

1 **Title:** Decoding Depression Severity from Intracranial Neural Activity

2 **Author list:** Jiayang Xiao<sup>1,2</sup>, Nicole R. Provenza<sup>1</sup>, Joseph Asfour<sup>3</sup>, John Myers<sup>1</sup>, Raissa K. Mathura<sup>1</sup>,  
3 Brian Metzger<sup>1</sup>, Joshua A. Adkinson<sup>1</sup>, Anusha B. Allawala<sup>4</sup>, Victoria Pirtle<sup>1</sup>, Denise Oswald<sup>1</sup>, Ben  
4 Shofty<sup>1</sup>, Meghan E. Robinson<sup>1</sup>, Sanjay J. Mathew<sup>5</sup>, Wayne K. Goodman<sup>5</sup>, Nader Pouratian<sup>6</sup>, Paul R.  
5 Schrater<sup>7,8</sup>, Ankit B. Patel<sup>2,3,9</sup>, Andreas S. Tolias<sup>2,3,9</sup>, Kelly R. Bijanki<sup>1</sup>, Xaq Pitkow<sup>2,3,9</sup>, Sameer A.  
6 Sheth<sup>1\*</sup>

7 **Affiliations:**

8 <sup>1</sup>Department of Neurosurgery, Baylor College of Medicine, Houston, TX, United States.

9 <sup>2</sup>Department of Neuroscience, Baylor College of Medicine, Houston, TX, United States.

10 <sup>3</sup>Department of Electrical and Computer Engineering, Rice University, Houston, TX, United States.

11 <sup>4</sup>School of Engineering, Brown University, Providence, RI, United States.

12 <sup>5</sup>Department of Psychiatry, Baylor College of Medicine, Houston, TX, United States.

13 <sup>6</sup>Department of Neurological Surgery, UT Southwestern Medical Center, Dallas, TX, United States.

14 <sup>7</sup>Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, United  
15 States.

16 <sup>8</sup>Department of Psychology, University of Minnesota, Minneapolis, MN, United States.

17 <sup>9</sup>Center for Neuroscience and Artificial Intelligence, Baylor College of Medicine, Houston, TX, United  
18 States.

19 \*Corresponding author: Sameer A. Sheth; [sasheth@bcm.edu](mailto:sasheth@bcm.edu)

20 **Abstract**

21 Disorders of mood and cognition are prevalent, disabling, and notoriously difficult to treat. Fueling this  
22 challenge in treatment is a significant gap in our understanding of their neurophysiological basis. Here,  
23 we used intracranial neural recordings in three patients with severe depression to investigate the neural  
24 substrates of this disorder. Across prefrontal regions, we found that reduced depression severity is  
25 associated with decreased low-frequency neural activity and increased high-frequency activity. When  
26 constraining our model to decode using a single region, spectral changes in the anterior cingulate cortex  
27 best predicted depression severity in all three subjects. Relaxing this constraint revealed unique,  
28 individual-specific sets of spatio-spectral features predictive of symptom severity, reflecting the  
29 heterogeneous nature of depression. The ability to decode depression severity from neural activity  
30 increases our fundamental understanding of how depression manifests in the human brain and provides a  
31 target neural signature for personalized neuromodulation therapies.

## 32 **Main text**

33 Psychiatric and cognitive disorders are among the most challenging ailments we face in terms of social,  
34 economic, and public health toll. This challenge derives in large part from their heterogeneity and  
35 complexity – heterogeneity in terms of the wide variance in manifestation of these disorders across  
36 individuals, complexity in terms of the dearth of objective biomarkers and limited understanding of  
37 underlying neurophysiological mechanisms. Adding to their complexity is the fact that these disorders  
38 arise from dysfunction not of isolated brain locations but rather of distributed, interconnected networks  
39 that span wide-ranging cortical and subcortical regions<sup>1,2</sup>. Networks implicated in psychiatric and  
40 cognitive disorders often include prefrontal regions<sup>3-5</sup>, which are the most evolutionarily evolved and  
41 are particularly challenging to model in non-human animals. To successfully engage and therapeutically  
42 modulate these dysfunctional circuits, we must attain a comprehensive understanding of their  
43 pathophysiology. The most precise tools available to accomplish this “circuit dissection” task in humans  
44 are electrophysiological recordings and stimulation with intracranial electrodes, given the relatively  
45 lower resolution and specificity of non-invasive modalities. Here we apply this approach to investigate  
46 the neurophysiological basis of a common and highly burdensome disorder—depression<sup>6</sup>.

47 Major depressive disorder (MDD) has a lifetime incidence of 10% to 15% and has immense  
48 social and economic consequences<sup>7,8</sup>. It is a major contributor to the overall global burden of disease<sup>9</sup>  
49 and in the U.S. alone accounts for more than \$200 billion in health care costs<sup>10</sup>. While many  
50 conventional treatments are available for depression, nearly one-third of patients are treatment-  
51 resistant<sup>11</sup>. A significant number of depressed patients do not respond to first-line medications even after  
52 multiple treatment trials<sup>12</sup>. Electroconvulsive therapy and transcranial magnetic stimulation are  
53 evidence-based treatments with short-term efficacy, but high rates of relapse are typical<sup>13-15</sup>.

54 One critical knowledge gap fueling the challenge of treating treatment-resistant depression  
55 (TRD) is an insufficient understanding of its neurophysiological basis. Most work to date has used non-  
56 invasive methods such as electroencephalography (EEG) and functional MRI<sup>16–20</sup>. These studies have  
57 described various alterations in brain activity patterns and suggested potential biomarkers, but precise  
58 neurophysiological understanding is still lacking<sup>21</sup>. This understanding, drilled down to the level of the  
59 individual, will be essential for treatment-resistant patients being considered for invasive  
60 neuromodulation such as deep brain stimulation (DBS)<sup>6,22,23</sup>.

61 Intracranial recordings such as those performed routinely for seizure monitoring<sup>24</sup> provide the  
62 requisite degree of spatial, temporal, and spectral specificity for this purpose. Whereas studying the  
63 neurophysiological basis of mood regulation is possible in the convenience sample of epilepsy  
64 patients<sup>25,26</sup>, doing so in a cohort of patients with severe depression and without co-morbid epilepsy  
65 would be advantageous. The location of recording electrodes can be targeted to depression-relevant  
66 regions, instead of being determined purely for seizure monitoring purposes. Activities during inpatient  
67 monitoring can prioritize dense sampling of depression severity measures without concern for  
68 interfering with seizure monitoring. The resulting features of interest from the ensuing analyses are more  
69 relevant to patients with TRD without contamination by processes related to epilepsy.

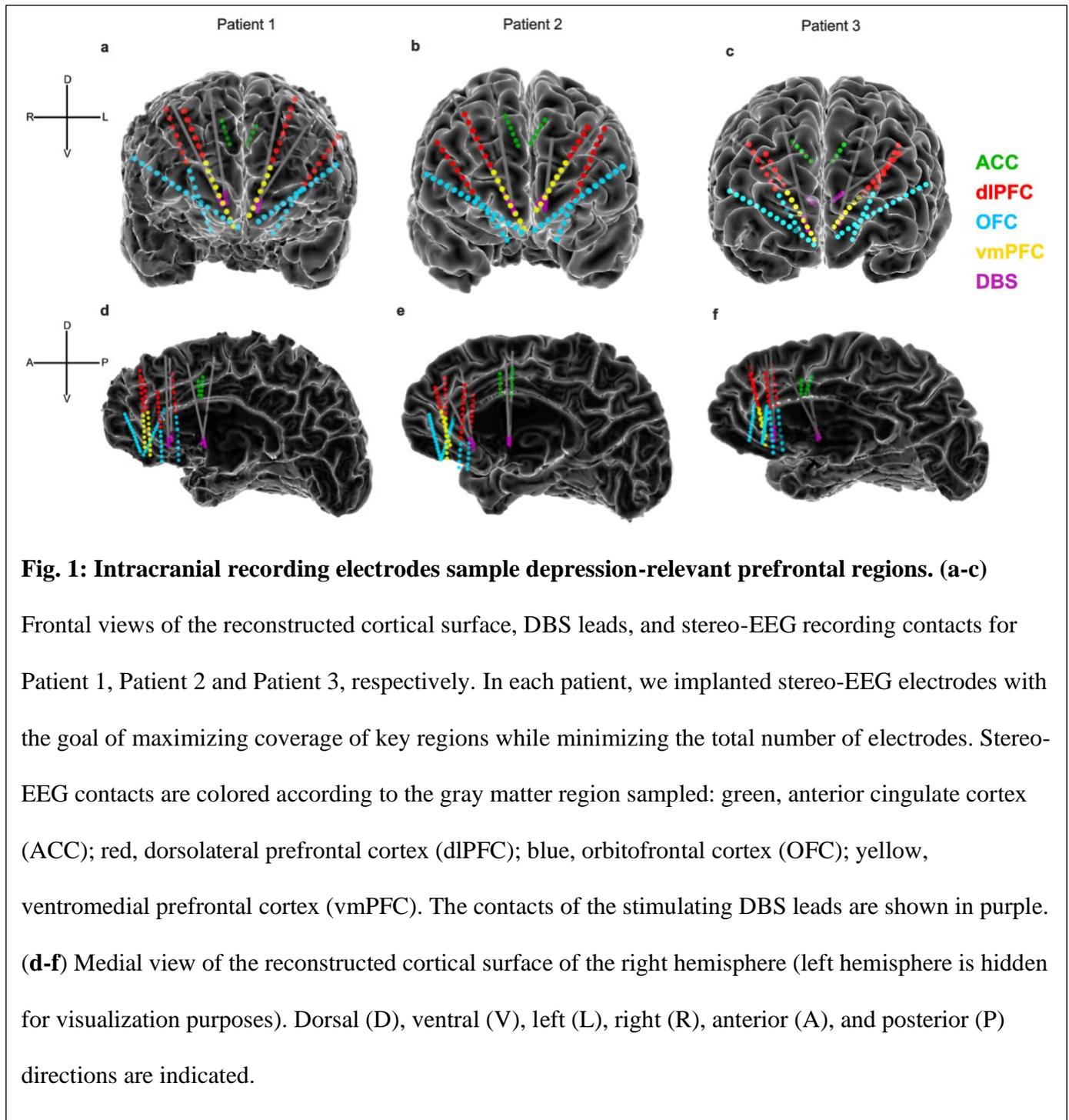
70 Here, our goal is to understand how depression is encoded in the brain by employing an  
71 intracranial EEG investigation platform incorporating dense behavioral assessments in TRD patients<sup>6</sup>. In  
72 doing so, we seek to address two critical questions: 1) what neurophysiological features characterize  
73 depression? 2) can we use these features to reliably predict depression severity? The first question  
74 entails finding neural correlates of symptom severity, while the second question addresses the more  
75 stringent requirement of finding truly predictive features. As our neuromodulatory therapies advance in  
76 sophistication, they will be able to incorporate these biomarkers of health and distress. Doing so will

77 hopefully allow therapeutic supply to more faithfully (spatially and temporally) match symptomatic  
78 demand and thereby improve outcomes.

## 79 **Results**

### 80 **Spectral activity across prefrontal regions correlates with depression severity**

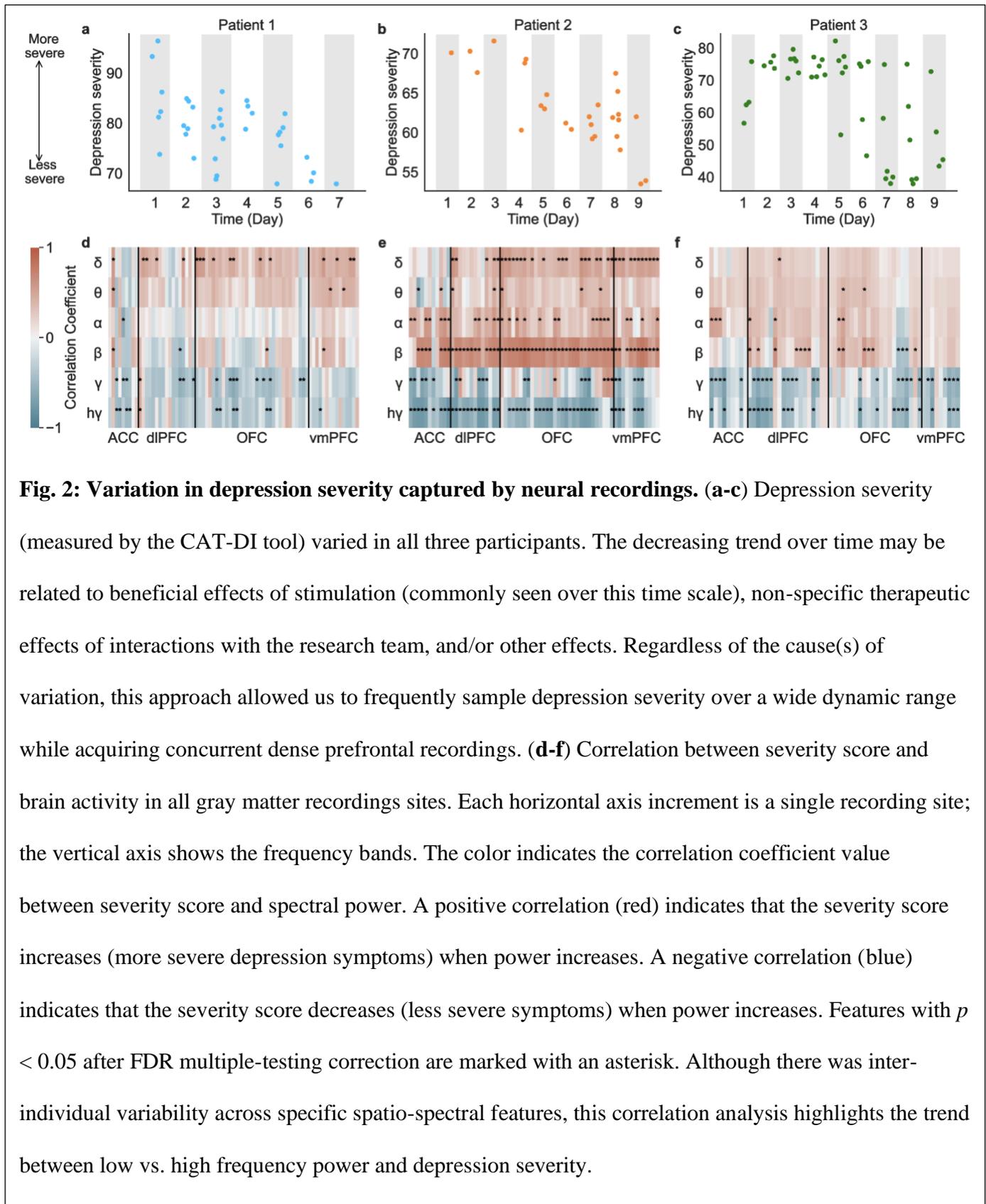
81 In our trial, three patients with severe TRD who met eligibility criteria and provided informed  
82 consent were implanted with two sets of intracranial electrodes during an initial surgery, one set  
83 primarily for stimulating and the other primarily for recording<sup>6,27</sup> (Fig. 1). The stimulation set consisted  
84 of two pairs of permanent DBS leads implanted bilaterally in two regions well-studied in DBS for TRD:  
85 the subcallosal cingulate (SCC)<sup>28</sup> and the ventral capsule/ventral striatum (VC/VS)<sup>29</sup>. The recording set  
86 consisted of percutaneously placed temporary “stereo-EEG”<sup>24</sup> electrodes implanted in brain regions  
87 involved in the regulation of mood and cognition: anterior cingulate cortex (ACC)<sup>18,30,31</sup>, dorsolateral  
88 prefrontal cortex (dlPFC)<sup>32,33</sup>, orbitofrontal cortex (OFC)<sup>34–36</sup>, and ventromedial prefrontal cortex  
89 (vmPFC)<sup>37,38</sup> (Fig. 1). Our clinical trial adapted the inpatient intracranial EEG investigation platform  
90 commonly used in epilepsy patients, re coined as the neurophysiology monitoring unit (NMU)<sup>6</sup>.  
91 Following the initial implant surgery, patients were kept in this inpatient monitoring unit for nine days.  
92 We assayed changes in depression severity while simultaneously recording from the implanted  
93 electrodes. Clinical outcomes from the first subject in this trial have recently been reported<sup>23</sup>. Here we  
94 report neural modeling results and include two additional subjects.



95 We measured symptom severity using the computerized adaptive test depression inventory  
96 (CAT-DI), a rapid assessment that correlates with standard depression severity scales<sup>39</sup>. Its adaptive  
97 nature allows each administration of this measure to take only 1-2 minutes to complete, and its use of a

98 variety of prompts prevents habituation and provides greater confidence with frequent sampling. We  
99 observed substantial variation in depression severity in all three participants (Patient 1: mean severity  
100 score= 78.9, std = 6.6; Patient 2: mean severity score = 63.0, std = 4.6; Patient 3: mean severity score =  
101 64.5, std = 14.1) over the 9-day monitoring period, with a trend of declining severity over the NMU stay  
102 (Fig. 2a-c).

103 To evaluate how neural activity varied throughout the NMU stay, we extracted spectral power  
104 from six frequency bands: delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz), gamma  
105 (35-50 Hz), and high-gamma (70-150 Hz), yielding six spectral power features per channel for each  
106 depression severity measurement. Spectral power features in prefrontal channels showed strong  
107 correlations with symptom severity scores after correcting for multiple comparisons. Although there was  
108 heterogeneity across patients, power in low frequencies including delta, theta, alpha, and beta were  
109 generally positively correlated with symptom severity (Fig. 2d-f, 92.0% of significant features in Patient  
110 1, 99.2% of significant features in Patient 2, and 88.5% of significant features in Patient 3), while power  
111 in high frequencies including gamma and high-gamma were generally negatively correlated with  
112 symptom severity (92.9% of significant features in Patient 1, 93.2% of significant features in Patient 2,  
113 and 100.0% of significant features in Patient 3). In all participants, a majority of brain regions  
114 demonstrated decreased low-frequency power and increased high-frequency power when symptoms  
115 were less severe.



## 116 **Prefrontal neural activity predicts depression severity**

117       The strength of the observed correlations in prefrontal cortex suggests that depression severity  
118 may reliably be predicted from spectral power features. To test this hypothesis, we fit regularized  
119 regression models to depression severity scores using neural activity recorded across prefrontal sites.  
120 Even with the relatively frequent sampling of depression severity, measurements were sparse (Patient 1:  
121 36 measurements; Patient 2: 27 measurements; Patient 3: 47 measurements) relative to the high  
122 dimensionality of neural features (approximately 400 features in each patient). In order to increase the  
123 generalizability of the model, we reduced the dimensionality of the neural data using automatic region  
124 selection and regularized regression. In particular, the model was constrained to use spectral power from  
125 a single region to predict depression severity. The selected region was chosen based on the analysis of  
126 only training data using a leave-one-out cross-validation strategy. After fitting the model, we used it to  
127 predict the depression severity score from a held-out test set and evaluated the prediction error using  
128 normalized root mean square error (NRMSE)<sup>26</sup>.

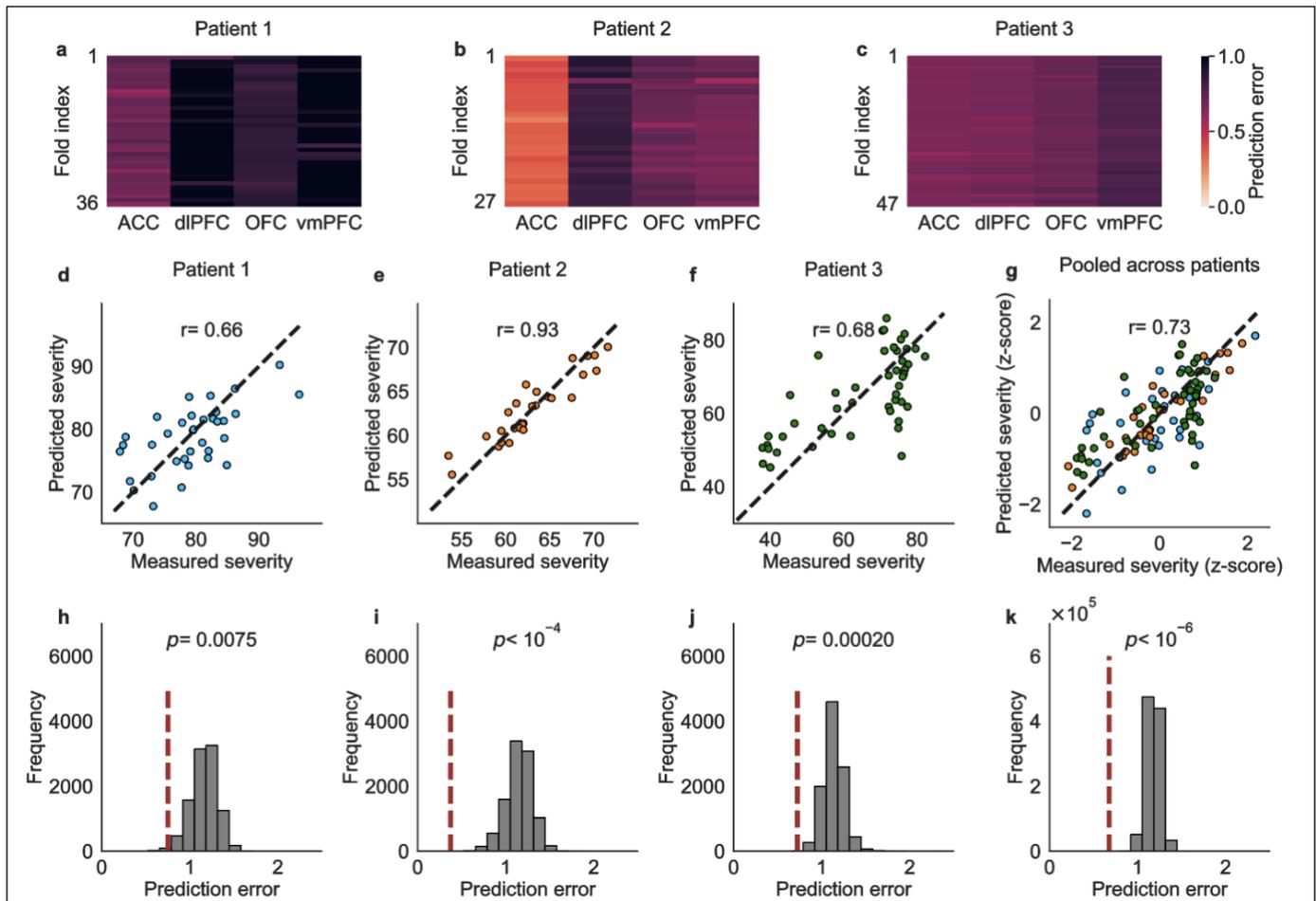
129       Using this approach, we were able to decode depression severity from neural signals in  
130 prefrontal cortex. Our model selected ACC across most folds of cross-validation (all folds in Patient 1  
131 and Patient 2, and 41 out of 47 folds in Patient 3), indicating that ACC was the most informative region  
132 for predicting depression severity in all three participants (Fig. 3a-c). Next, we explored relative feature  
133 importance within the ACC. Since the predictors were standardized in the model, the regression  
134 coefficients for each feature were indicators of feature importance, with larger coefficient magnitudes  
135 reflecting greater importance. A few features were consistently more important than others across all  
136 folds of the cross-validation and were significantly correlated with symptom severity score  
137 (Supplementary Fig. 1). Additionally, there was a strong and significant correlation between predicted  
138 and measured symptom score in each participant (Fig. 3d-f; Patient 1:  $r = 0.66$ ,  $p < 10^{-4}$ , Patient 2:  $r =$

139 0.93,  $p < 10^{-4}$ , Patient 3:  $r = 0.68$ ,  $p < 10^{-4}$ ), indicating high predictive performance. The decoder was  
140 also highly predictive when scores were standardized and pooled across all participants (Fig. 3g;  $r =$   
141  $0.73$ ,  $p < 10^{-4}$ ).

142 We assessed the significance of the NRMSE with a permutation test to evaluate decoder  
143 performance relative to chance. In each patient, we randomly permuted the time indices of severity  
144 scores and repeated the same leave-one-out cross-validation process. With the scores permuted and thus  
145 mismatched with the neural data, depression severity could no longer be accurately predicted  
146 (Supplementary Fig. 2). The prediction accuracies of the decoders were significantly greater than chance  
147 performance in all participants (Fig. 3h-j; Patient 1: NRMSE = 0.75,  $p < 0.01$ , Patient 2: NRMSE = 0.37,  
148  $p < 10^{-4}$ , Patient 3: NRMSE = 0.72,  $p < 10^{-3}$ ), and when the scores were pooled across participants (Fig  
149 3k; NRMSE = 0.68,  $p < 10^{-6}$ ).

150 Given our observation of decreasing depression severity over time (Fig. 2a-c), we tested whether  
151 our model was trivially identifying a temporal correlation. To do so, we fit a linear regression model to  
152 depression severity over time and computed the residuals (Supplementary Fig. 3a-c). This process  
153 effectively decorrelated depression severity with time. We then used the neural data to predict the  
154 residuals. The decoders accurately predicted the residuals in all patients (Supplementary Fig. 3d-k;  $p <$   
155  $0.05$  for all patients,  $p < 10^{-6}$  for the scores across patients), indicating that the neural features, not the  
156 progress of time, drove the accurate predictions of depression severity.

157 To add further confidence to our model, we performed 5-fold cross-validation in addition to our  
158 original leave-one-out cross-validation approach. The model was trained and selected on four folds of  
159 symptom severity scores, and the scores in the other fold were predicted. The decoder accurately  
160 predicted depression severity in 5-fold cross-validation as well (Supplementary Fig. 4;  $p < 0.01$  for all  
161 patients,  $p < 10^{-6}$  for the scores across patients), providing further confidence in the results.



**Fig. 3: Decoding depression severity from neural activity in prefrontal cortex.** (a-c) Prediction error of the training set in each fold of the leave-one-out cross-validation. The horizontal axis shows the brain regions, and the vertical axis shows the fold indices of leave-one-out cross-validation. The color indicates the prediction error. Each row includes the prediction error for all regions tested in the training set. Brighter red indicates a smaller error and therefore more accurate prediction. The ACC demonstrated greatest prediction accuracy across most folds of cross-validation, although its degree of supremacy varied across the three subjects. (d-g) The predicted score from leave-one-out cross-validation is plotted against the measured score for each depression severity measurement. Points closer to the diagonal indicate more accurate predictions. r values denote Pearson correlation coefficients. Scores from Patient 1 are shown in blue, scores from Patient 2 are shown in orange, and scores from

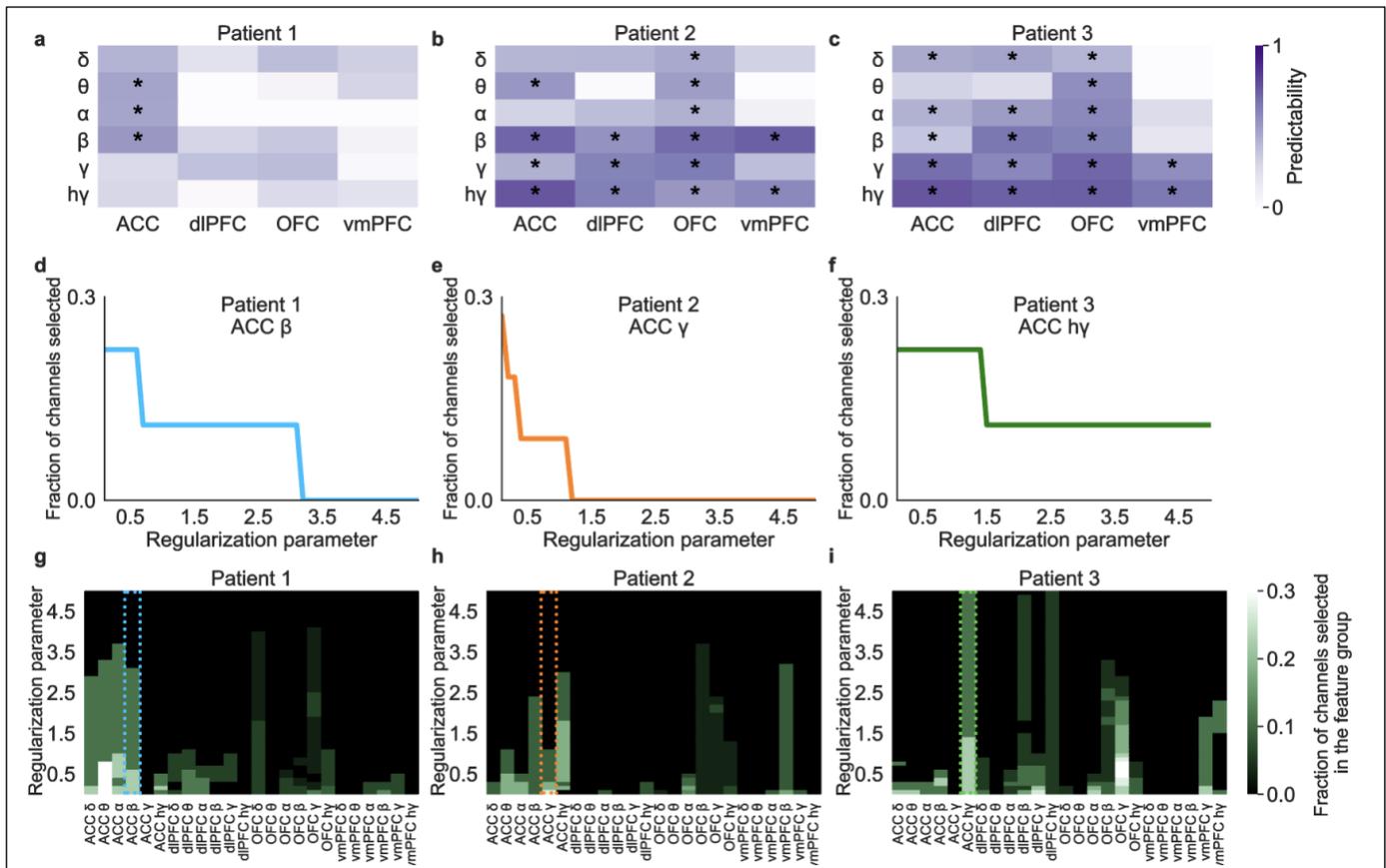
Patient 3 are shown in green. **(h-k)** Distribution of the chance level NRMSE computed from leave-one-out cross-validation for sets of permuted scores (gray,  $n = 10^4$  permuted tests for each patient and  $n = 10^6$  for standardized severity scores pooled across patients). NRMSE for the model trained with true measured severity scores is shown as a red vertical dashed line. Permutation testing shows that the true prediction error is significantly lower than chance.

162 To evaluate the performance of the region selection technique employed for dimensionality  
163 reduction, we fit the decoder without automatic region selection. These decoders can still predict  
164 depression severity at levels significantly better than chance (Supplementary Fig. 5a-h;  $p < 0.05$  for all  
165 patients,  $p < 10^{-6}$  for the scores across patients), but have larger prediction errors than decoders with  
166 region selection (Supplementary Fig. 5i-l). Thus inclusion of region selection improves the decoding  
167 accuracy of our model.

168 Next, we explored the predictability of spectral power from single brain regions and frequency  
169 bands (as opposed to all frequency bands) by using these individual features as inputs to the model.  
170 While many spectral power features showed significant correlations with depression severity (Fig. 2d-f),  
171 not all of these features may necessarily be predictive. We trained separate models for each region and  
172 frequency band combination and then performed multiple comparison corrections. For Patient 1, we  
173 found that theta, alpha, and beta power in ACC had significant predictability (Fig. 4a). For Patient 2 and  
174 Patient 3, we observed more widespread predictability across several regions and frequency bands (Fig.  
175 4b-c). ACC beta power was the overall most predictive feature in Patient 1, whereas ACC high-gamma  
176 was most predictive in Patient 2 and Patient 3.

177 To further focus on the neural features that are important for predicting depression severity, we  
178 fit the model using all spatial and spectral features without automatic region selection. This time,

179 however, we fit the model using a range of values of the regularization parameter alpha to explore the  
 180 relative importance of individual features. This parameter determines the penalty imposed on the model  
 181 for including more features. As alpha becomes larger, penalization for adding features increases, and  
 182 thus fewer features are used. Using ACC beta power, ACC gamma power and ACC high-gamma power  
 183 as examples, the fraction of recording channels selected in specific feature groups decreases as the  
 184 regularization parameter alpha increases (Fig. 4d-f). When fitting the model using all individual features,  
 185 we found that the greatest fraction of channels selected is in ACC, especially when penalization is  
 186 higher (Fig. 4g-i). As penalization decreases and more features are permitted, spectral features in dlPFC,  
 187 OFC and vmPFC are increasingly included, indicating that informative features for predicting  
 188 depression severity are not exclusive to the ACC, but instead are distributed across prefrontal cortical  
 189 regions in various frequency bands and in an individual-specific manner.



**Fig. 4: Spatio-spectral features for predicting depression severity are individual-specific. (a-c)**

Predictability of depression severity when using features from a single frequency band in a single brain region. The color indicates the correlation coefficient value between predicted and measured scores.

Correlation coefficients with  $p < 0.05$  (after FDR correction) are marked with an asterisk. Significantly predictive features are restricted to low-frequency bands in ACC in Patient 1 but are more distributed

across spatio-spectral features in Patients 2 and 3. **(d-i)** Fraction of recording channels selected in a

specific brain region and frequency band as the regularization parameter alpha varies. **(d-f)** As

illustrative examples, we show ACC beta power in Patient 1, ACC gamma power in Patient 2, and ACC

high-gamma power in Patient 3. At low values of alpha (more permissive of additional features), many

features are selected for inclusion in some fraction of recording channels, but a greater fraction is

included for features with most predictive power. As alpha increases, features with most predictive

power survive longer and in a greater fraction of channels. **(g-i)** The same information as in **d-f** is

shown, but as a heatmap and for all features. The three examples in **d-f** are shown with color-

corresponding dotted boxes around their respective column. Again, the fraction of recording channels for

particular features decreases with increasing alpha. ACC features are selected the most in these patients,

but feature distribution is highly individual-specific.

190 **Discussion**

191 Finding neurophysiological features that characterize and even predict symptom severity is

192 critical for improving our understanding of and developing precise treatments for depression. Here we

193 accurately decoded fluctuations in depression severity over time from intracranially recorded neural

194 activity in three TRD patients. We found that decreased low-frequency power and increased high-

195 frequency power in prefrontal cortex correlate with lower depression severity (i.e., healthier states).

196 Using these spatio-spectral features, we trained a model to predict depression severity and explored the  
197 features that were most important for prediction, rather than features that simply showed significant  
198 correlations. The transparency and explainability of our model allowed us to identify the ACC as the  
199 most influential subregion for decoding depression severity in all three patients. Beyond the ACC, we  
200 found various individual-specific neurophysiological features distributed across prefrontal cortex with  
201 predictive power. We consider these feature sets to be personalized neural biomarkers for depression  
202 severity.

203 The spectral pattern of correlation we observed, increased high-frequency activity and reduced  
204 low-frequency activity, has also been found to be important for decoding positive and negative affect in  
205 epilepsy patients<sup>40</sup>. While affective state is not identical to our depression-specific measures, it is likely  
206 related. The explainable model we employed was trained on intracranial neural data collected from a  
207 cohort of patients with depression, thus eliminating the confounds of epilepsy comorbidities and  
208 possible attendant neurophysiological abnormalities. The correlation we observed between neural  
209 features and improved depression symptoms is reminiscent of similar associations between this spectral  
210 pattern (in particular, increased high-frequency activity) and improved performance in several domains.  
211 In visual cortex, for example, increased gamma power predicts faster reaction times in a perception task,  
212 perhaps related to increased neural synchrony and resultant facilitated information transfer<sup>41</sup>. Intracranial  
213 investigations of human memory have shown that increased high frequency activity and decreased low  
214 frequency activity in left temporal cortex predict verbal memory encoding<sup>42-44</sup>, again perhaps reflecting  
215 the facilitatory effect of synchronized neuronal spiking<sup>45</sup>. In prefrontal cortex, we have previously  
216 shown that increased gamma power predicts improved performance on a cognitive interference task on a  
217 trial-by-trial basis<sup>46</sup>, potentially reflecting optimal allocation of cognitive control resources relative to  
218 demand<sup>47</sup>. These observations across brain regions mediating perception-to-action behaviors highlight

219 the importance of gamma power as an indicator of neural coherence and a candidate biomarker of  
220 performance.

221 In areas subserving cognitive processes, however, the relationship appears more complicated.  
222 Though insufficient application of cognitive resources leads to eroded performance in controlled  
223 decision-making tasks<sup>46,47</sup>, unconstrained attentive resources can also be maladaptive. For example,  
224 constitutively high activity in PFC regions including the ACC may be the physiological driver of the  
225 inappropriately sustained attention to the external environment (e.g., the ordering of objects in a desk  
226 drawer) or to internal feelings (e.g., certainty that the stove is off) characteristic of obsessive-compulsive  
227 disorder<sup>48,49</sup>. The same requirement for balance is true in affective domains. For example, consider  
228 reward sensitivity, one of the cardinal features of positive affect<sup>50</sup>. Whereas a hallmark feature of  
229 depression is insufficient reward sensitivity<sup>51</sup>, inappropriately elevated reward-seeking behavior is  
230 pathognomonic of addiction disorders<sup>51,52</sup>. Thus in disorders such as depression, which include  
231 dysfunction in cognitive and affective domains<sup>53</sup>, these competing forces must achieve balance in order  
232 to produce adaptive, productive, euthymic behavior. Consequently, neurophysiological biomarkers of  
233 such states will likely be more complex.

234 We therefore extended our investigation beyond the observed correlations and built a decoder to  
235 predict depression severity from spectral power features. Due to the sparse sampling of symptom state  
236 relative to that of the electrophysiological data, we began with a model formulation forced to use only  
237 the smallest subset of regions, thus producing a model that is more generalizable and less prone to  
238 overfitting. The consistent selection of ACC in all patients speaks to the importance of this region in  
239 mood and cognitive regulation<sup>40,54</sup>. Even as the region selectivity requirement was relaxed, the  
240 prominence of ACC remained (Fig. 4d-i). Although this brain region was commonly selected across  
241 patients, the predictive spectral features in ACC differed across participants. The features that were

242 highly predictive were a subset of those that showed significant correlations with depression severity. In  
243 Patient 1, beta power was most predictive of depression severity, whereas in Patients 2 and 3, high-  
244 gamma power was most predictive. Continued efforts can test the hypothesis that certain features are  
245 common predictors across individuals, whereas other features are individual-specific.

246 As seen in Figure 4, even though the ACC was the most predictive region for a reliable  
247 depression decoder in all patients, it was not the only one. Indeed, we found that features with  
248 significant predictive power were distributed across various prefrontal regions and frequency bands and  
249 were individual specific. This heterogeneity may reflect underlying differences in the involvement of  
250 depression-relevant regions across individuals. A growing body of work is attempting to associate  
251 differentially involved brain networks with the observed phenotypic diversity of depressive  
252 manifestations<sup>2,17</sup>. Within this context, our results further underscore the complex nature of this disorder  
253 and the need to account for inter-individual variability in order to optimally engage and treat  
254 symptomatic networks<sup>2,17,55,56</sup>.

255 Future efforts for decoding mood and affect will benefit from the development of continuous,  
256 objective behavioral markers. Currently, measuring these domains relies on administering behavioral  
257 assessments, a process that suffers from subjectivity and places a high burden on patients to frequently  
258 and accurately report their mental state. Contrast this situation with that of motor decoders, which enjoy  
259 the advantage of objective and highly temporally resolved measures of position, velocity, acceleration,  
260 and related variables<sup>57-60</sup>. Promising methods for affective measures with comparable characteristics  
261 include utilizing video and audio recordings<sup>61</sup>. Incorporating facial expression and speech analysis may  
262 provide more extensive affective/emotional measures and thereby create the opportunity to develop  
263 better mood decoders.

264 In conclusion, this study demonstrates the feasibility of accurately decoding depression severity  
265 based on intracranial prefrontal recordings in TRD patients. With this unique dataset of human  
266 intracranial recordings captured alongside measures of symptom severity, we also gain a deeper  
267 understanding of the pathophysiology of depression. Ultimately, our findings help to elucidate the  
268 neurophysiological underpinnings of depression and may lead to the design of more effective,  
269 personalized neuromodulatory interventions.

270 **Acknowledgments:** We thank our study participants for their commitment and trust. This study was  
271 supported by the National Institutes of Health grant no. UH3 NS103549 (SAS, KRB, JX, NRP, JM,  
272 RKM, BM, JAA, ABA, VP, DO, BS, MER, SJM, WKG and NP), grant no. K01 MH116364 (KRB and  
273 BM), grant no. R21 NS104953 (KRB and BM), grant no. UH3 NS100549 (WKG), grant no. R01  
274 MH114854 (WKG), and the McNair Foundation (SAS).

275 **Author contributions statement:** SAS and XP initiated the study. JX conceptualized analysis  
276 procedures, analyzed the data, and drafted the manuscript with support from NRP, JA, JM, RKM, PRS,  
277 ABP, AST, KRB, XP, and SAS. SAS, NP, WKG, SJM, and KRB oversaw the organization of the  
278 clinical trial, subject recruitment, and regulatory activities. JX, RKM, BM, JAA, ABA, VP, and DO  
279 performed data collection in the neurophysiology monitoring unit. RKM, BS, MER, and KRB  
280 performed MRI analysis. XP, SAS, AST, ABP, PRS, and KRB supervised and guided the data analysis.  
281 JX, NRP, and SAS wrote the manuscript and all authors contributed to the review and revision of the  
282 manuscript.

283 **Competing interests statement:** SAS has consulting agreements with Boston Scientific, Neuropace,  
284 Abbott, and Zimmer Biomet. WKG has received donated devices from Medtronic and has consulting  
285 agreements with Biohaven Pharmaceuticals. SJM is supported through the use of resources and facilities  
286 at the Michael E. DeBakey VA Medical Center, Houston, Texas and receives support from The  
287 Menninger Clinic. SJM has served as a consultant to Allergan, Alkermes, Axsome Therapeutics,  
288 BioXcel Therapeutics, Clexio Biosciences, Eleusis, EMA Wellness, Engrail Therapeutics, Greenwich  
289 Biosciences, Intra-Cellular Therapies, Janssen, Levo Therapeutics, Perception Neurosciences, Praxis  
290 Precision Medicines, Neumora, Neurocrine, Relmada Therapeutics, Sage Therapeutics, Seelos  
291 Therapeutics, Signant Health, and Worldwide Clinical Trials. SJM has received research support from  
292 Biohaven Pharmaceuticals, Janssen, Merck, NeuroRx, Sage Therapeutics, and VistaGen Therapeutics.  
293 The remaining authors declare no competing interests.

294 **References**

- 295 1. Gong, Q. & He, Y. Depression, Neuroimaging and Connectomics: A Selective Overview. *Biol. Psychiatry*  
296 **77**, 223–235 (2015).
- 297 2. Williams, L. M. Precision psychiatry: A neural circuit taxonomy for depression and anxiety. *The Lancet*  
298 *Psychiatry* **3**, 472–480 (2016).
- 299 3. Hare, B. D. & Duman, R. S. Prefrontal cortex circuits in depression and anxiety: contribution of discrete  
300 neuronal populations and target regions. *Mol. Psychiatry* **25**, 2742–2758 (2020).
- 301 4. Selemon, L. D. & Zecevic, N. Schizophrenia: a tale of two critical periods for prefrontal cortical  
302 development. *Transl. Psychiatry* **5**, e623–e623 (2015).
- 303 5. Manelis, A., Iyengar, S., Swartz, H. A. & Phillips, M. L. Prefrontal cortical activation during working  
304 memory task anticipation contributes to discrimination between bipolar and unipolar depression.  
305 *Neuropsychopharmacology* **45**, 956–963 (2020).
- 306 6. Allawala, A. *et al.* A Novel Framework for Network-Targeted Neuropsychiatric Deep Brain Stimulation.  
307 *Neurosurgery* **89**, E116–E121 (2021).
- 308 7. Pedersen, C. B. *et al.* A Comprehensive Nationwide Study of the Incidence Rate and Lifetime Risk for  
309 Treated Mental Disorders. *JAMA Psychiatry* **71**, 573–581 (2014).
- 310 8. Hasin, D. S. *et al.* Epidemiology of Adult DSM-5 Major Depressive Disorder and Its Specifiers in the  
311 United States. *JAMA Psychiatry* **75**, 336–346 (2018).
- 312 9. James, S. L. *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for  
313 354 Diseases and Injuries for 195 countries and territories, 1990–2017: A systematic analysis for the  
314 Global Burden of Disease Study 2017. *Lancet* **392**, 1789–1858 (2018).
- 315 10. Greenberg, P. E. *et al.* The Economic Burden of Adults with Major Depressive Disorder in the United  
316 States (2010 and 2018). *Pharmacoeconomics* **39**, 653–665 (2021).
- 317 11. Köhler-Forsberg, O., Cusin, C. & Nierenberg, A. A. Evolving Issues in the Treatment of Depression.  
318 *JAMA - J. Am. Med. Assoc.* **321**, 2401–2402 (2019).
- 319 12. Postorivo, D. & Tye, S. J. Novel Antidepressant Approaches for Refractory Depression. *Curr. Treat.*

- 320           *Options Psychiatry* **8**, 141–157 (2021).
- 321 13.   Jelovac, A., Kolshus, E. & McLoughlin, D. M. Relapse Following Successful Electroconvulsive Therapy  
322       for Major Depression: A Meta-Analysis. *Neuropsychopharmacology* **38**, 2467–2474 (2013).
- 323 14.   Carpenter, L. L. *et al.* Transcranial Magnetic Stimulation (TMS) for Major Depression: A multisite,  
324       naturalistic, observational study of acute treatment outcomes in clinical practice. *Depress. Anxiety* **29**, 587–  
325       596 (2012).
- 326 15.   Kellner, C. H. *et al.* Continuation Electroconvulsive Therapy vs Pharmacotherapy for Relapse Prevention  
327       in Major Depression: A Multisite Study From the Consortium for Research in Electroconvulsive Therapy  
328       (CORE). *Arch. Gen. Psychiatry* **63**, 1337–1344 (2006).
- 329 16.   de Aguiar Neto, F. S. & Rosa, J. L. G. Depression biomarkers using non-invasive EEG: A review.  
330       *Neurosci. Biobehav. Rev.* **105**, 83–93 (2019).
- 331 17.   Drysedale, A. T. *et al.* Resting-state connectivity biomarkers define neurophysiological subtypes of  
332       depression. *Nat. Med.* **23**, 28–38 (2017).
- 333 18.   Rolls, E. T. *et al.* Functional Connectivity of the Anterior Cingulate Cortex in Depression and in Health.  
334       *Cereb. Cortex* **29**, 3617–3630 (2019).
- 335 19.   Whitton, A. E. *et al.* Electroencephalography Source Functional Connectivity Reveals Abnormal High-  
336       Frequency Communication Among Large-Scale Functional Networks in Depression. *Biol. psychiatry.*  
337       *Cogn. Neurosci. neuroimaging* **3**, 50–58 (2018).
- 338 20.   Damborská, A. *et al.* EEG Resting-State Large-Scale Brain Network Dynamics Are Related to Depressive  
339       Symptoms . *Frontiers in Psychiatry* vol. 10 548 (2019).
- 340 21.   Sullivan, C. R. P., Olsen, S. & Widge, A. S. Deep brain stimulation for psychiatric disorders: From focal  
341       brain targets to cognitive networks. *Neuroimage* **225**, 117515 (2021).
- 342 22.   Scangos, K. W. *et al.* Closed-loop neuromodulation in an individual with treatment-resistant depression.  
343       *Nat. Med.* **27**, 1696–1700 (2021).
- 344 23.   Sheth, S. A. *et al.* Deep Brain Stimulation for Depression Informed by Intracranial Recordings. *Biol.*  
345       *Psychiatry* (2022) doi:10.1016/j.biopsych.2021.11.007.

- 346 24. González-Martínez, J. *et al.* Technique, Results, and Complications Related to Robot-Assisted  
347 Stereoelectroencephalography. *Neurosurgery* **78**, 169–180 (2016).
- 348 25. Kirkby, L. A. *et al.* An Amygdala-Hippocampus Subnetwork that Encodes Variation in Human Mood.  
349 *Cell* **175**, 1688-1700.e14 (2018).
- 350 26. Sani, O. G. *et al.* Mood variations decoded from multi-site intracranial human brain activity. *Nat.*  
351 *Biotechnol.* **36**, 954 (2018).
- 352 27. Sheth, S. A. *et al.* Deep brain stimulation for depression informed by intracranial recordings. *Biol.*  
353 *Psychiatry* (2021) doi:10.1016/j.biopsych.2021.11.007.
- 354 28. Holtzheimer, P. E. *et al.* Subcallosal cingulate deep brain stimulation for treatment-resistant depression: a  
355 multisite, randomised, sham-controlled trial. *The Lancet Psychiatry* **4**, 839–849 (2017).
- 356 29. Dougherty, D. D. *et al.* A Randomized Sham-Controlled Trial of Deep Brain Stimulation of the Ventral  
357 Capsule/Ventral Striatum for Chronic Treatment-Resistant Depression. *Biol. Psychiatry* **78**, 240–248  
358 (2015).
- 359 30. Späti, J. *et al.* Prefrontal Thinning Affects Functional Connectivity and Regional Homogeneity of the  
360 Anterior Cingulate Cortex in Depression. *Neuropsychopharmacology* **40**, 1640–1648 (2015).
- 361 31. Ho, T. C. *et al.* Inflexible Functional Connectivity of the Dorsal Anterior Cingulate Cortex in Adolescent  
362 Major Depressive Disorder. *Neuropsychopharmacology* **42**, 2434–2445 (2017).
- 363 32. Du, L. *et al.* Stimulated left DLPFC-nucleus accumbens functional connectivity predicts the anti-  
364 depression and anti-anxiety effects of rTMS for depression. *Transl. Psychiatry* **7**, 3 (2018).
- 365 33. Avissar, M. *et al.* Functional connectivity of the left DLPFC to striatum predicts treatment response of  
366 depression to TMS. *Brain Stimul.* **10**, 919–925 (2017).
- 367 34. Scangos, K. W., Makhoul, G. S., Sugrue, L. P., Chang, E. F. & Krystal, A. D. State-dependent responses  
368 to intracranial brain stimulation in a patient with depression. *Nat. Med.* **27**, 229–231 (2021).
- 369 35. Cheng, W. *et al.* Increased functional connectivity of the posterior cingulate cortex with the lateral  
370 orbitofrontal cortex in depression. *Transl. Psychiatry* **8**, 90 (2018).
- 371 36. Rao, V. R. *et al.* Direct Electrical Stimulation of Lateral Orbitofrontal Cortex Acutely Improves Mood in

- 372 Individuals with Symptoms of Depression. *Curr. Biol.* **28**, 3893-3902.e4 (2018).
- 373 37. Young, C. B. *et al.* Anhedonia and general distress show dissociable ventromedial prefrontal cortex  
374 connectivity in major depressive disorder. *Transl. Psychiatry* **6**, e810–e810 (2016).
- 375 38. Etkin, A., Büchel, C. & Gross, J. J. The neural bases of emotion regulation. *Nat. Rev. Neurosci.* **16**, 693–  
376 700 (2015).
- 377 39. Gibbons, R. D. *et al.* Development of a computerized adaptive test for depression. *Arch. Gen. Psychiatry*  
378 **69**, 1104–1112 (2012).
- 379 40. Bijanzadeh, M. *et al.* Decoding naturalistic affective behaviour from spectro-spatial features in multiday  
380 human iEEG. *Nat. Hum. Behav.* (2022) doi:10.1038/s41562-022-01310-0.
- 381 41. Womelsdorf, T., Fries, P., Mitra, P. P. & Desimone, R. Gamma-band synchronization in visual cortex  
382 predicts speed of change detection. *Nature* **439**, 733–736 (2006).
- 383 42. Burke, J. F. *et al.* Human intracranial high-frequency activity maps episodic memory formation in space  
384 and time. *Neuroimage* **85**, 834–843 (2014).
- 385 43. Long, N. M., Burke, J. F. & Kahana, M. J. Subsequent memory effect in intracranial and scalp EEG.  
386 *Neuroimage* **84**, 488–494 (2014).
- 387 44. Greenberg, J. A., Burke, J. F., Haque, R., Kahana, M. J. & Zaghoul, K. A. Decreases in theta and  
388 increases in high frequency activity underlie associative memory encoding. *Neuroimage* **114**, 257–263  
389 (2015).
- 390 45. Tong, A. P. S., Vaz, A. P., Wittig, J. H., Inati, S. K. & Zaghoul, K. A. Ripples reflect a spectrum of  
391 synchronous spiking activity in human anterior temporal lobe. *Elife* **10**, e68401 (2021).
- 392 46. Haegens, S. *et al.* Alpha and broadband high-frequency activity track task dynamics and predict  
393 performance in controlled decision-making. *Psychophysiology* **n/a**, e13901 (2021).
- 394 47. Shenhav, A., Botvinick, M. M. & Cohen, J. D. The Expected Value of Control: An Integrative Theory of  
395 Anterior Cingulate Cortex Function. *Neuron* **79**, 217–240 (2013).
- 396 48. Milad, M. R. & Rauch, S. L. Obsessive-compulsive disorder: beyond segregated cortico-striatal pathways.  
397 *Trends Cogn. Sci.* **16**, 43–51 (2012).

- 398 49. McGovern, R. A. & Sheth, S. A. Role of the dorsal anterior cingulate cortex in obsessive-compulsive  
399 disorder: converging evidence from cognitive neuroscience and psychiatric neurosurgery. *J. Neurosurg.*  
400 *JNS* **126**, 132–147 (2017).
- 401 50. Insel, T. R. The NIMH Research Domain Criteria (RDoC) Project: Precision Medicine for Psychiatry. *Am.*  
402 *J. Psychiatry* **171**, 395–397 (2014).
- 403 51. *Diagnostic and statistical manual of mental disorders : DSM-5™*. *DSM-5* (American Psychiatric  
404 Publishing, a division of American Psychiatric Association, 2013).
- 405 52. Volkow, N. D. & Morales, M. The Brain on Drugs: From Reward to Addiction. *Cell* **162**, 712–725 (2015).
- 406 53. Otte, C. *et al.* Major depressive disorder. *Nat. Rev. Dis. Prim.* **2**, 16065 (2016).
- 407 54. Rolls, E. T. The cingulate cortex and limbic systems for emotion, action, and memory. *Brain Struct. Funct.*  
408 **224**, 3001–3018 (2019).
- 409 55. Du, J. *et al.* Functional connectivity of the orbitofrontal cortex, anterior cingulate cortex, and inferior  
410 frontal gyrus in humans. *Cortex* **123**, 185–199 (2020).
- 411 56. Kupfer, D. J., Frank, E. & Phillips, M. L. Major depressive disorder: new clinical, neurobiological, and  
412 treatment perspectives. *Lancet (London, England)* **379**, 1045–1055 (2012).
- 413 57. Willett, F. R., Avansino, D. T., Hochberg, L. R., Henderson, J. M. & Shenoy, K. V. High-performance  
414 brain-to-text communication via handwriting. *Nature* **593**, 249–254 (2021).
- 415 58. Kao, J. C. *et al.* Single-trial dynamics of motor cortex and their applications to brain-machine interfaces.  
416 *Nat. Commun.* **6**, 1–12 (2015).
- 417 59. Chaudhary, U., Birbaumer, N. & Ramos-Murguialday, A. Brain–computer interfaces for communication  
418 and rehabilitation. *Nat. Rev. Neurol.* **12**, 513–525 (2016).
- 419 60. Abbaspourazad, H., Choudhury, M., Wong, Y. T., Pesaran, B. & Shanechi, M. M. Multiscale low-  
420 dimensional motor cortical state dynamics predict naturalistic reach-and-grasp behavior. *Nat. Commun.* **12**,  
421 607 (2021).
- 422 61. Dibeklioglu, H., Hammal, Z. & Cohn, J. F. Dynamic Multimodal Measurement of Depression Severity  
423 Using Deep Autoencoding. *IEEE J. Biomed. Heal. Informatics* **22**, 525–536 (2018).

424 **Methods**

425 **Study design**

426 An early feasibility trial (NCT03437928) of individualized deep brain stimulation (DBS) guided  
427 by intracranial recordings was conducted in adults (n=3) with treatment-resistant depression (TRD).  
428 Inclusion criteria include failure of pharmacological, cognitive/behavioral, and electroconvulsive  
429 therapies, severity of symptoms, and informed consent. Exclusion criteria include a history of psychosis,  
430 personality disorder, recent suicide attempt, or neurodegenerative disorder. All patients (Patient 1,  
431 Hispanic male in his 30s; Patient 2, White female in her 50s; Patient 3, White female in her 40s) met  
432 inclusion criteria and provided written and verbal consent to participate in the study. This trial is funded  
433 by the NIH BRAIN Initiative (UH3 NS103549) and approved by the FDA (IDE number G180300) and  
434 our single IRB (Baylor College of Medicine IRB number H-43036).

435 **Implant surgery**

436 Four DBS leads (Cartesia, Boston Scientific) were placed bilaterally in the ventral  
437 capsule/ventral striatum (VC/VS) and subcallosal cingulate (SCC) as previously described<sup>27</sup>. Ten  
438 stereoelectroencephalography (sEEG) electrodes were placed bilaterally in depression-relevant brain  
439 regions including anterior cingulate cortex (ACC), dorsolateral prefrontal cortex (dlPFC), orbitofrontal  
440 cortex (OFC), and ventromedial prefrontal cortex (vmPFC). A stereotactic robotic (ROSA, Zimmer  
441 Biomet) was used for the placement of both the DBS leads and sEEG electrodes<sup>6</sup>. Trajectories were  
442 carefully planned preoperatively to avoid sulci and blood vessels, and to maximize gray matter  
443 coverage. Accurate electrode location was verified intraoperatively using a fluoroscopic computerized  
444 tomography (CT) scanner and postoperatively with a true stereotactic CT. Following this implant  
445 surgery, patients were kept in the inpatient epilepsy monitoring unit (EMU) while we conducted a series  
446 of recording and stimulation studies.

447 **Depression severity measurements**

448 We acquired measurements of depression severity throughout the nine-day inpatient monitoring  
449 period using the computerized adaptive test for depression inventory (CAT-DI)<sup>39</sup>. Our choice of CAT-  
450 DI was motivated by our desire to capture symptom states that evolve over relatively short periods of  
451 time. Disease states related to mental illness are not static and symptom severity can fluctuate over  
452 minutes to days<sup>62</sup>. Depression assessment using standard scales such as the Hamilton Rating Scale for  
453 Depression (HAM-D)<sup>63</sup> and Montgomery-Åsberg Depression Rating Scale (MADRS)<sup>64</sup> are unwieldy  
454 and inappropriate for frequently repeated measurements because they take tens of minutes to administer  
455 and are designed for infrequent sampling (days to weeks). The CAT-DI satisfied our need for a dense  
456 sampling of depression severity.

457 Each CAT-DI administration typically includes approximately 12 question items selected from a  
458 bank of 389 possible items based on real-time feedback from previous items answered by the  
459 participant. Through CAT-DI, we collected rapid and relatively frequent measurements of depression  
460 severity. The CAT-DI score has been shown to exhibit a strong correlation with other established  
461 depression rating scales such as the Patient Health Questionnaire 9 (PHQ-9) and Hamilton Rating Scale  
462 for Depression (HAM-D)<sup>39</sup>. A higher CAT-DI score indicates more severe depression symptoms. A total  
463 of 37 CAT-DI tests were completed in Patient 1, 30 CAT-DI tests were completed in Patient 2, and 47  
464 CAT-DI tests were completed in Patient 3 with concomitant neural recordings (see below). For Patient  
465 1, we did not obtain CAT-DI surveys with neural recordings until day 3 of the 9-day period. Therefore,  
466 we refer to day 3 after surgery as day 1 in our visualization of CAT-DI scores over time (Fig. 2a).

467 **MRI and CT imaging protocols**

468 Prior to surgical implant, we conducted a preoperative MRI on a Siemens Prisma 3T scanner  
469 with a 64-channel head-neck coil. High-resolution (0.8 mm isotropic) T1-weighted anatomical images

470 (Magnetization-Prepared Rapid Acquisition with Gradient Echo, MPRAGE; repetition time (TR) of  
471 2,400 ms, time echo (TE) of 2.24 ms, an inversion time (TI) of 1,160 ms, a flip angle of 8°, and an  
472 acquisition time (TA) of ~7 min) were acquired. T2-weighted images (SPACE; 0.8 mm isotropic; TR of  
473 3,200 ms, TE of 563 ms, and a TA of ~6 min) were acquired in the same session. In addition,  
474 participants underwent postoperative, high-resolution clinical CT scans to capture electrode placement.

#### 475 **Cortical reconstructions**

476 FreeSurfer v6.0.0 (<https://surfer.nmr.mgh.harvard.edu/>)<sup>65</sup> was used to perform an automatic  
477 cortical reconstruction on the preoperative T1-weighted MRI. The T2-weighted MRI was used to  
478 improve reconstruction of the pial surfaces. Functional Magnetic Resonance Imaging for the Brain  
479 Software Library's Linear Image Registration Tool (FLIRT) (v6.0)<sup>66,67</sup> was used to align the  
480 postoperative CT data to the preoperative T1-weighted MRI. The postoperative CT data were used to  
481 determine the contact positions relative to local neuroanatomy.

#### 482 **Electrode localization**

483 Electrode coordinates were determined manually from the co-registered CT data in BioImage  
484 Suite v3.5b1<sup>68</sup> and placed into the native MRI space. An expert rater (BS) examined the images and  
485 determined whether the contact was in white or gray matter based on where it was plotted on the brain  
486 slice. Contacts that were determined to be in white matter were excluded from further analysis.  
487 Electrodes were also manually labeled to brain regions according to their anatomical location by the  
488 same rater. The cortical surface (pial surface) and electrode locations were reconstructed using the  
489 Multi-Modal Visualization Tool<sup>69</sup>.

#### 490 **Intracranial recordings**

491 Neural signals were recorded during the administration of CAT-DI test. Herein we will refer to  
492 each CAT-DI timepoint and its associated neural data as a 'block'. Signals were recorded with sEEG

493 electrodes at 2 kHz using a Cerebus data acquisition system (BlackRock Microsystems, UT, USA). All  
494 signals were amplified and bandpass filtered from 0.3-500 Hz (4th order Butterworth filter). DBS was  
495 off during all recordings.

#### 496 **Neural data preprocessing and signal conditioning**

497 Raw signals were visually inspected for the presence of recording artifacts. Channels that were  
498 found to have excessive noise were excluded to prevent noise from spreading to other channels through  
499 re-reference. Blocks with poor quality of the neural recording in more than half of the channels were  
500 excluded from further analysis. One block out of 37 was removed from Patient 1, three blocks out of 30  
501 were removed from Patient 2, and no block was removed from Patient 3. Each channel was notch  
502 filtered (60 Hz and its harmonics) to reduce line-noise artifacts. To reduce the effects of volume  
503 conduction, signals were then re-referenced through bipolar reference by subtracting the voltage of the  
504 neighboring contact on each sEEG electrode<sup>70</sup>. A total of 75 sEEG gray matter channels in Patient 1, 66  
505 channels in Patient 2, and 59 channels in Patient 3 remained after bipolar referencing were used for  
506 further analysis.

#### 507 **Feature extraction**

508 After signals were down-sampled to 1000 Hz, we performed a Hilbert transform to estimate  
509 spectral power features in six different frequency bands: 1-4 Hz (delta), 4-8 Hz (theta), 8-12 Hz (alpha),  
510 12-30 Hz (beta), 35-50 Hz (gamma) and 70-150 Hz (high-gamma). Then for each channel, we log-  
511 transformed the average power during the CAT-DI test for each frequency band.

#### 512 **Spectral activity analysis**

513 Pearson correlation coefficient and corresponding *p*-value were calculated between depression  
514 severity score and spectral power features from six frequency bands in all gray matter channels. FDR

515 multiple-testing correction was performed across the frequency bands and channels to correct  $p$ -values.  
516 A Pearson correlation coefficient with corrected  $p$ -value smaller than 0.05 was defined as significant.

### 517 **Automatic region selection**

518 The recording channels per patient were distributed across four main prefrontal regions: ACC,  
519 dlPFC, OFC, and vmPFC. To increase the generalizability of the model and avoid overfitting, we greatly  
520 reduced the number of model parameters by using a region selection technique<sup>26</sup>. Automatic region  
521 selection was employed to minimize the number of regions used in the decoder. Using training data  
522 only, we implemented model selection through inner cross-validation. The optimal brain region for  
523 decoding was determined by comparing independent models, each fit with the channels in one specific  
524 region. In the first stage of candidate model selection, independent decoders were fit with a single  
525 region, and one was selected. If the selected decoder achieved significance, automatic region selection  
526 stopped. Otherwise, the next stage consisted of fitting models using channels from two regions. The  
527 process continued until a decoder was found that achieved significance. In all three patients, significance  
528 was achieved after only the first stage, using features from a single region. Thus, this process largely  
529 decreased the number of model parameters.

### 530 **Model fitting**

531 We used LASSO regression to fit neural data to the CAT-DI scores as described in (1). LASSO  
532 regularization was used to minimize the objective function:

$$533 \min_w \frac{1}{2n_{samples}} \|Xw - y\|_2^2 + \alpha \|w\|_1 \quad (1)$$

534 where  $y$  is the measured CAT-DI score and  $X$  includes all features in the optimal brain region, which is  
535 selected by automatic region selection with inner-level cross-validation.  $w$  is the weight vector.  $\|Xw -$   
536  $y\|_2^2$  is the residual sum of squares, which measures the discrepancy between the predicted scores and  
537 the measured scores.  $\|w\|_1$  is the  $\ell_1$  term, which is the sum of the magnitudes of weights in the weight

538 vector.  $\alpha$  is a constant that multiplies the  $\ell_1$  term and is selected using inner cross-validation. By adding  
539  $\ell_1$  penalty, this regularization encourages sparsity since coefficients for many features are likely to  
540 become zero, hence eliminating these features from the model. Coordinate descent was used to fit the  
541 weights<sup>71</sup>. This regression method further reduces the dimensionality of the model.

## 542 **Measurement of prediction error**

543 We quantified the prediction error using normalized root mean square error (NRMSE)<sup>26</sup> as  
544 described in equation (2). NRMSE is defined by

$$545 \text{NRMSE} = \frac{\sqrt{\sum_{j=1}^n (\hat{y}_j - y_j)^2}}{\sqrt{\sum_{j=1}^n (y_j^* - y_j)^2}} \quad (2)$$

546 where  $n$  is the total number of CAT-DI scores,  $\hat{y}_j$  is the  $j$ th predicted score,  $y_j$  is the  $j$ th measured score,  
547 and  $y_j^*$  is the mean of other measured scores except for the  $j$ th measured score. The NRMSE quantifies  
548 how much more accurately the model predicts a test score relative to a model that predicts a test score  
549 simply as the average of the other scores. A lower NRMSE indicates that the prediction is more  
550 accurate. NRMSE was used as the prediction error criterion for selecting candidate models and for the  
551 final evaluation of the selected model when testing model significance.

## 552 **Cross-validation**

553 To assess the performance of our decoder, we used both leave-one-out (Fig. 3) and 5-fold cross-  
554 validation (Supplementary Fig. 4).

555 For leave-one-out cross-validation, we left out one CAT-DI score as test data and used the rest of  
556 the scores as training data to select the region and build the model. Through inner cross-validation using  
557 only the training data, we obtained the prediction error for all candidate models. To make the model  
558 more robust to noise, we defined a feature value as an outlier when it was more than four standard

559 deviations away from the mean value of that feature in the training set. We then automatically  
560 substituted all outlier values with the mean values of the corresponding feature in the training data. Less  
561 than 2% of feature values were identified as outliers for all participants.

562 The model with the lowest prediction error was then selected as the best model. This model  
563 fitting and selection process had no knowledge of the test CAT-DI score. We then used the best  
564 candidate model from each fold to predict the corresponding test score in the outer cross-validation. We  
565 repeated this leave-one-out procedure for all CAT-DI scores and then computed the NRMSE of the  
566 decoder using equation (2).

567 For 5-fold cross-validation, we split the data into five folds, trained the model on four folds of  
568 CAT-DI scores, and predicted the scores in the other fold. We repeated this procedure five times to  
569 obtain predicted scores for all CAT-DI points and then computed NRMSE to evaluate the performance  
570 of the decoder. In both leave-one-out and 5-fold cross-validation, model training and selection were  
571 conducted using only CAT-DI scores and neural activity within the training set.

## 572 **Model assessment**

573 We first calculated the true prediction error using the measured scores and predicted scores from  
574 each patient. In order to estimate decoding performance due to chance, we randomly permuted each  
575 symptom severity score on the neural data and then repeated the cross-validation process in the same  
576 manner as before. We did  $10^4$  random permutations to build a distribution of prediction errors expected  
577 due to chance. Next, we counted the number of samples in the distribution for which the prediction error  
578 due to chance was lower than the true prediction error. In this way, the  $p$ -value was defined as the  
579 probability that a model fit to a permuted set of symptom severity scores had a higher prediction  
580 accuracy than a model fit to the true scores.

581 Next, we evaluated the model across all patients by pooling their data. CAT-DI scores were z-  
582 scored based on the measured scores within each patient. These standardized scores were then pooled  
583 across patients. We repeated the permutation process  $10^6$  times to estimate the chance distribution of  
584 pooled prediction errors and calculated the corresponding  $p$ -value in the same way as described above.  
585 Here, the significance of the prediction error was the probability that decoding using the permuted set of  
586 CAT-DI scores (z-scored) has a lower NRMSE than using the true set of CAT-DI scores (z-scored).

### 587 **Evaluation of time as a potential confound of model performance**

588 CAT-DI scores for all three participants decreased over days in the EMU (Fig. 2a-c). In order to  
589 ensure that neural features were truly capturing fluctuations in depression severity rather than the  
590 passage of time, we fit a linear regression model to depression severity over time using the least-squares  
591 approach. For each data point, we calculated the residual by subtracting the predicted depression  
592 severity score using the line of best fit from the actual severity score. In this way, the resulting residuals  
593 were decorrelated with time (Supplementary Fig. 3). We then used the neural activity to predict the  
594 residuals using the same leave-one-out cross-validation technique as we have previously described.

### 595 **Evaluation of feature importance**

596 In order to evaluate the relative importance of features corresponding to individual frequency  
597 bands and subregions, we fit the model using features from single subregions and single frequency  
598 bands. We then calculated the Pearson correlation coefficient between predicted and true severity scores,  
599 and computed the corresponding  $p$ -values with FDR multiple-testing correction.

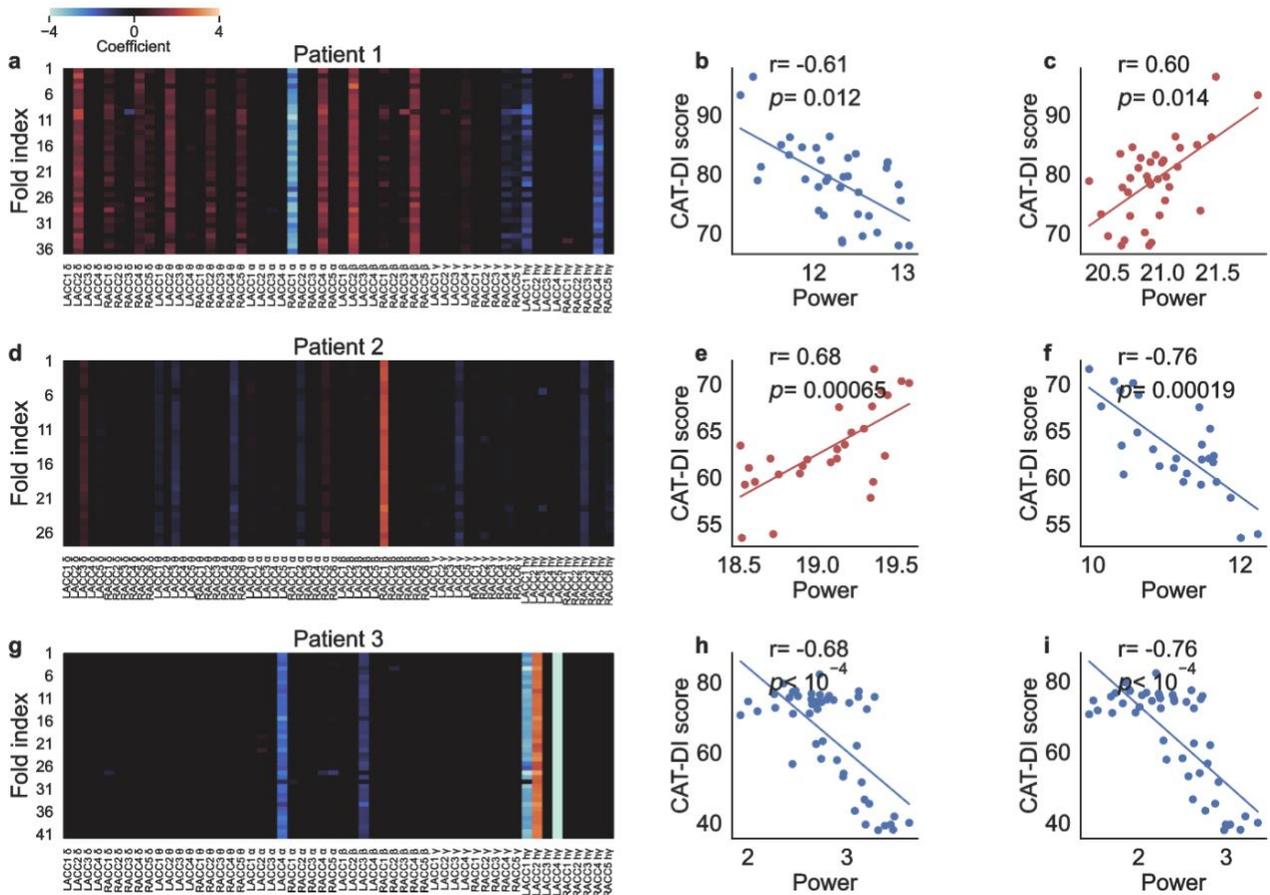
600 Lastly, to investigate feature importance without automatic region selection, we fit all features  
601 with LASSO regression at a range of different regularization parameter values. The regularization  
602 parameter alpha determines the penalty for using more features. As alpha decreases, the number of  
603 features used in the model increases. We fit the model using 50 different alpha values ranging from 0.1

604 to 5.0 in steps of 0.1. At each step, we calculated the fraction of channels in each feature group (defined  
605 by the region and frequency band) selected by the model. We completed these analyses using all data, as  
606 our goal was to better understand the effects of variation in regularization parameter alpha on the  
607 features that were selected by the model.

## 608 **References**

- 609 62. Provenza, N. R. *et al.* The Case for Adaptive Neuromodulation to Treat Severe Intractable Mental  
610 Disorders. *Front. Neurosci.* **13**, 152 (2019).
- 611 63. Hamilton, M. A rating scale for depression. *J. Neurol. Neurosurg. Psychiatry* **23**, 56 LP – 62 (1960).
- 612 64. Montgomery, S. A. & Åsberg, M. A New Depression Scale Designed to be Sensitive to Change. *Br. J.*  
613 *Psychiatry* **134**, 382–389 (1979).
- 614 65. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).
- 615 66. Jenkinson, M. & Smith, S. A global optimisation method for robust affine registration of brain images.  
616 *Med. Image Anal.* **5**, 143–156 (2001).
- 617 67. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved Optimization for the Robust and Accurate  
618 Linear Registration and Motion Correction of Brain Images. *Neuroimage* **17**, 825–841 (2002).
- 619 68. Joshi, A. *et al.* Unified framework for development, deployment and robust testing of neuroimaging  
620 algorithms. *Neuroinformatics* **9**, 69–84 (2011).
- 621 69. Felsenstein, O. *et al.* Multi-Modal Neuroimaging Analysis and Visualization Tool (MMVT). *arXiv Prepr.*  
622 **1**, 1–29 (2019).
- 623 70. Bastos, A. M. & Schoffelen, J.-M. A Tutorial Review of Functional Connectivity Analysis Methods and  
624 Their Interpretational Pitfalls . *Frontiers in Systems Neuroscience* vol. 9 (2016).
- 625 71. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via  
626 Coordinate Descent. *J. Stat. Softw.* **33**, 1–22 (2010).

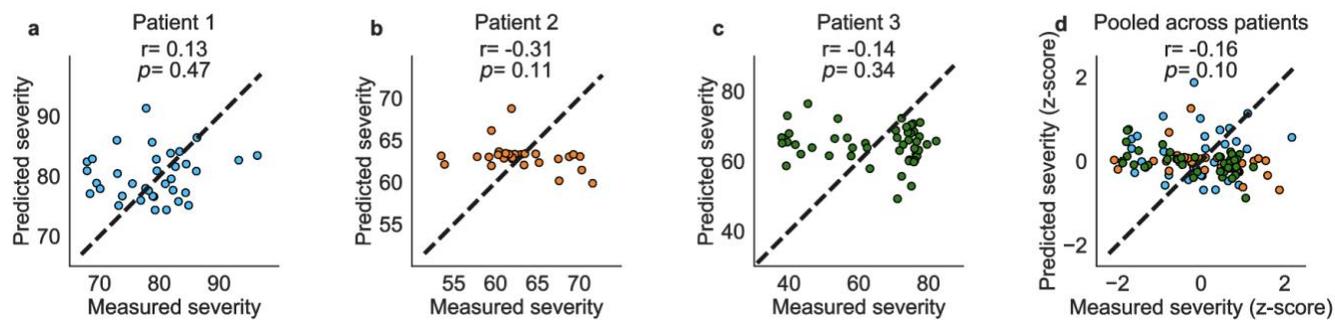
627 **Supplementary Figures**



628

629 **Supplementary Fig. 1: Training data is used to select features that are predictive of depression**

630 **severity.** (a, d, g) LASSO regression coefficients for input features in each fold of the leave-one-out  
 631 cross-validation selecting ACC. Horizontal axis shows the feature names. Vertical axis shows the fold  
 632 indices of leave-one-out cross-validation. The color indicates the coefficient value. Each row includes  
 633 the coefficients for all features in the selected region from that fold. Most coefficients become zero in  
 634 the LASSO regression, thus the color in most areas is black. Red indicates that the coefficient in the  
 635 LASSO regression is positive, and blue indicates that the coefficient is negative. (b-c, e-f, h-i) The  
 636 correlation coefficient and corresponding  $p$ -value (after FDR multiple-testing correction) between the  
 637 CAT-DI score and the value of example features. Red indicates that the correlation coefficient is  
 638 positive, and blue indicates that the correlation coefficient is negative.

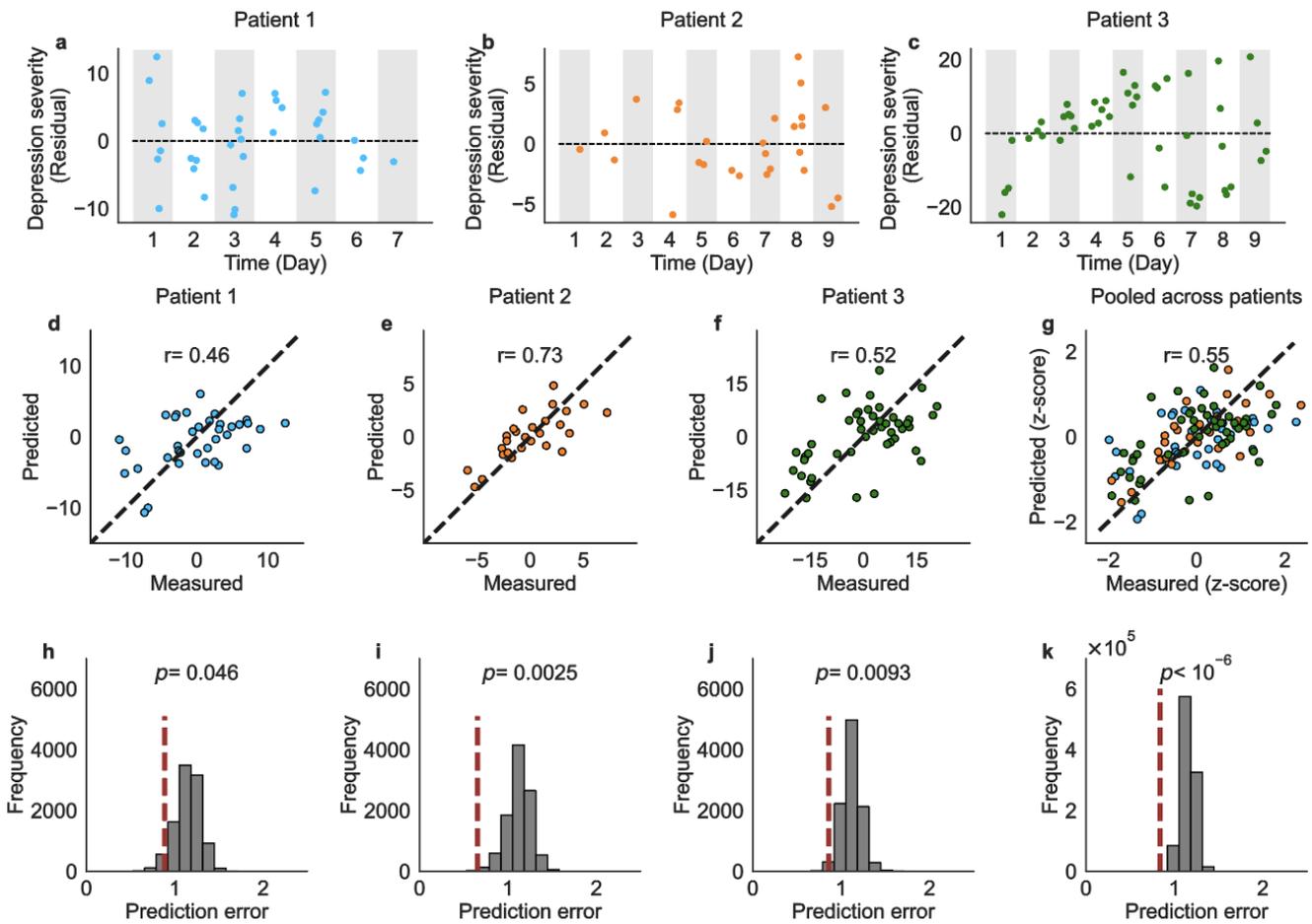


639

640 **Supplementary Fig. 2: Permuted CAT-DI scores yield low predictive performance.** Predictions for

641 an example set of permuted CAT-DI scores in Patient 1 (a), Patient 2 (b), Patient 3 (c), and pooled

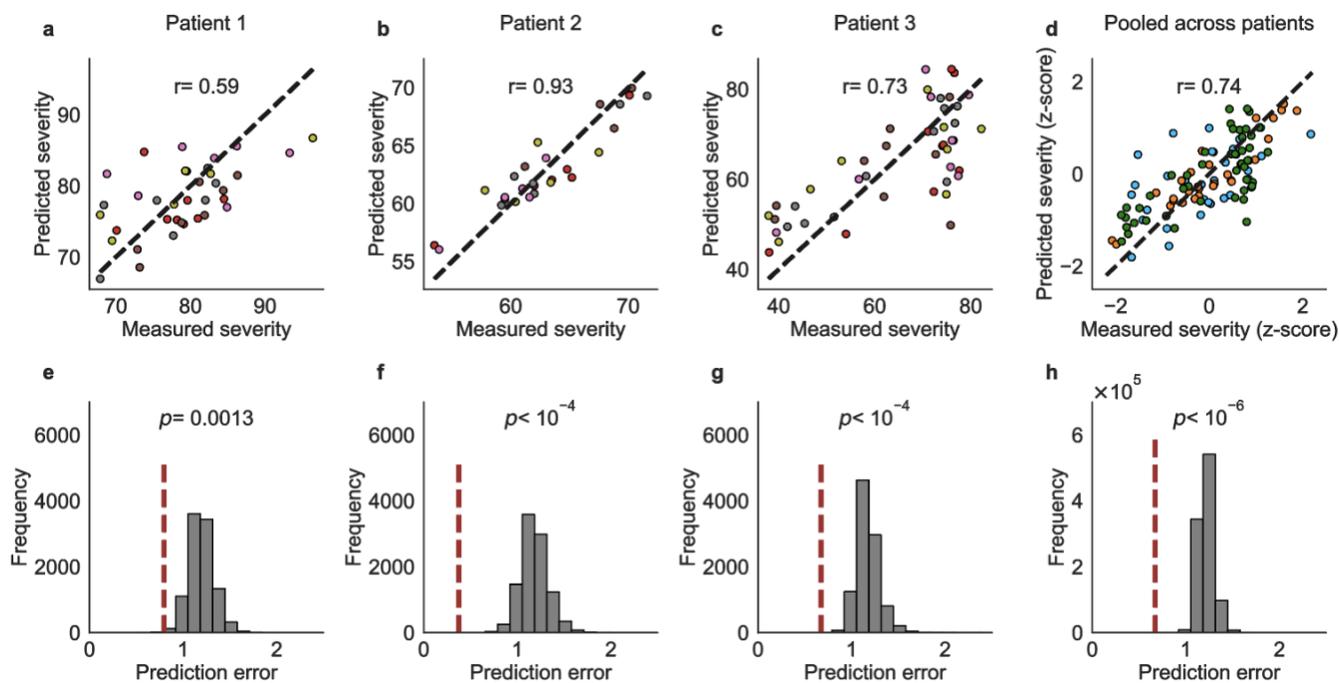
642 across patients (d).



643

644 **Supplementary Fig. 3: Depression severity can be predicted after removing the effect of time.**

645 (a-c) We fit a linear regression model between depression severity and time using the least-squares  
 646 approach. For each data point, we calculated the residual by taking the difference between the actual  
 647 depression severity and the predicted severity from the line of best fit. The residuals were then predicted  
 648 by the neural activity. (d-g) The predicted residual from leave-one-out cross-validation is plotted against  
 649 the true residual. (h-k) Distribution of the NRMSE for sets of permuted residuals (gray,  $n = 10^4$   
 650 permuted tests for each patient and  $n = 10^6$  for z-scores pooled across patients). NRMSE for the model  
 651 trained with true residuals is shown as a red vertical dotted line.



652

653 **Supplementary Fig. 4: Depression severity can be decoded using 5-fold cross-validation.** (a-d) The

654 predicted score from each fold of 5-fold cross-validation against the measured test score corresponding

655 to that fold. Points with the same color come from the same fold. Scores from Patient 1 are shown in

656 blue, scores from Patient 2 are shown in orange, scores from Patient 3 are shown in green in d. The

657 Pearson correlation coefficients are shown on the plots. The corresponding  $p$ -value is less than  $10^{-4}$  for

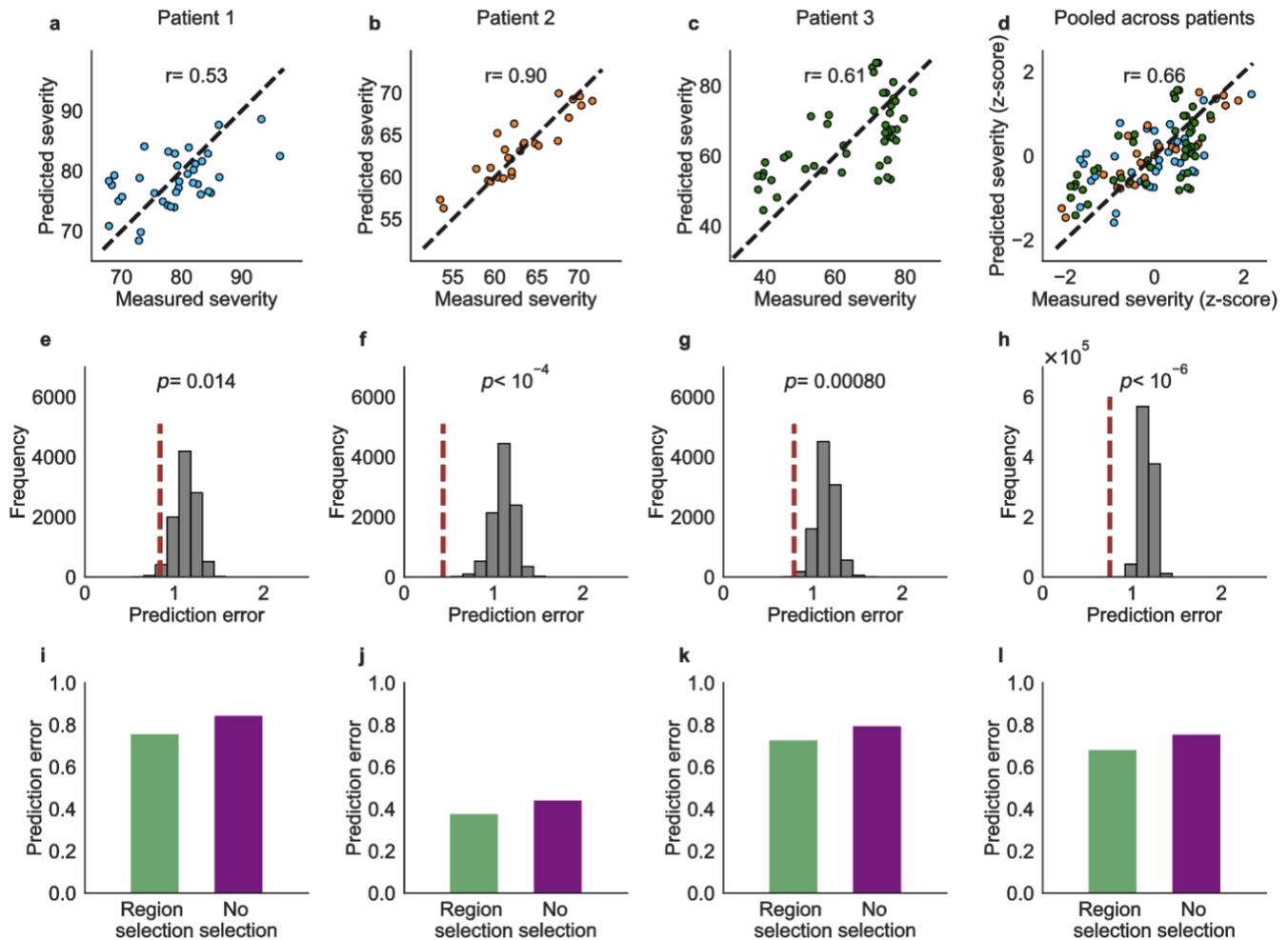
658 each patient and pooled across patients. (e-h) Distribution of the NRMSE for sets of permuted scores

659 (gray,  $n = 10^4$  permuted tests for each patient in 5-fold cross-validation and  $n = 10^6$  for z-scores pooled

660 across patients). NRMSE of the model trained using true measured CAT-DI scores is shown as a red

661 vertical dotted line. Permutation testing shows a prediction error significantly lower than chance in 5-

662 fold cross-validation in Patient 1, Patient 2, Patient 3, and when the scores were pooled across patients.



663

664 **Supplementary Fig. 5: Predictability when decoding without automatic region selection.** (a-d) The  
665 predicted score from leave-one-out cross-validation against the measured score when using LASSO  
666 regression without automatic region selection. (e-h) NRMSE of the LASSO regression model trained  
667 using true measured CAT-DI scores (red vertical dotted line) and distribution of the NRMSE for sets of  
668 permuted scores (gray bars,  $n = 10^4$  permuted tests for each patient,  $n = 10^6$  permuted tests for z-scores  
669 pooled across patients). (i-l) Comparison between prediction error in models with and without region  
670 selection technique. A lower prediction error indicates a more accurate prediction.