

# SMRI Biomarkers Predict Electroconvulsive Treatment Outcomes: Accuracy with Independent Data Sets

Rongtao Jiang<sup>1,2,13</sup>, Christopher C Abbott<sup>3,13</sup>, Tianzi Jiang<sup>1,2,4</sup>, Yuhui Du<sup>5,6</sup>, Randall Espinoza<sup>7</sup>, Katherine L Narr<sup>7,8</sup>, Benjamin Wade<sup>8</sup>, Qingbao Yu<sup>5</sup>, Ming Song<sup>1</sup>, Dongdong Lin<sup>5</sup>, Jiayu Chen<sup>5</sup>, Thomas Jones<sup>3</sup>, Miklos Argyelan<sup>9,10,11</sup>, Georgios Petrides<sup>9,10,11</sup>, Jing Sui<sup>\*,1,2,4,5</sup> and Vince D Calhoun<sup>3,5,12</sup>

<sup>1</sup>Brainnetome Center and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China; <sup>2</sup>University of Chinese Academy of Sciences, Beijing, China; <sup>3</sup>Department of Psychiatry, University of New Mexico, Albuquerque, NM, USA; <sup>4</sup>Chinese Academy of Sciences Center for Excellence in Brain Science, Institute of Automation, Beijing, China; <sup>5</sup>The Mind Research Network and Lovelace Biomedical and Environmental Research Institute, Albuquerque, NM, USA; <sup>6</sup>School of Computer and Information Technology, Shanxi University, Taiyuan, China; <sup>7</sup>Department of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles, Los Angeles, CA, USA; <sup>8</sup>Ahmanson-Lovelace Brain Mapping Center, Department of Neurology, University of California at Los Angeles, Los Angeles, CA, USA; <sup>9</sup>Center for Psychiatric Neuroscience, The Feinstein Institute for Medical Research, Manhasset, NY, USA; <sup>10</sup>Division of Psychiatry Research, Zucker Hillside Hospital, Northwell System, Glen Oaks, NY, USA; <sup>11</sup>Departments of Psychiatry and Molecular Medicine, Hofstra Northwell School of Medicine, Hempstead, NY, USA; <sup>12</sup>Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM, USA

Owing to the rapid and robust clinical effects, electroconvulsive therapy (ECT) represents an optimal model to develop and test treatment predictors for major depressive disorders (MDDs), whereas imaging markers can be informative in identifying MDD patients who will respond to a specific antidepressant treatment or not. Here we aim to predict post-ECT depressive rating changes and remission status using pre-ECT gray matter (GM) in 38 MDD patients and validate in two independent data sets. Six GM regions including the right hippocampus/parahippocampus, right orbitofrontal gyrus, right inferior temporal gyrus (ITG), left postcentral gyrus/precuneus, left supplementary motor area, and left lingual gyrus were identified as predictors of ECT response, achieving accuracy of 89, 90 and 86% for remission prediction in three independent, age-matched data sets, respectively. For MDD patients, GM density increases only in the left supplementary motor cortex and left postcentral gyrus/precuneus after ECT. These results suggest that treatment-predictive and treatment-responsive regions may be anatomically different but functionally related in the context of ECT response. To the best of our knowledge, this is the first attempt to quantitatively identify and validate the ECT treatment biomarkers using multi-site GM data. We address a major clinical challenge and provide potential opportunities for more effective and timely interventions for electroconvulsive treatment.

*Neuropsychopharmacology* (2018) **43**, 1078–1087; doi:10.1038/npp.2017.165; published online 23 August 2017

## INTRODUCTION

Existing as the world's primary cause of disability, major depression is highly prevalent, affecting > 350 million lives and contributing to ~1 million suicides each year (WHO). The economic costs of depression outrank major health conditions such as cancer and diabetes, and are estimated near \$200 billion in the United States alone (Mrazek *et al*, 2014). Although treatable, current interventions are only moderately successful. Two-thirds of patients require two or

more antidepressant drug trials and a third remain unresponsive to multiple medication trials (McGrath *et al*, 2006). Electroconvulsive therapy (ECT) has an efficacy about triple that of pharmacotherapy. ECT has several first-line indications for depressive episodes including acute suicidality, catatonia, and psychosis (Weiner *et al*, 2001). In contrast, treatment unresponsiveness to antidepressant pharmacotherapy, a second-line indication for ECT presents a more challenging clinical decision. Opting for another antidepressant drug trial risks extending the depressive episode in the context of non-response. However, due to cost, potential side effects and stigma, ECT is typically used only after numerous medication failures.

If the mental health clinician had an accurate prognostic test of response, more personalized treatment algorithms could improve clinical outcome. Such biomarkers could inform both clinician and the individual patient about the risks and benefits of an optimal treatment at a specific time in the depressive episode. Past research has focused on

\*Correspondence: Professor J Sui, National Laboratory of Pattern Recognition and Brainnetome Center, Institute of Automation, Chinese Academy of Sciences, 95 Zhongguancun East Road, Beijing 100190, China, Tel: +86 82544518, Fax: +86 82544777, E-mail: jing.sui@nlpr.ia.ac.cn

<sup>13</sup>Rongtao Jiang and Christopher C. Abbott contributed equally as the first authors.

Received 26 April 2017; revised 18 July 2017; accepted 21 July 2017; accepted article preview online 31 July 2017

clinical and demographic factors, but these biomarkers have not achieved the accuracy required for translation (Dombrovski *et al*, 2005; Haq *et al*, 2015). Some prior neuroimaging studies have attempted to determine whether treatment selection biomarkers (ie, assessing variations in baseline brain measures) associate with ECT outcomes (Joshi *et al*, 2016; Oudega *et al*, 2014; Ten Doerschate *et al*, 2014), but most could be described more as post correlation rather than prediction (Whelan and Garavan, 2014). Two key factors motivate the current study. First, machine learning and data mining techniques (eg, logistic regression and support vector machines) provide a means for classifying treatment outcome and identifying new prediction-related networks. Second, ECT provides an optimal laboratory to identify prediction and treatment response networks, due to the magnitude and speed of response. Several recent ECT-neuroimaging investigations have demonstrated the promise of such methods for isolating prognostic biomarkers of ECT response with accuracy between 78 and 89% (Redlich *et al*, 2016; van Waarde *et al*, 2015; Wade *et al*, 2016). However, these investigations either used binary classification or lacked validation with independent data sets. Identified brain areas, which may serve as potential biomarkers, vary significantly and sometimes remain contradictory because of the heterogeneity of depression or partially the differences in the clinical characteristics of samples (Bora *et al*, 2012). Therefore, if biomarkers could be used to guide clinical practice, they should be able to predict outcomes for new individuals and independent cohorts based on the identified features obtained from models developed previously with training data (Wachinger *et al*, 2016). Furthermore, as reported by American Psychiatric Association (2012), predictive biomarkers should have an accuracy greater than 80% to be clinically useful.

In this study, we have three aims: (1) predict clinical outcome of MDD after ECT (both depression rating changes and remission status) with a data-driven individualized prediction framework (Meng *et al*, 2017) by using pre-ECT gray matter (GM) data; (2) verify the effectiveness of a set of identified GM regions on ECT prediction by using independent samples; and (3) assess the longitudinal changes of these GM regions among MDD remitters, MDD non-remitters, and demographically matched healthy controls (HCs). Based on above analyses, we expect to discover a set of GM regions of interests (ROIs) that are able to better predict individual clinical outcomes after ECT treatment quantitatively for multi-site data, which is in accordance with the Precision Medicine Initiative (Shah *et al*, 2016) released by NIMH.

## MATERIALS AND METHODS

### Inclusion/Exclusion Criteria

Three data sets from existing ECT studies that followed similar protocols were used, including University of New Mexico (UNM; 38 MDD and 27 demographically matched HCs), University of California at Los Angeles (UCLA; 42 MDD), and the Northwell Health-Long Island Jewish Health System (LIJ; 17 MDD). All MDD patients were scanned within 72 h of the first ECT treatment and after the ECT series; HCs were scanned at two time points between 4- and

6-week intervals. The human research protections office at each site approved the investigation and data-sharing agreements. Inclusion criteria consisted of the following: (1) diagnosis of MDD (with or without psychotic features); (2) clinical indications for ECT; and (3) Hamilton Depression Rating Scale 24-item (HDRS-24) > 21 (Kellner *et al*, 2006); and (4) age range between 50 and 80 years for the UNM sample. Exclusion criteria consisted of the following: (1) defined neurological or neurodegenerative disorder (eg, history of head injury with loss of consciousness > 5 min, epilepsy, and Alzheimer's disease); (2) other psychiatric conditions (eg, schizophrenia, schizoaffective disorder, and bipolar disorder); (3) current drug or alcohol-use disorder, except for nicotine; and (4) contraindications to MRI. Additional exclusion criteria for the HC group included any current or past psychiatric history as confirmed by the Structured Clinical Interview for DSM-IV-TR Disorders, Research Version, Non-patient Edition (First *et al*, 2002).

### Clinical Assessments

The primary clinical assessment was the HDRS before and after the ECT sessions. UNM and LIJ used the HDRS-24 and UCLA used the HDRS-17. ECT remission was defined as > 50% reduction in HDRS and final HDRS-24 score  $\leq$  10 (UNM, LIJ) or HDRS-17  $\leq$  7 (UCLA) (Heijnen *et al*, 2010). In UNM site, all MDD patients continued with antidepressant and antipsychotic medications with minimal medication changes during the ECT series (dose-titration only). UCLA and LIJ both discontinued antidepressant medications before and during the ECT series, whereas some LIJ subjects received lorazepam as needed for anxiety during the ECT series.

### Demographics and Clinical Outcomes

For the UNM site, 38 subjects with MDD (age  $63.8 \pm 9.0$  years, 24 males) and 27 HCs (age  $60.0 \pm 8.3$  years, 11 males) completed the study (Table 1). MDD subjects had both psychotic ( $n=16$ ) and non-psychotic ( $n=22$ ) depressive episodes. Most UNM MDD subjects (36/38, 95%) were treated with antidepressant and concurrent antipsychotic (21/38, 55%) medications. Antidepressant medications included select serotonin reuptake inhibitors ( $n=15$ ), serotonin norepinephrine reuptake inhibitors ( $n=16$ ), norepinephrine dopamine reuptake inhibitors ( $n=1$ ), and tricyclic antidepressants ( $n=4$ ). MDD subjects received an average of  $11.1 \pm 3.2$  ECT treatments with both right unilateral ( $n=30$ ) and bitemporal ( $n=8$ ) electrode placements. UNM MDD subjects improved from a baseline (pre-ECT) HDRS-24 of  $32.8 \pm 7.5$  to a post-ECT HDRS-24 of  $8.6 \pm 9.6$  ( $T(37)=11.8$ ,  $p<0.01$ , 71.1% remitters). The majority of UNM subjects received longitudinal neuropsychological screening with the Repeatable Battery of Neuropsychological Status (Randolph *et al*, 1998), which demonstrated a modest improvement in total score (scaled score mean of  $100 \pm 15$ , pre-ECT:  $84.9 \pm 19.8$ , post-ECT:  $89.2 \pm 20.0$ ,  $T(29)=2.0$ ,  $p=0.06$ ). Among all five subscales, only the immediate memory demonstrated a significant improvement (pre-ECT:  $79.1 \pm 21.4$ , post-ECT:  $89.9 \pm 19.6$ ,  $T(29)=4.3$ ,  $p=1.8e-4$ ) in patients, see Supplementary Table S2.

The validation data sets include a wider age range from 20 to 74 years. UCLA included 42 MDD subjects (age  $42.8 \pm 14.1$

**Table 1** Clinical and Demographic Information

	UNM-MDD mean (SD)	UCLA-MDD mean (SD)	LIJ-MDD mean (SD)	UNM-HC mean (SD)
Sample size ( <i>n</i> )	38	10	7	27
Age: years	63.8 (9.0)	57.6 (7.8)	62.0 (7.5)	60.0 (8.3)
Gender: M/F	24/14	5/5	5/2	11/16
Total number of ECT treatments	11.1 (3.2)	10.7 (3.6)	7.3 (3.5)	
Pre-ECT HDRS <sup>a</sup>	32.8 (7.5)	25.5 (4.0)	30.7 (4.8)	
Post-ECT HDRS <sup>a</sup>	8.6 (9.6)	10.8 (6.8)	16.7 (11.8)	
ΔHDRS (pre–post)	24.1 (12.6)	14.7 (7.7)	14.0 (11.4)	
Responder (%)	30/38 (78.9)	2/10 (20)	4/7 (57.1)	
Remitter (%)	27/38 (71.1)	2/10 (20)	3/7 (42.9)	

Abbreviations: ECT, electroconvulsive therapy; HDRS, Hamilton Depression Rating Scale; LIJ, Long Island Jewish Health System; UCLA, University of California at Los Angeles; UNM, University of New Mexico.

<sup>a</sup>UNM and LIJ used HDRS-24, UCLA used HDRS-17.

years, 18 males) and LIJ included 17 MDD subjects (age  $47.2 \pm 15.1$  years, 10 males). For the age-matched subjects with UNM (ie, age > 50 years), 10 and 7 subjects were retained for UCLA and LIJ, respectively. UCLA subjects received an average of  $10.7 \pm 3.6$  ECT treatments with most subjects receiving right unilateral electrode placement and achieved an improvement from a pre-ECT HDRS-17 of  $25.5 \pm 4.0$  to a post-ECT HDRS-17 of  $10.8 \pm 6.8$  ( $p < 0.01$ , 20.0% remitters). The lower remitter rates in the UCLA sample are related to the more stringent remission criteria for the HDRS-17 ( $< 7$ ). LIJ included 7 MDD subjects (age  $62.0 \pm 7.5$  years, 5 males) that received fewer ECT treatments,  $7.3 \pm 3.5$ , with bi-frontal electrode placement. LIJ MDD subjects had a response from a pre-ECT HDRS-24 of  $30.7 \pm 4.8$  to a post-ECT  $16.7 \pm 11.8$  ( $p < 0.01$ , 42.9% remitters).

### ECT Procedure

The UNM ECT service used a Thymatron System IV (Somatics, Lake Bluff, IL, USA) and started with a right unilateral d'Elia (ultra-brief pulse width (0.3 ms), stimulus dosage at  $6 \times$  threshold) electrode placement unless bitemporal (brief pulse width (1 ms), stimulus dosage at  $2 \times$  threshold) electrode placement was clinically indicated. UCLA used a Mecta 5000Q (MECTA Corp., Tualatin, OR, USA) with similar electrode placement and stimulus dosages. LIJ used Thymatron System IV with bifrontal electrode placement (brief pulse width (1 ms), stimulus dosage at  $1.5 \times$  threshold). Further adjustments to energy occurred as needed for inadequate seizure duration, defined as  $< 25$  s of electroencephalogram seizure activity. Treatments occurred three times weekly until adequate clinical response or clinical decision to stop treatment for non-responders. ECT management was not manipulated for purposes of this investigation and followed accepted clinical standards as promulgated by the APA ECT Task Force Report. Patients were oxygenated throughout the procedure with a disposable bag and mask. Blood pressure, pulse and oxygen saturation were monitored throughout the procedure.

### Data Acquisition and Preprocessing

Structural magnetic resonance imaging data were collected at UNM on a 3 Tesla Siemens Trio scanner (repetition time = 2.53 s (s), echo time = 1.64, 3.50, 5.32, 7.22, 9.08 ms, inversion time = 1.20 s, flip angle = 7°, number of excitations = 1, and voxel size =  $1 \times 1 \times 1$  mm<sup>3</sup>), LIJ on a 3T GE HDx scanner (Argyelan *et al*, 2016), and UCLA on a 3-T Allegra scanner (Joshi *et al*, 2016). Using the unified segmentation methods of SPM8, the sMRI data were normalized to Montreal Neurologic Institute (MNI) space, resliced to  $3 \times 3 \times 3$  mm<sup>3</sup> and segmented into GM, white matter and cerebral spinal fluid. The GM images were smoothed with a full width at half maximum  $8 \times 8 \times 8$  mm<sup>3</sup> Gaussian filter. Subject outlier detection was performed with a spatial Pearson's correlation with the MNI template image, to ensure that all subjects were properly segmented (Gupta *et al*, 2015) and no outliers were found.

### Individualized Prediction

We previously developed a generalized, data-driven prediction framework, which has successfully predicted cognitive scores for individuals with a high accuracy, by whole-brain voxel-wise searching on MRI data (Meng *et al*, 2017). Here we adopted this algorithm to select baseline (pre-ECT) structural imaging features to predict eventual change in depression rating scores ( $\Delta$ HDRS = pre-ECT HDRS – post-ECT HDRS) using UNM data set. Rigorous nested cross-validation (10-fold and leave-one-out) was performed, to ensure the validity of the detected GM features, as shown in Supplementary Figure S1.

In the leave-one-out loop, one subject was left out as the test subject, the remaining 37 subjects were used as training set, which was repeated 38 times (equal to the subject length) to test through all subjects. For the training subjects, we first utilized a whole-brain, voxel-wise feature selection method called ReliefF (Stokes and Visweswaran, 2012), to identify the most relevant voxel features to the predicted measure ( $\Delta$ HDRS); see more method details in Supplementary File S3. Then spatial clustering was performed on the above selected voxels, to obtain a reduced number of GM ROI clusters, which were further averaged to result in ROI-based features. Compared with a fixed brain atlas, this data-driven

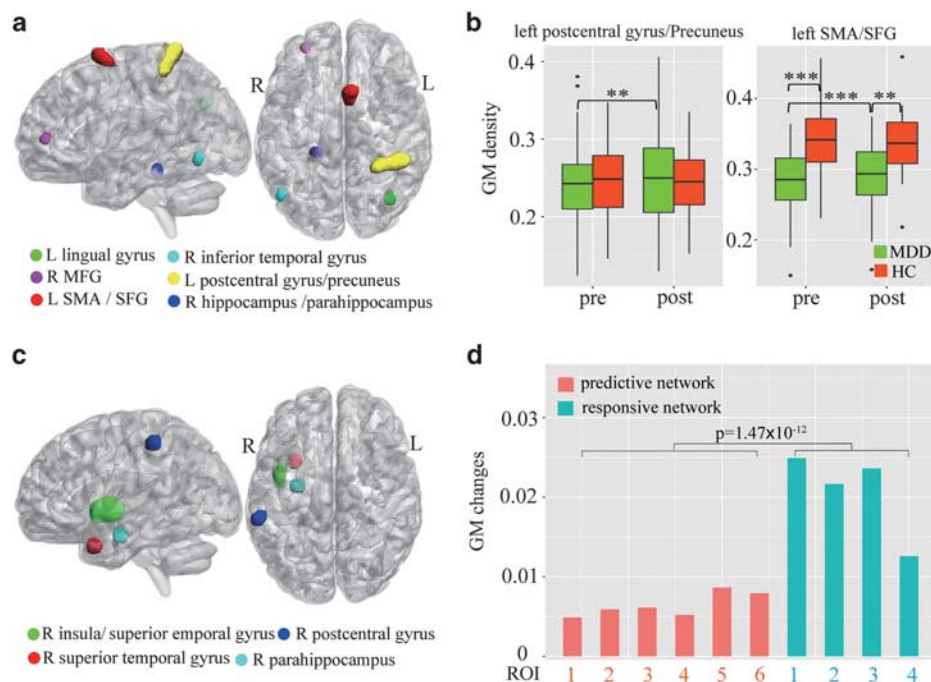


clustering is more flexible and may increase the accuracy of the prediction network as reported in Meng *et al* (2017). Third, the cluster-based features were further refined by correlation-based feature selection (Tripoliti *et al*, 2010), producing an optimal feature subset that is highly correlated with the predicted measure ( $\Delta$ HDRS) but has low redundancy within itself. This step is necessary, as the derived ROI-based features (GM regions) may indicate decreased relevance to the target measure compared with initial voxel-wise features and may contain redundant information. Finally, a set of GM ROIs were combined in a 10-fold linear regression analysis to predict the continuous value of  $\Delta$ HDRS. Here, both spatial clustering and correlation-based feature selection can be viewed as a feature reduction strategy, with no parameters needed to be tuned in our framework. At each loop, we calculated the predicted  $\Delta$ HDRS for each left-out subject and obtained the identified GM ROIs, which can be common or different loop by loop. After performing nested cross-validation, the unbiased prediction of  $\Delta$ HDRS for all the UNM subjects was achieved. By selecting the GM regions that were most frequently identified in all 38 loops (occurrence more than half of the loop times,  $38/2 = 19$ ), we finally determined six GM ROIs as a predictive network of the clinical target ( $\Delta$ HDRS), as shown in Figure 1a. Based on the predicted  $\Delta$ HDRS and the baseline HDRS, we can calculate the post-ECT HDRS for each patient, who can be further classified as remitter or

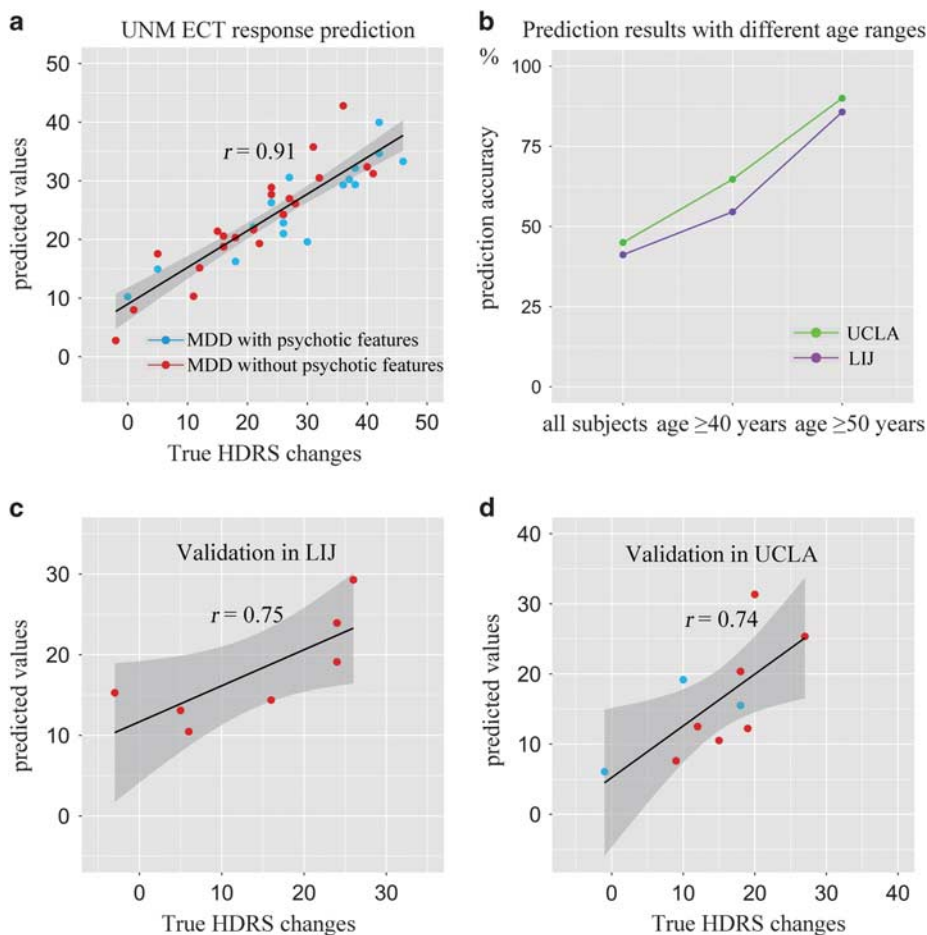
non-remitter according to the clinical assessment (Heijnen *et al*, 2010). Finally, compared with true treatment outcome, the sensitivity, specificity, positive predictive value, negative predictive value, and the prediction accuracy for remitter were calculated; please see more details in the Supplementary Methods Section.

### Independent Cohort Validation of Identified GM Regions

We further used two independent datasets (LIJ and UCLA), to evaluate the validity of the identified GM ROIs for predicting ECT treatment outcome. Namely, whether individual symptom changes can be predicted solely based on these identified ROIs in new data sets. Same ROI features (as shown in Figure 1a) were extracted from UCLA and LIJ data sets. Then we ran linear regression using a leave-one-out cross-validation for each site, in which the six GM ROIs were used as regressors. The  $\Delta$ HDRS for these two sites were then estimated with site-specific  $\beta$ -weights and the dichotomous remitter criterion was further adopted to predict the post-ECT treatment outcome. Owing to the heterogeneity of age between UNM (50 ~ 80 years old) and UCLA/LIJ (20 ~ 74 years old), we separate the validation data set into different age groups and applied the leave-one-out linear regression to (1) entire validation subjects, (2) subjects with an age > 40 years, and (3) subjects with an age > 50 years, respectively. In



**Figure 1** (a) The six identified pre-electroconvulsive therapy (ECT) gray matter (GM) regions of interest (ROIs) as predictors of  $\Delta$ HDRS in sagittal and axial view. The six extracted GM ROIs contributing to the prediction of ECT treatment outcome, include right hippocampus/ parahippocampus (Brodmann area (BA) 30), right inferior temporal gyrus (ITG; BA 37), right middle frontal gyrus (MFG; BA 10, 11), left lingual gyrus (BA 19, 39), left postcentral gyrus/precuneus (BA 1,2,3,7), and left supplementary motor area (SMA)/superior frontal gyrus (SFG) (BA 6). (b) Repeated-measures analysis of variance (RM-ANOVA) identified GM changes before and after ECT in only two predictive regions: the left postcentral gyrus/precuneus and left SMA/SFG. *Post-hoc t*-tests on these regions demonstrated increased GM density for both regions with ECT. Relative to HCs, the left SMA/SFG had less GM density in major depressive disorder (MDD) both at pre-ECT (pre-ECT/healthy control (HC)  $T(36) = 4.45$ ,  $p < 0.001$ ) and post-ECT (post-ECT/HC  $T(36) = 3.35$ ,  $p < 0.001$ ). (c) Four treatment-responsive regions identified by voxel-wise RM-ANOVA with group  $\times$  time interaction, including right insula, right superior temporal gyrus, parahippocampal gyrus, and postcentral gyrus. (d) Longitudinal GM changes for six predictive ROIs and four treatment-responsive regions. The magnitude of longitudinal GM change is much greater in treatment-responsive network than in treatment-predictive regions ( $p < 0.001$ ). See more details in Supplementary Table S1.



**Figure 2** Scatter plot of the predicted  $\Delta$ HDRS (pre–post) with respect to their true values for three sites. (a) A Spearman's correlation of  $r = 0.91$  was achieved between the estimated  $\Delta$ HDRS and its true values in University of New Mexico (UNM). For sub-group depressed patients with (blue dots) or without (red dots) psychotic features, the Spearman's correlations are 0.89 and 0.93, respectively (not listed in figure), indicating no significant differences between two major depressive disorder (MDD) subgroups. Based on the estimated  $\Delta$ HDRS and the pre- electroconvulsive therapy (ECT) Hamilton Depression Rating Scale (HDRS), patients were further classified as remitter or non-remitter. By extracting the same six gray matter (GM) regions as identified in UNM and using them as regressors for two independent cohorts: Long Island Jewish Health System (LIJ) and University of California at Los Angeles (UCLA), we performed leave-one-out linear regression analyses. (b) Prediction accuracies in validation sites with different age ranges. In the two validation sites, the derived classification accuracy for remission status range from 40% to ~65% for the entire validation sample and for subjects with age  $> 40$  years. However, when only include the age-matched subjects with UNM (ie, age  $> 50$  years), we achieved Spearman correlation of  $r = 0.75$  (c) and  $r = 0.74$  (d) for LIJ and UCLA, respectively. Furthermore, high prediction precision for remission status were also achieved (UCLA: sensitivity 100%, specificity 87.5%, and accuracy 90.0%; LIJ: sensitivity 66.7%, specificity 100%, and accuracy 95.7%).

addition, to demonstrate the validity of the selected 6 predictive GM ROIs, we also randomly selected 6 GM ROIs from 116 atlas-based ROIs derived from AAL (Automated Anatomical Labeling) atlas segmentation and performed the same linear regression with leave-one-out cross-validation. The process was repeated for 100 times in each of the two validation data sets and the results were compared with what we obtained using the 6 identified GM ROIs.

### Longitudinal Analyses and Group Comparison

For the UNM data set, repeated-measures analysis of variance (RM-ANOVA) assessed group (MDD and HC)  $\times$  time interaction for GM density extracted from each of the six GM regions. *Post-hoc* two-sample (pre-ECT/HC, post-ECT/HC), paired (pre-/post-ECT) *t*-tests, and correlations between  $\Delta$ HDRS and  $\Delta$ GM were performed on the

identified predictive network to determine the directionality of change for significant interactions. Paired two-sample *t*-tests were also performed in each of the six GM regions (pre- and post-ECT) for MDD remitters and non-remitters to detect the influence of ECT on GM densities. In addition, group comparison among remitters, non-remitters, and HCs at baseline and at the second timepoint were conducted with two-sample *t*-tests. Finally, RM-ANOVA was used to assess group (MDD and HC)  $\times$  time interaction for the whole brain voxel-wise GM density with FDR correction for multiple comparisons.

## RESULTS

### Identified Predictive GM Regions

Six GM ROIs were identified as a set of predictors for ECT treatment response in UNM data (Figure 1(a)): (1) right

**Table 2** Prediction Accuracy of  $\Delta$ HDRS and ECT Remitters for Three Sites

	UNM		UCLA > 50 years		LIJ > 50 years	
No. of subjects	38		10		7	
Correlation $r/p$	0.91/2e-15		0.75/0.01		0.74/0.05	
RMSE	6.0		5.8		7.0	
	Predicted remitter	Predicted non-remitter	Predicted remitter	Predicted non-remitter	Predicted remitter	Predicted non-remitter
True remitter	24	3	2	0	2	1
True non-remitter	1	10	1	7	0	4
Sensitivity <sup>a</sup>	88.9%		100%		66.7%	
Specificity <sup>a</sup>	90.9%		87.5%		100%	
PPV <sup>a</sup>	96.0%		66.7%		100%	
NPV <sup>a</sup>	76.9%		100%		80.0%	
Accuracy <sup>a</sup>	89.5%		90.0%		85.7%	

Abbreviations: ECT, electroconvulsive therapy; HDRS, Hamilton Depression Rating Scale; LIJ, Long Island Jewish Health System; MDD, major depressive disorder; NPV, negative predictive value; PPV, positive predictive value; RMSE, root mean squared prediction error; UCLA, University of California at Los Angeles; UNM, University of New Mexico.

<sup>a</sup>Based on the predicted  $\Delta$ HDRS and the pre-ECT HDRS, the post-ECT HDRS could be estimated and each MDD patient can further be classified as remitter or non-remitter according to (Heijnen *et al*, 2010). Here, sensitivity is defined as number of correctly predicted remitters divided by number of true remitters; specificity is defined as number of correctly predicted non-remitters divided by number of true non-remitters; PPV is defined as number of correctly predicted remitters divided by number of predicted remitters; NPV is defined as number of correctly predicted non-remitters divided by number of predicted remitters; accuracy is defined as number of correctly predicted remitters and non-remitters divided by number of all MDD.

middle frontal gyrus (MFG; Brodmann area (BA) 10,11); (2) right hippocampus and parahippocampus (BA 30); (3) right ITG (BA 37); (4) left postcentral gyrus/precuneus (BA 1, 2, 3, and 7); (5) left supplementary motor area (SMA) and superior frontal gyrus (SFG) (BA 6); and (6) left lingual gyrus (BA 39, 19).

### Prediction Accuracy (UNM Cohort)

The predicted  $\Delta$ HDRS from the UNM pre-ECT sMRI data had a Spearman's correlation of  $r=0.91$  with true  $\Delta$ HDRS (Figure 2a), root mean squared prediction error (RMSE) = 6.0. MDD subjects either with psychotic features (blue dots,  $n=16$ ) or without psychotic features (red dots,  $n=22$ ) had similar results ( $r=0.89$  and  $0.93$ , respectively). In addition, based on the predicted  $\Delta$ HDRS, binary classification of MDD patients as remitters/non-remitters achieved a high degree of precision (sensitivity 88.9%, specificity 90.9%, and accuracy 89.5%) (Table 2). Prediction of remission status from these six ROIs with linear regression via leave-one-out cross-validation also resulted in high precision (sensitivity 81.5%, specificity 90.9%, and 84.2% accuracy).

### Independent Cohort Validation Using Six Identified GM Regions

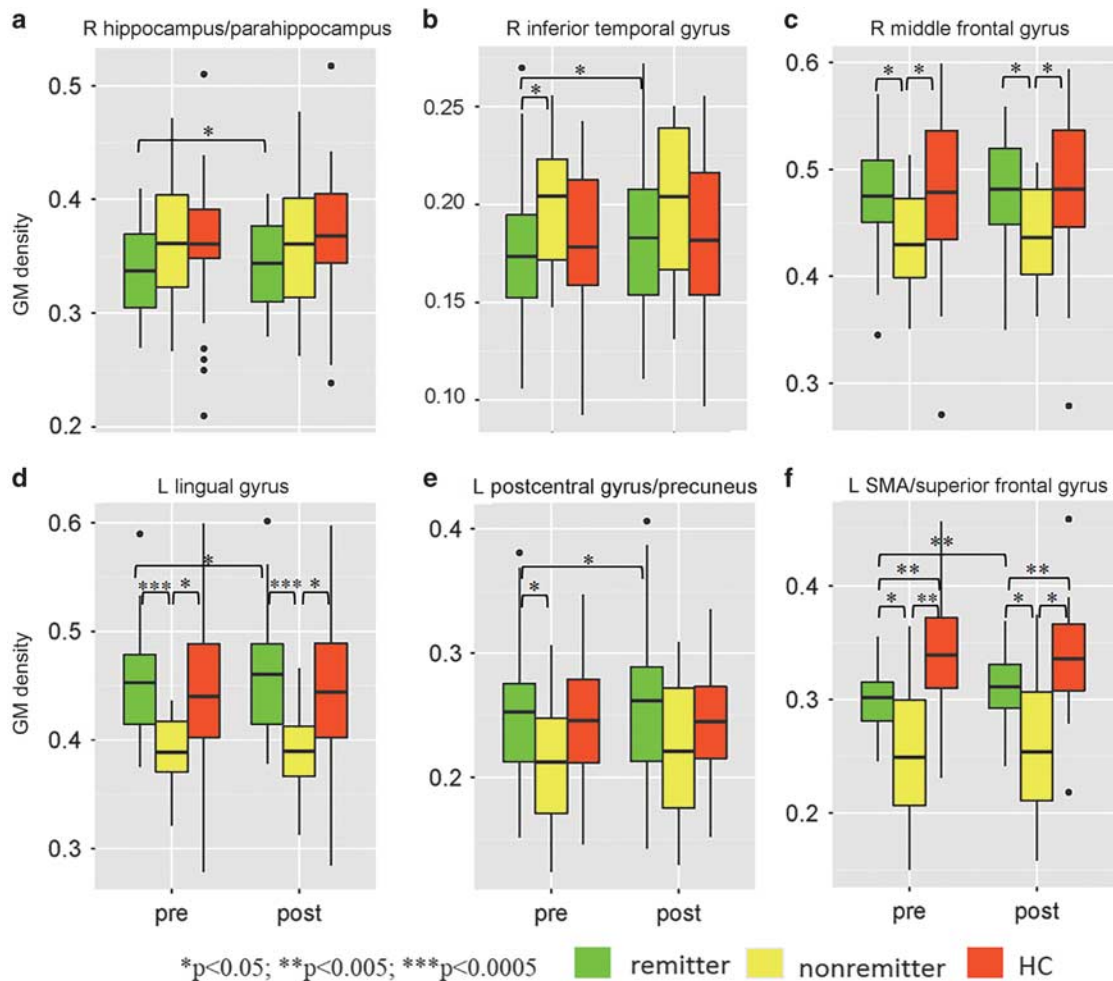
In the two validation sites, as shown in Figure 2b, the derived classification accuracy for remission status ranged from 40% to ~65% for the entire validation sample and for subjects with age > 40 years. However, when including only age-matched subjects with UNM (ie, age > 50 years), more accurate results were obtained. For UCLA, 10 subjects were retained and the six GM ROIs predicted  $\Delta$ HDRS at Spearman's correlation of  $r=0.75$ , RMSE=5.8, with a 90.0% classification accuracy for remitters vs non-remitters (sensitivity 100% and specificity 87.5%) and thus all remitters

were correctly predicted. For LIJ > 50 years ( $n=7$ ),  $\Delta$ HDRS prediction reached a Spearman's correlation of  $r=0.74$  between true and predicted values, RMSE = 7.0, achieving a remission classification at an accuracy of 85.7% (sensitivity 66.7% and specificity 100%), see Figure 2c and d, and Table 2.

In contrast, predictions from six randomly selected AAL-based ROIs achieved a mean correlation of  $r=0.29 \pm 0.21$  and  $0.31 \pm 0.18$  for LIJ and UCLA data set, respectively, for all three age ranges, significantly lower than results derived via the six identified GM ROIs ( $p < 0.0001$ ). Consequently, for the randomly selected ROIs, the remission classification accuracy remained low in every permutation (accuracy < 70% in all cases), further confirming the effectiveness of the identified six GM ROIs for predicting treatment outcome in a specific age range.

### Longitudinal Analysis and Group Comparison

To examine longitudinal changes, RM-ANOVA was applied to each of six predictive regions and we identified two regions with group (MDD/HC)-by-time interactions for the UNM data, ie, the left postcentral gyrus/precuneus ROI ( $F(3, 64)=7.8$ ,  $p=0.007$ ) and the left SMA/SFG ROI ( $F(3, 64)=13.4$ ,  $p < 0.001$ , see Figure 1(b), Supplementary Table S1. *Post-hoc t*-tests on these regions demonstrated increased GM density for both left postcentral gyrus/precuneus ( $T(37)=3.24$ ,  $p=0.003$ ) as well as left SMA/SFG ( $T(37)=3.68$ ,  $p < 0.001$ ). Relative to HCs, the left SMA/SFG had less GM density in MDD both for pre-ECT ( $T(36)=4.45$ ,  $p < 0.001$ ) and post-ECT ( $T(36)=3.35$ ,  $p < 0.001$ ). However, none of the six GM ROIs showed significant correlations between changes of GM ( $\Delta$ GM) and depression ratings ( $\Delta$ HDRS). Figure 1c indicated the voxel-wise RM-ANOVA results of group  $\times$  time interaction, four regions including right insula, right superior temporal gyrus, parahippocampal gyrus, and



**Figure 3** Longitudinal gray matter (GM) changes among remitters, non-remitters, and healthy controls on the six identified GM regions. Subplot a-f denotes each of the six regions, see title of the subplots. Longitudinal GM density increases were observed for remitters in all five of the six GM regions (the right orbitofrontal gyrus did not demonstrate longitudinal change). None of these regions showed significant differences for non-remitters. Five of the six GM regions had baseline (pre- electroconvulsive therapy (ECT)) GM density differences between remitters and non-remitters ( $p < 0.05$ ) (the right hippocampus/ parahippocampus did not reveal pre-ECT remitter/non-remitter differences), suggesting structural heterogeneity exists between remitters and non-remitters.

postcentral gyrus were identified as treatment-responsive network. Figure 1d demonstrated the longitudinal GM changes for six predictive ROIs and four treatment-responsive regions. Remarkably, the magnitude of longitudinal GM change is much greater in treatment-responsive network than in treatment-predictive regions.

Figure 3 demonstrated GM changes before and after ECT among three groups: MDD remitters, MDD non-remitters, and HCs. Interestingly, longitudinal GM density increases were observed for remitters in all five of the six GM regions (right MFG did not demonstrate longitudinal change), whereas none of these regions showed significant increases or reduction for non-remitters. Five of the six GM regions had baseline (pre-ECT) GM density differences between remitters and non-remitters ( $p < 0.05$ ) (right hippocampus/ parahippocampus region did not demonstrate this difference) with remitters and HC group demonstrating higher (Figure 3f) or lower (only for ITG, Figure 3b) GM density than non-remitters at baseline.

## DISCUSSION

The goal of developing biomarkers for prognosis is to perform individualized predictions of health outcome (Gabrieli *et al*, 2015). In this work, we adopted pre-ECT segmented GM to predict the treatment outcome based on a generalized prediction framework, which has been used to predict individual cognitive scores successfully. Six GM ROIs were identified as predictors of eventual remission status after ECT, achieving accuracy of 89, 90 and 86% for three independent, age-matched data sets, respectively, significantly higher than using randomly selected six atlas-based ROIs. Although prior investigations in mental disorders typically use binary (categorical) classification, we achieved quantitative estimation of post-ECT changes in depression ratings for each individual patient, which are highly correlated with true  $\Delta$ HDRS ( $r > 0.74$  in three sites, RMSE  $< 7.0$ ), and are robust to the presence of psychosis in MDD. More importantly, effectiveness of the identified six GM predictors was verified on two independent datasets, in which relatively higher sensitivity and specificity was



achieved on prediction of ECT remission status, especially when compared with results from randomly selected six AAL-based ROIs. Results suggest that individual depressive symptom changes can be predicted solely based on the six regions identified from the unbiased prediction framework, and may serve as ECT treatment predictors specifically for MDD patients who are older than 50 years.

A unique aspect of our work is that instead of focusing on priori presumed regions, we employed whole-brain voxel-wise data mining, followed by spatial clustering and feature refining. Compared with an atlas-based feature selection and alternative LASSO (least absolute shrinkage and selection operator) method (Bunea *et al*, 2011), our technique enables more flexibility and higher precision as reported in Meng *et al* (2017) (see Supplementary Method Section), acquiring a set of treatment-predictive GM regions which cannot be identified by simple longitudinal analysis. In addition, previous studies deriving predictive brain regions at one site may not work well for other data sets (Wachinger *et al*, 2016). This may be related to over-fitting of the training data at one site, especially when the training samples are limited, while features are high dimensional. In the current investigation, we used the source data (UNM) to identify a predictive network (six ROIs) of depression rating changes, which were further adopted as regressors in two independent test sets (UCLA and LIJ) with a leave-one-out cross-validation and achieved a significantly higher prediction accuracy than using randomly selected six atlas-based ROIs. This enables more flexibility to derive site-specific weights on the fixed GM predictors. We can also incorporate site-specific weights for each regression lines, representing a type of transfer learning, similar to a recent transfer learning method called domain adaptation (Wachinger *et al*, 2016), which may serve as a future research direction. Finally, other types of neuroimaging (functional MRI and diffusion weighted imaging) and genomics can be employed for prediction either separately or in the context of multimodal fusion (Sui *et al*, 2014), which may detect potentially important variations or relationships that may only be partially revealed by single modality (Calhoun and Sui, 2016). A long-term goal of this line of research is to develop tools that can facilitate clinical judgement.

Previous machine-learning ECT studies have investigated imaging predictors of response (Redlich *et al*, 2016; van Waarde *et al*, 2015; Wade *et al*, 2016). These investigations have used resting state fMRI or structural measures from one single data set. Similar to the current investigation, Redlich *et al* (2016) applied classification using support vector machines to pre-ECT sMRI data. The results demonstrated that the subcallosal cingulate gyrus was a key region for ECT response prediction with 84% accuracy, although other sMRI regions also contributed. We identified six GM regions in total, referred to as a 'predictive network,' which included both cortical and subcortical regions that may predict eventual clinical response with higher accuracy. Our prediction network did not include the subcallosal cingulate gyrus, perhaps due to different age ranges (the current investigation included an older age range). However, our identified GM ROIs did overlap with a previous report (Costafreda *et al*, 2009), in which increased GM density in MFG of MDD was predictive of eventual response to pharmacotherapy (fluoxetine) (Costafreda *et al*, 2009).

In addition, the UNM predictive network failed to maintain the response accuracy outside of the demographically matched external validation samples from UCLA and LIJ. Age-related differences in response rates have been a focus of clinical ECT prediction studies with older age (>50 years) often associated with more favorable ECT response rates (Nordenskjold *et al*, 2012). Although not completed in the present investigation, demographic, and clinical variables (duration of depressive episode, number of depressive episodes, age of depression onset, etc.) may be incorporated into predictive imaging algorithms to improve response accuracy. Alternatively, predictive networks may be identified for specific demographic categories (ie, predictive network for age range 20–40 years) or depression phenotypes. Consistent with the latter, Drysdale *et al* (2017) used hierarchical clustering analysis to identify depression biotypes from baseline neuroimaging data. Anxiety- and anhedonia-related depression biotypes improved prediction accuracy for transcranial magnetic stimulation for depressive episodes.

Analysis demonstrated longitudinal GM changes in two of the six identified ROIs: the left SMA/SFG and left post central gyrus/precuneus. The locations of neuroplasticity change are proximal to the site of maximal current density for subjects that received right unilateral electrode placement. Interestingly, the remaining four identified GM regions of the prediction network did not demonstrate any significant change in GM density. The lack of longitudinal change in predictive ROIs is consistent with a previous data-driven ECT investigation. Wade *et al* (2016) used baseline GM volume and surface-based shape metrics of the caudate, putamen, pallidum, and nucleus accumbens to predict overall MDD response to ECT with up to 89% accuracy, whereas only the left putamen volume showed significant increase after ECT.

In contrast to the modest longitudinal GM density changes in predictive biomarkers, whole brain longitudinal analysis revealed changes in right medial temporal lobe neuroplasticity. Medial temporal lobe neuroplasticity, particularly of the hippocampus, is now a widely replicated finding in the ECT neuroimaging literature (Wilkinson *et al*, 2017). The treatment-responsive region had little overlap with the six predictive ROIs and more robust (order of magnitude) increase in GM density relative to longitudinal changes with the predictive ROIs. Similarly, Redlich *et al* (2016) used longitudinal analysis to demonstrate similar changes in medial temporal lobe neuroplasticity, which were separate from their predictive biomarkers within the subcallosal cingulate gyrus. Collectively, these results suggest that predictive and treatment-responsive regions may be anatomically separate but functionally related in the context of ECT response. Clarification of the relationship between predictive and treatment-responsive biomarkers will inform novel (and more focal) electrode placement development to improve clinical outcomes (improved efficacy, reduced cognitive impairment). fMRI effective connectivity analysis may elucidate the relationship between predictive and treatment-responsive biomarkers and further inform the optimal biomarker(s) for targeted engagement.

Despite a high prediction accuracy and verification of the use of potential biomarkers, limitations of this study should be acknowledged when interpreting our results. First, UNM



subjects were treated with antidepressant medications during the baseline (pre-ECT) scan and throughout the ECT series. Both UCLA and LIJ tapered and discontinued antidepressant medications before the baseline assessment. Concurrent treatment with some antidepressants may work synergistically with ECT (Sackeim et al, 2009) and share a similar but less potent mechanism of action (Malberg et al, 2000). However, our identified predictive ROIs were robust to these clinical differences as well as other site-specific difference such as electrode placement, scanner inhomogeneity, number of treatments, and the inclusion or exclusion of psychotic subjects. Second, important clinical characteristics such as age of onset and duration of depressive episode were not assessed during this investigation and remain a focus of subsequent investigations. Another potential limitation is that the GM images were resliced to a voxel size of  $3 \times 3 \times 3$  and smoothed by an  $8 \times 8 \times 8$  kernel, which may tend to render the results as more conservative.

To the best of our knowledge, this is the first attempt to quantitatively identify and validate ECT treatment biomarkers based on multi-site pre-ECT sMRI data for individual MDD patients, which could potentially discover a set of new imaging features critical in treatment remission, and thus guide personalized clinical care. Future applications of machine learning may inform treatment-specific variables such as electrode placement or optimal current amplitude. Importantly, this same method may be applied to ECT relapse too, which lacks any clinical or demographic biomarker to guide clinical practice. Moreover, this investigation addressed a major clinical challenge and our identified predictive regions may have broad impact on translational medicine, by providing opportunities for more effective and timely interventions. In sum, the methods developed in this and other investigations on ECT prediction may be generalized to other depressed samples and different clinical populations, further informing the precision medicine initiative (Shah et al, 2016).

## FUNDING AND DISCLOSURE

The authors declare no conflicts of interest.

## ACKNOWLEDGMENTS

This work was supported in part by the National High Tech Program (863, number 2015AA020513) and China National Natural Science Foundation (number 81471367), the Strategic Priority Research Program of the Chinese Academy of Sciences (grant number XDB02060005), Natural Science Foundation of Shanxi Province (grant number 2016021077), and National Institute of Health (1R01EB005846, 1R01MH094524, and P20GM103472).

## REFERENCES

- American Psychiatric Association (2012). In: *Consensus Report of the APA Work Group on Neuroimaging Markers of Psychiatric Disorders*. Association AP (ed): Arlington, VA, USA.
- Argyelan M, Lencz T, Kaliora S, Sarpal DK, Weissman N, Kingsley PB et al (2016). Subgenual cingulate cortical activity predicts the efficacy of electroconvulsive therapy. *Transl Psychiatry* 6: e789.
- Bora E, Fornito A, Pantelis C, Yucel M (2012). Gray matter abnormalities in major depressive disorder: a meta-analysis of voxel based morphometry studies. *J Affect Disord* 138: 9–18.
- Bunea F, She Y, Ombao H, Gongvatana A, Devlin K, Cohen R (2011). Penalized least squares regression methods and applications to neuroimaging. *Neuroimage* 55: 1519–1527.
- Calhoun VD, Sui J (2016). Multimodal fusion of brain imaging data: a key to finding the missing link(s) in complex mental illness. *Biol Psychiatry Cogn Neurosci Neuroimaging* 1: 230–244.
- Costafreda SG, Chu C, Ashburner J, Fu CH (2009). Prognostic and diagnostic potential of the structural neuroanatomy of depression. *PLoS ONE* 4: e6353.
- Dombrowski AY, Mulsant BH, Haskett RF, Prudic J, Begley AE, Sackeim HA (2005). Predictors of remission after electroconvulsive therapy in unipolar major depression. *J Clin Psychiatry* 66: 1043–1049.
- Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y et al (2017). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat Med* 23: 28–38.
- First MB, Spitzer RL, Gibbon M, Williams JBW (2002). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition*. New York State Psychiatric Institute, Biomedical Research: New York.
- Gabrieli JD, Ghosh SS, Whitfield-Gabrieli S (2015). Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron* 85: 11–26.
- Gupta CN, Calhoun VD, Rachakonda S, Chen J, Patel V, Liu J et al (2015). Patterns of gray matter abnormalities in schizophrenia based on an international mega-analysis. *Schizophr Bull* 41: 1133–1142.
- Haq AU, Sitzmann AF, Goldman ML, Maixner DF, Mickey BJ (2015). Response of depression to electroconvulsive therapy: a meta-analysis of clinical predictors. *J Clin Psychiatry* 76: 1374–1384.
- Heijnen WT, Birkenhager TK, Wierdsma AI, van den Broek WW (2010). Antidepressant pharmacotherapy failure and response to subsequent electroconvulsive therapy: a meta-analysis. *J Clin Psychopharmacol* 30: 616–619.
- Joshi SH, Espinoza RT, Pirnia T, Shi J, Wang Y, Ayers B et al (2016). Structural plasticity of the hippocampus and amygdala induced by electroconvulsive therapy in major depression. *Biol Psychiatry* 79: 282–292.
- Kellner CH, Knapp RG, Petrides G, Rummans TA, Husain MM, Rasmussen K et al (2006). Continuation electroconvulsive therapy vs pharmacotherapy for relapse prevention in major depression: a multisite study from the Consortium for Research in Electroconvulsive Therapy (CORE). *Arch Gen Psychiatry* 63: 1337–1344.
- Malberg JE, Eisch AJ, Nestler EJ, Duman RS (2000). Chronic antidepressant treatment increases neurogenesis in adult rat hippocampus. *J Neurosci* 20: 9104–9110.
- McGrath PJ, Stewart JW, Fava M, Trivedi MH, Wisniewski SR, Nierenberg AA et al (2006). Tranylcypromine versus venlafaxine plus mirtazapine following three failed antidepressant medication trials for depression: a STAR\*D report. *Am J Psychiatry* 163: 1531–1541 quiz 1666.
- Meng X, Jiang R, Lin D, Bustillo J, Jones T, Chen J et al (2017). Predicting individualized clinical measures by a generalized prediction framework and multimodal fusion of MRI data. *Neuroimage* 145(Pt B): 218–229.
- Mrazek DA, Hornberger JC, Altar CA, Degtjar I (2014). A review of the clinical, economic, and societal burden of treatment-resistant depression: 1996–2013. *Psychiatr Serv* 65: 977–987.
- Nordenskjold A, von Knorring L, Engstrom I (2012). Predictors of the short-term responder rate of electroconvulsive therapy in depressive disorders—a population based study. *BMC Psychiatry* 12: 115.
- Oudega ML, van Exel E, Stek ML, Wattjes MP, van der Flier WM, Comijs HC et al (2014). The structure of the geriatric depressed

- brain and response to electroconvulsive therapy. *Psychiatry Res* **222**: 1–9.
- Randolph C, Tierney MC, Mohr E, Chase TN (1998). The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): preliminary clinical validity. *J Clin Exp Neuropsychol* **20**: 310–319.
- Redlich R, Opel N, Grotegerd D, Dohm K, Zaremba D, Burger C *et al* (2016). Prediction of individual response to electroconvulsive therapy via machine learning on structural magnetic resonance imaging data. *JAMA Psychiatry* **73**: 557–564.
- Sackeim HA, Dillingham EM, Prudic J, Cooper T, McCall WV, Rosenquist P *et al* (2009). Effect of concomitant pharmacotherapy on electroconvulsive therapy outcomes: short-term efficacy and adverse effects. *Arch Gen Psychiatry* **66**: 729–737.
- Shah SH, Arnett D, Houser SR, Ginsburg GS, MacRae C, Mital S *et al* (2016). Opportunities for the cardiovascular community in the precision medicine initiative. *Circulation* **133**: 226–231.
- Stokes ME, Visweswaran S (2012). Application of a spatially-weighted relief algorithm for ranking genetic predictors of disease. *BioData Mining* **5**: 20.
- Sui J, Huster R, Yu Q, Segall JM, Calhoun VD (2014). Function-structure associations of the brain: evidence from multimodal connectivity and covariance studies. *Neuroimage* **102**(Pt 1): 11–23.
- Ten Doesschate F, van Eijndhoven P, Tendolkar I, van Wingen GA, van Waarde JA (2014). Pre-treatment amygdala volume predicts electroconvulsive therapy response. *Front Psychiatry* **5**: 169.
- Tripoliti EE, Fotiadis DI, Argyropoulou M, Manis G (2010). A six stage approach for the diagnosis of the Alzheimer's disease based on fMRI data. *J Biomed Inform* **43**: 307–320.
- van Waarde JA, Scholte HS, van Oudheusden LJ, Verwey B, Denys D, van Wingen GA (2015). A functional MRI marker may predict the outcome of electroconvulsive therapy in severe and treatment-resistant depression. *Mol Psychiatry* **20**: 609–614.
- Wachinger C, Reuter M, Alzheimer's Disease Neuroimaging I, Australian Imaging B, Lifestyle Flagship Study of Ageing (2016). Domain adaptation for Alzheimer's disease diagnostics. *Neuroimage* **139**: 470–479.
- Wade BS, Joshi SH, Njau S, Leaver AM, Vasavada M, Woods RP *et al* (2016). Effect of electroconvulsive therapy on striatal morphometry in major depressive disorder. *Neuropsychopharmacology* **41**: 2481–2491.
- Weiner RD, Coffey CE, Fochtmann LJ, Greenberg RM, Isenberg KE, Kellner CH *et al* (2001). *The Practice of Electroconvulsive Therapy: Recommendations for Treatment, Training, and Privileging*, Second edn. American Psychiatric Association: Washington, DC.
- Whelan R, Garavan H (2014). When optimism hurts: inflated predictions in psychiatric neuroimaging. *Biol Psychiatry* **75**: 746–748.
- WHO World Health Organization, Depression (2017). [http://www.who.int/mental\\_health/management/depression/en/](http://www.who.int/mental_health/management/depression/en/).
- Wilkinson ST, Sanacora G, Bloch MH (2017). Hippocampal volume changes following electroconvulsive therapy: a systematic review and meta-analysis. *Biol Psychiatry Clin Neurosci Neuroimaging* **327**–335.

Supplementary Information accompanies the paper on the Neuropsychopharmacology website (<http://www.nature.com/npp>)