive effect partially overlapped with the cluster demonstrating non-specific prediction in the EEG analyses[31].

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 (as patients here were medication-free). Importantly, as EMBARC recruited individuals without a history of antidepressant treatment failures to maximize power to detect treatment moderation effects, these findings may not necessarily generalize to all individuals with major depression, especially those with a history of failing one or more adequate antidepressant trials. Regarding methodological rigour, additional work should be done to assess and optimize test–retest reliability of emotional conflict regulation signals in patients with depression, 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[38–41]. 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 versus 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 affect 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 (for example, repetitive transcranial magnetic stimulation (rTMS), electroconvulsive therapy or psychotherapy). Typically, patients undergo iterative rounds of pharmacological treatment before advancing to a treatment modality such as rTMS[42,43], 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[42,43], 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 enroled medication-resistant patients and showed benefit of real over sham rTMS[44,45]. 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[46].

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[47].

Methods

Participants and treatment. The study methods and participant characteristics are presented here in brief, and for more detail we refer the reader to 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 previous 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[48], at least moderate depression severity with a score ≥ 14 on the Quick Inventory of Depressive Symptomatology-Self Report at screening and randomization, a major depressive episode beginning before age 30, either a chronic recurrent episode (duration ≥ 2 years) or recurrent major depressive disorder (at least two 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 6 months or substance abuse in the past 2 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, vagal nerve 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 symptom severity and chronicity with a block randomization procedure. Sertraline dosing began at 50 mg with 50-mg capsules and was increased as tolerated if the patient did not respond until a maximum of 200 mg (ref. [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)', NCT#01407094.

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 and Friesen[49], 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 1,000 ms in a fast event-related design. Interstimulus intervals were 3,000–5,000 ms in a pseudo-randomized order counterbalanced for facial expression, gender, word and response button. Stimuli were either congruent (for example, fearful face with 'FEAR') or incongruent (for example, fearful face with 'HAPPY'), and stimuli were furthermore balanced to achieve an equal fraction of current and previous trial congruency, while ensuring no direct stimulus

repetitions. Before 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 min and 14 s.

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, although emotional conflict results in slowing of reaction times, this effect can be mitigated in iI trials, compared to 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 previous 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[36,50,51]. Critically, this contrast between postincongruent incongruent and postcongruent incongruent trial compares brain responses to physically identical stimuli (that is, incongruent trials) that differ only on the relative emotional conflict regulatory context in which they come due to previous trial congruency and is furthermore independent of the I–C trial conflict response contrast. Neuroimaging acquisition parameters are shown in Supplementary Table 2.

fMRI preprocessing and first-level modelling. FSL tools were used to preprocess imaging data[52,53]. Functional images were first realigned to structural images with an affine registration matrix and boundary-based registration based upon tissue segmentation as implemented in FSL's FLIRT, which was concatenated with a nonlinear normalization of each participant's T1 image to the Montreal Neurological Institute 152-person 1 mm^3 T1 template with 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 cerebrospinal fluid were regressed out of motion-corrected functional images. A 6-mm full-width at half max 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 modelling trials of interest were convolved with the hemodynamic response function. First-level general linear models were estimated in SPM 8 (ref. [54]). Regressors corresponded to zero-duration markers set at the onset of stimuli, which were explicitly categorized by congruency (incongruent or congruent) and previous trial type (postincongruent or postcongruent) to model conflict response and regulation effects. This process resulted in four different trial types in total, in addition to nuisance regressors for error trials and posterror trials (when applicable) and six motion parameters.

Participant selection and quality-control procedures. To ensure the quality of imaging measures in testing moderation and prediction effects, consistent with our previous work in imaging treatment effects[55,56], we instituted cutoffs for absolute level of motion (root mean square of the absolute level of movement <4 mm across the mean of the squared maximum displacements in each of the six translational and rotational parameters estimated during realignment) during the baseline scans. In addition, to ensure that brain activation measures reflect task-relevant metrics, we also instituted a minimum level of behavioural accuracy during completion of the emotional conflict task as an additional quality-control metric (total accuracy $\geq 80\%$ 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 versus 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 $\alpha = 0.05$ (i.e., tolerating the rejection of the null hypothesis when the null hypothesisis true 5% of the time) 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 (that is, 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 versus placebo effect on a continuous outcome by a continuous measure is largely unknown. However, previous studies[24–27] examining brain activation-based predictors of treatment response in depression (but without a placebo control group, in smaller samples, and/or using 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 used 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 (for example, 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[57,58] (linear relationship(s) between predictor(s) and outcomes, model residuals normally distributed with equivalent variances across distribution of fitted values, and so on). When this was not possible (for example, for voxel-level moderation analyses), adherence to model assumptions was assessed with 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: (1) for two-sample mean comparisons, d was calculated as the difference in means divided by the pooled standard deviation; (2) 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: (1) for two-sample mean comparisons, the 95% confidence intervals for Cohen's d were calculated with the cohen.d.ci function from the psych library[59] in R; (2) for one-sample t-tests, the 95% confidence intervals for Cohen's d were calculated according to the following formula: $d \pm$ s.e.m. $\times 1.96$ (critical z value), where s.e.m. $= \sqrt{((1/n) + (d^2/2n))}$.

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 v.21.0 (ref. [60]) along with variables coding for participant, treatment arm, site (four-centred dummy-coded variables, each specifying one of the non-UT Southwestern study sites/scanners (Columbia used two separate scanners) versus UT Southwestern) and time (scaled from 0 to 1) to assess the effect of sertraline versus 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 individual variation in baseline depressive symptom severity. A fixed intercept as well as a fixed effect of time, effect of Columbia scanner 1 versus UT Southwestern site, effect of Columbia scanner 2 versus UT Southwestern site, effect of Massachusetts General Hospital versus UT Southwestern site, effect of University of Michigan versus UT Southwestern site, interaction of time with each of the aforementioned four site variables, treatment arm × time, and the interaction of treatment arm × time with each of the four aforementioned site variables were modelled, with the treatment arm × time interaction effect of interest specifying differential trajectories in symptom change over time for sertraline versus 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 centred to yield an interpretable (that is, grand average) overall treatment arm × time effect.

Assessing task effects on brain activity. The entire randomized sample was subjected to a voxel-wise 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 (postincongruent incongruent trials (iI) versus postcongruent incongruent trials (cI); iI–cI). We constrained all imaging analyses to a search space defined by a whole-brain mask constructed from (1) a 30% probabilistic voxel-wise threshold derived from spatially normalized, segmented binary grey matter images from an independent sample of healthy comparison participants[55,56] and (2) 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 grey 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 with a whole-brain (constrained within the search space) 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 (ref. [60]).

Assessing brain-behaviour relationships for emotional conflict regulation. Individual contrast maps specifying the difference in activation for iI–cI trials were subjected to a robust regression[61,62] implemented in R[63] using the fmri package[64]. 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-centred 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 modelled as a function of an intercept, the grand mean-centred 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 used 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 v.21.0 (ref. [60]).

Assessing baseline brain activation moderation of the sertraline versus 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[63] with the fmri package[64] (see example R syntax available in Supplementary Software). We constrained analyses to a search space defined by a whole-brain mask constructed from (1) a 30% probabilistic voxel-wise threshold derived from spatially normalized, segmented binary grey matter images from an independent sample of healthy comparison participants[55,56] and (2) 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 grey matter. At each voxel within this search space, the moderating effect of baseline brain activation (mean-centred) was examined with a longitudinal linear mixed model implemented in the nlme R package[65]. Modelled effects included a random and fixed intercept, fixed effects of time, time × treatment arm, four mean-centred, dummy-coded scanner variables corresponding to: (1) Columbia University scanner 1 versus UT Southwestern scanner; (2) Columbia University scanner 2 versus UT Southwestern scanner; (3) Massachusetts General Hospital scanner versus UT Southwestern scanner; and (4) University of Michigan scanner versus UT Southwestern scanner, treatment arm × time × brain activation and all model variables in interaction with each of the four 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 (that is, 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 on the number of individuals displaying adequate signal coverage at that voxel (which tended to be less in areas prone to echoplanar imaging 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 with the fmri package, and voxel-level FDR correction within the whole-brain specified search space (see above) was conducted 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 v.21.0 (ref. [60]).

Machine learning analyses. RVMs 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 before penalizing unnecessary complexity in the model, RVM is able to automatically determine the feature sparsity under a Bayesian evidence framework[66]. The effectiveness of the popular L1 regularization-based sparse learning algorithms[67] depends on the selection of regularization hyperparameter to a large extent. Cross-validation on training sets 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 because fewer training samples can be used for model calibration[68]. Different from the L1 regularization approaches, all model parameters in RVMs can be efficiently estimated with all available training data without the need of cross-validation. RVMs therefore provide a more accurate estimation of the sparse solution than other L1 regularization-based sparse-learning algorithms[69], especially when a small training set is available. In the past decades, RVMs have demonstrated their strength in various fields including EEG classification for brain-computer interface[70] and bioinformatics analysis of gene expression data[71,72].

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 had sufficient whole-brain coverage such that data were available from almost all regions of interest (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[63] using the package mice[73]. The following observed variables were used 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 QIDS total score; baseline Mood and Symptom Questionnaire subscale scores for Anxious Arousal, Anhedonic Depression, and General Distress; Snaith-Hamilton Pleasure Scale total score; age; years of education; gender; and Wechsler Abbreviated Scale of Intelligence t-scores for Vocabulary and Matrix Reasoning. Extractions were conducted on cortical ROIs, defined by 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[74]. 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 to limit parcellation-related specificity. ROIs were mapped to seven previously identified functional networks according to the spatial overlap between each ROI and each network[74]. In addition to these cortical ROIs, subcortical ROIs included striatal[75] and cerebellar[76] parcellations based on the same seven functional networks, amygdala ROIs[77], anterior and posterior hippocampal ROIs[78] and the thalamus[79]. We then regressed imaging sites 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 with RVM learning to predict each individual's before minus after 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)-fold cross-validation (see Supplementary Fig. 6). Specifically, for ten repetitions, all participants were randomly divided into ten 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 tenfold 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 1,235 ROI features. After repeating tenfold cross-validation ten times, we determined each participant's predicted symptom change by taking the median of the predicted values across each of the ten 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 modelling 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 ten rounds of cross-validation. We conducted similar RVM analyses on item-level clinical, historical, demographic and behavioural 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[33], the Quick Inventory of Depressive Symptoms[32], the Mood and Anxiety Questionnaire[34], the Childhood Trauma Questionnaire[35], age, education and I–C and iI–cI reaction time and accuracy difference scores.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

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

Code availability

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

Received: 25 May 2019; Accepted: 16 August 2019;
Published online: 23 September 2019

References

1. Hasin, D. S. et al. Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. *JAMA Psychiat.* **75**, 336–346 (2018).
2. Lopez-Munoz, F. & Alamo, C. Monoaminergic neurotransmission: the history of the discovery of antidepressants from 1950s until today. *Curr. Pharm. Des.* **15**, 1563–1586 (2009).
3. Moore, T. J. & Mattison, D. R. Adult utilization of psychiatric drugs and differences by sex, age, and race. *JAMA Intern. Med.* **177**, 274–275 (2017).
4. Kirsch, I. *The Emperor's New Drugs: Exploding the Antidepressant Myth* (Random House, 2009).
5. Khan, A. & Brown, W. A. Antidepressants versus placebo in major depression: an overview. *World Psychiatry* **14**, 294–300 (2015).

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 (2008).
7. Fournier, J. C. et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA* **303**, 47–53 (2010).
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* **391**, 1357–1366 (2018).
9. Drysdale, A. T. et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat. Med.* **23**, 28–38 (2017).
10. Kraemer, H. C. Messages for clinicians: moderators and mediators of treatment outcome in randomized clinical trials. *Am. J. Psychiatry* **173**, 672–679 (2016).
11. Etkin, A., Buchel, C. & Gross, J. J. The neural bases of emotion regulation. *Nat. Rev. Neurosci.* **16**, 693–700 (2015).
12. Gross, J. J. *Handbook of Emotion Regulation* (Guilford Press, 2014).
13. Gyurak, A., Gross, J. J. & Etkin, A. Explicit and implicit emotion regulation: a dual-process framework. *Cogn. Emot.* **25**, 400–412 (2011).
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 (2016).
15. Egner, T., Etkin, A., Gale, S. & Hirsch, J. Dissociable neural systems resolve conflict from emotional versus nonemotional distracters. *Cereb. Cortex* **18**, 1475–1484 (2008).
16. Etkin, A., Egner, T., Peraza, D. M., Kandel, E. R. & Hirsch, J. Resolving emotional conflict: a role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron* **51**, 871–882 (2006).
17. Maier, M. E. & di Pellegrino, G. Impaired conflict adaptation in an emotional task context following rostral anterior cingulate cortex lesions in humans. *J. Cogn. Neurosci.* **24**, 2070–2079 (2012).
18. Chechko, N. et al. Brain circuitries involved in emotional interference task in major depression disorder. *J. Affect. Disord.* **149**, 136–145 (2013).
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).
20. Etkin, A., Prater, K. E., Hoeft, F., Menon, V. & Schatzberg, A. F. 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 (2010).
21. Etkin, A. & Schatzberg, A. F. 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 (2011).
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 (2017).
23. Widge, A. S. et al. Treating refractory mental illness with closed-loop brain stimulation: progress towards a patient-specific transdiagnostic approach. *Exp. Neurol.* **287**, 461–472 (2017).
24. Gyurak, A. et al. Frontoparietal activation during response inhibition predicts remission to antidepressants in patients with major depression. *Biol. Psychiatry* **79**, 274–281 (2016).
25. Williams, L. M. 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 (2015).
26. Etkin, A. et al. A cognitive-emotional biomarker for predicting remission with antidepressant medications: a report from the iSPOT-D trial. *Neuropsychopharmacology* **40**, 1332–1342 (2015).
27. Langenecker, S. A. et al. Frontal and limbic activation during inhibitory control predicts treatment response in major depressive disorder. *Biol. Psychiatry* **62**, 1272–1280 (2007).
28. Pizzagalli, D. A. Frontocingulate dysfunction in depression: toward biomarkers of treatment response. *Neuropsychopharmacology* **36**, 183–206 (2011).
29. Trivedi, M. H. et al. Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): rationale and design. *J. Psychiatr. Res.* **78**, 11–23 (2016).
30. Goldstein-Piekarski, A. N. et al. Intrinsic functional connectivity predicts remission on antidepressants: a randomized controlled trial to identify clinically applicable imaging biomarkers. *Transl. Psychiatry* **8**, 57 (2018).
31. 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 Psychiat.* **75**, 547–554 (2018).
32. Rush, A. J. et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol. Psychiatry* **54**, 573–583 (2003).
33. Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R. & Jacobs, G. A. *Manual for the State-Trait Anxiety Inventory* (Consulting Psychologists Press, 1983).
34. Wardenaar, K. J. et al. Development and validation of a 30-item short adaptation of the Mood and Anxiety Symptoms Questionnaire (MASQ). *Psychiatry Res.* **179**, 101–106 (2010).
35. Bernstein, D. P. et al. Initial reliability and validity of a new retrospective measure of child abuse and neglect. *Am. J. Psychiatry* **151**, 1132–1136 (1994).
36. Kerns, J. G. et al. Anterior cingulate conflict monitoring and adjustments in control. *Science* **303**, 1023–1026 (2004).
37. Korb, A. S., Hunter, A. M., Cook, I. A. & Leuchter, A. F. Rostral anterior cingulate cortex theta current density and response to antidepressants and placebo in major depression. *Clin. Neurophysiol.* **120**, 1313–1319 (2009).
38. Clayson, P. E. & Larson, M. J. Adaptation to emotional conflict: evidence from a novel face emotion paradigm. *PLoS One* **8**, e75776 (2013).
39. Tang, D., Hu, L., Chen, A., Clayson, P. E. & Larson, M. J. The neural oscillations of conflict adaptation in the human frontal region. *Biol. Psychol.* **93**, 364–372 (2013).
40. Larson, M. J., Clawson, A., Clayson, P. E. & Baldwin, S. A. Cognitive conflict adaptation in generalized anxiety disorder. *Biol. Psychol.* **94**, 408–418 (2013).
41. Suzuki, K. & Shinoda, H. Transition from reactive control to proactive control across conflict adaptation: an sLORETA study. *Brain Cogn.* **100**, 7–14 (2015).
42. 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 (2017).
43. Nguyen, K. H. & Gordon, L. G. Cost-effectiveness of repetitive transcranial magnetic stimulation versus antidepressant therapy for treatment-resistant depression. *Value Health* **18**, 597–604 (2015).
44. O'Reardon, J. P. 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 (2007).
45. George, M. S. et al. Daily left prefrontal transcranial magnetic stimulation therapy for major depressive disorder: a sham-controlled randomized trial. *Arch. Gen. Psychiatry* **67**, 507–516 (2010).
46. Pescosolido, B. A. 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 (2010).
47. Insel, T. R. & Cuthbert, B. N. Medicine. Brain disorders? Precisely. *Science* **348**, 499–500 (2015).
48. First, M., Spitzer, R., Gibbon, M. & William, J. *Structured Clinical Interview forDSM-IV-TR Axis 1 Disorders,* Research Version, Patient Edition (SCID-I/P) (New York State Psychiatric Institute Press, 2002).
49. Ekman, P. & Friesen, W. V. *Pictures of Facial Affect* (Consulting Psychologists, 1976).
50. Gratton, G., Coles, M. G. & Donchin, E. Optimizing the use of information: strategic control of activation of responses. *J. Exp. Psychol. Gen.* **121**, 480–506 (1992).
51. Egner, T. & Hirsch, J. Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nat. Neurosci.* **8**, 1784–1790 (2005).
52. Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W. & Smith, S. M. FSL. *Neuroimage* **62**, 782–790 (2012).
53. Smith, S. M. et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23**, S208–S219 (2004).
54. Friston, K. J. et al. Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* **2**, 189–210 (1995).
55. Fonzo, G. A. et al. PTSD psychotherapy outcome predicted by brain activation during emotional reactivity and regulation. *Am. J. Psychiatry* **174**, 1163–1174 (2017).
56. Fonzo, G. A. et al. Selective effects of psychotherapy on frontopolar cortical function in PTSD. *Am. J. Psychiatry* **174**, 1175–1184 (2017).
57. Fox, J. *Applied Regression Analysis and Generalized Linear Models* 2nd edn (Sage Publications, 2008).
58. Loy, A., Hofmann, H. & Cook, D. Model choice and diagnostics for linear mixed-effects models using statistics on street corners. *J. Comput. Graph. Stat.* **26**, 478–492 (2017).
59. Revelle, W. psych: Procedures for Psychological, Psychometric, and Personality Research. R package version 1.8.12. https://CRAN.R-project.org/package=psych (2018).
60. IBM SPSS Statistics for Macintosh v.21.0 (IBM Corp, 2012).
61. Huber, P. J. Robust regression: asymptotics, conjectures and monte carlo. *Ann. Stat.* **1**, 799–821 (1973).
62. Wager, T. D., Keller, M. C., Lacey, S. C. & Jonides, J. Increased sensitivity in neuroimaging analyses using robust regression. *Neuroimage* **26**, 99–113 (2005).
63. R: a language and environment for statistical computing v. 3.2.3 (R Foundation for Statistical Computing, 2015).
64. Tabelow, K. & Polzehl, J. Statistical parametric maps for functional MRI experiments in R: the package fMRI. **44**, 21 (2011).

65. Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. nlme: linear and nonlinear mixed effects models. R package version 3.1-141. https://CRAN.R-project.org/package=nlme (2019).
66. Tipping, M. E. Sparse bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **1**, 211–244 (2001).
67. Tibshirani, R. Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc. B* **58**, 267–288 (1996).
68. Cawley, G. C. & Talbot, N. L. Preventing over-fitting during model selection via Bayesian regularisation of the hyper-parameters. *J. Mach. Learn. Res.* **8**, 841–861 (2007).
69. Wipf, D. P. & Rao, B. D. Sparse bayesian learning for basis selection. *IEEE Trans. Signal Process.* **52**, 2153–2164 (2004).
70. Zhang, Y. et al. Sparse bayesian classification of EEG for brain-computer interface. *IEEE Trans. Neural Netw. Learn. Syst.* **27**, 2256–2267 (2016).
71. Cawley, G. C. & Talbot, N. L. Gene selection in cancer classification using sparse logistic regression with Bayesian regularization. *Bioinformatics* **22**, 2348–2355 (2006).
72. Li, Y., Campbell, C. & Tipping, M. Bayesian automatic relevance determination algorithms for classifying gene expression data. *Bioinformatics* **18**, 1332–1339 (2002).
73. van Buuren, S. & Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained equations in R. *J. Stat. Soft.* **45**, 1–67 (2011).
74. Schaefer, A. et al. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).
75. Choi, E. Y., Yeo, B. T. & Buckner, R. L. The organization of the human striatum estimated by intrinsic functional connectivity. *J. Neurophysiol.* **108**, 2242–2263 (2012).
76. Buckner, R. L., Krienen, F. M., Castellanos, A., Diaz, J. C. & Yeo, B. T. The organization of the human cerebellum estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 2322–2345 (2011).
77. Patenaude, B., Smith, S. M., Kennedy, D. N. & Jenkinson, M. A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* **56**, 907–922 (2011).
78. Chen, A. C. & Etkin, A. Hippocampal network connectivity and activation differentiates post-traumatic stress disorder from generalized anxiety disorder. *Neuropsychopharmacology* **38**, 1889–1898 (2013).
79. Behrens, T. E. et al. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nat. Neurosci.* **6**, 750–757 (2003).

Acknowledgements

The EMBARC study was supported by the National Institute of Mental Health of the National Institutes of Health under award numbers U01MH092221 (M.H.T.) and U01MH092250 (P.J.M., M.M.W.). This work was also funded in part by the Hersh Foundation (M.H.T., principal investigator). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

G.A.F. contributed the analysis and interpretation of the data and the drafting and revision of the manuscript. A.E. contributed to the design of the study, the analysis and interpretation of the data, and the drafting and revision of the manuscript. Y.Z. contributed to the analysis and interpretation of the data and the drafting and revision of the manuscript. W.W. contributed to the analysis and interpretation of the data. C.C., C.C.F., M.K.J. and J.T. contributed to the conduct of the study, analysis and interpretation of the data, and revision of the manuscript. T.D., P.A., M.M., P.J.M. and M.F. contributed to the design and conduct of the study. M.M.W. contributed to the design and conduct of the study, and revision of the manuscript. M.H.T. was the study PI and contributed to study design and funding, conduct of the study, analysis and interpretation of the data, and the drafting and revision of the manuscript.

Competing interests

A.E. (lifetime disclosure) holds equity in Mindstrong Health and Akili Interactive for unrelated work, has received research funding from the National Institute of Mental Health, Department of Veterans Affairs, Cohen Veterans Bioscience, Brain and Behaviour Research Foundation, Dana Foundation, Brain Resource Inc, and the Stanford Neurosciences Institute, and consulted for Cervel, Takaeda, Posit Science, Acadia, Otsuka, Lundbeck and Janssen. G.A.F. received research support from the National Institute of Mental Health and the Department of Veterans Affairs. T.D.'s research has been funded by NIH, NIMH, NARSAD, TSA, IOCDF, Tufts University, DBDAT and Otsuka Pharmaceuticals, he has received honoraria, consultation fees and/or royalties from the MGH Psychiatry Academy, BrainCells Inc., Clintara, LLC, Inc., Systems Research and Applications Corporation, Boston University, the Catalan Agency for Health Technology Assessment and Research, the National Association of Social Workers Massachusetts, the Massachusetts Medical Society, Tufts University, NIDA, NIMH, Oxford University Press, Guilford Press and Rutledge. He has also participated in research funded by DARPA, NIH, NIA, AHRQ, PCORI, Janssen Pharmaceuticals, The Forest Research Institute, Shire Development Inc., Medtronic, Cyberonics, Northstar, and Takeda. P.J.M. has received funding from the National Institute of Mental Health, New York State Department of Mental Hygiene, Research Foundation for Mental Hygiene (New York State), Forest Research Laboratories, Sunovion Pharmaceuticals, and Naurex Pharmaceuticals (now Allergan). In the past two years, M.M.W. received funding from the National Institute of Mental Health (NIMH), the National Institute on Drug Abuse (NIDA), the National Alliance for Research on Schizophrenia and Depression (NARSAD), the Sackler Foundation, the Templeton Foundation; and receives royalties from the Oxford University Press, Perseus Press, the American Psychiatric Association Press, and MultiHealth Systems. M.F. has received research support from Abbot Laboratories; Alkermes, Inc.; American Cyanamid; Aspect Medical Systems; AstraZeneca; Avanir Pharmaceuticals; BioResearch; BrainCells Inc.; Bristol-Myers Squibb; CeNeRx BioPharma; Cephalon; Clintara, LLC; Cerecor; Covance; Covidien; Eli Lilly and Company; EnVivo Pharmaceuticals, Inc.; Euthymics Bioscience, Inc.; Forest Pharmaceuticals, Inc.; Ganeden Biotech, Inc.; GlaxoSmithKline; Harvard Clinical Research Institute; Hoffman-LaRoche; Icon Clinical Research; i3 Innovus/Ingenix; Janssen R&D, LLC; Jed Foundation; Johnson & Johnson Pharmaceutical Research & Development; Lichtwer Pharma GmbH; Lorex Pharmaceuticals; Lundbeck Inc.; MedAvante; Methylation Sciences Inc.; National Alliance for Research on Schizophrenia & Depression (NARSAD); National Center for Complementary and Alternative Medicine (NCCAM); National Institute of Drug Abuse (NIDA); National Institute of Mental Health (NIMH); Neuralstem, Inc.; Novartis AG; Organon Pharmaceuticals; PamLab, LLC.; Pfizer Inc.; Pharmacia-Upjohn; Pharmaceutical Research Associates., Inc.; Pharmavite® LLC; PharmoRx Therapeutics; Photothera; Reckitt Benckiser; Roche Pharmaceuticals; RCT Logic, LLC (formerly Clinical Trials Solutions, LLC); Sanofi-Aventis US LLC; Shire; Solvay Pharmaceuticals, Inc.; Stanley Medical Research Institute (SMRI); Synthelabo; Tal Medical; Wyeth-Ayerst Laboratories; he has served as advisor or consultant to Abbott Laboratories; Acadia; Affectis Pharmaceuticals AG; Alkermes, Inc.; Amarin Pharma Inc.; Aspect Medical Systems; AstraZeneca; Auspex Pharmaceuticals; Avanir Pharmaceuticals; AXSOME Therapeutics; Bayer AG; Best Practice Project Management, Inc.; Biogen; BioMarin Pharmaceuticals, Inc.; Biovail Corporation; BrainCells Inc; Bristol- Myers Squibb; CeNeRx BioPharma; Cephalon, Inc.; Cerecor; CNS Response, Inc.; Compellis Pharmaceuticals; Cypress Pharmaceutical, Inc.; DiagnoSearch Life Sciences (P) Ltd.; Dinippon Sumitomo Pharma Co. Inc.; Dov Pharmaceuticals, Inc.; Edgemont Pharmaceuticals, Inc.; Eisai Inc.; Eli Lilly and Company; EnVivo Pharmaceuticals, Inc.; ePharmaSolutions; EPIX Pharmaceuticals, Inc.; Euthymics Bioscience, Inc.; Fabre-Kramer Pharmaceuticals, Inc.; Forest Pharmaceuticals, Inc.; Forum Pharmaceuticals; GenOmind, LLC; GlaxoSmithKline; Grunenthal GmbH; i3 Innovus/Ingenis; Intracellular; Janssen Pharmaceutica; Jazz Pharmaceuticals, Inc.; Johnson & Johnson Pharmaceutical Research & Development, LLC; Knoll Pharmaceuticals Corp.; Labopharm Inc.; Lorex Pharmaceuticals; Lundbeck Inc.; MedAvante, Inc.; Merck & Co., Inc.; MSI Methylation Sciences, Inc.; Naurex, Inc.; Nestle Health Sciences; Neuralstem, Inc.; Neuronetics, Inc.; NextWave Pharmaceuticals; Novartis AG; Nutrition 21; Orexigen Therapeutics, Inc.; Organon Pharmaceuticals; Osmotica; Otsuka Pharmaceuticals; Pamlab, LLC.; Pfizer Inc.; PharmaStar; Pharmavite LLC.; PharmoRx Therapeutics; Precision Human Biolaboratory; Prexa Pharmaceuticals, Inc.; Puretech Ventures; PsychoGenics; Psylin Neurosciences, Inc.; RCT Logic, LLC Formerly Clinical Trials Solutions, LLC; Rexahn Pharmaceuticals, Inc.; Ridge Diagnostics, Inc.; Roche; Sanofi-Aventis US LLC.; Sepracor Inc.; Servier Laboratories; Schering-Plough Corporation; Solvay Pharmaceuticals, Inc.; Somaxon Pharmaceuticals, Inc.; Somerset Pharmaceuticals, Inc.; Sunovion Pharmaceuticals; Supernus Pharmaceuticals, Inc.; Synthelabo; Taisho Pharmaceutical; Takeda Pharmaceutical Company Limited; Tal Medical, Inc.; Tetragenex Pharmaceuticals, Inc.; TransForm Pharmaceuticals, Inc.; Transcept Pharmaceuticals, Inc.; Vanda Pharmaceuticals, Inc.; VistaGen; he has received speaking or publishing fees from Adamed, Co; Advanced Meeting Partners; American Psychiatric Association; American Society of Clinical Psychopharmacology; AstraZeneca; Belvoir Media Group; Boehringer Ingelheim GmbH; Bristol-Myers Squibb; Cephalon, Inc.; CME Institute/Physicians Postgraduate Press, Inc.; Eli Lilly and Company; Forest Pharmaceuticals, Inc.; GlaxoSmithKline; Imedex, LLC; MGH Psychiatry Academy/Primedia; MGH Psychiatry Academy/Reed Elsevier; Novartis AG; Organon Pharmaceuticals; Pfizer Inc.; PharmaStar; United BioSource, Corp.; Wyeth-Ayerst Laboratories; he has equity holdings in Compellis and PsyBrain, Inc.; he has a patent for Sequential Parallel Comparison Design (SPCD), which are licensed by MGH to Pharmaceutical Product Development, LLC (PPD); and patent application for a combination of Ketamine plus Scopolamine in Major Depressive Disorder (MDD), licensed by MGH to Biohaven; and he receives copyright royalties for the MGH Cognitive & Physical Functioning Questionnaire (CPFQ), Sexual Functioning Inventory (SFI), Antidepressant Treatment Response Questionnaire (ATRQ), Discontinuation-Emergent Signs & Symptoms (DESS), Symptoms of Depression Questionnaire (SDQ), and SAFER; Lippincott, Williams & Wilkins; Wolkers Kluwer; World Scientific Publishing Co. Pte. Ltd. M.H.T. is or has been an advisor/consultant and received fee from (lifetime disclosure): Abbott Laboratories, Inc., Abdi Ibrahim, Akzo (Organon Pharmaceuticals Inc.), Alkermes, AstraZeneca, Axon Advisors, Bristol-Myers Squibb Company, Cephalon, Inc., Cerecor, CME Institute of Physicians, Concert Pharmaceuticals, Inc., Eli Lilly & Company, Evotec, Fabre-Kramer Pharmaceuticals, Inc., Forest Pharmaceuticals, GlaxoSmithKline, Janssen Global Services, LLC, Janssen Pharmaceutica Products, LP, Johnson & Johnson PRD, Libby, Lundbeck, Meade

Johnson, MedAvante, Medtronic, Merck, Mitsubishi Tanabe Pharma Development America, Inc., Naurex, Neuronetics, Otsuka Pharmaceuticals, Pamlab, Parke-Davis Pharmaceuticals, Inc., Pfizer Inc., PgxHealth, Phoenix Marketing Solutions, Rexahn Pharmaceuticals, Ridge Diagnostics, Roche Products Ltd., Sepracor, SHIRE Development, Sierra, SK Life and Science, Sunovion, Takeda, Tal Medical/Puretech Venture, Targacept, Transcept, VantagePoint, Vivus, and Wyeth-Ayerst Laboratories. In addition, he has received grants/research support from: Agency for Healthcare Research and Quality (AHRQ), Cyberonics, Inc., National Alliance for Research in Schizophrenia and Depression, National Institute of Mental Health and National Institute on Drug Abuse. G.A.F., Y.Z., W.W., C.C., C.C.F., M.K., P.A. and J.T. report no competing interests. M.M. has no conflicts of interest with respect to this paper.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-019-0732-1.

Correspondence and requests for materials should be addressed to A.E.

Peer review information Primary handling editor: Mary Elizabeth Sutherland

nature research

Corresponding author(s): Amit Etkin, M.D., Ph.D.

Last updated by author(s): Aug 1, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	MRI data were collected using conventional imaging acquisition software on each of the MRI platforms noted. Behavioral data were collected using e-prime software (which presented the stimuli). Clinical and self-report data were collected by web and paper forms into the StudyTrax database
Data analysis	Imaging data were analyzed using tools from FSL 5, SPM8, and R, as well as custom Matlab code for the Sparse Bayesian Learning. Other statistical analyses used SPSS v21.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

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

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Final analyzable sample size: 251 patients and 40 healthy controls
Data exclusions	Data were excluded for excessive head motion, poor task accuracy and acquisition errors or artifacts (see supplemental figure 1)
Replication	Strict voxelwise wholebrain FDR was conducted on data analyzed in a full intent-to-treat manner (thus incorporating all eligible randomized participants with sufficient data quality, thereby minimizing risk of bias due to small sample sizes or only analyzing completers). The predictive multivariable regression models used a 10-fold cross-validation procedure.
Randomization	Patients were randomized 1:1 into the sertraline versus placebo treatment arms
Blinding	All MRI/behavioral data acquisition occurred prior to randomization, and was thus blind to treatment arm. Clinical assessors were also blind to treatment arm

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☐	☒ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	The study methods and participant characteristics are presented here in brief, as well as in a publication detailing the rationale and design of EMBARC. 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). 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 , 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 healthy individuals who were medically and psychiatrically healthy, were recruited across the four sites, and assessed in a similar manner as the patients.
Recruitment	See above.

Ethics oversight	Institutional review boards for the four participating institutions: University of Texas Southwestern Medical Center, Massachusetts General Hospital, Columbia University, and University of Michigan.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care for Depression (EMBARC), NCT#01407094. https://clinicaltrials.gov/ct2/show/NCT01407094
Study protocol	Study protocol is published here: Trivedi, M. H. 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).
Data collection	Data was located across four sites in academic medical centers in Texas, Massachusetts, New York, and Michigan. Data was collected from July 2011 to April 2016.
Outcomes	Outcomes were pre-defined during the design stage and included emotional conflict regulation-related brain activation moderators of the effect of sertraline vs. placebo on the 17-item Hamilton Rating Depression Scale from baseline to 8 weeks after the start of treatment.

Magnetic resonance imaging

Experimental design

Design type	Event-related task-based functional magnetic resonance imaging
Design specifications	One run of 148 trials, with duration 1s and jittered 3-5s inter-trial intervals
Behavioral performance measures	Button presses were recorded and used to calculate accuracy and reaction time

Acquisition

Imaging type(s)	Functional
Field strength	3 Tesla
Sequence & imaging parameters	Gradient Echo EPI. Acquisition parameters are in supplemental table 2
Area of acquisition	Whole brain

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	FSL tools were used to preprocess imaging data . 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 mm3 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.
Normalization	Non-linear normalization with FSL's fnirt
Normalization template	MNI152
Noise and artifact removal	Six motion parameters were utilized as covariate nuisance regressors in first-level models.
Volume censoring	N/A

Statistical modeling & inference

Model type and settings	Mass-univariate GLM analyses were performed. 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 . 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.

Effect(s) tested

For each contrast of interest (il-cl and inc-con), first-level, spatially-normalized activation contrast maps from each individual were loaded into R using the fmri package. 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; 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. Modeled effects included a random and fixed intercept, fixed effects of time, time x 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 x time x brain activation, and all model variables in interaction with each of the 4 site variables. See supplemental table 3 for all variables in the mixed models. The effect of interest was the treatment arm x time x 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 NIFTI 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 (q < 0.05) were then thresholded and clustered for subsequent individual-level extraction and visualization of effects in IBM SPSS version 21.0.

Specify type of analysis: ☒ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference
(See Eklund et al. 2016)

Voxel-wise

Correction

false discovery rate (FDR)-corrected p<0.05

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☐	☒ Multivariate modeling or predictive analysis

Multivariate modeling and predictive analysis

Relevance Vector Machines (RVM) were used to build a regression model for the prediction of pre- minus post-treatment change in HAMD17. 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 45. The effectiveness of the popular L1-regularization based sparse learning algorithms 46 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 47. 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 48, 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 49 and bioinformatics analysis of gene expression data 50,51.

RVM learning was conducted on data from patients who had baseline HAMD17, 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 HAMD17, multiple imputation by chained equations was conducted in R42 using the package mice52. The following observed variables were utilized in order to impute endpoint HAMD17 values for missing data via Bayesian regression: baseline HAMD17, week 1 HAMD17, week 2 HAMD17, week 3 HAMD17, week 4 HAMD17, week 6 HAMD17, 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 53. 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 53. In addition to these cortical ROIs, subcortical ROIs included striatal54 and cerebellar55 parcellations based on the same seven functional networks, amygdala ROIs 56, anterior and posterior hippocampal ROIs57 and the thalamus58. 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 HAMD17 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 10x10 cross-validation (see supplemental 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 HAMD17 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 59, the Quick Inventory of Depressive Symptoms 60, the Mood and Anxiety Questionnaire 61, the Childhood Trauma Questionnaire 62, age, education and I-C and il-cl reaction time and accuracy difference scores.