Online supplement for Siddiqi et al., Distinct Symptom-Specific Treatment Targets for Circuit-Based Neuromodulation. Am J Psychiatry (doi: 10.1176/appi.ajp.2019.19090915)

Supplementary methods

Subjects and data collection

Discovery dataset

30 patients with treatment-resistant depression received MRI scans prior to a routine course of clinical TMS at the Berenson-Allen Center for Noninvasive Brain Stimulation in Boston, MA. Patients were referred for clinical treatment due to treatment-resistant major depression. 3000 pulses of high-frequency TMS were delivered at 120% of resting motor threshold (RMT) in 4-second trains with a 26-second inter-train interval using a NeuroStar clinical stimulator (Neuronetics Inc, Malvern, PA) or a MagStim Super Rapid stimulator (Magstim Company Ltd, UK)[1]. Treatment was targeted to a scalp location 5.5 cm anterior to the site that induced a contraction in the right abductor pollicis brevis muscle. The stimulation site was recorded using stereotactic neuronavigation.

Self-report Beck Depression Inventory (BDI) and clinician-report 24-item Hamilton Rating Scale for Depression (HAMD) were collected before and after the treatment course. This protocol was approved by the Committee on Clinical Investigations at Beth Israel Deaconess Medical Center. The first 25 of these 30 patients were used in a prior publication[1].

Replication dataset

168 subjects with treatment-resistant depression received MRI scans as part of the OPT-TMS trial, a multi-site randomized clinical trial of TMS for major depression[2,3]. Data were collected at the Medical University of South Carolina, Columbia University, the University of Washington, and Emory University. 81 subjects received active treatment, while 87 received sham. 3000 pulses of high-frequency 10 Hz TMS were delivered at 120% of RMT in 4-second trains with a 26-second inter-train interval using a Neuronetics Model 2100 Therapy System (Neuronetics Inc, Malvern, PA). For 67% of subjects, treatment was delivered at a scalp location 5 cm anterior to the site that induced a contraction in the contralateral abductor pollicis brevis muscle. The stimulation site was recorded by placing a Vitamin E capsule over the stimulation site during the MRI scan. When this stimulation site overlapped with premotor cortex (33% of subjects), the stimulation site was moved an additional 1 cm anteriorly.

Self-report Inventory of Depressive Symptoms (IDS) and clinician-report 28-item Hamilton Rating Scale for Depression (HAMD) were collected before and after the treatment course. Montgomery-Asberg Depression Rating Scale (MADRS) was also collected as a secondary clinician-report measure. Patients who demonstrated at least 30% improvement after 3 weeks (active or sham groups) received additional blinded treatment for up to 3 weeks or until remission was reached. Nonremitters were then transferred to an open-label extension phase for up to 6 weeks or until remission was reached[2]. For simplicity and to allow direct comparison to sham, the 3-week time point was used as our primary outcome measure. However, we also repeated our analysis using the timepoint immediately after the blinded phase, which could be anywhere between 3 and 6 weeks depending on the patient. The protocol was approved by the institutional review board at each participating institution.

Seed-based connectivity analysis

Generation of stimulation site connectivity maps

For each subject in all three datasets, a region of interest (ROI) representing the stimulation volume was defined using concentric spheres of progressively decreasing intensity with a maximum radius of 12 mm as described previously[4]. Resting-state functional connectivity between this ROI and all other brain voxels was computed using fMRI data from a large cohort of normal subjects (n=1000)[5] as described previously[1]. The resulting connectivity maps were used for all subsequent analyses. Except where otherwise specified, all subsequent analyses were conducted using MATLAB R2018b.

Calculation of symptom-response maps across all subjects

These stimulation site connectivity maps were compared with improvement in each of the 21 symptoms on the Beck Depression Inventory (BDI). At each voxel, correlation was computed between stimulation site connectivity and clinical improvement. Clinical improvement was defined as absolute change in each symptom between the pre-treatment and post-treatment assessment. This analysis was repeated for each symptom at each voxel in the brain. At each voxel, this yielded a correlation coefficient between that voxel's stimulation site connectivity and improvement in each symptom. Across all voxels, this yielded 21 maps depicting the degree to which each brain voxel predicted improvement in each of the 21 symptoms. We refer to these voxel-wise maps as "symptom-response maps" (Figure 1C, main text).

These maps were not individually tested for significance due to the risk for multiple comparisons artifact, as there were a total of over 1,000,000 analyses (21 symptoms x 65,536 voxels in each symptom map). The maps were also not thresholded in order to avoid bias introduced by arbitrary threshold selection. Instead, all subsequent analyses were conducted based on spatial correlation between pairs of maps, as this approach reduces all of the comparisons into a single analysis.

Clustering

Clustering of circuit maps

Connections correlated with improvement in each depression symptom (symptom-response maps) were generated as above. The similarity between these 21 maps was determined by computing spatial correlations between each pair of maps. This yielded a 21 x 21 correlation matrix that quantifies the similarity between each pair of maps. To identify clusters in this matrix, the individual symptoms were treated as nodes and the cross-correlations were treated as distance metrics. The values were clustered using Ward's minimum variance method for hierarchical clustering. The optimal cluster solution was determined using the gap criterion, which identifies the minimum number of clusters required to approach an asymptote in error measurement[6]. This approach was chosen because it provides a single metric for the strength of the clustering solution (percent variance explained, measured by the gap statistic), facilitating control analyses and comparison to randomly permuted data. This also eliminates the risk of multiple comparisons artifact, as the maps are all reduced to a single metric of clustering strength. The cluster solutions were depicted using a force-directed graph visualization algorithm in Gephi 0.9.2.

Control analyses

Control analyses were conducted to confirm that the clustering results were not driven by the following four factors:

1. *Characteristics of the stimulation site alone:* The clustering analysis was repeated using baseline symptom severity rather than symptom improvement. Baseline symptom severity should be unrelated to stimulation site, thereby isolating the effects of stimulation site alone. Similarity between each pair of symptom-response maps was calculated (absolute value of the spatial cross-correlation) and statistically compared to the real data using an unpaired t-test (Fig. S2b).
2. *Covariance in symptom response alone:* The clustering analysis was repeated based on correlation between symptom change alone, without considering the stimulation sites or their connectivity. The strength of clustering was compared to our primary result using the gap statistic (Fig. S2c).
3. *Random chance:* The clustering analysis was repeated after randomly assigning each patient's stimulation site to a different patient's clinical response. A two-cluster solution was forced in this analysis. Similarity between each pair of symptom-response maps was calculated as above, and the mean was compared to real data using a permutation test with 10,000 iterations. The strength of clustering was compared to real data using the gap statistic (Fig. S2d). The analysis was also repeated without forcing a two-cluster solution in order to determine the frequency with which a two-cluster solution would arise spontaneously.
4. *Artificial binarization of a continuous distribution:* To test whether our clustering could have artificially binarized a normal distribution, we tested for normality using a Kolomgorov-Smirnov test.
5. *Influence of confounders:* To determine whether our results may have been driven by relevant clinical confounders, we repeated the circuit mapping procedure after including age, sex, and number of extant psychotropic medications as covariates. The resulting maps were compared with the original cluster maps to determine the influence of these confounders.

Replication of clustering model

The full clustering analysis was repeated for both symptom inventories in both datasets. This resulted in a pair of cluster-response maps for each dataset and symptom inventory, including the discovery BDI, discovery HAMD, replication HAMD, and replication IDS. In the replication dataset, maps were computed independently for active and sham data.

To test whether our clustering model replicated across symptom scales, we compared the self-report maps from our discovery dataset to the clinician-report maps for our discovery dataset, and the self-report maps from our replication dataset to the clinician-report maps from our replication dataset. Similarity between maps was assed using spatial correlation, producing four r values (2 cluster response maps x 2 datasets). These four r values were converted to a normal distribution using Fisher's r to z transform, then averaged to produce a single number reflecting the reproducibility of our maps across symptom scales. Significance was assessed via permutation testing (repeating the above analysis 10,000 times with random shuffling of stimulation sites and symptom responses in both datasets). In the randomly permuted data, the cluster containing the "sadness" item was labeled as "dysphoric," and the other cluster was labeled as "anxiosomatic."

To test whether our clustering model replicated across datasets, we compared the discovery cluster-response maps to the replication cluster-response maps. For simplicity, we first averaged the self-report and clinician-report maps within each dataset to generate two dysphoric cluster response maps and two anxiosomatic maps (one for each dataset). Similarity between maps was assed using spatial correlation, producing two r values (one for the dysphoric cluster and one for the anxiosomatic cluster). These two r values were converted to a normal distribution using Fishers r-to-z transform, then averaged to produce a single number reflecting the reproducibility of our maps across datasets. Significance was assessed via permutation testing as above, permuting only the replication dataset while leaving the discovery dataset unchanged.

To test whether the replication was specific to active versus sham stimulation, we repeated the above analysis using the sham arm of the replication dataset. As above, a single spatial correlation value was computed to represent the reproducibility of our maps across datasets. We hypothesized that this value would be higher for the active arm of the replication dataset compared the sham arm of the replication dataset and computed the difference between these two values. To determine whether this difference was larger than expected by chance, we re-computed this difference 10,000 times after randomly permuting the active and sham data from the replication dataset while leaving the discovery dataset unchanged.

Generation of conglomerate cluster maps across multiple datasets

Symptom-response maps were generated for all 97 symptoms in all four scales across both datasets. This yielded a total of 97 symptom response maps, which were clustered using the same methods described above. This conglomerate cluster solution was depicted using a force-directed graph visualization. The size of nodes was proportional to the normalized PageRank score, which quantifies the importance of a network node based on its connectedness[7]. This score was used to assess the contribution of each symptom to the clustering solution. For each scale in each dataset, the two most-contributory symptoms were identified for each cluster.

For the conglomerated clusters in each dataset, the symptom-response maps in that cluster were averaged to create an overall conglomerate pair of cluster maps.

Prediction of clinical utility

Identification of potential treatment targets

By definition, TMS to brain voxels whose connectivity is similar to our dysphoric cluster response map should improve dysphoric symptoms while TMS to brain voxels whose connectivity is similar to our anxiosomatic map should improve anxiosomatic symptoms. We therefore identified potential TMS treatment targets by identifying brain voxels that best matched these connectivity patterns. For each voxel in the brain, the connectivity profile of that voxel was compared with each cluster-response map using spatial correlations. Each voxel was labeled with this spatial r-value, which represents the similarity between that voxel's connectivity and our dysphoric and anxiosomatic circuits. This yielded a map of targets expected to modulate each symptom cluster. Because these maps showed minimal overlap, the difference between the two maps was computed to produce an overall "targeting atlas." The voxel values on this targeting atlas should predict relative improvement in dysphoric versus anxiosomatic symptoms.

Using the same procedure described above, a separate targeting atlas was also computed for each dataset alone, yielding a distinct "discovery" and "replication" targeting atlas. The discovery targeting atlas was used to predict clinical improvement in the replication active and sham datasets, while the replication targeting atlas was used to predict clinical improvement in the discovery dataset.

Validation across studies

A systematic literature review was conducted to identify therapeutic TMS studies that measured distinct mood and anxiety rating scales, which were considered as proxies for dysphoric and anxiosomatic clusters. This followed an approach similar to what was used in prior work involving retrospective analysis of multiple studies for identification of optimal TMS targets[8]. Studies were included if they satisfied the following criteria:

1. At least one mood scale and one anxiety scale were reported.
2. At least 3 weeks of daily therapeutic repetitive TMS were administered to the prefrontal cortex.
3. Treatment intensity was at least 5 Hz, as lower frequencies likely have inhibitory effects which were not included in our predictive model.
4. Stimulation site was clearly reported with adequate detail (or citation) to determine the location of the stimulation site with respect to our targeting atlas.
5. All subjects included in the study carried a primary DSM-IV or DSM-5 psychiatric diagnosis; for instance, studies of stroke rehabilitation or chronic pain were not included.

For each of these studies (listed in Table S3), the stimulation site was identified and converted to MNI coordinates. MNI coordinates for the EEG F3 target were estimated based on the report in Fox *et al.*, 2012[8]. The Beam F3 target was estimated based on Fried *et al.*, 2014[9]. The left-sided "5 cm" target was determined empirically as the mean location of all stimulation sites in the replication dataset, while the right-sided 5cm target was identified at the corresponding contralateral site. The left-sided "5.5 cm" target was determined empirically based on the mean of all stimulation sites in the discovery dataset. The "6 cm" target was estimated by extrapolating based on the distance between the "5 cm" and "5.5 cm" targets. The location of the task fMRI-based target was calculated empirically based on the mean of all neuronavigated stimulation sites in the dataset. The published dmPFC target coordinates included a superficial coordinate (0,60,60) and a deep brain coordinate (0,30,30)[10-12], neither of which was on the surface of the brain; as a result, the stimulated volume was estimated as the mean of these two coordinates, which was on the cortical surface. The anti-sgACC target coordinate[13] was reported directly in the original paper.

For studies that reported multiple mood scales, the analysis was based on the mood scale that was most similar to the anxiety scale. For instance, if the study's main anxiety metric was the Beck Anxiety Inventory, the mood analysis was based on the Beck Depression Inventory. If the study's main anxiety metric was the Hamilton Anxiety Rating Scale, the mood analysis was based on the Hamilton Depression Rating Scale. If there was no mood scale to directly correspond with the anxiety scale, then the study's pre-defined primary mood scale was used for the analysis. There were no studies reporting multiple scales that directly measure anxiety.

For each study, the mean stimulation site was plotted on the overall TMS targeting atlas. The voxel value of the stimulation site was compared with the ratio of percentage improvement in anxiety symptoms to percentage improvement in mood symptoms. A Pearson correlation was used to quantify this relationship. The success rate of the target map was also calculated based on the percentage of studies

in which the TMS targeting atlas map successfully predicted which symptom type would preferentially improve. A single-proportion z-test was used to determine whether this percentage was significantly different from 50%.

References (supplementary methods)

1. Weigand A, Horn A, Caballero R, et al. Prospective Validation That Subgenual Connectivity Predicts Antidepressant Efficacy of Transcranial Magnetic Stimulation Sites. *Biol Psychiatry.* 2018;84(1):28-37.
2. George MS, Lisanby SH, Avery D, et al. Daily left prefrontal transcranial magnetic stimulation therapy for major depressive disorder: a sham-controlled randomized trial. *Arch Gen Psychiatry.* 2010;67(5):507-516.
3. Johnson KA, Baig M, Ramsey D, et al. Prefrontal rTMS for treating depression: location and intensity results from the OPT-TMS multi-site clinical trial. *Brain Stimul.* 2013;6(2):108-117.
4. Fox MD, Liu H, Pascual-Leone A. Identification of reproducible individualized targets for treatment of depression with TMS based on intrinsic connectivity. *Neuroimage.* 2013;66:151-160.
5. Yeo BT, Krienen FM, Sepulcre J, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol.* 2011;106(3):1125-1165.
6. Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology).* 2001;63(2):411-423.
7. Ghoshal G, Barabasi AL. Ranking stability and super-stable nodes in complex networks. *Nat Commun.* 2011;2:394.
8. Fox MD, Buckner RL, White MP, Greicius MD, Pascual-Leone A. Efficacy of transcranial magnetic stimulation targets for depression is related to intrinsic functional connectivity with the subgenual cingulate. *Biol Psychiatry.* 2012;72(7):595-603.
9. Fried PJ, Rushmore RJ, 3rd, Moss MB, Valero-Cabre A, Pascual-Leone A. Causal evidence supporting functional dissociation of verbal and spatial working memory in the human dorsolateral prefrontal cortex. *Eur J Neurosci.* 2014;39(11):1973-1981.
10. Downar J, Geraci J, Salomons TV, et al. Anhedonia and reward-circuit connectivity distinguish nonresponders from responders to dorsomedial prefrontal repetitive transcranial magnetic stimulation in major depression. *Biol Psychiatry.* 2014;76(3):176-185.
11. Dunlop K, Woodside B, Lam E, et al. Increases in frontostriatal connectivity are associated with response to dorsomedial repetitive transcranial magnetic stimulation in refractory binge/purge behaviors. *Neuroimage Clin.* 2015;8:611-618.
12. Dunlop K, Woodside B, Olmsted M, Colton P, Giacobbe P, Downar J. Reductions in Cortico-Striatal Hyperconnectivity Accompany Successful Treatment of Obsessive-Compulsive Disorder with Dorsomedial Prefrontal rTMS. *Neuropsychopharmacology.* 2016;41(5):1395-1403.
13. Blumberger DM, Vila-Rodriguez F, Thorpe KE, et al. Effectiveness of theta burst versus high-frequency repetitive transcranial magnetic stimulation in patients with depression (THREE-D): a randomised non-inferiority trial. *Lancet.* 2018;391(10131):1683-1692.

Supplementary Results

Figure S1: All 21 symptom-response maps derived from the discovery BDI dataset

Figure S2: Clustering was significantly better for real data than for three control analyses. Left panels depict Fisher-transformed cross correlogram to show the spatial correlation between symptom-based circuit maps (diagonals are depicted in black). Middle panels depict a force-directed graph visualization produced in Gephi 0.9.2. In this algorithm, the correlation between nodes is treated as an attractive force, so highly correlated nodes are in close proximity to one another. Node sizes are proportional to the normalized PageRank score(2), a metric of the degree to which that node contributes to the solution. Distinct colors represent distinct clusters. Right panels represent variance explained by the clustering solution, as quantified by the gap statistic(4). **(a)** Individual symptom-based circuit maps were strongly correlated or anti-correlated with one another (left panel). Symptoms thus separated into two distinct clusters which explained 73% of the variance. **(b)** When repeating the analysis with baseline symptoms instead of symptom change, the cross-correlogram revealed a continuous pattern rather than two discrete clusters. A two-cluster solution explained only 25% of the variance. **(c)** Clustering is not evident based on symptom improvement alone. **(d)** Permutation testing showed that clusters generated by random chance are weaker than those generated from the actual data.

Figure S3: Distributions of cross-correlations between symptom maps. (a) Across 100 permutations, randomly-shuffled data showed cross-correlations that followed a normal distribution with a peak near zero. (b) The real data followed a skewed distribution with a trough near zero.

Figure S4: Force-directed graph visualizations depicting the clustering solutions for each dataset. Visualization follows the same parameters described in Fig. S1.

Figure S5: Cluster-response maps across different datasets (top: lateral view, bottom: medial view). Cluster-response maps were reproducible across different symptom scales and independent cohorts.

Figure S6: Regions of overlap between the two cluster maps.

Figure S7: Circuit maps for two-cluster solution generated when using a connectome database of 38 subjects with major depression rather than 1000 healthy controls.

Figure S8: Alignment of optimal targets with consensus cortical parcellation schemes.

Table S1: Dataset characteristics and patient demographics

Table S2: Clustering is not driven by baseline symptoms or overall clinical trajectory.
In the discovery dataset and the active arm of the replication dataset, clinical improvement was approximately equal between the two symptom clusters. In the sham dataset, dysphoric symptoms improved significantly more than anxiosomatic symptoms. Anxiosomatic symptom improvement was significantly greater in the active replication dataset than in the sham replication dataset. These results are consistent with the fact that the majority of patients in the replication dataset were stimulated at relatively anxiosomatic stimulation sites.
Clinical change in each cluster was not significantly correlated with baseline severity of that cluster in either the discovery dataset or the active arm of the replication dataset. In the sham dataset, clinical improvement was significantly related to baseline severity in the corresponding symptom cluster.

Table S3: Index of specific symptoms in figure 3a.

Table S4: Details of the studies included in the exploratory meta-analysis.

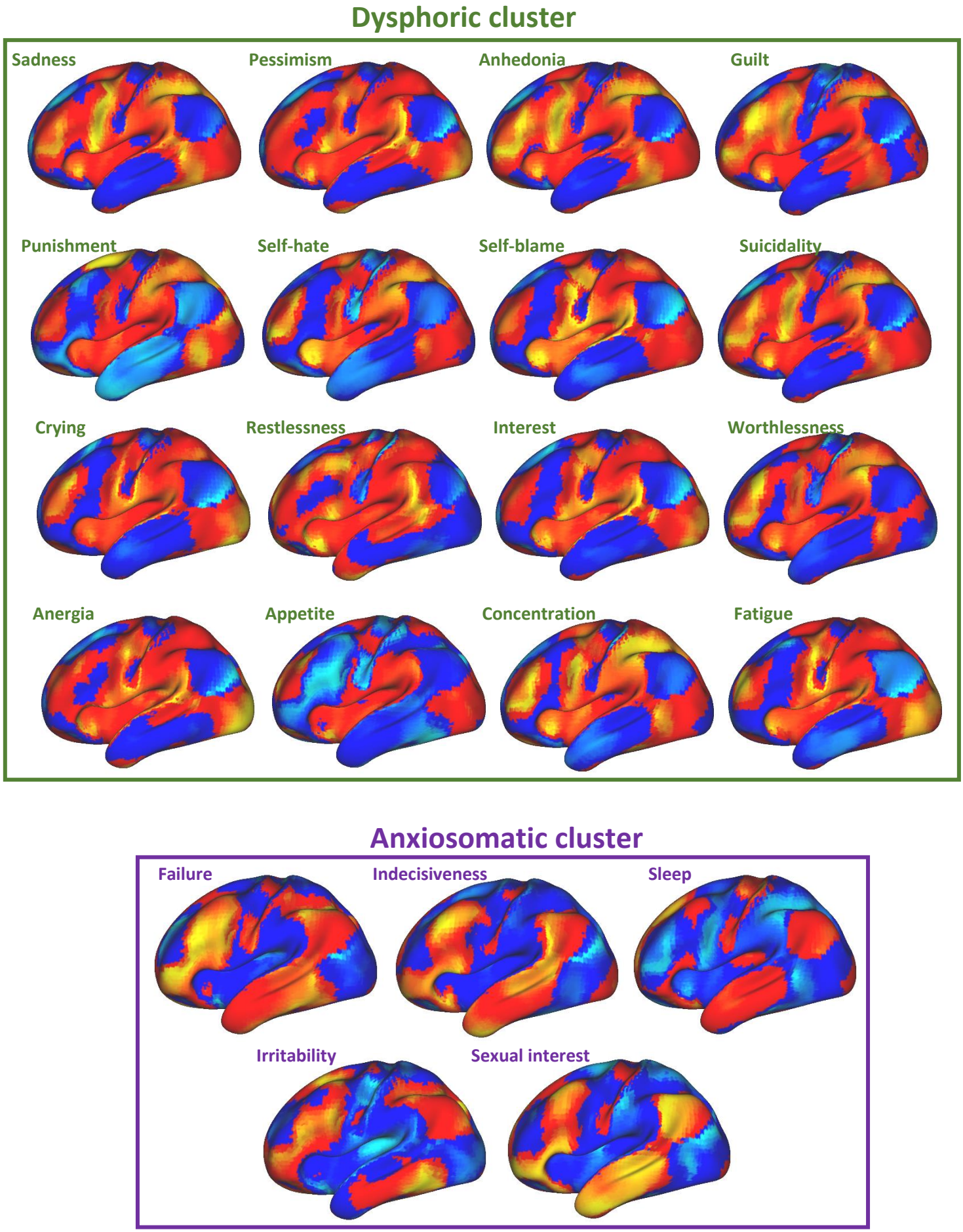
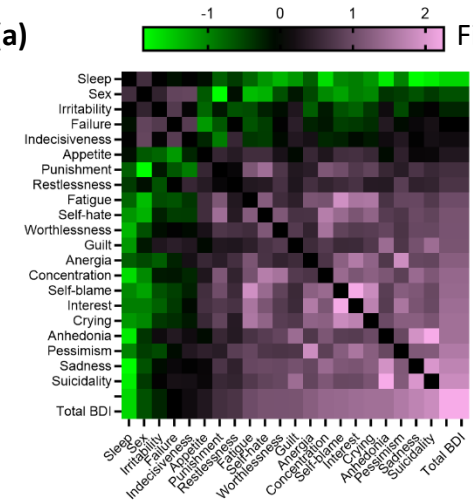
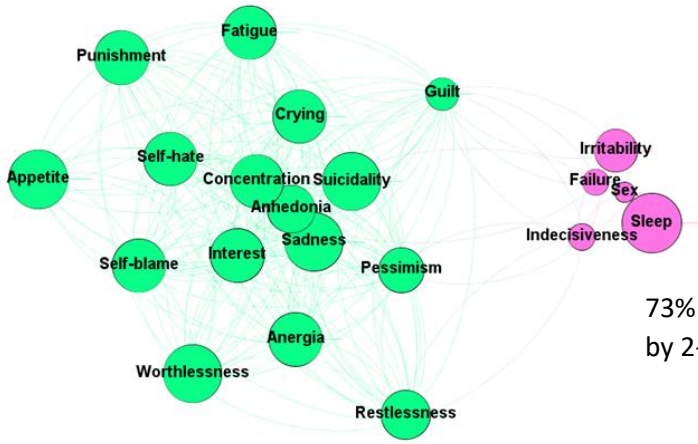


Figure S1: All 21 symptom-response maps derived from the discovery BDI dataset.

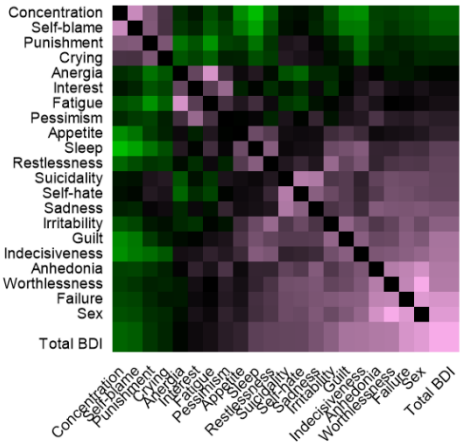
(a) Fisher *z*

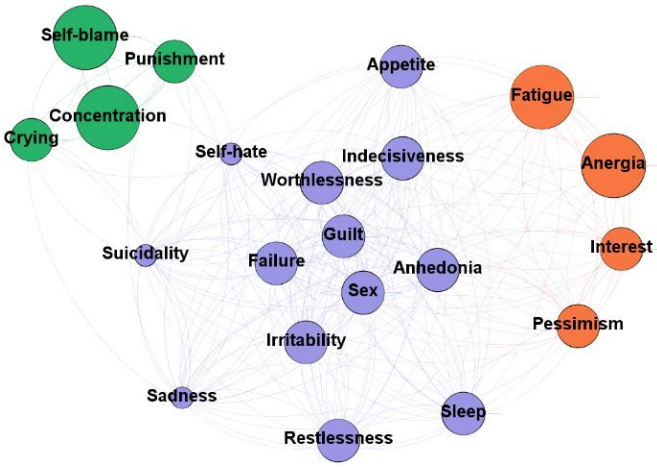




73% variance explained by 2-cluster solution

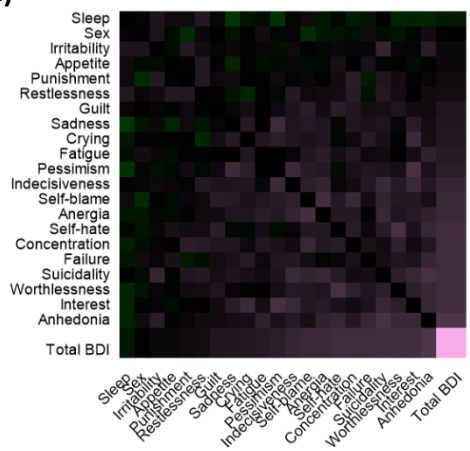
(b) Unpaired t-test: $p = 1.0 \times 10^{-12}$

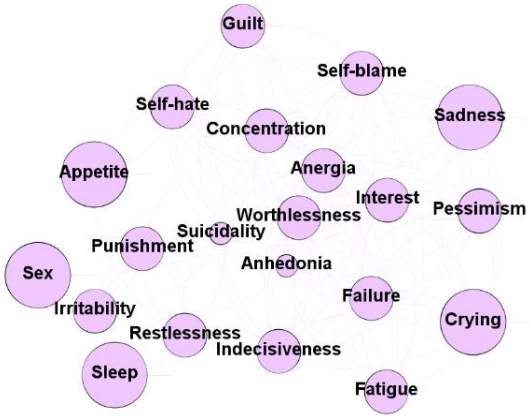




25% variance explained by 2-cluster solution

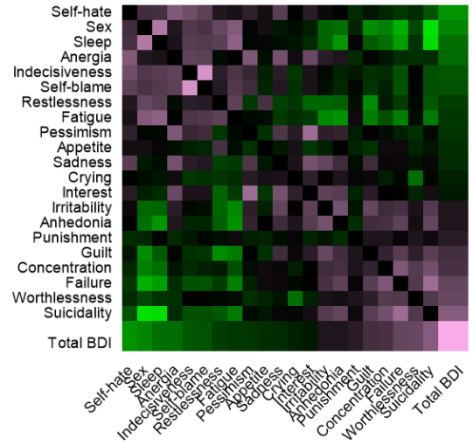
(c)

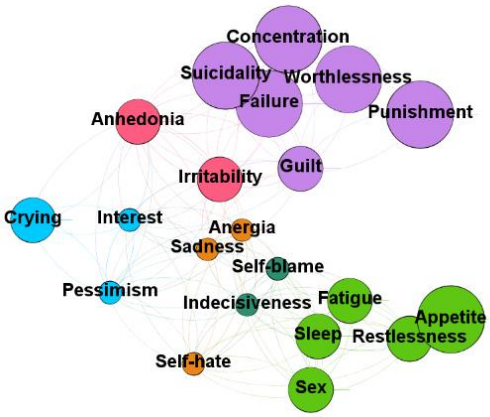




No clustering evident

(d) Permutation test for cluster maps: $p = 0.005$



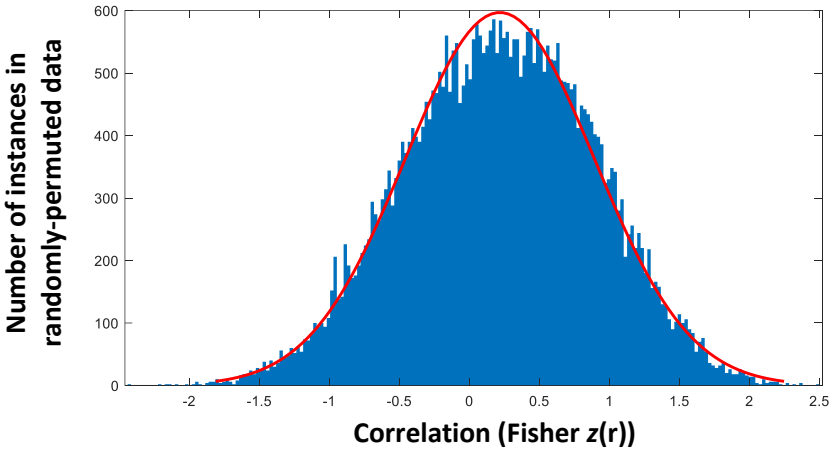


Mean 28% variance explained by 2-cluster solutions ($p = 0.02$)

Mean 40% variance explained by "optimal" solutions ($p = 0.03$)

Figure S2

(a)



(b)

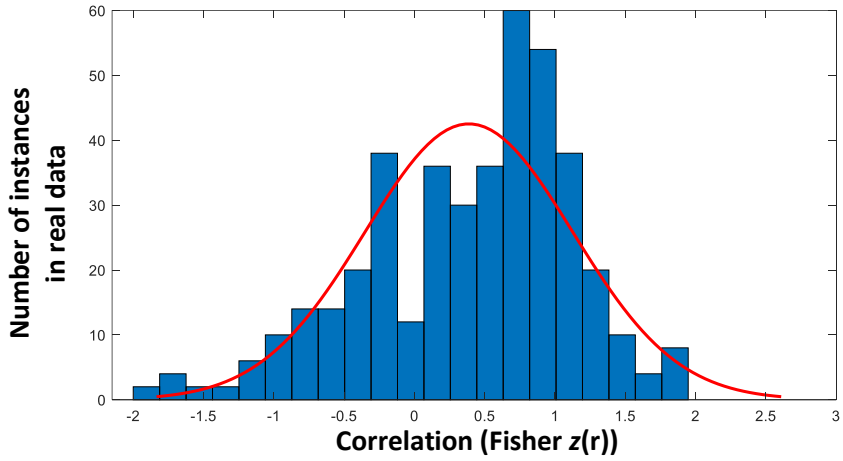


Figure S3: Distributions of cross-correlations between symptom maps. (a) Across 100 permutations, randomly-permuted data showed cross-correlations that followed a normal distribution with a peak near zero. (b) The real data followed a skewed distribution with a trough near zero.

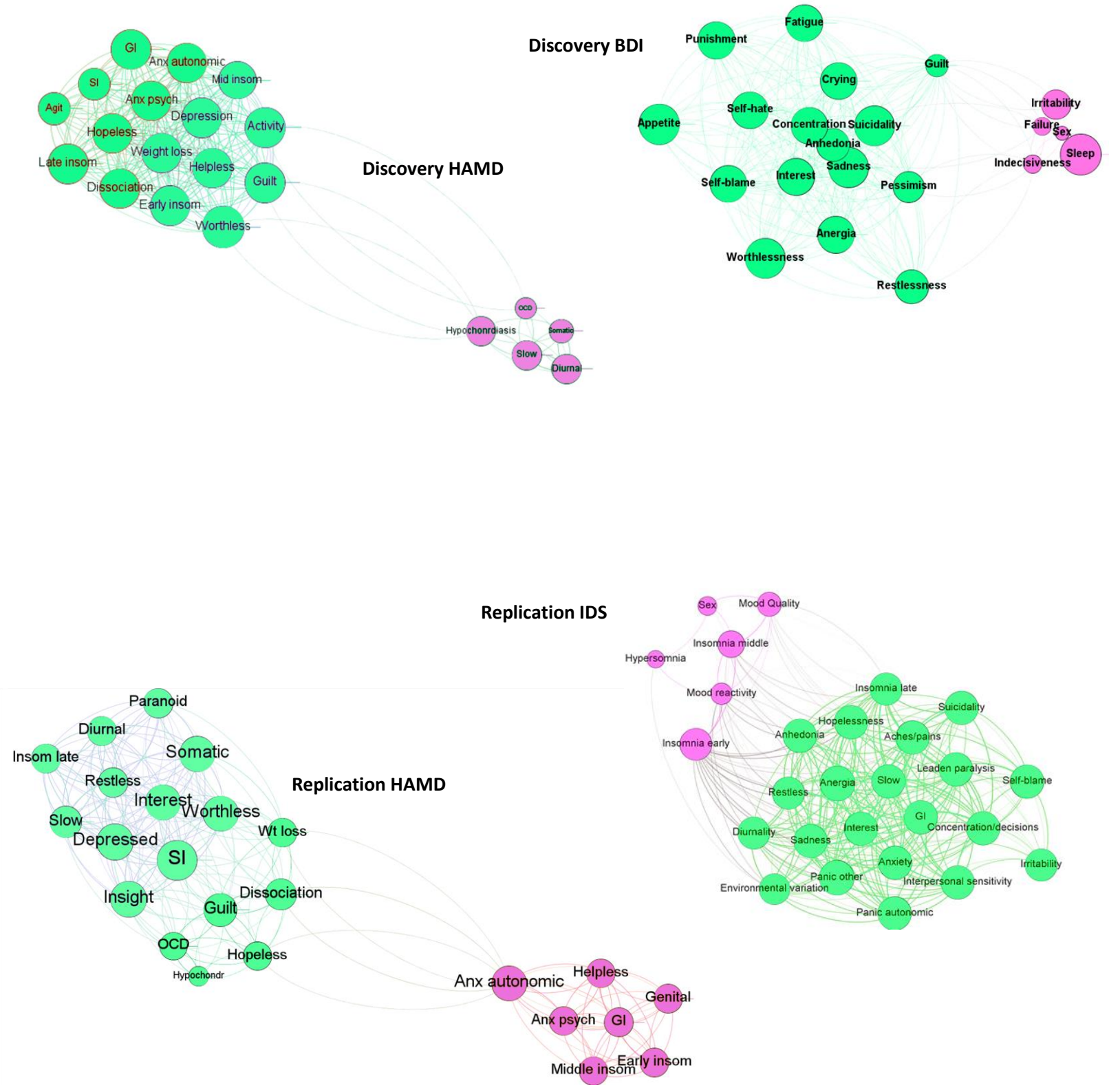
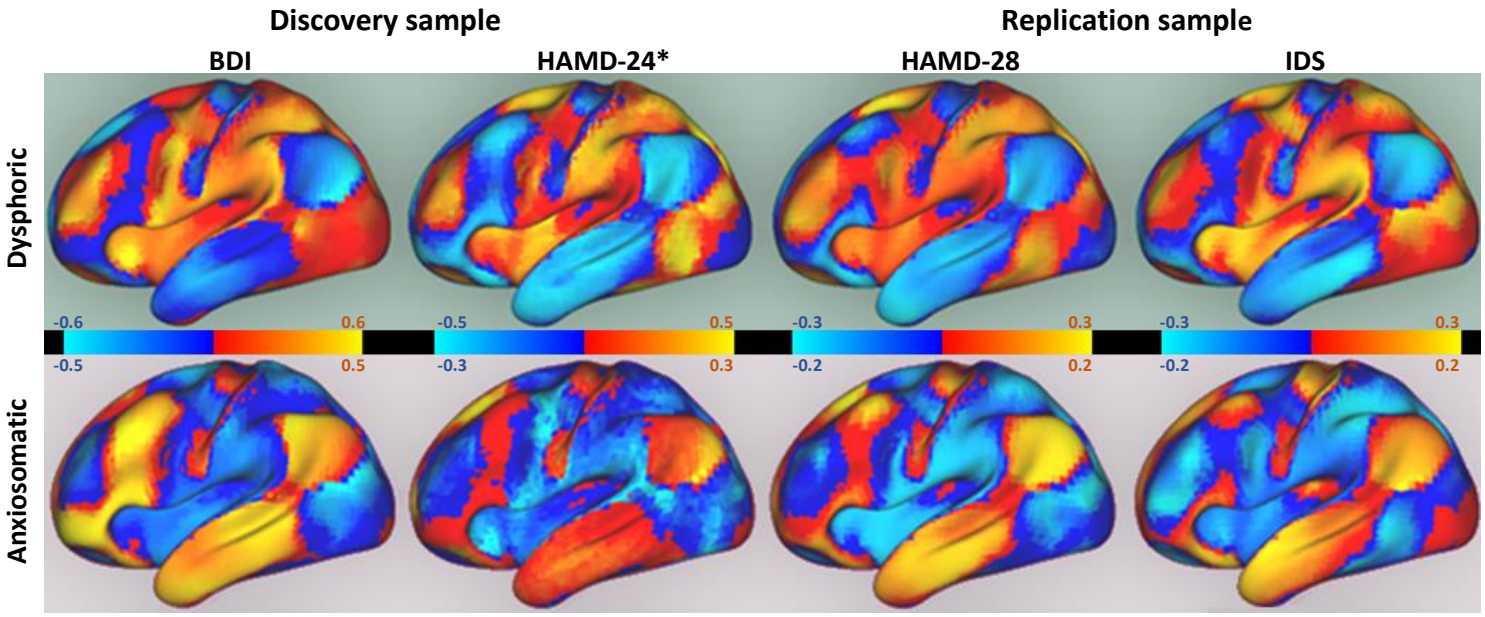


Figure S4: Force-directed graph visualizations depicting the clustering solutions for each dataset. Visualizations follow the same parameters described in Fig. S1.



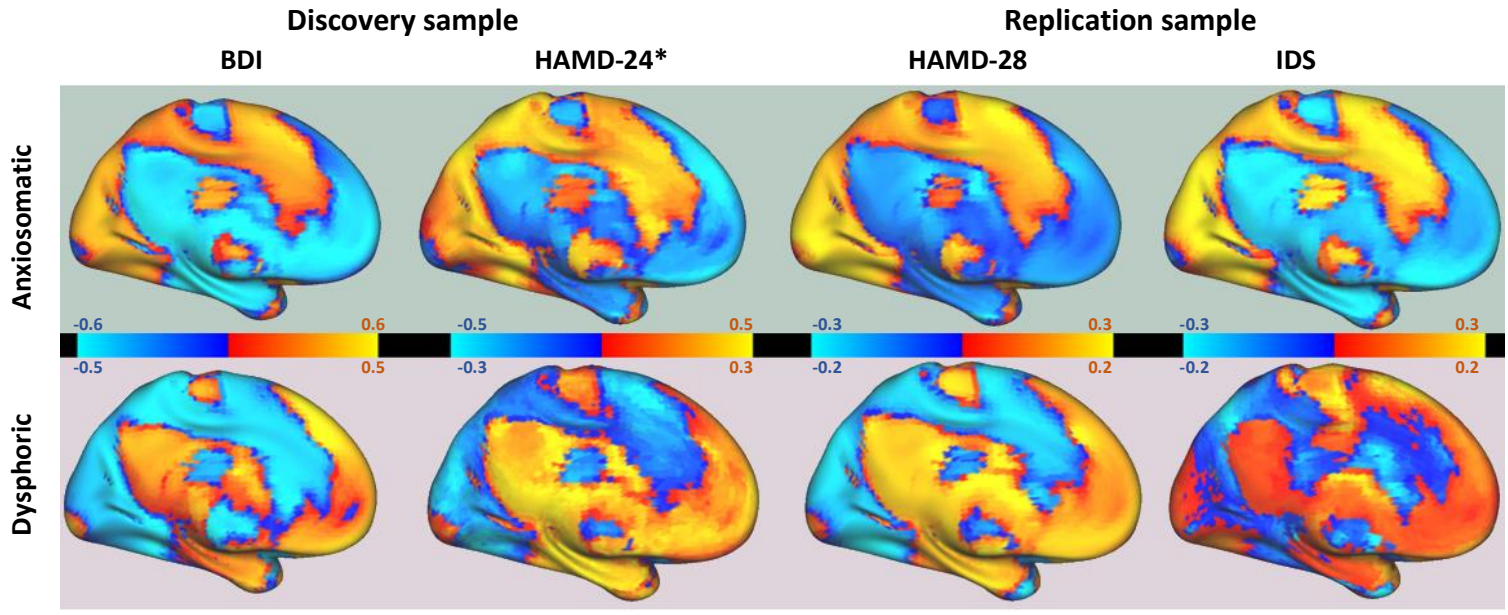


Figure S5: Cluster-response maps across different datasets (top: lateral view, bottom: medial view). Cluster-response maps were reproducible across different symptom scales and independent cohorts.

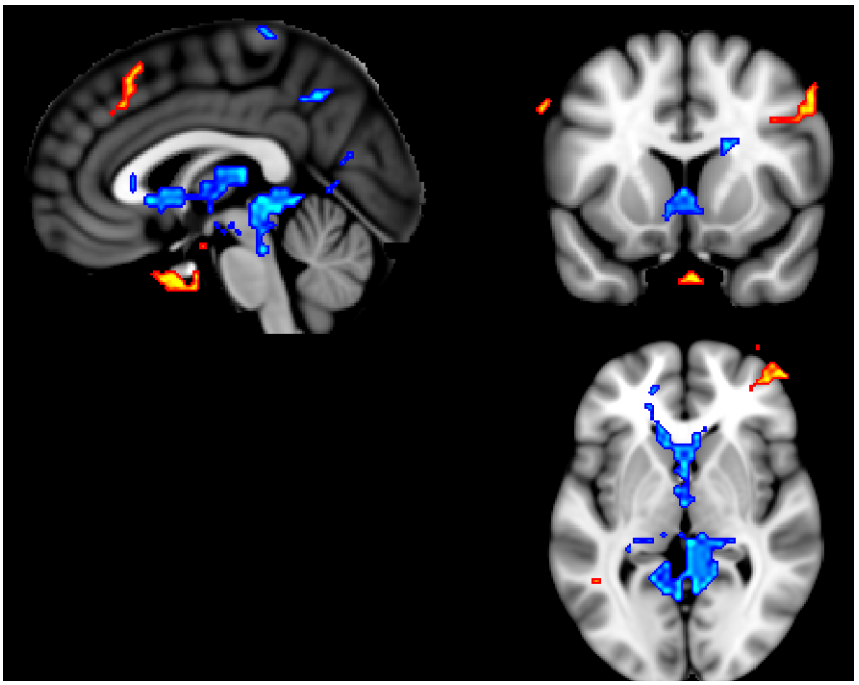
*Three items in the HAMD-24 (discovery sample, secondary analysis) were omitted from the standard clinical assessment due to clinician judgment. These included Item 14 (genital symptoms), item 17 (insight), and Item 20 (paranoia).

Positive peaks

Region	Coordinate
Right orbitofrontal cortex	(17, 48, -12)
Left DLPFC	(-47, 30, 36)
Left anterior insula	(-27, 21, -6)

Negative peaks

Region	Coordinate
Right fusiform gyrus	(22, -36, -18)
Right extrastriate cortex	(52, -60, 12)
Left extrastriate cortex	(-40, -60, 15)
Periaqueductal gray	(0, -32, -6)



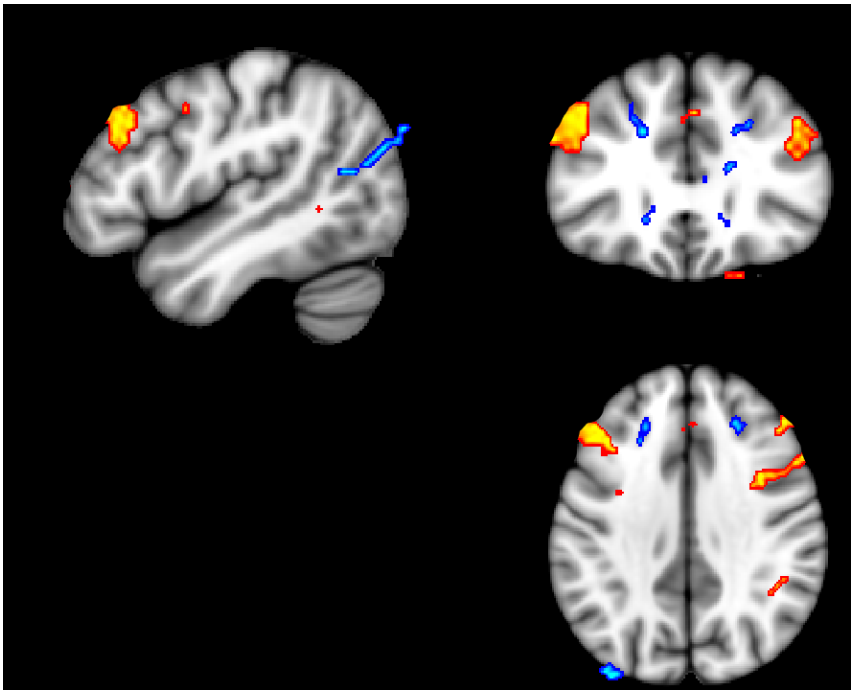


Figure S6: Regions of overlap between the two cluster maps.

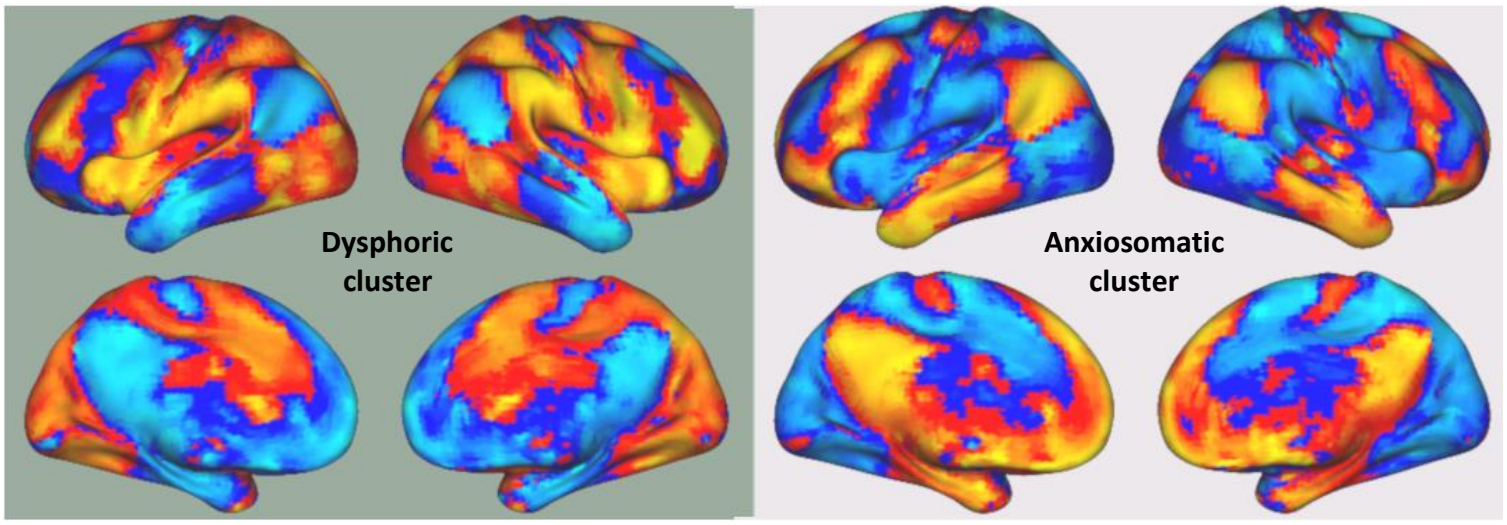


Figure S7: Two-cluster solution generated when using a connectome database of 38 subjects with major depression rather than 1000 healthy controls.

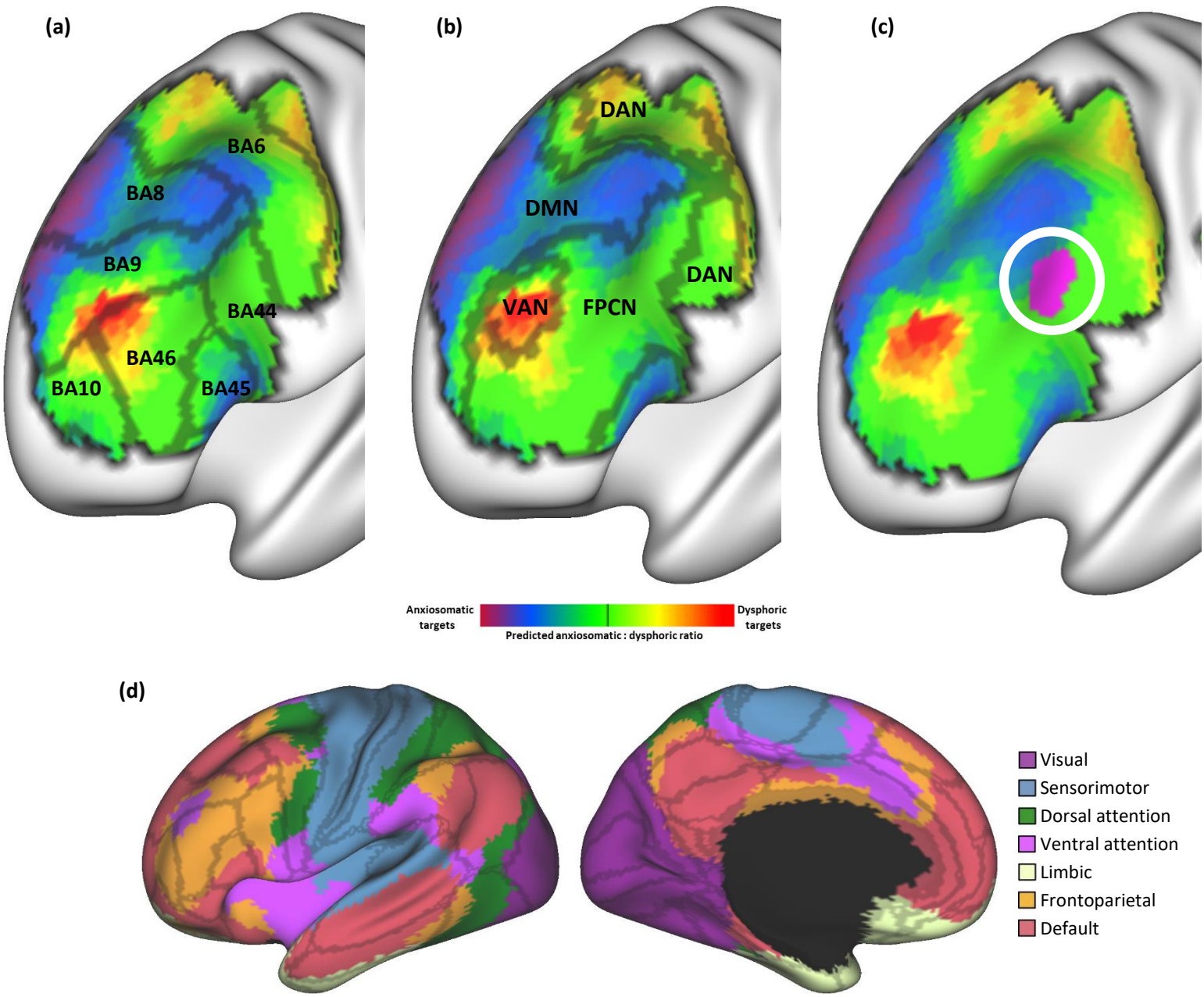


Figure S8: Alignment of optimal targets with consensus cortical parcellation schemes.
(a) Brodmann areas: The dysphoric target lies at the intersection of Brodmann areas 9, 10, and 46. The anxiosomatic targets lie in Brodmann area 8.
(b) Yeo parcels(5)**:** The dysphoric target aligns with the Ventral Attention Network (VAN) parcel, also known as the "cingulo-opercular network" or the "salience network." Other parts of the dysphoric network also align with the dorsal attention network (DAN). The anxiosomatic target aligns with the default mode network (DMN).
(c) Lesion network map of depression(6): A dorsolateral prefrontal site that has been shown to be connected to depression-causing lesions is depicted in magenta. This site was not preferentially connected to either symptom-specific circuit.
(d) Surface projection of Yeo parcellation (colors) and Brodmann parcellation (gray lines) for reference.

Table S1: Dataset characteristics and patient demographics

	Discovery	Replication
Sample size	30	81 active, 87 sham
Treatment device	47% Neuronetics, 53% Magstim	Neuronetics 2100
Setting	Naturalistic	Multi-site trial
Targeting method	"5.5 cm"	"5 cm"
Clinical outcomes	BDI (primary), HAMD	HAMD (primary), IDS
Stimulation site recording procedure	Stim sites recorded using neuronavigation	Stim site marked during MRI
Mean age (range)	53 (24-67)	47 (22-69)
Gender	67% female	57% female
Concomitant antidepressant use	100%	0%
Mean number of concomitant medications	3.0	0

		Dysphoric cluster	Anxiosomatic cluster
Clinical improvement (Mean ± SD)	**Discovery**	47% ± 24%	49% ± 28%
	Replication (active)	14% ± 36%	16% ± 52%*
	Replication (sham)	13% ± 26%**	-4% ± 77%**
Correlation between baseline and change (Spearman rho)	**Discovery**	0.07 (p = 0.75)	0.19 (p = 0.36)
	Replication (active)	0.02 (p = 0.84)	0.11 (p = 0.32)
	Replication (sham)	0.24 (p = 0.03)	0.25 (p = 0.02)

*p<0.05 in comparison with sham (unpaired t-test)
**p<0.05 in comparison with the other cluster (paired t-test)

Table S2: Clustering is not driven by differences in baseline symptoms or overall clinical trajectory of patients in each dataset.

In the discovery dataset and the active arm of the replication dataset, clinical improvement was approximately equal between the two symptom clusters. In the sham dataset, dysphoric symptoms improved significantly more than anxiosomatic symptoms. Anxiosomatic symptom improvement was significantly greater in the active replication dataset than in the sham replication dataset. These results are consistent with the fact that the majority of patients in the replication dataset were stimulated at relatively anxiosomatic stimulation sites.

Clinical change in each cluster was not significantly correlated with baseline severity of that cluster in either the discovery dataset or the active arm of the replication dataset. In the sham dataset, clinical improvement was significantly related to baseline severity in the corresponding symptom cluster.

Discovery cohort		Replication cohort	
1	BDI Sadness	43	IDS Insomnia early
2	BDI Pessimism	44	IDS Insomnia middle
3	BDI Failure	45	IDS Insomnia late
4	BDI Anhedonia	46	IDS Hypersomnia
5	BDI Guilt	47	IDS Sadness
6	BDI Punishment	48	IDS Irritability
7	BDI Self-hate	49	IDS Anxiety
8	BDI Self-blame	50	IDS Mood reactivity
9	BDI Suicidality	51	IDS Diurnality
10	BDI Crying	52	IDS Environmental variation
11	BDI Restlessness	53	IDS Mood Quality
12	BDI Interest	54	IDS Concentration/decisions
13	BDI Indecisiveness	55	IDS Self-blame
14	BDI Worthlessness	56	IDS Hopelessness
15	BDI Anergia	57	IDS Suicidality
16	BDI Sleep	58	IDS Interest
17	BDI Irritability	59	IDS Anergia
18	BDI Appetite	60	IDS Anhedonia
19	BDI Concentration	61	IDS Sex
20	BDI Fatigue	62	IDS Slow
21	BDI Sex	63	IDS Restless
22	HAMD Depression	64	IDS Aches/pains
23	HAMD guilt	65	IDS Panic autonomic
24	HAMD Suicide	66	IDS Panic other
25	HAMD Insomnia early	67	IDS GI
26	HAMD Insomnia middle	68	IDS Interpersonal sensitivity
27	HAMD Insomnia late	69	IDS Leaden paralysis
28	HAMD Activities	70	HAMD Depression
29	HAMD Slowing	71	HAMD Guilt
30	HAMD Restlessness	72	HAMD Suicide
31	HAMD Anxiety psychic	73	HAMD Insomnia early
32	HAMD Anxiety autonomic	74	HAMD Insomnia middle
33	HAMD Somatic GI	75	HAMD Insomnia late
34	HAMD Somatic general	76	HAMD Activities
35	HAMD Hypochondriasis	77	HAMD Slowing
36	HAMD Weight loss	78	HAMD Restlessness
37	HAMD Diurnal	79	HAMD Anxiety psychic
38	HAMD Dissociation	80	HAMD Anxiety autonomic
39	HAMD Obsessionality	81	HAMD Somatic GI
40	HAMD Helplessness	82	HAMD Somatic general
41	HAMD Hopelessness	83	HAMD Genital
42	HAMD Worthlessness	84	HAMD Hypochondriasis
		85	HAMD Weight loss
		86	HAMD Insight
		87	HAMD Diurnality
		88	HAMD Dissociation
		89	HAMD Paranoia
		90	HAMD Obsessionality
		91	HAMD Helplessness
		92	HAMD Hopelessness
		93	HAMD Worthlessness
		94	HAMD Anergia
		95	HAMD Hypersomnia
		96	HAMD Increased appetite
		97	HAMD Rejection sensitivity

Table S3: Index of specific symptoms in figure 3a. Green symptoms fell into the dysphoric cluster, while purple symptoms fell into the anxiosomatic cluster.

	Study	Target	Diagnosis/ population	*n*	Mood Scale	Anxiety Scale
Superficial TMS	Blumberger(7)	Anti-sgACC	MDD	177	HAMD	BSI-A
	Carpenter(8)	Beam F3	PTSD	35	IDS-SR	PSS
	Berlim(9)	EEG F3	MDD	15	HAMD	HAMA
	Taylor(10)	Functional	MDD	16	MADRS	GAD-7
	Leong(11)	Left 6cm	MDD	32	HAMD	GAD-7
	Downar(12)	dmPFC	MDD	47	BDI	BAI
	Dunlop 1(13)	dmPFC	AN/BN	28	BDI	BAI
	Dunlop 2(14)	dmPFC	OCD	20	BDI	BAI
	Yesavage(15)	Left 6cm	MDD	73	HAMD	PCL-M
	Discovery data	Left 5.5cm	MDD	30	Clusters	Clusters
	Replication data	Left 5cm	MDD	81	Clusters	Clusters
	Mansur(16)	Right 5cm	OCD	30	HAMD	HAMA
	Dilkov(17)	Right 5cm	GAD	15	HAMD	HAMA
	Tovar-Perdomo(18)	Beam F3	MDD	24	QIDS-C	BAI
Deep TMS	Levkovitz(19)	Left 5.5cm	MDD	65	HAMD	HAMA
	Tavares(20)	Left 6cm	BPAD	25	HAMD	HAMA
	Berlim(21)	Left 6cm	MDD	17	HAMD	HAMA
	Kaster(22)	Left 5.5cm	Geriatric MDD	27	HAMD	BSI-A
	Isserles(23)	mPFC	PTSD	9	HAMD	CAPS
	Rosenberg*(24, 25)	Left 5.5cm	MDD	8	HAMD	HAMA

Diagnoses:
MDD = Major Depressive Disorder
PTSD = Post-traumatic stress disorder
AN/BN = Anorexia nervosa and bulimia nervosa
OCD = Obsessive-compulsive disorder
GAD = Generalized Anxiety Disorder
BPAD = Bipolar affective disorder (current episode depressed)

Dysphoric scales:
HAMD = Hamilton Rating Scale for Depression
IDS-SR = Inventory of Depressive Symptoms (self-report)
MADRS = Montgomery-Asberg Depression Rating Scale
BDI = Beck Depression Inventory
Clusters = data-driven clustering
QIDS-C = Quick Inventory of Depressive Symptoms (clinician-report)

Anxiosomatic scales:
BSI-A = Brief Symptom Inventory for Anxiety
PSS = Perceived Stress Scale
PCL-M = PTSD Checklist for Military
CAPS = Clinician-Administered PTSD Scale
HAMA = Hamilton Rating Scale for Anxiety
BAI = Beck Anxiety Inventory
GAD-7 = Generalized Anxiety Disorder 7-item Scale

Treatment targets:
Anti-sgACC: MRI neuronavigated coordinate with maximal normative sgACC anti-correlation (Fox et al, 2012)(1)
"5cm": 5cm anterior to motor cotex
"5.5cm": 5.5cm anterior to motor cortex
"6cm": 6cm anterior to motor cortex
EEG F3: F3 coordinate on standard 10-20 EEG system
Beam F3: Scalp-based heuristic to estimate location of F3 (Beam et al, 2009)(3)
dmPFC: Neuronavigated dorsomedial prefrontal coordinate
mPFC: Scalp-based medial prefrontal target

*This dataset included two publications from the same center with the same treatment protocol. Data from individual subjects were reported in both publications. Due to the small sample sizes, the studies were combined into a single dataset. Subjects were included in this analysis if they completed the full 4-week treatment protocol.

Table S4: Details of the studies included in the exploratory meta-analysis.

Supplementary references

1. Fox MD, Buckner RL, White MP, Greicius MD, Pascual-Leone A. Efficacy of transcranial magnetic stimulation targets for depression is related to intrinsic functional connectivity with the subgenual cingulate. Biol Psychiatry. 2012;72(7):595-603.
2. Ghoshal G, Barabasi AL. Ranking stability and super-stable nodes in complex networks. Nat Commun. 2011;2:394.
3. Beam W, Borckardt JJ, Reeves ST, George MS. An efficient and accurate new method for locating the F3 position for prefrontal TMS applications. Brain Stimul. 2009;2(1):50-4.
4. Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2001;63(2):411-23.
5. Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J Neurophysiol. 2011;106(3):1125-65.
6. Padmanabhan JL, Cooke D, Joutsa J, Siddiqi SH, Ferguson M, Darby RR, et al. A human depression circuit derived from focal brain lesions. Biological Psychiatry. 2019.
7. Blumberger DM, Vila-Rodriguez F, Thorpe KE, Feffer K, Noda Y, Giacobbe P, et al. Effectiveness of theta burst versus high-frequency repetitive transcranial magnetic stimulation in patients with depression (THREE-D): a randomised non-inferiority trial. Lancet. 2018;391(10131):1683-92.
8. Carpenter LL, Conelea C, Tyrka AR, Welch ES, Greenberg BD, Price LH, et al. 5Hz Repetitive transcranial magnetic stimulation for posttraumatic stress disorder comorbid with major depressive disorder. J Affect Disord. 2018;235:414-20.
9. Berlim MT, McGirr A, Beaulieu MM, Turecki G. High frequency repetitive transcranial magnetic stimulation as an augmenting strategy in severe treatment-resistant major depression: a prospective 4-week naturalistic trial. J Affect Disord. 2011;130(1-2):312-7.
10. Taylor SF, Ho SS, Abagis T, Angstadt M, Maixner DF, Welsh RC, et al. Changes in brain connectivity during a sham-controlled, transcranial magnetic stimulation trial for depression. J Affect Disord. 2018;232:143-51.
11. Leong K, Chan P, Grabovac A, Wilkins-Ho M, Perri M. Changes in mindfulness following repetitive transcranial magnetic stimulation for mood disorders. Can J Psychiatry. 2013;58(12):687-91.
12. Downar J, Geraci J, Salomons TV, Dunlop K, Wheeler S, McAndrews MP, et al. Anhedonia and reward-circuit connectivity distinguish nonresponders from responders to dorsomedial prefrontal repetitive transcranial magnetic stimulation in major depression. Biol Psychiatry. 2014;76(3):176-85.
13. Dunlop K, Woodside B, Lam E, Olmsted M, Colton P, Giacobbe P, et al. Increases in frontostriatal connectivity are associated with response to dorsomedial repetitive transcranial magnetic stimulation in refractory binge/purge behaviors. Neuroimage Clin. 2015;8:611-8.
14. Dunlop K, Woodside B, Olmsted M, Colton P, Giacobbe P, Downar J. Reductions in Cortico-Striatal Hyperconnectivity Accompany Successful Treatment of Obsessive-Compulsive Disorder with Dorsomedial Prefrontal rTMS. Neuropsychopharmacology. 2016;41(5):1395-403.
15. Yesavage JA, Fairchild JK, Mi Z, Biswas K, Davis-Karim A, Phibbs CS, et al. Effect of Repetitive Transcranial Magnetic Stimulation on Treatment-Resistant Major Depression in US Veterans: A Randomized Clinical Trial. JAMA Psychiatry. 2018.
16. Mansur CG, Myczkowki ML, de Barros Cabral S, Sartorelli Mdo C, Bellini BB, Dias AM, et al. Placebo effect after prefrontal magnetic stimulation in the treatment of resistant obsessive-compulsive disorder: a randomized controlled trial. Int J Neuropsychopharmacol. 2011;14(10):1389-97.
17. Dilkov D, Hawken ER, Kaludiev E, Milev R. Repetitive transcranial magnetic stimulation of the right dorsal lateral prefrontal cortex in the treatment of generalized anxiety disorder: A randomized, double-blind sham controlled clinical trial. Prog Neuropsychopharmacol Biol Psychiatry. 2017;78:61-5.

18. Tovar-Perdomo S, McGirr A, Van den Eynde F, Rodrigues Dos Santos N, Berlim MT. High frequency repetitive transcranial magnetic stimulation treatment for major depression: Dissociated effects on psychopathology and neurocognition. J Affect Disord. 2017;217:112-7.
19. Levkovitz Y, Harel EV, Roth Y, Braw Y, Most D, Katz LN, et al. Deep transcranial magnetic stimulation over the prefrontal cortex: evaluation of antidepressant and cognitive effects in depressive patients. Brain Stimul. 2009;2(4):188-200.
20. Tavares DF, Myczkowski ML, Alberto RL, Valiengo L, Rios RM, Gordon P, et al. Treatment of Bipolar Depression with Deep TMS: Results from a Double-Blind, Randomized, Parallel Group, Sham-Controlled Clinical Trial. Neuropsychopharmacology. 2017;42(13):2593-601.
21. Berlim MT, Van den Eynde F, Tovar-Perdomo S, Chachamovich E, Zangen A, Turecki G. Augmenting antidepressants with deep transcranial magnetic stimulation (DTMS) in treatment-resistant major depression. World J Biol Psychiatry. 2014;15(7):570-8.
22. Kaster TS, Daskalakis ZJ, Noda Y, Knyahnytska Y, Downar J, Rajji TK, et al. Efficacy, tolerability, and cognitive effects of deep transcranial magnetic stimulation for late-life depression: a prospective randomized controlled trial. Neuropsychopharmacology. 2018;43(11):2231-8.
23. Isserles M, Shalev AY, Roth Y, Peri T, Kutz I, Zlotnick E, et al. Effectiveness of deep transcranial magnetic stimulation combined with a brief exposure procedure in post-traumatic stress disorder--a pilot study. Brain Stimul. 2013;6(3):377-83.
24. Rosenberg O, Shoenfeld N, Zangen A, Kotler M, Dannon PN. Deep TMS in a resistant major depressive disorder: a brief report. Depress Anxiety. 2010;27(5):465-9.
25. Rosenberg O, Zangen A, Stryjer R, Kotler M, Dannon PN. Response to deep TMS in depressive patients with previous electroconvulsive treatment. Brain Stimul. 2010;3(4):211-7.