NeuroImage: Clinical 19 (2018) 831–839

Contents lists available at ScienceDirect

NeuroImage: Clinical

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

Acute trajectories of neural activation predict remission to pharmacotherapy in late-life depression

Helmet T. Karim[a,1], Maxwell Wang[b,1], Carmen Andreescu[a], Dana Tudorascu[a,c,d], Meryl A. Butters[a], Jordan F. Karp[a], Charles F. Reynolds III[a], Howard J. Aizenstein[a,e,*]

[a] Department of Psychiatry, University of Pittsburgh, Pittsburgh, USA
[b] Medical Scientist Training Program, University of Pittsburgh School of Medicine and Carnegie Mellon University, University of Pittsburgh, Pittsburgh, USA
[c] Department of Biostatistics, University of Pittsburgh, Pittsburgh, USA
[d] Department of Internal Medicine, University of Pittsburgh, Pittsburgh, USA
[e] Department of Bioengineering, University of Pittsburgh, Pittsburgh, USA

ARTICLE INFO

Keywords:
LLD
fMRI
Prediction

ABSTRACT

Pharmacological treatment of major depressive disorder (MDD) typically involves a lengthy trial and error process to identify an effective intervention. This lengthy period prolongs suffering and worsens all-cause mortality, including from suicide, and is typically longer in late-life depression (LLD). Our group has recently demonstrated that during an open-label venlafaxine (serotonin-norepinephrine reuptake inhibitor) trial, significant changes in functional resting state connectivity occurred following a single dose of treatment, which persisted until the end of the trial. In this work, we propose an analysis framework to translate these perturbations in functional networks into predictors of clinical remission. Participants with LLD (N = 49) completed 12-weeks of treatment with venlafaxine and underwent functional magnetic resonance imaging (fMRI) at baseline and a day following a single dose of venlafaxine. Data was collected at rest as well as during an emotion reactivity task and an emotion regulation task. Remission was defined as a Montgomery-Asberg Depression Rating Scale (MADRS) ≤ 10 for two weeks. We computed eigenvector centrality (whole brain connectivity) and activation during the emotion regulation and emotion reactivity tasks. We employed principal components analysis, Tikhonov-regularized logistic classification, and least angle regression feature selection to predict remission by the end of the 12-week trial. We utilized ten-fold cross-validation and Receiver Operator Curves (ROC) curve analysis. To determine task-region pairs that significantly contributed to the algorithm's ability to predict remission, we used permutation testing. Using the fMRI data at both baseline and after the first dose of treatment yielded a sensitivity of 72% and a specificity of 68% (AUC = 0.77), a 15% increase in accuracy over baseline MADRS. In general, the accuracy at baseline was further improved by using the change in activation following a single dose. Activation of the frontal cortex, hippocampus, parahippocampus, caudate, thalamus, medial temporal cortex, middle cingulate, and visual cortex predicted treatment remission. Acute, dynamic trajectories of functional imaging metrics in response to a pharmacological intervention are a valuable tool for predicting treatment response in late-life depression and elucidating the mechanism of pharmacological therapies in the context of the brain's functional architecture.

1. Introduction

In late-life depression (LLD), the time between initiating treatment and clinical response generally takes 4–6 weeks. This delayed clinical effect is associated with prolonged suffering, exacerbated medical comorbidities, and increased risk of suicide (Andreescu and Reynolds, 2011). While many studies report changes in neural activation in depressed older adults during emotional reactivity or regulation tasks, relatively few studies have focused on the predictive utility of the identified changes or described changes within a timeframe that permits this information to be used clinically.

Unlike neuroimaging literature, there exists a rich literature on the clinical, genetic, and social factors that affect treatment response and remission. Past studies have found that factors like: allelic variation in serotonin transporters, endogenous depression status (genetic risk or family history of depression), sex, baseline depression severity as well

* Corresponding author at: Western Psychiatric Institute and Clinic, 3811 O'Hara Street, Pittsburgh, PA 15213, USA.
E-mail address: aizen@pitt.edu (H.J. Aizenstein).
[1] Shared first author contribution.

https://doi.org/10.1016/j.nicl.2018.06.006
Received 23 February 2018; Received in revised form 1 June 2018; Accepted 4 June 2018
Available online 08 June 2018

as change within 2–4 weeks of treatment, duration of current episode, age of onset of depression, response and adherence to past treatments, pain symptoms, baseline anxiety, suicide ideation, rapid eye movement (REM) latency, and baseline sleep quality have all been associated with treatment response, but even other factors like family support, social inequalities, expectations and perceived quality of care, heavy drinking, presence of stress-provoking agents, as well as eccentric personalities (paranoid, schizoid, or cluster C personality disorders) are important (Andreescu et al., 2007, 2008; Cohen et al., 2006; Dew et al., 1997; Gallagher and Thompson, 1983; Gildengers et al., 2005; Joel et al., 2014; Karp et al., 1993, 2005; Lotrich et al., 2001; Marmar et al., 1989; Martire et al., 2008; Morse et al., 2005; Mulsant et al., 2006; Oslin, 2005; Pollock et al., 2000; Reynolds et al., 1991; Smith et al., 1999; Szanto et al., 2003; Tew et al., 2006). In general, the most commonly used clinical indicators of treatment response include baseline symptom severity, previous response to treatment, and symptom change within the first two weeks (Andreescu et al., 2008; Joel et al., 2014).

There is emerging evidence that neural markers may have predictive capacity toward reducing the number of trialed antidepressants and possibly improving antidepressant outcomes (Canli et al., 2005; Siegle et al., 2006). In a subset of the current sample in this analysis, we identified neural changes within a day of the first dose of antidepressants utilizing functional magnetic resonance imaging (fMRI), which was dependent on remission status (Karim et al., 2016; Khalaf et al., 2016). These early responses indicate that while behavioral changes often take weeks to manifest, a patient's underlying neural activity is quickly changed by antidepressant treatment in a detectable fashion. Additionally, in a subset of the current sample we showed that these early neural changes are also correlated with the chronic neural changes observed at the end of the antidepressant treatment trial (Karim et al., 2016; Khalaf et al., 2016). There is evidence that these fast changes are not only detectable but could also be potentially used to predict whether an antidepressant at a given dosage will be able to engage a neural target to facilitate long-term behavioral changes.

There exists a larger literature in mid-life major depressive disorder (MDD) compared to LLD, and these studies suggest that prediction of treatment response is feasible (Costafreda et al., 2009; Liu et al., 2012; Marquand et al., 2008; Nouretdinov et al., 2011). On average, these studies report accuracies ranging from 69 to 90% (mean 83%), sensitivities from 78 to 89% (mean 85%), and specificities from 52 to 90% (mean 80%). They utilize several algorithms including support vector machines, weighted regressions, logistic classifiers, and decision trees, and mainly employ leave-one-out cross-validation (LOOCV) due to limited sample size. One study in late-life demonstrated that resting state connectivity predicted treatment response, while other structural features predicted diagnosis (Patel et al., 2015, 2016). Patel et al. (2015) identified that the integrity of white matter in the anterior salience network and the functional connectivity of the dorsal default mode network predicted treatment response. In general, these studies focus on either structural or functional imaging data collected at baseline and are potentially limited due to sample size and cross-validation concerns. While there is a much larger literature that suggests that several features are predictive of treatment response in LLD, comparatively fewer studies have investigated the predictive capacity of neural engagement during neural activation and connectivity and far fewer have investigated acute changes in neural activation in response to treatment.

In this study, in a sample of LLD participants (N = 49) receiving venlafaxine, we investigated the treatment response predictive capacity of three neural markers: activation during an emotion reactivity task, activation during an explicit emotion regulation task, and whole brain voxel-wise connectivity (eigenvector centrality) at rest. These metrics were chosen based on past literature, which identifies abnormal neural activation in MDD – mainly hyperlimbic activation coupled with hypofrontal activation (related to emotion reactivity and executive function that is critical for explicit emotion regulation). Further, it is also well established that depressive symptoms are associated with resting state network dysfunction (Alexopoulos, 2005; Andreescu et al., 2013; Karim et al., 2016; Wu et al., 2011). We investigated the predictive capacity of pre-treatment neural activation and of the change in neural activation following a single dose of venlafaxine. We compared the predictive capacity of these markers (both separately and in unison) to the predictive capacity of baseline depression severity. This is one of few studies that have investigated the predictive utility of acute change in neural activation with respect to treatment remission in LLD. We hypothesized that these markers would have greater predictive capacity than baseline depression severity and that early changes in these tasks would improve our predictive capacity than pre-treatment neural markers alone. We further hypothesized that neural and clinical markers together would further improve response prediction.

2. Methods

2.1. Study design and participants

We collected neuroimaging data as part of a larger 5-year multi-site study of treatment in LLD that collected neuroimaging data at one site (Pittsburgh, USA). Participants were recruited and prescribed with open-label venlafaxine (a serotonin and norepinephrine reuptake inhibitor). Participants were included if they were at least 50 years old, met Diagnostic and Statistical Manual for Mental Disorders-IV (SCID-IV) criteria for MDD and had a Montgomery-Asberg Depression Rating Scale (MADRS) score of 15 or higher at baseline. Participants were excluded if they had a history of mania or psychosis, alcohol or substance abuse (within last 3 months), dementia or neurodegenerative disease as well as conditions with known effects on mood and cognition (e.g. stroke, multiple sclerosis, vasculitis, significant head trauma, and/or unstable hypertension). Informed consent was obtained from all participants prior to engaging in any research procedures, and the University of Pittsburgh Institutional Review Board approved this study.

All MRI scanning was conducted in the morning. Five MRI scans were collected during the treatment trial. Participants came in on the first day for a baseline scan (no medication). In the same evening, they were given a placebo, after which they returned the next day for another scan (placebo scan). The evening of that scan, they were given their first dose of venlafaxine (37.5 mg), after which they returned the next day for another scan (single dose scan). They continued their medication for one week and returned for another scan (week one scan). They returned a final time after the end of the treatment trial (12 weeks, end scan). This analysis does not utilize the week one or end of trial scans as we intended to understand the predictive capacity of the neuroimaging data over an acute period. Henceforth, we only describe the relevant scanning procedures and analyses. During the trial, participants returned for weekly or bi-weekly clinical visits and the venlafaxine dosage was increased as necessary (up to a maximum of 150 mg/day by week 6). Participants who did not show signs of response by week 6 had the dosage increased up to a maximum of 300 mg/day. At the end of the study, participants were classified as remitters if they had a MADRS $\leq$ 10 for at least two weeks during the trial (and remained so until the end of the trial). Fig. 1 summarizes the study timeline.

A total of 62 participants signed consent. Eleven were excluded due to: side effects of medication (N = 2), non-adherence to protocol (N = 2), inaccurate diagnosis of MDD (N = 1), and inability to determine remission status due to lost/missing data (N = 6). Among the remaining participants (N = 51), two participants did not complete all MRI scanning but did complete the treatment trial. In summary, 49 participants were included in this analysis.

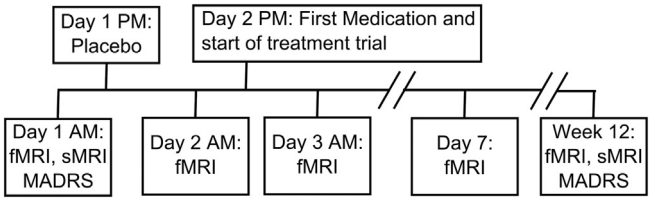

Fig. 1. The study design protocol: Functional and structural magnetic resonance imaging (fMRI and sMRI, respectively) was performed in the morning during various times through the trial. On the first day, participants came in for an fMRI scan (baseline) and then were given a placebo following the scan. On the second day, they returned for another fMRI scan and then were started on venlafaxine following the scan. They returned the next day (~12 h later) for another fMRI scan (1st dose change). They continued their medication as normal and came in for scans at the end of the first week and at the end of the trial. Only the fMRI scans at baseline and 1st dose change were used in this paper.

2.2. MRI protocols

All scanning was conducted at the University of Pittsburgh Medical Center on a Research dedicated 3 T Siemens Trio TIM scanner (Munich, Germany) using a 12-channel head coil. The baseline and end scan protocol included both a structural and functional image, while other scans collected only functional sequences. In this manuscript, we limited our analysis to the functional sequences, which were a resting state sequence, an explicit emotion regulation task sequence, and an emotional reactivity (faces/shapes) sequence.

An axial, whole brain 3D magnetization prepared rapid gradient echo (MPRAGE) was collected with repetition time (TR) = 2300 ms, echo time (TE) = 3.43 ms, flip angle (FA) = 9°, inversion time (TI) = 900 ms, field of view (FOV) = 256 × 224, 176 slices, 1 mm isotropic resolution and with Generalized Autocalibrating Partial Parallel Acquisition (GRAPPA) factor = 2. An axial, whole brain 2D fluid attenuated inversion recovery (FLAIR) was collected with TR = 9160 ms, TE = 90 ms, FA = 150°, TI = 2500 ms, FOV = 256 × 212, 48 slices, and 1 × 1 × 3 mm resolution.

An axial, whole brain (excluding cerebellum) echo planar (EPI) T2*-weighted functional image was collected to measure the blood oxygen level dependent (BOLD) response with TR = 2000 ms, TE = 34 ms, FA = 90°, FOV = 128 × 128, 28 slices, 2 × 2 × 4 mm resolution. The duration of the face/shapes task (see *Functional Imaging Metrics*) was 117 volumes (~4 min), the explicit emotion regulation task (see *Functional Imaging Metrics*) was 270 volumes (~9 min), and the resting state was 150 volumes (~5 min). Due to variability in placement by MR technicians the coverage of the functional scans was in general limited to above the cerebellum and below the top aspect of the motor cortex (though this varied slightly between functional sequences). Participants were instructed to lie awake and view a cross hair during resting state.

2.2.1. Emotional reactivity task

The face/shapes task is widely used and has been found to robustly activate the amygdala (Hariri et al., 2002, 2003). Participants were instructed to match either a face cue or a shape cue. A cue was shown on the top center of the screen and they were instructed to respond with an MR-compatible glove (left or right index finger) by matching to one of two simultaneously presented faces. The facial expressions shown were either angry or fearful. During the shapes, they match a shape to one of simultaneously presented shapes. The shapes task (5 blocks) was interleaved with the faces task (4 blocks) and each block lasted 24 s containing 6 trials (4 s each). Before the beginning of each block participants were instructed visually to “match emotion” or “match form” (2 s). The face images are presented from a set of 12 different images (six per block, three of each sex) and are all derived from a standard set of pictures of facial affect. Stimulus presentation and responses were controlled using E-prime software (Psychology Software Tools, Inc., Pittsburgh).

2.2.2. Explicit emotional regulation task

Participants were shown emotionally neutral or negative images from the standardized International Affective Picture System (IAPS) (Lang, 2005) and were instructed to either “Look” or “Decrease.” This task has been described previously (Khalaf et al., 2016) and has been used to activate prefrontal cortex (especially the dorsolateral prefrontal cortex) as a means of explicitly regulating limbic reactivity. During the look instruction, participants were to view content naturally. During the decrease instruction, participants were instructed to reappraise negative images to actively alter the elicited emotion. A master level staff member instructed participants on how to reappraise prior to entering the scanner. After each image they were asked to rate how negatively they felt from 1 to 5. The neutral (11 events), negative (15 events), and negative regulate (15 events) conditions were interleaved and each event lasted 6 s. The inter-trial interval was 13 s with no jitter (though they were not locked to a TR). This allowed for modeling of each individual response by allowing for enough time in between each stimulus, but likely resulted in lower power to detect each individual effect. The images are presented from a set of images and stimulus presentation and responses were controlled using E-prime software (Psychology Software Tools, Inc., Pittsburgh).

2.3. Structural processing

All processing was conducted using statistical parametric mapping (SPM12; http://www.fil.ion.ucl.ac.uk/spm/) in MATLAB (MATLAB 2016b, The MathWorks, Natick, 2016). Interpolation was conducted using 4th degree B-spline interpolation, normalized mutual information similarity metric for coregistration between images of different types, and mutual information similarity metric for motion correction unless otherwise stated. The FLAIR was coregistered to the MPRAGE (affine transform). Both images were input into a multi-spectral segmentation (Ashburner and Friston, 2005), which (after bias correction) segmented them into gray, white matter, cerebrospinal fluid, skull, soft-tissue, and air. Due to high white matter hyperintensity burden the number of Gaussians used to identify white matter was set to two (which improves the segmentation). This process generates a deformation field that can be used to normalize other images to a standard anatomic space (Montreal Neurological Institute, MNI) (Ashburner and Friston, 2005). An automatic mask for the intracranial volume was generated by thresholding the intracranial tissues with a probability of 0.1, filling the mask (imfill), and then performing a morphological closing operation (imclose, sphere of one voxel) in MATLAB. This mask (intracranial volume, ICV) was applied to the MPRAGE to remove non-brain tissues (which improves functional-structural coregistration).

2.4. BOLD pre-processing

The explicit emotion regulation task and the resting state data were slice time corrected (temporally middle slice was used as reference) prior to performing motion correction. All functional BOLD data was motion corrected (rigid coregistration to the mean), coregistered to the skull-stripped MPRAGE (mean functional image used to calculate affine transformation), normalized to MNI space using the deformation field calculated previously (2 mm isotropic resolution), and smoothed using a Gaussian kernel with FWHM of 8 mm. All images were investigated by human eye to confirm that coregistration and normalization steps were accurate.

Motion was evaluated using ArtRepair toolbox (Mazaika et al., 2007). During the emotional faces reactivity task, participants had low maximum translations [mean = 1.26 mm (std = 1.21)], low root mean squared (RMS) [1.11 mm (0.81)], and low percentage of volumes displaying head jerks above 0.5 mm [6.2% (10.7%)]. During the resting

state, participants had low maximum translations [1.27 mm (1.26)], low root mean squared (RMS) [1.04 mm (0.85)], and slightly higher percentage of volumes displaying head jerks above 0.5 mm [10.9% (19.9%)] that were corrected for using wavelet-despiking in later stages. During the explicit emotion regulation task, participants had low maximum translations [1.87 mm (1.91)], low root mean squared (RMS) [1.40 mm (1.08)], but slightly elevated percentage of volumes displaying head jerks above 0.5 mm [9.4% (30.8%)], with a few particularly bad cases that were removed. There were no group differences in any of these motion metrics between remitters and non-remitters between any time points.

For resting state BOLD, spike artifacts were removed using a previously established method that uses wavelets to filter spike artifacts (Patel et al., 2014). Five principal components of white matter and cerebrospinal fluid were extracted as well as 6 motion parameters and a vector to model the mean of the time series (Whitfield-Gabrieli and Nieto-Castanon, 2012). Band-pass filtering was conducted by including several regressors that represented cosines with all discrete frequencies except those within the standard expected resting state frequencies (0.008 to 0.15 Hz).

2.5. Modeling task activation: emotion reactivity and emotion regulation tasks

Mass-univariate general linear modeling (i.e. each voxel is independently modeled) was performed to model the mean, faces task, shapes task, and six parameters of motion (from motion correction). The canonical hemodynamic response function was used to convolve the faces and shapes tasks to expected hemodynamic responses. A high-pass filter of 1/128 Hz was utilized to account for low frequency noise. An autoregressive [AR(1)] filter was used to account for serial correlations due to aliased biorhythms and unmodelled activation. The contrast faces minus shapes was used to perform all voxel-wise analyses.

Similarly, the explicit emotion regulation task included similar parameters however it modeled the activation during the neutral and negative viewing tasks as well as the reappraisal task (during viewing of negative images). The contrast of interest was negative reappraise minus negative viewing, which modeled the activation during reappraisal adjusting for activation during the negative viewing task.

2.6. Resting state BOLD: eigenvector centrality

Eigenvector centrality was calculated using the fastECM toolbox (Lohmann et al., 2010). Briefly, centrality is a measure of connectedness of a voxel or region. Mean centrality is a related measure that calculates the mean voxel-wise connectivity of a single voxel to all other voxels where a greater centrality would imply that a voxel is more widely connected. FastECM uses singular value decomposition to circumvent the calculation of large correlation matrices.

2.7. Dimensionality reduction and machine learning

We used a combination of Principal Component Analysis (PCA), Least Angle Regression, and Logistic Classification to identify differences on the individual level in our fMRI features that could be linked to remission. Using SPM12, each individual's fMRI maps (resting state eigenvector centrality, emotional regulation, and emotional reactivity) were averaged across 116 regions in MNI152 space outlined in the Automatic Anatomical Labeling Atlas (Tzourio-Mazoyer et al., 2002), resulting in a 348-length feature vector for each individual for a given time point. While we could have estimated regions that were significantly associated with each task, without an independent sample we felt that this may bias our estimates (as we would have to select those regions using this sample). PCA allows for the estimation of principal component vectors across participants that vary together, thus it is likely that regions that activated similarly were combined into a single vector.

We tested two major fMRI feature vectors. One was the fMRI features (116 regions of activation during emotion reactivity, emotion regulation, or centrality) at baseline, prior to treatment. The other was the change in fMRI features following a single dose (or placebo), which was defined as the difference between the feature vector after a single dose (or placebo) and baseline. Due to small-number error concerns, we chose this method over a percentage change metric.

All analyses were performed within a ten-fold cross-validation scheme to address over-fitting and multiple comparisons concerns. To avoid biasing our estimates, all data demeaning, dimensionality reduction, feature selection, and hyper-parameter optimization were performed via nested cross-validation loops. To establish bounds on the accuracy of our algorithm, we repeated the cross-validation scheme 30 times, each time redrawing the cross-validation folds.

We used a combination of PCA and Least Angle Regression to selectively reduce the dimensionality of the dataset to components that were relevant to remission (Wold et al., 1987). Least Angle Regression has been proposed as a “less greedy” alternative to the popular LASSO (Grandvalet, 1998) algorithm that favors the net contribution by multiple features simultaneously over identifying single features independently (Efron et al., 2004). Using the components selected by these two algorithms, a logistic classifier was then trained on these components and used to predict remission on the test fold of the cross-validation scheme. Accuracy was assessed using Receiver Operator Curves (ROC) analysis.

To determine predictors that utilized information from multiple scanning time points or the baseline MADRS score, we averaged the predictors from each individual algorithm to generate an averaged predictor, a concept commonly referred to as an unweighted voting algorithm (Dietterich, 2000). This procedure is meant to combine the predictive power of several feature sets without suffering from over-fitting concerns.

2.8. Permutation testing

To determine which anatomical regions of the fMRI metric maps leant themselves to accurate and reliable predictions of remission, we utilized permutation testing. More specifically, we randomly shuffled the remitter/non-remitter labels within our dataset and recomputed the entire cross-validated classification pipeline. Each feature (a single region for a given fMRI feature) was then ranked on its relative importance to the classifier compared to all other regions across all metrics. This ranking procedure was repeated 1000 times and compared to the rankings found when using the “true” remission labels. A region/task pair is significantly associated with remission if its true ranking is within the top 5% of random permutation ranking trials.

To understand the relative contribution of the fMRI metrics to each prediction used in this paper, we also applied an additional permutation test to the feature sets themselves (i.e., instead of permuting remitter/non-remitter labels, the images themselves were permuted). For a given time point, we permuted the features from a single fMRI feature and repeated the cross-validation and training/evaluation procedure. This was repeated 1000 times for each fMRI feature at both time points to assess the individual contribution from each fMRI map.

3. Results

The differences between remitters and non-remitters at baseline regarding age, sex, education, race, Cumulative Illness Rating Scale for Geriatrics (CIRS-G), depression type, Mini-Mental State Examination (MMSE), and Montgomery-Asberg Depression Rating Scale (MADRS) are shown in Table 1. Only baseline and end of trial MADRS showed significant differences between the two groups. Non-remitters had higher baseline MADRS scores indicating higher depression severity.

Table 1
Differences in various demographic and clinical factors between remitters and non-remitters are shown above. Apart from the baseline MADRS, there are no significant differences between the remitter and non-remitter group. Abbreviations: CIRSG—Cumulative Illness Rating Scale for Geriatrics, MMSE—Mini-Mental State Examination, MADRS—Montgomery-Asberg Depression Rating Scale, IQR—interquartile range.

	Remitters (N = 25)	Non-remitters (N = 24)	Comparison
Age (years): median ± IQR	66 ± 10.5	64.5 ± 7.5	t (47), p = 0.33
Sex	7 Male, 18 female	11 Male, 23 female	X^2 (48), p = 0.20
Race	21 Caucasian, 4 African-American	20 Caucasian, 4 African-American	X^2 (48), p = 0.95
Education (years): median ± IQR	13 ± 6	15.5 ± 2.5	Mann-Whitney *U* test, p = 0.37
Depression type	15 Recurrent, 8 single (N = 23)	14 Recurrent, 10 single	X^2 (46), p = 0.63
CIRS-G: median ± IQR	10 ± 4	8 ± 4.5 (N = 23)	t (46), p = 0.72
Baseline MMSE: median ± IQR	30 ± 2	29 ± 1.75	Mann-Whitney U test, p = 0.54
Baseline MADRS: median ± IQR	23 ± 11.25	27 ± 7.75	t (47), p = 0.04
Ending MADRS: median ± IQR	3 ± 6 (N = 24)	19.5 ± 8 (N = 23)	*t*-Test (45), p = 4.5×10^{-11}

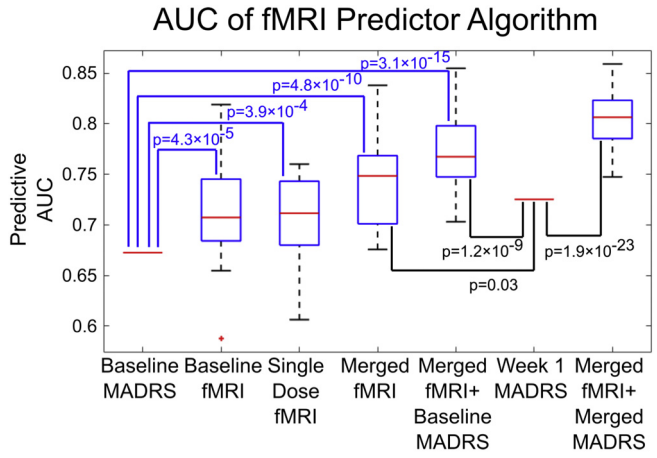

Fig. 2. The predictive accuracy of remission among 49 subjects was determined using 30 trials of repeating a 10-fold cross-validation scheme and is shown via interquartile range boxplots. The second and third column represent the accuracy of using the classification algorithm on only the functional imaging data (resting state centrality, emotional reactivity task, and emotional regulation task) available at baseline or the change in imaging metric a day after the first dose of venlafaxine. The fourth column represents averaging the predictions from the second and third column, while the fifth column shows the accuracy from averaging the predictions from the first four columns. We find that utilizing functional imaging along with our proposed algorithms improves the predictive power of the MADRS questionnaire by 15% (other demographic variables such as age, sex, education level, and race had no significant predictive power and thus were not included). The last two columns show the accuracy of utilizing the MADRS at one week (change in MADRS was less accurate) and using that value in combination with the fMRI data at baseline and post-first dosage. p-Values were calculated as one-sample *t*-tests with a null hypothesis that the accuracy of the algorithm was equal to that to the MADRS at baseline or at one week.

We also report each of these measures for the entire sample (intent-to-treat) in Supplemental Table 1.

3.1. Accuracy of fMRI classification algorithm

Fig. 2 illustrates the accuracy of the fMRI classification algorithm using the fMRI scans collected at different time points using the Area under Curve (AUC) metric. Within this context, AUC represents the probability that given two participants, one remitter and one non-remitter, that the algorithm of interest will correctly classify the remitter as being more likely to have a positive treatment response. This ranges from 50% for a random-guessing algorithm to 100% for perfect accuracy. We found that utilizing our fMRI procedures and classification algorithm yielded an approximate 15% increase in AUC over that of simply using the MADRS alone. In general, we show that while there is no significant difference in AUC between baseline and change in fMRI features, the two features together significantly improve the overall AUC and adding baseline depression severity further improves it. For comparison, we also show the accuracies when utilizing the MADRS at one or two weeks after the trial start, as well as the composite accuracy of these values when used with the fMRI results. We find that our imaging approach significantly outperforms both these values by approximately 7% in AUC. Fig. 3 shows the ROC curve and individual predictors from the median trial (i.e., trial with median AUC) from the fMRI/MADRS algorithm shown in the fifth column in Fig. 2. The placebo minus baseline scan did not predict remission [median AUC of 0.56 (IQR, 0.52–0.6)] better than MADRS (p = 2.8e−11). Placebo minus baseline combined with the baseline prediction [median AUC of 0.63 (IQR, 0.68–0.71)] was not significantly better than MADRS (p = 0.688). Thus, the placebo results were excluded from further analysis.

Fig. 4 shows the region/task pairs that passed permutation significance testing. As only one region from the 1st dose change in the fMRI resting state centrality metric passed permutation testing, maps of that region (left superior temporal gyrus) were not shown. Table 2 shows the list of significant region/task pairs along with their exact weights assigned by the classification algorithm, where positive signs indicate a positive association with remission (e.g., higher baseline activation or greater increase in activation is predictive of remission). These regions included frontal cortex, parahippocampus, hippocampus, caudate, thalamus, medial temporal cortex, middle cingulate, and visual cortex.

Fig. 5 illustrates the effects of permuting the feature set for a given fMRI metric at a single time point. We found that the largest drop in AUC occurred when permuting the emotional reactivity feature map, indicating the relative utility of this probe in finding metrics that can be used for remission prediction.

4. Discussion

In this paper, we present a novel method of predicting treatment response in late-life-depression by utilizing the functional imaging metrics in response to a single-dose of a pharmacological intervention. By leveraging a combination of resting and task-based functional imaging scans (emotional reactivity and regulation), both before treatment and after the first dose of venlafaxine, we found that we could improve the predictive capacity of baseline clinical factors by 15% in a cohort of 49 participants. Within this cohort, we define baseline clinical factors as the total baseline Montgomery-Asberg Depression Rating Scale (MADRS) as other demographic information (age, sex, education level, and race) had no significant predictive power. We found that while both the baseline scan and the scan following a single dose were able to reliably predict remission with a greater accuracy than the baseline MADRS, utilizing them together yielded significantly higher accuracy.

We further demonstrated that the change in activation following a single dose significantly improves the overall predictive capacity

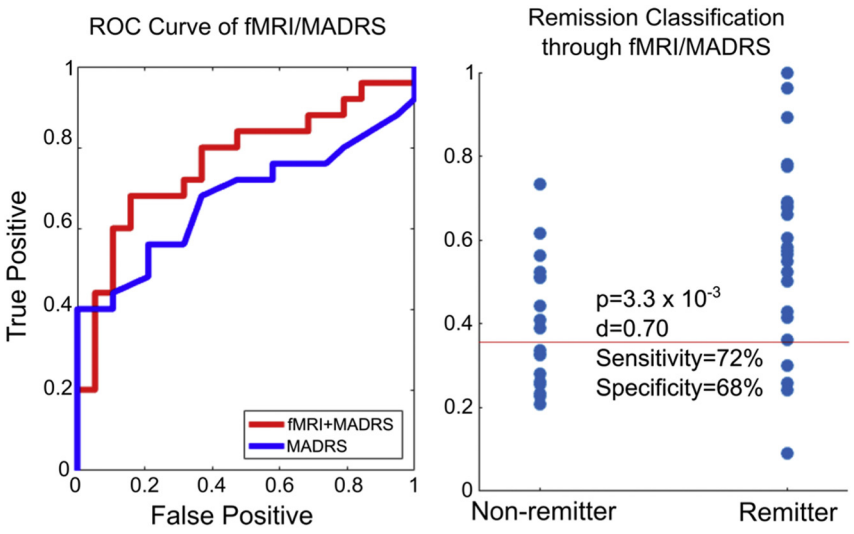

Fig. 3. Left) The ROC curve of remission-prediction accuracy using functional imaging data before treatment, functional imaging 24 h after the first treatment dose, and the baseline MADRS score. Here true positive denotes the percentage of remitting subjects that were correctly predicted as such, while false positive indicates the percentage of non-remitters incorrectly classified as remitters. The ROC curve of MADRS alone is shown for comparison. Right). The predicted remission probabilities used to generate the ROC curve on the left. "p" is calculated as a two-sample *t*-test while "d" is the Cohen's effect size. The red line represents the cutoff probability that gives the sensitivity and specificity values shown.

compared to pre-treatment prediction. One possible application of these results is facilitating neural target engagement – where we would both find a neural target and engage it to significantly improve depressive symptoms. Given the rising popularity of innovative methods to target neural markers (using interventions such as transcranial magnetic stimulation, TMS, or transcranial direct-current stimulation, tDCS), determining the appropriate neural targets to engage is a field of rising interest (George et al., 2000). Future studies investigating differential target engagement are needed, as past studies seem to suggest that different therapies result in differential engagement of similar neural targets (Frodl et al., 2011). This would allow for identification of a pre-treatment neural target and matching with an appropriate antidepressant or therapy to engage or alter that target.

There is a small, but significant past literature in LLD that focuses on pre-treatment activation as well as connectivity associated with treatment response (Aizenstein et al., 2009; Brassen et al., 2008; Khalaf et al., 2016). These studies have identified hypoactive executive function during emotion reactivity (Aizenstein et al., 2009; Brassen et al., 2008), changes in resting state connectivity (including one using a subset of this study, Karim et al.) (Alexopoulos, 2005; Andreescu et al., 2013; Karim et al., 2016; Wu et al., 2011), including in a subset of this study we found lower pre-treatment centrality in the inferior frontal

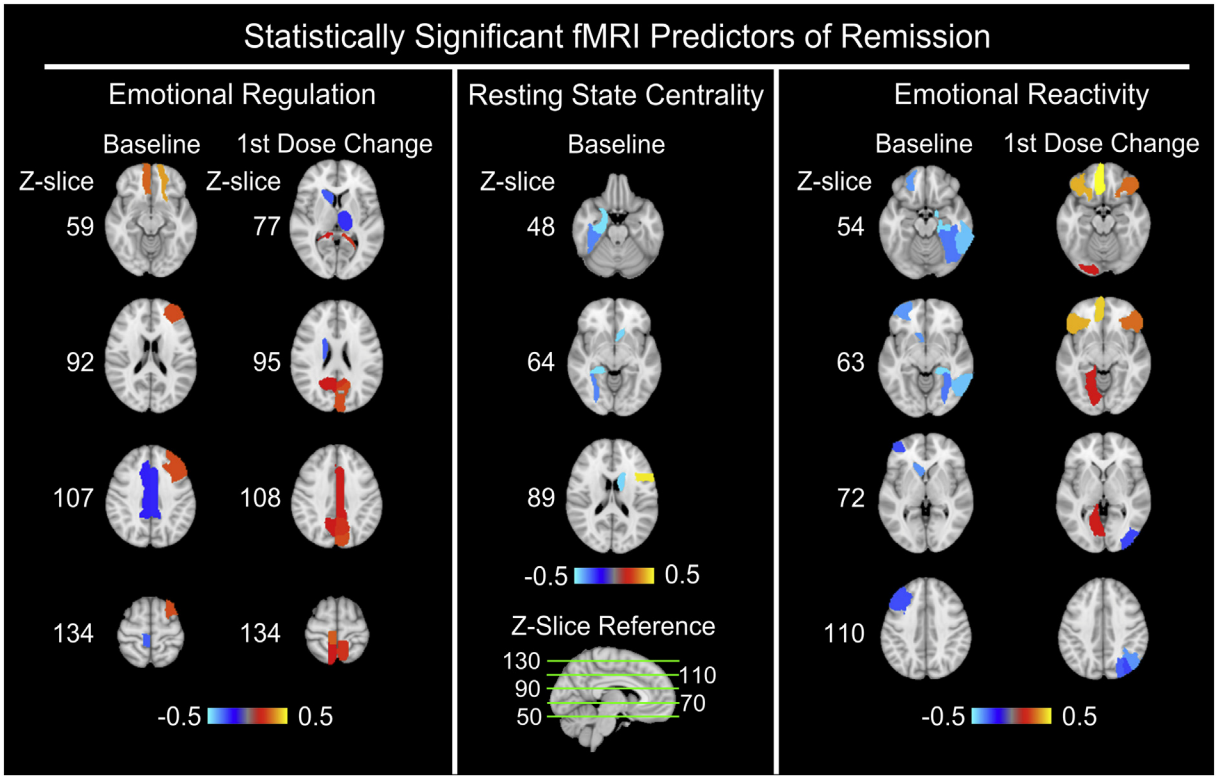

Fig. 4. Axial slices of the relative importance of region/task pairs that passed statistical permutation significance testing (p = 0.05) are shown above, with the z-coordinates in MNI152 space shown for reference. Here, bright yellow shades indicate that the region/task pair is positively associated with remission (i.e., higher baseline activation or greater increase in activation is predictive of remission), whereas bright blue shades indicate a negative association. As the 1st dose change of resting state centrality only displayed one region that passed permutation testing, maps of that region (left superior temporal gyrus) are not shown. These results were calculated by averaging the predictor importance weights assigned by the classification across all ten folds of cross-validation and over all thirty trials.

Table 2
The importance weights of the region/task pairs that passed permutation significance testing (p = 0.05). Here, positive signs indicate a positive association with remission. These results were calculated by averaging the predictor importance weights assigned by the classification across all ten folds of cross-validation and over all thirty trials.

	Baseline	1st dose change
Emotional reactivity	Left middle frontal gyrus: −0.29 Left lateral orbitofrontal cortex: −0.38 Right hippocampus: −0.58 Right parahippocampus −0.45 Right fusiform: −0.34 Left caudate: −0.37 Right inferior temporal gyrus: −0.43	Left/right inferior frontal gyrus, pars orbitalis: +0.40/+0.31 Left middle frontal gyrus, orbital part: +0.43 Left rectus: +0.61 Left lingual: +0.19 Right superior occipital gyrus: −0.32 Right middle occipital gyrus: −0.28 Right angular gyrus: −0.37
Emotional regulation	Left superior frontal gyrus, orbital part: +0.37 Left middle frontal gyrus, orbital part: +0.30 Left middle frontal gyrus: +0.27 Left/Right middle cingulate area: −0.25/−0.25 Right paracentral lobule: −0.31	Left middle cingulate area: +0.19 Left cuneus (increased): +0.27 Left/right precuneus: +0.23/+0.18 Right paracentral lobule: +0.29 Right caudate: −0.31 Left thalamus: −0.26
Resting state centrality	Left inferior frontal gyrus, pars opercularis: +0.46 Right parahippocampus: −0.46 Right fusiform: −0.36 Left caudate: −0.45	Left superior temporal gyrus: −0.54

gyrus (IFG) as well as greater pre-treatment MeFG centrality (Karim et al., 2016). One large study identified pre-treatment subtypes of depression, specifically that there exist four major subtypes that have distinct abnormalities in resting state connectivity (Drysdale et al., 2017), that also demonstrated a specificity for different treatments. In general, our results and previous studies support the use of pre-treatment fMRI for improving treatment outcomes.

Previous studies have identified that there exist neural changes following treatment in LLD (Aizenstein et al., 2009; Brassen et al., 2008). In a subset of this study, we have demonstrated that there not only exist changes in both activation during explicit emotion regulation and resting state connectivity, but also that these changes occur following a single dose of an antidepressant (Karim et al., 2016; Khalaf et al., 2016). Our past work also identified that the early changes in activation and connectivity reflected later changes at the end of the study, in several instances the connectivity or activation after a single dose did not differ from the end of the trial (Aizenstein et al., 2009; Brassen et al., 2008).

The most critical component among the tasks (emotion reactivity, explicit emotion regulation) and resting state (eigenvector centrality) seems to be emotion reactivity. It may be that the emotion reactivity task also involves implicit processing and consequently implicit regulation. Past work has shown that these tasks evoke limbic reactivity, and the consequent neural response is to regulate this signal through further implicit processing and possibly regulation. These processes may be more acutely altered compared to centrality and explicit emotion regulation. For example, previous work in a subset of this sample has demonstrated that changes in explicit emotion regulation do not manifest acutely in response to treatment, likely since such a change would require an alteration of higher level cognitive networks which are hypothesized to not occur until much later on – though there do exist some changes (during negative viewing conditions) in some regions (Khalaf et al., 2016). More studies are needed to truly understand this process; however, we could speculate that this may be because the changes in activation early on are associated with implicit rather than explicit emotion processing (i.e., emotion reactivity vs. explicit regulation task, respectively). This may be how these processes (changes in activation during emotion reactivity task) are altered without explicit awareness.

We have shown that several regions have a predictive capacity for remission, the majority of which are in the frontal cortex. The PFC works to regulate limbic activation and thus is a critical part of treatment response. Consequently, other regions like the parahippocampus, hippocampus, caudate, thalamus, and middle cingulate as well as the medial temporal cortex are also implicated. In general, these may be related to contextual processing, reward, and explicit emotion regulation. Without understanding the changes in activation in each region, it is unclear how each independent region contributes, however the regions that are predictive at baseline are not necessarily predictive in the change following a single dose. It is possible that the baseline factors may predict treatment resistance (rather than treatment response to a particular therapy), whereas the change is predictive for response to a particular antidepressant (though we cannot test this). For instance, low executive network recruitment may predict a poor response in general, but an increase in activation following an antidepressant may also predict, however these are not necessarily the same (e.g., the

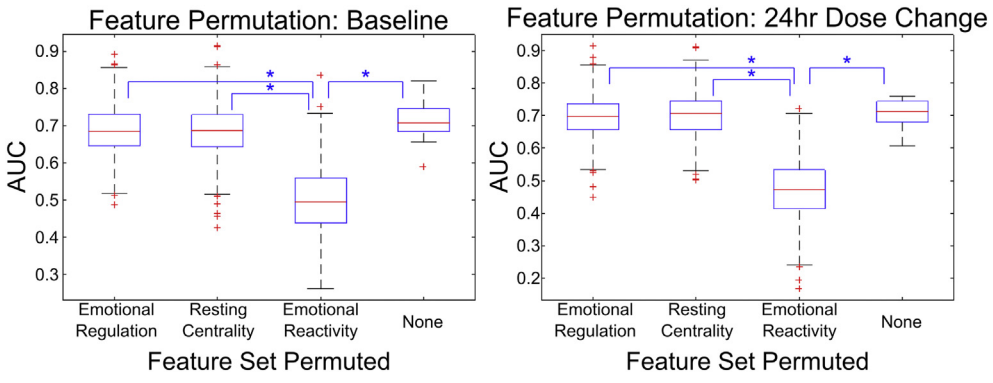

Fig. 5. The impact to the classification accuracy of the fMRI predictor algorithm when certain feature map sets are permuted between subjects is shown via interquartile boxplots. The first three columns represent the result of shuffling the features of a given fMRI metric map between subjects, while the last column represents the accuracy when no features are permuted. Note that the largest drop in accuracy occurs when the emotional reactivity features are permuted, indicating the utility of using a task to probe specific features of neural activity. Statistical significance was determined via two-sample *t*-tests. All the pairs marked by asterisks have a p-value bound below 10^{-35}.

antidepressant could have no effects on executive network activation). Further, this may reflect individuals that do not necessarily fit into the monoamine theory of depression. Future studies should investigate not only remission to particular pharmacotherapies compared to placebo, but also remission to other treatments (like transcranial magnetic stimulation or cognitive behavioral therapy). One caveat is that the direct activation of a region may not be necessary for remission but occur as an indirect effect of perturbations to other regions, however there is no way to determine this in our study. Specifically, the actual changes in activation and connectivity observed after a single dose may occur as a result of only a small perturbation that was initially caused by the venlafaxine. Only a study that directly injects a dose in the scanner and continuously monitored the changes in activation could directly answer such a hypothesis. We view these regions as potential neural targets for a more comprehensive perturbation study.

Traditional modeling approaches understand neural differences in the context of remission (outcome is traditionally defined as the neural markers while remission is a predictor variable), however this approach reverses this modeling by attempting to understand remission instead. While the clinical utility of such approaches is not yet well understood, this type of modeling helps improve our understanding of treatment response variability (since not all remitters or non-remitters are alike). For example, our work demonstrates the greater utility of the signal in the emotion reactivity task compared to the centrality or the emotion regulation task. Interestingly, the current approach and standard statistical analytical approaches are not necessarily convergent and the major factor that contributes to this is that traditional approaches make the assumption that all voxels are independent and for each voxel asks how the variance in activation is explained by remission. This approach instead assumes that the voxels are highly interdependent (using PCA allows independent components to be generated), and instead asks how the variance in remission is explained by these components.

There are several limitations in this study. While the sample size is comparatively large, it is limited from a machine learning perspective. This limitation is especially important in several ways. In particular, past literature has shown that depression is a highly heterogeneous disorder and thus likely contributes to the relatively modest improvement in prediction in our study as well as past work. Further, treatment response itself is highly heterogeneous and it is likely that there are many paths to remission even within a single antidepressant. Finally, there are other factors that may play an important role, for instance individual variability in antidepressant metabolism may contribute to the fMRI response. We employed a 10-fold cross-validation as well as ensuring that any data reduction was done in fold (to avoid bias), to affirm the viability of the algorithm and address over-fitting concerns. These approaches help ensure that the improvement in prediction (above MADRS alone) is stable, and by performing all data reduction in fold we further avoid biasing our model (as this would improve our estimate of the different components within a sample). We also conducted the cross-validation multiple times (redrawing the folds) as it is possible that certain folds are more predictive than others. While these may address some over-fitting, it is not a replacement for independent validation and future studies should include data on larger cohorts and allow for independent samples for verification. Ultimately, models that are well-validated should be shared to be tested in studies from multiple sites to test their predictive utility. This study was done in LLD; thus, it is unclear how well this would generalize to mid-life depression. Past studies have identified that there exist several clinical features that are predictive of treatment response, such as anxiety, cognitive impairment, and history of treatment response. In this study, we found that only baseline depression severity was predictive (though we did not have cognitive status, besides the Mini-Mental State Exam, or treatment history). Our only measure of anxiety was the single item on the MADRS – thus, a more complex measure of anxiety may be better suited to predicting treatment response. It remains unclear how our algorithm would do compared to one using only clinical features (like anxiety, but also cognitive measures). This may explain our limited specificity in our results, and by adding in more robust measurements of anxiety, cognitive measures, as well as other features of their depression (single vs. recurrent, life-time burden, etc.) we can improve the overall specificity of the model. However, from a statistical perspective, possessing multiple methods to assess remission probability will allow these methods to cross-verify each other to increase the final predictive power. Further, we have only predicted remission, however there are several aspects of remission and response that are important like the time to remit (early vs. late) as well as the stability of the remission. Due to the small sample size, we are unable to adequately investigate time to remission and without follow-up data are unable to look at the durability of this remission. Future studies should investigate not only these aspects but also other qualities of remission and response (i.e., relapse and recurrence – both long term qualities of response). Furthermore, it is unclear whether these changes following a single dose are a direct cause of the drug itself or indirect effects – however our past work has shown that the change in activation/connectivity reflects the changes at the end of the trial (Karim et al., 2017; Khalaf et al., 2016). We failed to predict remission using changes following a placebo, which further supports that this prediction is specific to the change following venlafaxine. It is possible that there may be order effects (e.g., practice effect on the tasks) that could have affected the predictive utility of the neural changes following the third session (single dose venlafaxine) and not the second session (placebo). Finally, past studies have demonstrated that there are several structural features that are predictive of treatment response: notably white matter hyperintensities (Bella et al., 2010; Gunning-Dixon et al., 2010; Sneed et al., 2011). We have decided against combining both structural and functional features to understand the predictive capacity of functional features. While structural features do not directly observe target engagement or differential effects of various treatments, they are likely important in understanding the subtype of depression – and thus should still be investigated in the future. Despite this, we believe that a combination of structural and functional features will allow greater insight into understanding remission in depression.

Our study has demonstrated that neural activation pre-treatment as well as following a single dose increases the predictive capacity for remission and further that emotion reactivity paradigms are the most useful predictors (compared to resting state EVC or explicit emotion regulation). This builds on an already existing literature that has identified neural and behavioral subtypes – as our pre-treatment markers predicted remission. We also identified a single dose engagement effect – building on a sparse literature that seems to suggest that the neural activation occurs acutely. Thus, measuring this engagement is likely an important part of improving the overall efficacy of these treatments. While the predictive utility is modest, this may largely be due to high heterogeneity in depression and high treatment response variability coupled with a small sample size and strict cross-validation approaches. In general, we demonstrate that the prediction may be useful, but future studies are needed to demonstrate the true clinical utility of neuroimaging in this field. Utilizing computational psychiatric approaches will allow for patients to be classified not only by their clinical symptoms, but also a set of neural targets that may need to be engaged. By engaging each target in a systematic manner, we may be able to improve overall response rates for depression treatment.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nicl.2018.06.006.

Acknowledgements

This study was funded by: NIMH R01 MH076079, 5R01 AG033575, K23 MH086686, P30 MH090333, P50 AG05133, 5R01 MH083660 and T32 MH019986. HTK had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. JFK has previously been given medication supplies for

investigator-initiated trials from Indivior and Pfizer. All other authors declare no conflicts of interest.

References

Aizenstein, H.J., Butters, M.A., Wu, M., Mazurkewicz, L.M., Stenger, V.A., Gianaros, P.J., Becker, J.T., Reynolds III, C.F., Carter, C.S., 2009. Altered functioning of the executive control circuit in late-life depression: episodic and persistent phenomena. Am. J. Geriatr. Psychiatry 17, 30–42.

Alexopoulos, G.S., 2005. Depression in the elderly. Lancet 365, 1961–1970.

Andreescu, C., Reynolds III, C.F., 2011. Late-life depression: evidence-based treatment and promising new directions for research and clinical practice. Psychiatr. Clin. N. Am. 34 (335–355, vii-iii).

Andreescu, C., Lenze, E.J., Dew, M.A., Begley, A.E., Mulsant, B.H., Dombrovski, A.Y., Pollock, B.G., Stack, J., Miller, M.D., Reynolds, C.F., 2007. Effect of comorbid anxiety on treatment response and relapse risk in late-life depression: controlled study. Br. J. Psychiatry 190, 344–349.

Andreescu, C., Mulsant, B.H., Houck, P.R., Whyte, E.M., Mazumdar, S., Dombrovski, A.Y., Pollock, B.G., Reynolds III, C.F., 2008. Empirically derived decision trees for the treatment of late-life depression. Am. J. Psychiatry 165, 855–862.

Andreescu, C., Tudorascu, D.L., Butters, M.A., Tamburo, E., Patel, M., Price, J., Karp, J.F., Reynolds III, C.F., Aizenstein, H., 2013. Resting state functional connectivity and treatment response in late-life depression. Psychiatry Res. 214, 313–321.

Ashburner, J., Friston, K.J., 2005. Unified segmentation. NeuroImage 26, 839–851.

Bella, R., Pennisi, G., Cantone, M., Palermo, F., Pennisi, M., Lanza, G., Zappia, M., Paolucci, S., 2010. Clinical presentation and outcome of geriatric depression in subcortical ischemic vascular disease. Gerontology 56, 298–302.

Brassen, S., Kalisch, R., Weber-Fahr, W., Braus, D.F., Buchel, C., 2008. Ventromedial prefrontal cortex processing during emotional evaluation in late-life depression: a longitudinal functional magnetic resonance imaging study. Biol. Psychiatry 64, 349–355.

Canli, T., Cooney, R.E., Goldin, P., Shah, M., Sivers, H., Thomason, M.E., Whitfield-Gabrieli, S., Gabrieli, J.D., Gotlib, I.H., 2005. Amygdala reactivity to emotional faces predicts improvement in major depression. Neuroreport 16, 1267–1270.

Cohen, A., Houck, P.R., Szanto, K., Dew, M.A., Gilman, S.E., Reynolds III, C.F., 2006. Social inequalities in response to antidepressant treatment in older adults. Arch. Gen. Psychiatry 63, 50–56.

Costafreda, S.G., Chu, C., Ashburner, J., Fu, C.H., 2009. Prognostic and diagnostic potential of the structural neuroanatomy of depression. PLoS One 4, e6353.

Dew, M.A., Reynolds III, C.F., Houck, P.R., Hall, M., Buysse, D.J., Frank, E., Kupfer, D.J., 1997. Temporal profiles of the course of depression during treatment. Predictors of pathways toward recovery in the elderly. Arch. Gen. Psychiatry 54, 1016–1024.

Dietterich, T.G., 2000. Ensemble methods in machine learning. In: Multiple Classifier Systems 1857, pp. 1–15.

Drysdale, A.T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., Fetcho, R.N., Zebley, B., Oathes, D.J., Etkin, A., Schatzberg, A.F., Sudheimer, K., Keller, J., Mayberg, H.S., Gunning, F.M., Alexopoulos, G.S., Fox, M.D., Pascual-Leone, A., Voss, H.U., Casey, B.J., Dubin, M.J., Liston, C., 2017. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nat. Med. 23, 28–38.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression. Ann. Stat. 32, 407–499.

Frodl, T., Scheuerecker, J., Schoepf, V., Linn, J., Koutsouleris, N., Bokde, A.L., Hampel, H., Moller, H.J., Bruckmann, H., Wiesmann, M., Meisenzahl, E., 2011. Different effects of mirtazapine and venlafaxine on brain activation: an open randomized controlled fMRI study. J. Clin. Psychiatry 72, 448–457.

Gallagher, D.E., Thompson, L.W., 1983. Effectiveness of psychotherapy for both endogenous and nonendogenous depression in older adult outpatients. J. Gerontol. 38, 707–712.

George, M.S., Nahas, Z., Molloy, M., Speer, A.M., Oliver, N.C., Li, X.B., Arana, G.W., Risch, S.C., Ballenger, J.C., 2000. A controlled trial of daily left prefrontal cortex TMS for treating depression. Biol. Psychiatry 48, 962–970.

Gildengers, A.G., Houck, P.R., Mulsant, B.H., Dew, M.A., Aizenstein, H.J., Jones, B.L., Greenhouse, J., Pollock, B.G., Reynolds III, C.F., 2005. Trajectories of treatment response in late-life depression: psychosocial and clinical correlates. J. Clin. Psychopharmacol. 25, S8–13.

Grandvalet, Y., 1998. Least absolute shrinkage is equivalent to quadratic penalization. In: ICANN 98. Springer, pp. 201–206.

Gunning-Dixon, F.M., Walton, M., Cheng, J., Acuna, J., Klimstra, S., Zimmerman, M.E., Brickman, A.M., Hoptman, M.J., Young, R.C., Alexopoulos, G.S., 2010. MRI signal hyperintensities and treatment remission of geriatric depression. J. Affect. Disord. 126, 395–401.

Hariri, A.R., Tessitore, A., Mattay, V.S., Fera, F., Weinberger, D.R., 2002. The amygdala response to emotional stimuli: a comparison of faces and scenes. NeuroImage 17, 317–323.

Hariri, A.R., Mattay, V.S., Tessitore, A., Fera, F., Weinberger, D.R., 2003. Neocortical modulation of the amygdala response to fearful stimuli. Biol. Psychiatry 53, 494–501.

Joel, I., Begley, A.E., Mulsant, B.H., Lenze, E.J., Mazumdar, S., Dew, M.A., Blumberger, D., Butters, M., Reynolds III, C.F., Team, I.G.I., 2014. Dynamic prediction of treatment response in late-life depression. Am. J. Geriatr. Psychiatry 22, 167–176.

Karim, H.T., Andreescu, C., Tudorascu, D., Smagula, S.F., Butters, M.A., Karp, J.F., Reynolds, C., Aizenstein, H.J., 2016. Intrinsic functional connectivity in late-life depression: trajectories over the course of pharmacotherapy in remitters and non-remitters. Mol. Psychiatry 22 (3), 450–457.

Karim, H., Andreescu, C., Tudorascu, D., Smagula, S., Butters, M., Karp, J., Reynolds, C., Aizenstein, H., 2017. Intrinsic functional connectivity in late-life depression: trajectories over the course of pharmacotherapy in remitters and non-remitters. Mol. Psychiatry 22, 450.

Karp, J.F., Frank, E., Anderson, B., George, C.J., Reynolds, C.F., Mazumdar, S., Kupfer, D.J., 1993. Time to remission in late-life depression: analysis of effects of demographic, treatment, and life-events measures. Depression 1, 250–256.

Karp, J.F., Weiner, D., Seligman, K., Butters, M., Miller, M., Frank, E., Stack, J., Mulsant, B.H., Pollock, B., Dew, M.A., Kupfer, D.J., Reynolds III, C.F., 2005. Body pain and treatment response in late-life depression. Am. J. Geriatr. Psychiatry 13, 188–194.

Khalaf, A., Karim, H., Berkout, O.V., Andreescu, C., Tudorascu, D., Reynolds, C.F., Aizenstein, H., 2016. Altered functional magnetic resonance imaging markers of affective processing during treatment of late-life depression. Am. J. Geriatr. Psychiatry 24, 791–801.

Lang, P.J., 2005. International affective picture system (IAPS): affective ratings of pictures and instruction manual. In: Technical Report.

Liu, F., Guo, W., Yu, D., Gao, Q., Gao, K., Xue, Z., Du, H., Zhang, J., Tan, C., Liu, Z., Zhao, J., Chen, H., 2012. Classification of different therapeutic responses of major depressive disorder with multivariate pattern analysis method based on structural MR scans. PLoS One 7, e40968.

Lohmann, G., Margulies, D.S., Horstmann, A., Pleger, B., Lepsien, J., Goldhahn, D., Schloegl, H., Stumvoll, M., Villringer, A., Turner, R., 2010. Eigenvector centrality mapping for analyzing connectivity patterns in fMRI data of the human brain. PLoS One 5, e10232.

Lotrich, F.E., Pollock, B.G., Ferrell, R.E., 2001. Polymorphism of the serotonin transporter: implications for the use of selective serotonin reuptake inhibitors. Am. J. Pharmacogenomics 1, 153–164.

Marmar, C.R., Gaston, L., Gallagher, D., Thompson, L.W., 1989. Alliance and outcome in late-life depression. J. Nerv. Ment. Dis. 177, 464–472.

Marquand, A.F., Mourao-Miranda, J., Brammer, M.J., Cleare, A.J., Fu, C.H., 2008. Neuroanatomy of verbal working memory as a diagnostic biomarker for depression. Neuroreport 19, 1507–1511.

Martire, L.M., Schulz, R., Reynolds, C.F., Morse, J.Q., Butters, M.A., Hinrichsen, G.A., 2008. Impact of close family members on older adults' early response to depression treatment. Psychol. Aging 23, 447–452.

Mazaika, P., Whitfield-Gabrieli, S., Reiss, A., Glover, G., 2007. Artifact repair for fMRI data from high motion clinical subjects. Hum. Brain Mapp. 2007.

Morse, J.Q., Pilkonis, P.A., Houck, P.R., Frank, E., Reynolds III, C.F., 2005. Impact of cluster C personality disorders on outcomes of acute and maintenance treatment in late-life depression. Am. J. Geriatr. Psychiatry 13, 808–814.

Mulsant, B.H., Houck, P.R., Gildengers, A.G., Andreescu, C., Dew, M.A., Pollock, B.G., Miller, M.D., Stack, J.A., Mazumdar, S., Reynolds III, C.F., 2006. What is the optimal duration of a short-term antidepressant trial when treating geriatric depression? J. Clin. Psychopharmacol. 26, 113–120.

Nouretdinov, I., Costafreda, S.G., Gammerman, A., Chervonenkis, A., Vovk, V., Vapnik, V., Fu, C.H., 2011. Machine learning classification with confidence: application of transductive conformal predictors to MRI-based diagnostic and prognostic markers in depression. NeuroImage 56, 809–813.

Oslin, D.W., 2005. Treatment of late-life depression complicated by alcohol dependence. Am. J. Geriatr. Psychiatry 13, 491–500.

Patel, A.X., Kundu, P., Rubinov, M., Jones, P.S., Vertes, P.E., Ersche, K.D., Suckling, J., Bullmore, E.T., 2014. A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. NeuroImage 95, 287–304.

Patel, M.J., Andreescu, C., Price, J.C., Edelman, K.L., Reynolds III, C.F., Aizenstein, H.J., 2015. Machine learning approaches for integrating clinical and imaging features in late-life depression classification and response prediction. Int. J. Geriatr. Psychiatry 30, 1056–1067.

Patel, M.J., Khalaf, A., Aizenstein, H.J., 2016. Studying depression using imaging and machine learning methods. Neuroimage Clin. 10, 115–123.

Pollock, B.G., Ferrell, R.E., Mulsant, B.H., Mazumdar, S., Miller, M., Sweet, R.A., Davis, S., Kirshner, M.A., Houck, P.R., Stack, J.A., Reynolds, C.F., Kupfer, D.J., 2000. Allelic variation in the serotonin transporter promoter affects onset of paroxetine treatment response in late-life depression. Neuropsychopharmacology 23, 587–590.

Reynolds III, C.F., Hoch, C.C., Buysse, D.J., George, C.J., Houck, P.R., Mazumdar, S., Miller, M., Pollock, B.G., Rifai, H., Frank, E., et al., 1991. Sleep in late-life recurrent depression. Changes during early continuation therapy with nortriptyline. Neuropsychopharmacology 5, 85–96.

Siegle, G.J., Carter, C.S., Thase, M.E., 2006. Use of FMRI to predict recovery from unipolar depression with cognitive behavior therapy. Am. J. Psychiatry 163, 735–738.

Smith, G.S., Reynolds III, C.F., Pollock, B., Derbyshire, S., Nofzinger, E., Dew, M.A., Houck, P.R., Milko, D., Meltzer, C.C., Kupfer, D.J., 1999. Cerebral glucose metabolic response to combined total sleep deprivation and antidepressant treatment in geriatric depression. Am. J. Psychiatry 156, 683–689.

Sneed, J.R., Culang-Reinlieb, M.E., Brickman, A.M., Gunning-Dixon, F.M., Johnert, L., Garcon, E., Roose, S.P., 2011. MRI signal hyperintensities and failure to remit following antidepressant treatment. J. Affect. Disord. 135, 315–320.

Szanto, K., Mulsant, B.H., Houck, P., Dew, M.A., Reynolds III, C.F., 2003. Occurrence and course of suicidality during short-term treatment of late-life depression. Arch. Gen. Psychiatry 60, 610–617.

Tew Jr., J.D., Mulsant, B.H., Houck, P.R., Lenze, E.J., Whyte, E.M., Miller, M.D., Stack, J.A., Bensasi, S., Reynolds III, C.F., 2006. Impact of prior treatment exposure on response to antidepressant treatment in late life. Am. J. Geriatr. Psychiatry 14, 957–965.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage 15, 273–289.

Whitfield-Gabrieli, S., Nieto-Castanon, A., 2012. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. Brain Connect. 2, 125–141.

Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. Chemom. Intell. Lab. Syst. 2, 37–52.

Wu, M., Andreescu, C., Butters, M.A., Tamburo, R., Reynolds III, C.F., Aizenstein, H., 2011. Default-mode network connectivity and white matter burden in late-life depression. Psychiatry Res. 194, 39–46.