

Supplemental Methods

Validation of the Weight-Map Approach to Imputing subgenual anterior cingulate cortex

(sgACC) Signals

Subjects: Midnight Scan Club (MSC) Sample. The Midnight Scan Club dataset (1) was obtained from OpenNeuro.org (<https://openneuro.org/datasets/ds000224/versions/1.0.1>). This dataset includes 5 hours of (preprocessed, denoised, and surface registered) single-echo resting-state functional magnetic resonance imaging (fMRI) data (10 x 30-minute scans acquired over two months) collected from ten participants aged 24-34 years (mean age = 29.1 ± 3.3 years, 5F/5M). The data was analyzed “as is” (no additional preprocessing or denoising was performed).

Subjects: Weill Cornell Multi-Echo (ME) Sample. Two participants (“ME01”, 29-year-old male; “ME02”, 38-year-old male) in this investigation underwent 44 x 14.5-minute (10.5 hours total) and 64 x 14.5-minute (15.5 hours total) multi-echo fMRI scans, respectively. Three other participants (“ME03”, 24-year-old male; “ME04”, 31-year-old male; “ME05”, male in his mid-fifties) underwent 12 x 14.5-minute (2.9 hours) multi-echo fMRI scans. Participants ME01 and ME02 were study authors C.J.L., and J.D.P., respectively. The details of the acquisition parameters, preprocessing and denoising steps performed on the Weill Cornell Multi-echo dataset can be found in (2). For the analysis reported here, optimally combined multi-echo images were denoised using multi-echo independent component analysis followed by mean grey matter time-series regression.

Data Analysis: Subgenual Anterior Cingulate Cortex (sgACC) Functional Connectivity (FC) via Weight-Map Method. We validated a previously described method to measure sgACC FC in individual patients (3, 4). This method was previously referred to as the “seed

map” method. However, since this might be easily confused with simple seed-based FC, we adopted the term “weight-map”. To explain concisely, we first constructed an average sgACC FC map in 1200 Human Connectome Project (HCP) subjects (‘HCP1200 mask’), using a 10-mm diameter spherical sgACC seed, as used in prior reports (5). The resulting Pearson r values—both positive and negative—of this average FC map were then used as weights. To impute a subject’s sgACC time course, all their cortical node’s time courses were first weighted by the corresponding node’s r values in the HCP group average map. Strongly positive r values will amplify a node’s time course, strongly negative r values will invert and amplify, and r values around 0 will muffle a node’s time course. The mean of all weighted time courses was then used as a regressor for a brain-wide FC analysis. By doing so the sgACC’s time course is being substituted by a weighted average of all time courses commonly correlated with the sgACC, most of which have better signal quality than the sgACC itself (see Figure S1 for a graphical explanation). This approach has been used in previous reports (6–9), and in a recent study involving a large cohort of subjects, it was shown to have markedly improved test-retest reliability at two time-points relative to using a conventional sgACC seed (4). The dorsolateral prefrontal cortex (DLPFC) is masked out during the averaging procedure.

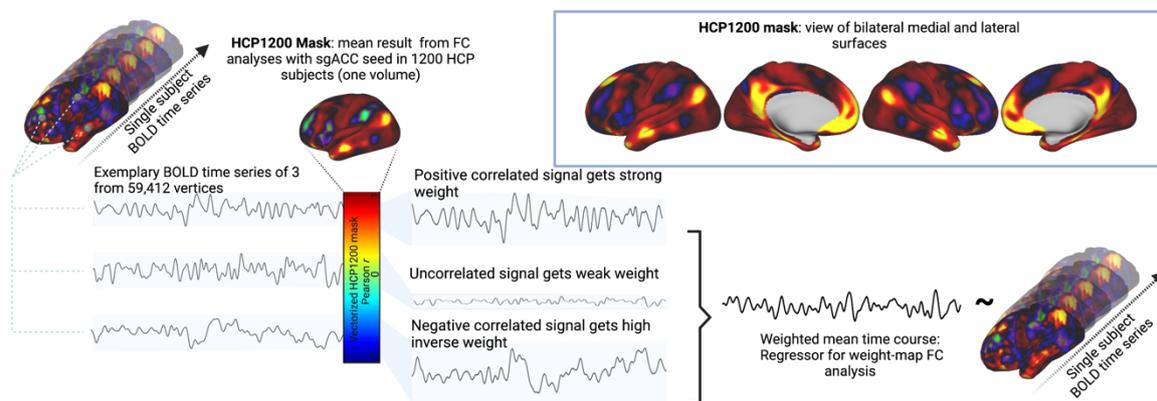


Figure S1. The weight-map method. First, vertex-wise time courses are extracted from a subject’s four-dimensional BOLD time series. Here, three exemplary BOLD time series are depicted. These time series are weighted by an functional connectivity (FC) map that serves as a prior: a group level FC result map from a brain-wide subgenual anterior cingulate cortex (sgACC) FC analysis in 1200 Human Connectome Project (HCP) subjects, here termed the ‘HCP1200 mask’ (see inlay for full view). BOLD-time courses corresponding to vertices that showed strong FC to the sgACC in the HCP1200 mask, get weighed stronger than BOLD time courses corresponding to vertices with low FC in the HCP1200 mask. For illustrative purposes, a schematic vectorized HCP1200 weighting mask is shown and the effect that low vs. strong positive and negative weights have on the weighted BOLD time series is schematically

depicted. Finally, the mean of the weighted BOLD time series is calculated and used as a regressor in a vertex-wise FC analysis in the same single subject's four-dimensional data set. Not depicted here is that for each subject vertices within the left DLPFC were replaced by NaN values before vertex-wise BOLD time courses were extracted, to not overly bias resulting FC values by HCP1200 mask values in this area.

Data Analysis: Reliability and Accuracy of Weight-Map Method. Test-retest reliability of FC within the DLPFC derived from either a simple sgACC seed or a sgACC weight-map was analyzed in the MSC single-echo data. Test-retest reliability was operationalized as the mean of all Pearson r values derived from spatial correlations between all pairs of the 10 runs for each subject (upper triangle of a 10 x 10 matrix with each entry representing one run). The accuracy with which the weight-map approach approximates 'true' sgACC FC was investigated in the ME data. This was possible because the sgACC signal quality in ME data is sufficiently high to yield reliable FC maps even when single nodes within the sgACC are used as seeds (2). Accuracy of sgACC weight-map functional connectivity was operationalized as the spatial correlation between the resulting FC maps from the simple sgACC vs. weight-map method in each subject's full concatenated dataset.

Modeling of Stimulation Sites in the DLPFC

Electric Field Modeling. A high-resolution (0.50 vertices per mm²) tetrahedral headmesh was created from the average T1-weighted anatomical image using headreco (10). Electric field (E-field) modeling was then performed with the MagVenture B70 coil file in SimNIBS (11). The coil center position was defined as the point on the scalp directly above the stimulation site (which was identified by converting the stimulation target in MNI152 stereotaxic space to native volume space using FSL's `img2imgcoord` tool). The coil was oriented posteriorly in the direction of the CP1 landmark from the 10-20 EEG system. The Norm of the E-field (a measure of E-field strength at a given cortical node, see [FAQ — SimNIBS 3.2.6 documentation](#) for a detailed explanation) was mapped to the midthickness surface using Connectome Workbench command-line utilities. Importantly, the strength of the E-field varies linearly with dI/dt (the

speed of variation of the current throughout the coil) and, for this reason, has no effect on its spatial distribution, i.e. the E-field distribution is the same irrespective of the stimulation strength. As it doesn't affect the E-field distribution, a fixed intensity (the default value in SimNIBS of $dI/dt = 1 \text{ A}/\mu\text{s}$) was used for all the simulations performed in our paper.

Weighted Cone. For comparison, a generic E-field approximation termed “weighted cone” was used that has been described in detail in (3) and was used in most prior reports (3, 6–9, 12). This method consists of weighting the signal of all nodes within a specified radius (12 mm) for proximity to the stimulation coordinates (linear weight) and then taking the weighted average within this radius (or hemisphere in volumetric space). Importantly, although the analyses here were all performed on surface space, depth information of nodes due to where they were located along a gyrus or sulcus was incorporated into the weighting scheme.

Modeling Clinical Outcome in THREE-D Patients as a Function of the Stimulation Site's FC with the sgACC (sgACC-StimFC):

THREE-D Sample. The THREE-D study was a randomized non-inferiority trial comparing the efficacy of 10 Hz repetitive transcranial magnetic stimulation (rTMS) versus intermittent theta-burst stimulation (iTBS) in 414 patients with major depression, enrolled between 2013 and 2016. THREE-D Study procedures are detailed elsewhere (13). Briefly, patients were eligible for inclusion if they met the following criteria: HDRS-17 score of at least 18, non-response to at least one antidepressant (based on a standardized assessment of medication history) or unable to tolerate at least two separate trials of antidepressants, stable antidepressant regimen for at least 4 weeks before treatment, which continued during treatment. Patients were excluded from participation based on the following criteria: current diagnosis of substance abuse or dependence, active suicidal intent, pregnancy, diagnosis of bipolar disorder or any psychotic disorder, previous rTMS treatment, history of ECT non-response,

personality disorder as primary pathology, unstable medical or substantial neurological illness, abnormal serology, presence of a cardiac pacemaker, intracranial implant, or metal in the cranium or more than three failed adequate antidepressant trials. In contrast to the analysis conducted in (13), intake of anticonvulsants or benzodiazepines at any dose was not treated as an exclusion criterion. Instead, these variables were included as covariates (see below). 250 patients (85%) included in this report were receiving any concurrent pharmacotherapy during treatment. A breakdown of concurrent medication per treatment group is listed in Table 1. All participants were actively depressed with at least moderately severe symptoms at the start of the trial (mean Quick Inventory of Depression Symptomatology Self Report [QIDS-SR] score = 17.1, mean Hamilton Depression Rating Scale-17 [HDRS-17] = 23.5).

The analyses in the present study included all participants whose data met the following additional criteria: two complete fMRI scans—before and after treatment—available (312 participants), complete clinical outcome data for the QIDS-SR or the HDRS-17, and successful cortical surface reconstruction (see below). 295 patients met these criteria. The analyses here contain subjects from the two major sites (CAMH, Toronto, and UHN, Toronto) that were both scanned on the same physical scanner and protocol (see details below). Forty subjects were scanned on a third site on a different scanner model and sequence and were not included here.

Additional demographic characteristics, as well as disease- and treatment-related features of the sample of 295 patients included in our analyses, are detailed in Table 1.

Clinical Outcome Measures. Details on outcome measures and sampling procedures within the THREE-D trial can be found elsewhere (13). Our variable of interest was the percent improvement in depressive symptoms after completion of the final TMS treatment session as defined in THREE-D. Depression severity was assessed with the QIDS-SR and the HDRS-17. Although the HDRS-17 was the primary outcome measure in the THREE-D study, for the results reported in the main text, we focused on the QIDS-SR, which we favored because of

1) more complete end-point data (n=295 vs n=293); 2) a self-rating questionnaire may have benefits in a scenario where raters were aware a treatment was delivered (although they were blinded to condition); and 3) self-report scales have their own limitations but they do not suffer from biases related to interrater reliability, which can be a significant issue for clinician-rated scales. However, we also report effects for HDRS-17 in the result section of the online supplement.

rTMS Treatment. Details on TMS treatment protocols can be found in (13). In short, patients in the THREE-D trial were randomly allocated (1:1) to receive either repeated 10 Hz rTMS or iTBS to the left DLPFC. Both treatments were delivered with a MagPro X100 or R30 stimulator, equipped with a Cool-B70 fluid cooled coil and high-performance cooler (MagVenture, Farum, Denmark). For 10 Hz rTMS conventional FDA cleared parameters were used (120% resting motor threshold (RMT) stimulation intensity; 10 Hz frequency; 4 s on and 26 s off; 3000 pulses per session; total duration of 37.5 min). iTBS was delivered at the same site and intensity (120% RMT), differing only in stimulation pattern and a total number of pulses (triplet 50 Hz bursts, repeated at 5 Hz; 2 s on and 8 s off; 600 pulses per session; total duration of 3 min 9 s). Initial treatment comprised 20 sessions in total, which consisted of once daily sessions (on weekdays; ie, five sessions a week).

rTMS Treatment Target. Treatment targeting was based on subject-specific T1w structural scans and a predefined stereotactic coordinate ($x = -38$, $y = +44$, $z = +26$) in MNI space, which led to significant scattering of individual target coordinates due to back transformation to each person's native brain space (Figure S2). Coil positioning over the obtained coordinates was ensured using real-time MRI-guided neuronavigation with a Visor neuronavigation system (ANT Neuro, Enschede, Netherlands). These coordinates were based on a prior report where these target coordinates exhibited peak anticorrelated sgACC FC at the group level (9). Target coordinates for all patients in native space are depicted in

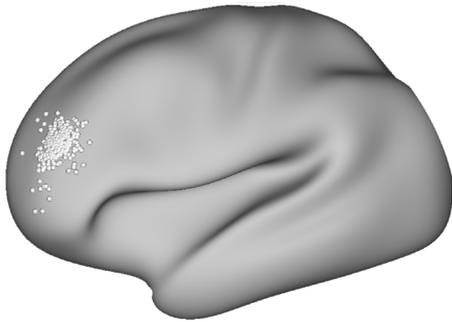


Figure S2. Distribution of target sites across the dorsolateral prefrontal cortex (DLPFC). White foci represent each subject's recorded target site in native space. For illustration purposes, these target sites are projected onto an inflated average surface rendering of all 295 patients.

MRI Data Acquisition. Imaging data in the THREE-D sample were acquired on a 3T GE HDx MRI system equipped with an 8-channel phased-array head coil. Patients were instructed to keep still and stay awake, with their eyes closed and to avoid thinking about anything in particular. For each participant, two datasets were acquired: a high resolution, T1-weighted fast spoiled gradient-echo structural dataset (TE = 12 ms, TI = 300 ms, flip angle = 20°, 116 sagittal slices, thickness = 1.5 mm, no gap, 256 x 256 matrix, FOV = 240 mm), and a 10-minute resting-state functional MRI dataset (T2*-weighted EPI, TE = 30 ms, TR = 2000 ms, flip angle = 85°, 32 axial slices, thickness = 5 mm, no gap, 64x64 matrix, FOV = 220mm).

MRI Preprocessing: Cortical Surface Generation. Anatomical data were preprocessed, and cortical surfaces generated using the “legacy” mode of the Human Connectome Project (HCP) PreFreeSurfer, FreeSurfer, and PostFreeSurfer pipelines (version 4.3).

MRI Preprocessing: Functional Data Preprocessing and Denoising of THREE-D Data. *Preprocessing of functional data with minimized spatial interpolation and volumetric smoothing.* The first volume of each scan was aligned, averaged, co-registered to the ACPC aligned T1-weighted anatomical image, and simultaneously corrected for spatial distortions

using FSL's topup and epi_reg programs. FreeSurfer's bbregister algorithm (14) was used to refine this co-registration. A unique 6 DOF registration (one per volume) to the target image was estimated at each time point of each scan using FSL's MCFLIRT tool (15), using a 4-stage (sinc) optimization. All of these steps (co-registration to the average target image, ACPC alignment, and correcting for spatial distortions) were concatenated using FSL's convertwarp tool and applied as a single spline warp after correcting for slice time differences using FSL's slicetimer program. The functional images underwent a brain extraction using the co-registered, skull-stripped T1-weighted anatomical image as a mask. Functional images were also corrected for signal intensity inhomogeneities using ANT's N4BiasFieldCorrection tool. All denoising was performed on these preprocessed, ACPC-aligned images.

Functional data denoising. Preprocessed functional data were denoised using ICA-AROMA (16), and mean gray matter signal regression was performed to remove spatially diffuse noise. Temporal masks were generated for censoring high motion time-points using a framewise displacement (FD) threshold of 0.3 mm.

Surface processing and CIFTI generation of fMRI data. The denoised fMRI time-series was mapped to the midthickness surfaces (using the "-ribbon-constrained" method), combined into the Connectivity Informatics Technology Initiative (CIFTI) format, and spatially smoothed with geodesic (for surface data) and Euclidean (for volumetric data) Gaussian kernels ($\sigma = 2.55$ mm) using Connectome Workbench command line utilities (17). Signals were normalized (z-scored). This yielded time courses representative of the entire cortical surface, subcortex (accumbens, amygdala, caudate, hippocampus, pallidum, putamen, and thalamus), and cerebellum, but excluding non-gray matter tissue.

Data Analysis: Statistics. All statistical analyses were performed in MATLAB R2020b/R2021a (MathWorks, Inc., Natick, Massachusetts, United States). Functional connectivity between each patient's rTMS target site and the sgACC (sgACC-StimFC) was calculated based on the above-described weight-map approach. FC features corresponding to the rTMS target coordinates were averaged using two approaches: either within the patch

of cortex receiving maximum percentile E-field strength or within a 'weighted cone' of a given radius. Data reported in the main manuscript were obtained from a weighted cone with a radius of 12 mm or an E-field threshold of maximum 99th percentile, respectively. However, all analyses were repeated across a range of radius thresholds (2, 4, 7, 9, 12 mm) and percentile thresholds (99.9th, 99.5th, 99th, 98.5th, 98th), see Figure S5. We repeated the above-described correlational analyses using multiple linear regression, with treatment condition, age, gender, and use of benzodiazepines or anticonvulsants as covariates. This resulted in a set of four linear models with either percent improvement in QIDS-SR or HDRS-17 as the outcome variable, and either sgACC-StimFC features from the weighted cone method or from E-field modeling as the predictor variable. Reported are the F-statistic and p-value of the full model vs. the degenerate model and the t-statistic and p-value of the predictor variable's beta-estimate.

The effect of sample size on effect-size variability was estimated using bootstrapped subsamples of different sample sizes (5, 15, 25, 50, 85, 130, 280). For each sample size, 10,000 sub-samples were drawn uniformly at random without replacement from the full dataset to correlate sgACC-StimFC strength and QIDS-SR percent improvement, using the exact same methods as for the results shown in Figure 2C.

The spatial distance between the target site and the most strongly anticorrelated site ('Optimal Site') in the DLPFC was derived as follows: A 12 mm weighted-cone was created across all nodes within the DLPFC and the center coordinate with the resulting most negative sgACC-StimFC was recorded, and its Euclidean distance to the target site derived. This distance in mm was then correlated with clinical improvement.

Data Analysis: Whole-Brain Functional Connectivity. We extended our analyses to the entire connectome using a well-validated multi-modal parcellation of the human brain (5). Applying the same weight-map approach as for the analysis presented in Figure 2 and described above, we computed weight-map-based FC maps for each parcel of the Glasser

brain parcellation. For each parcel, this involved first deriving an average FC map in the HCP1200 data using that parcel as a seed. The resulting node-wise Pearson r values were then applied as weights to a subject's fMRI time-course. The weighted average time course was then used as a regressor to obtain a brain-wide FC map. As for the sgACC-StimFC analysis, we operationalized each parcel's functional connectivity with the stimulated cortex in the DLPFC as the mean of all FC features within the surface area corresponding to the distribution of the 99th percentile strongest E-field. Using a prior established network assignment based on a large data set of healthy individuals and a clustering algorithm that was calibrated upon its performance in well-delineated sensory-motor networks (18), we illustrated the correlation between the E-field's FC to each parcel with clinical improvement, per network.

Investigating the Impact of Signal Properties on sgACC-StimFC

Stratification of THREE-D by Elementary Data Properties. We investigated the impact of data quality on the association between sgACC-StimFC and clinical improvement using two metrics that have been shown to influence fMRI findings (19): Frame-wise displacement (FD) as a measure of movement and temporal signal-to-noise-ratio (tSNR) as a measure of signal variance. For each measure, all subjects were first ranked by the mean value of both rsfMRI scans. Then sgACC-StimFC was correlated with clinical improvement in sequential overlapping samples of 50 subjects, from the subject with the lowest to the highest mean value. Departure from chance level was tested against the 95% range of the null distribution of Pearson r values derived from 10,000 bootstrapped sub-samples (uniformly drawn at random, without replacement). This allowed us to assess the impact of tSNR and FD respectively on our main finding.

Blind Rating of fMRI Signals According to Suspected Breathing Patterns. Three authors (J.D.P., I.G.E and C.J.L.) independently rated all 590 scans on an ordinal scale for certain presence (2), likely presence (1) or absence (0) of burst-like fMRI signal patterns. From this, an ordinal score was created for each subject, ranging from 0, all raters saw no evidence for bursting in both scans to 12, all raters saw strong evidence for bursting in both scans of a subject. For the analyses presented in Figure 4 of the main manuscript, a cut-off value of 6 was used (e.g. three raters voted at least 1 on both of a subject's scans); however, results are presented for the entire range of cut-off values in Figure S10. For the results presented in Figure S12 the rating scores were used in a continuous manner: 100,000 boot-strapped samples of 50 subjects across all percentile ranges of burst scores were randomly drawn. For each draw, the effect size (Pearson correlation between sgACC-StimFC and % QIDS-SR improvement) and mean bursting score were recorded, and the resulting bursting-scores were finally plotted against effect sizes. The likelihood of obtaining sex differences in each breathing pattern was determined by Chi-squared tests in each rater, and all three raters obtained sex differences for bursts but not for deep breaths, concordant with a prior report (20).

Supplemental Results

Validation of Imputed sgACC Signal via Weight Maps

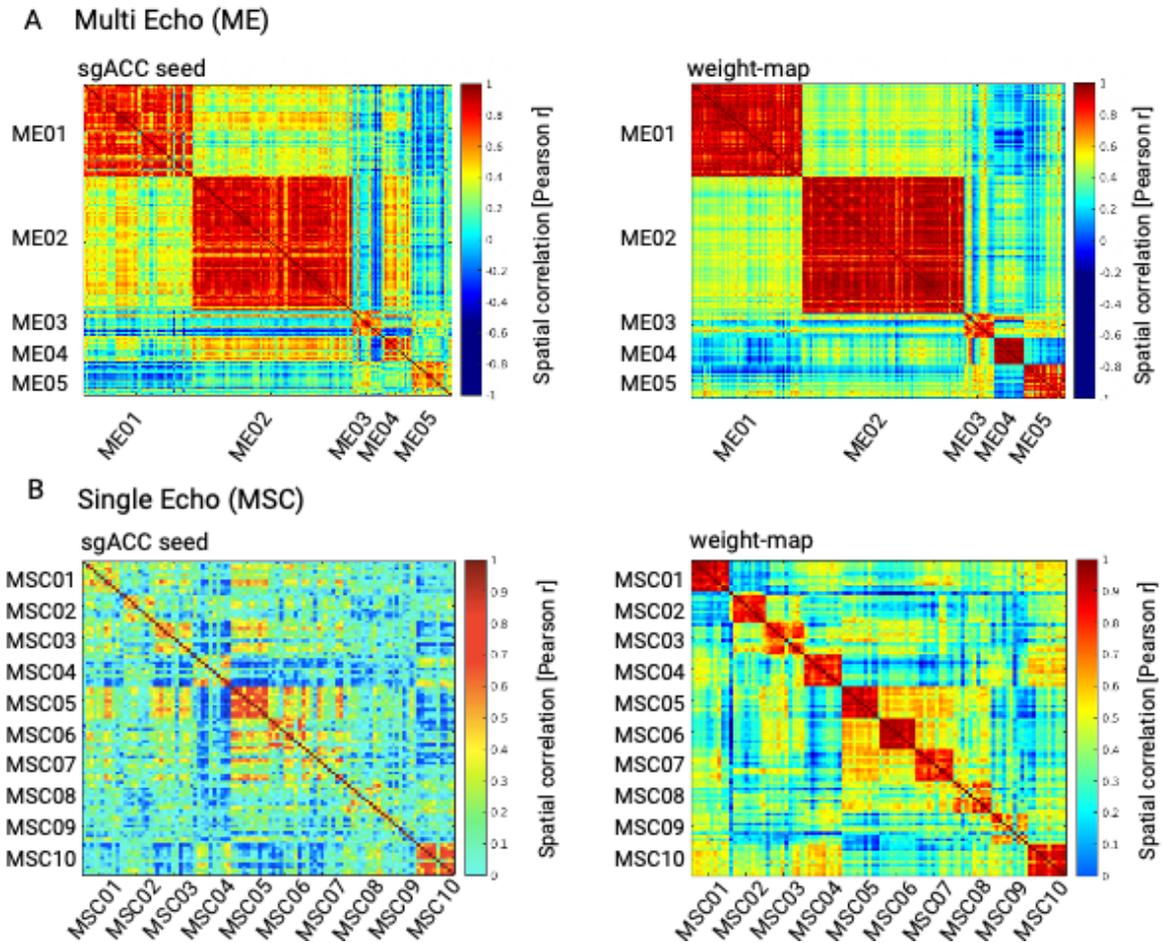


Figure S3. Spatial correlation of functional connectivity across runs. Heatmaps depict the spatial correlation of functional connectivity (FC) maps across runs, both within and between subjects when either a conventional simple subgenual anterior cingulate (sgACC) seed (left panel) or the sgACC weight-map method (right panel) was used. (A) In the multi-echo fMRI ‘ME’ data intra-subject correlation is high both when using a simple sgACC seed (left panel) or the weight-map method (right panel) with mostly marginal gain in session-to-session correlation with the weight-map method. Note, that in the ME data, subjects had different numbers of scanning sessions. (B) in the single-echo fMRI ‘MSC’ data, within-subject run-to-run correlations are low when using a simple sgACC seed. The weight-map method yielded considerably higher within-subject correlations, indicating high test-retest reliability. Low between-subject correlations are preserved, indicating preservation of interindividual variability.

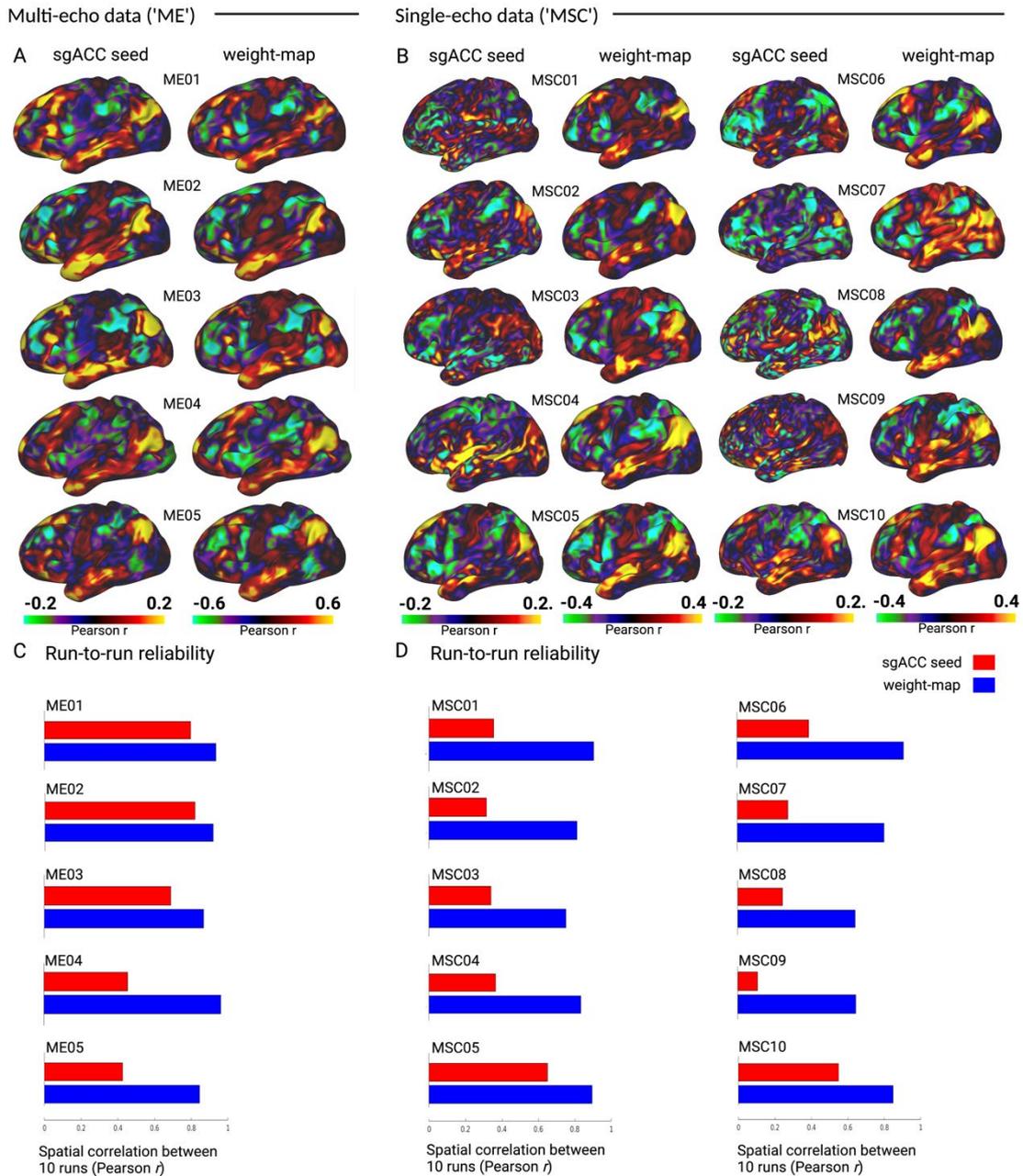


Figure S4. Reliability and accuracy of determining subgenual anterior cingulate cortex (sgACC) functional connectivity (FC) with the weight-map method. We used multi-echo fMRI data (ME) to investigate how accurately the weight-map method approximates true sgACC FC. (A) Cortical surface maps depicting sgACC FC for five densely sampled subjects (ME01-05), each with up to 15 hours of concatenated multi-band multi-echo fMRI data. Because the simple sgACC seed yielded reliable FC maps in the concatenated multi-echo fMRI data (2), we were able to test how accurately the FC maps derived from the weight-map approach (right column) approximated the ‘ground truth’ sgACC seed-derived FC maps (left column). Qualitatively, there was a high correspondence between the FC maps generated by both methods. The median spatial correlation (Pearson r) between the resulting maps for the two methods was 0.95. (B) We next tested the performance of the weight-map method in single-echo fMRI data. For this, we used the publicly available MSC data (1) that consists of ten densely sampled individuals, each with ten 30-minute single-echo fMRI scans. Depicted are cortical surfaces with FC maps from all MSC subjects’ concatenated time series (5 hours), derived either from a simple sgACC seed (left column) or the weight-map method (right column). Here, maps derived from the

sgACC seed were highly noisy whereas the weight-map method yielded FC maps with a consistent default mode network (DMN)-like network configuration resembling the multi-echo results. Note, that the weight-map method produces higher absolute FC values in both, MSC and ME data. (C) and (D) shows the test-retest reliability in terms of the mean spatial correlation (Pearson r) across each subject's total fMRI sessions for a simple sgACC seed (red) and the weight-map method (blue), for ME and MSC subjects respectively. Note the consistent improvement in reliability across MSC subjects, with a mean improvement from $r = 0.38$ to $r = 0.81$ in the total sample.

The effect of Burst-Breathing on the Accuracy and Reliability of the Weight-Map Method. Two of the five ME subjects (ME02, ME03) showed clear evidence for bursting on most of their scans. Comparing the results from Figure 1A between these two and the other ME subjects, we found that 1) the accuracy of the weight-map method in both bursting subjects (Pearson $r = 0.95$ for ME02, and $r = 0.97$ for ME03) were close to the sample's median ($r = 0.95$); 2) their run-to-run reliability ($r = 0.93$ for ME02, and $r = 0.89$ for ME03) was also highly similar to the sample's median ($r = 0.93$).

For MSC, burst-breathing could not be assessed, as no breathing trace exist, and the data is distributed in its fully denoised form (including global-signal regression), such that the fMRI signature of bursting is no longer detectable on a carpet-plot. However, the median run-to-run correlation in male MSC subjects (bursting is more frequent among males) was Pearson $r = 0.86$ vs. $r = 0.82$ in females.

Taken together, these findings do not suggest that burst-breathing meaningfully impacts the validity of the weight-map method. The slightly stronger run-to-run reliability in male subjects could be compatible with burst-breathing increasing the weigh-maps reliability or could be due to other factors. Sex was included as a covariate in the main analysis presented in Figure 2.

DLPFC Stimulation Modeling Detects Correlation With Treatment Outcomes

In the main text we report results from a correlation between sgACC-StimFC and QIDS-SR. This association remained significant in a linear model containing treatment condition, age,

sex, and benzodiazepine or anticonvulsant treatment as covariates ($t(294) = -3.11, p = 0.002$; $F[4,289]=3.23, p=0.007$). The results were weaker but still significant when the HDRS-17 was used as the clinical outcome measure ($t(292) = -1.99, p=0.047$; $F[4,287] = 3.06, p = 0.01$).

Correlations between sgACC-StimFC and clinical improvements (QIDS-SR) for a range of percentiles and radius diameters are shown in Figure S5.

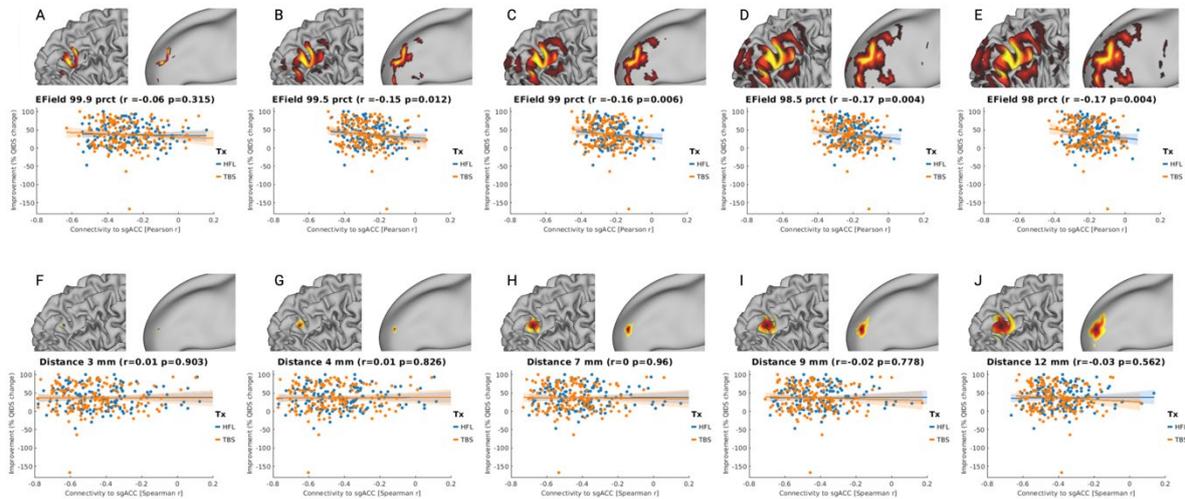


Figure S5. Correlations across range of thresholds. (A)-(E) shows the electric field (E-field) topology and the resulting correlations of sgACC-StimFC and clinical improvement for different E-field thresholds: (A) 99.9th percentile, (B) 99.5th percentile, (C) 99th percentile, (D) 98.5th percentile, (E) 98th percentile strongest E-field values. (F)-(J) shows the extent of covered cortical surface and the resulting correlations of sgACC-StimFC and clinical improvement for 5 different radius thresholds using the weighted cone method as used in prior studies. (F) 3mm, (G) 4mm, (H) 7mm, (I) 9mm, (J) 12mm. Note the wider range of subgenual anterior cingulate cortex functional connectivity (sgACC F FC) values resulting from the weighed-cone method as compared to E-field modeling (all plots have the same x-axis range). This is due to the more distributed, multi-focal nature of the predicted E-field, which results in the averaging of sgACC FC values across a wider portion of the DLPFC. Hence the E-field derived sgACC-StimFC maps are less likely to be biased by local minima or maxima of sgACC FC values.

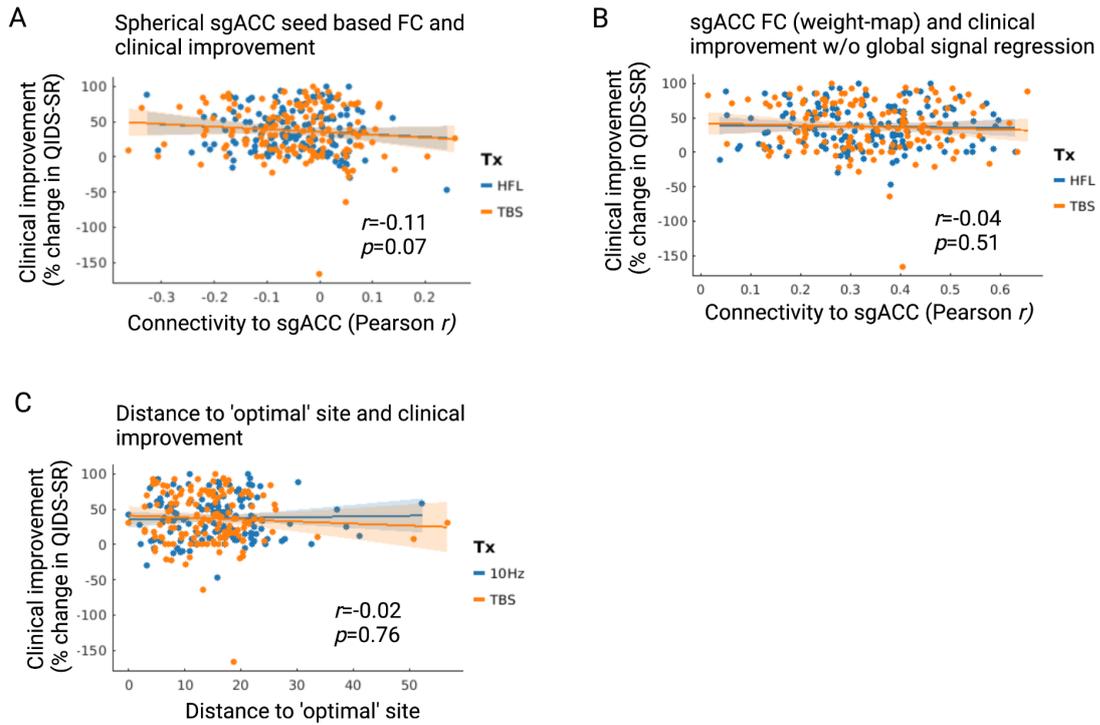
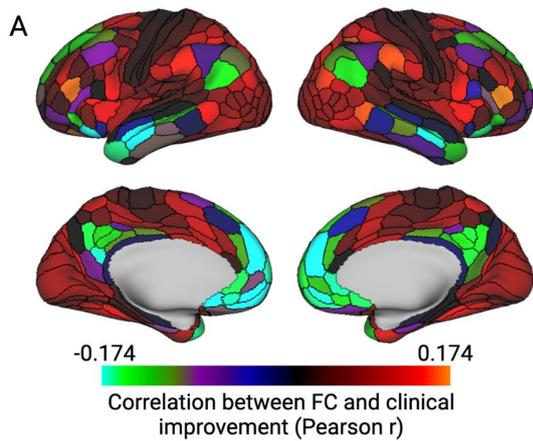


Figure S6. (A) Depicts the same analysis as in Figure 2C, but when a simple left subgenual anterior cingulate cortex (sgACC) seed (Glasser parcel 164) was used to estimate functional connectivity (FC) with the stimulation site (99th percentile strongest E-field) instead of the weight-map method. There is a weak correlation in the same direction as in Figure 2C, however, it doesn't reach significance (Pearson $r = -0.11$, $p = 0.07$). (B) depicts the same correlation as Figure 2C, but without applying mean gray matter time series regression, in which case no effect is detectable. (C) Scatterplot of clinical improvement (percent improvement on QIDS-SR) as a function of the distance of the recorded stimulation site to the site with the strongest negative FC with the sgACC within the dorsolateral prefrontal cortex ('optimal' site). Note that there was no significant correlation between distance to this assumed 'optimal site' and clinical improvement ($r = -0.02$, $p = 0.76$).

Anatomical distribution of FC correlations with clinical improvement



Network affiliations of FC correlations with clinical improvement

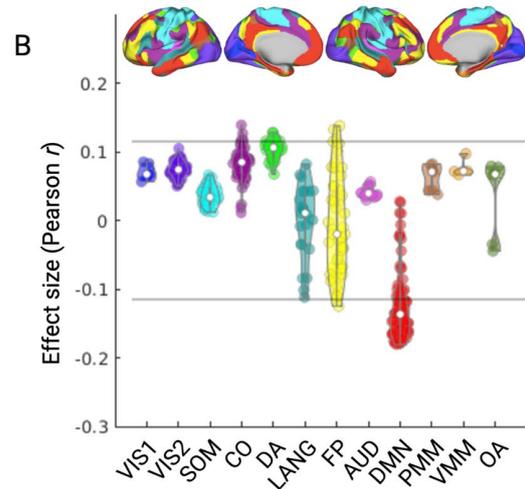


Figure S7. Whole-brain correlations between clinical improvement and functional connectivity (FC) with the dorsolateral prefrontal cortex (DLPFC) target site. (A) Cortical surface maps showing the correlation (Pearson r) of each Glasser (5) parcel's FC strength with clinical improvements (percent improvement on the QIDS-SR). FC strength was operationalized as the correlation between the BOLD signal time course of the DLPFC stimulation site (area receiving the 99th percentile strongest electric field [E-field]) and the BOLD signal time course for each parcel (weighted average time course using the weight-map approach described in the supplemental methods). Note the preponderance of stronger negative correlations among DMN parcels in the medial prefrontal cortex, posterior cingulate cortex, lateral parietal cortex, and temporal pole, among which the left sgACC shows the 4th strongest negative correlation with clinical improvement ($r = -0.18$). Overall, 70 parcels reached nominal statistical significance ($p < 0.05$). (B) Violin plots depicting the same data organized by functional network. Each parcel's network affiliation is defined as in (18). Vertical bars represent the nominal significance threshold. VIS1, visual 1; VIS2, visual 2; SOM, somatosensory; CO, cingulo-opercular; DA, dorsal-attention; LANG, language; FP, fronto-parietal; AUD, auditory; DMN, default-mode; PMM, posterior-multimodal; VMM, ventral-multimodal; OA, orbito-affective.

Main effect does not depend on movement in a subject's scan

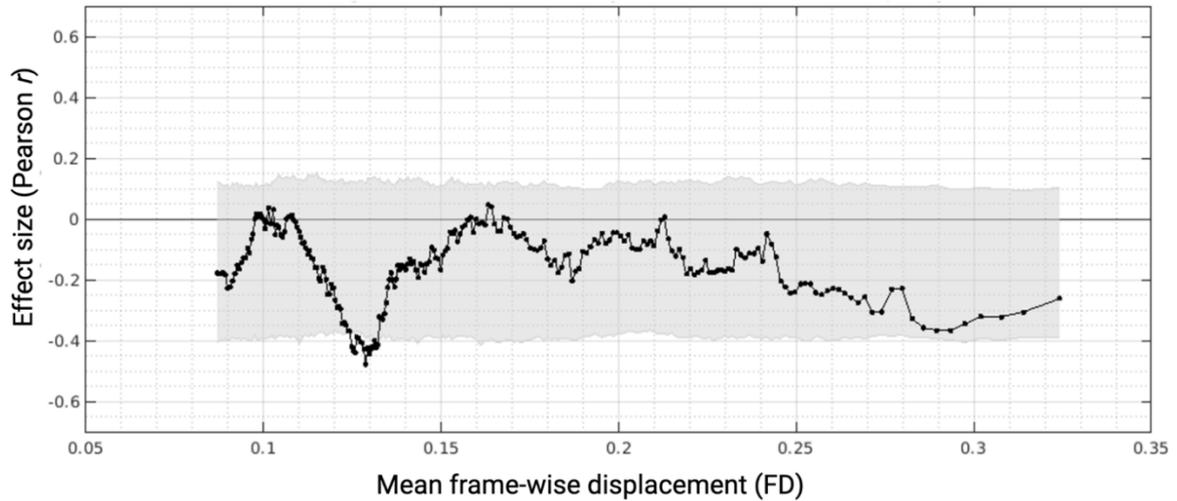


Figure S8. Main effect does not depend on movement in a subject's scan. Each dot represents the Pearson r correlation strength between the stimulation site's functional connectivity with the subgenual anterior cingulate cortex (sgACC-StimFC) and clinical improvement (% improvement in QIDS-SR) in a subsample of 50 subjects. Samples were drawn in ascending order, based on frame-wise displacement (FD) values in an overlapping sliding window manner. Gray shading represents 95% of the null distribution of Pearson r values in 10,000 boot-strapped sub-samples of 50 subjects.

sgACC-StimFC Predicts Treatment Outcomes in Patients With Large Global Signal

Fluctuations

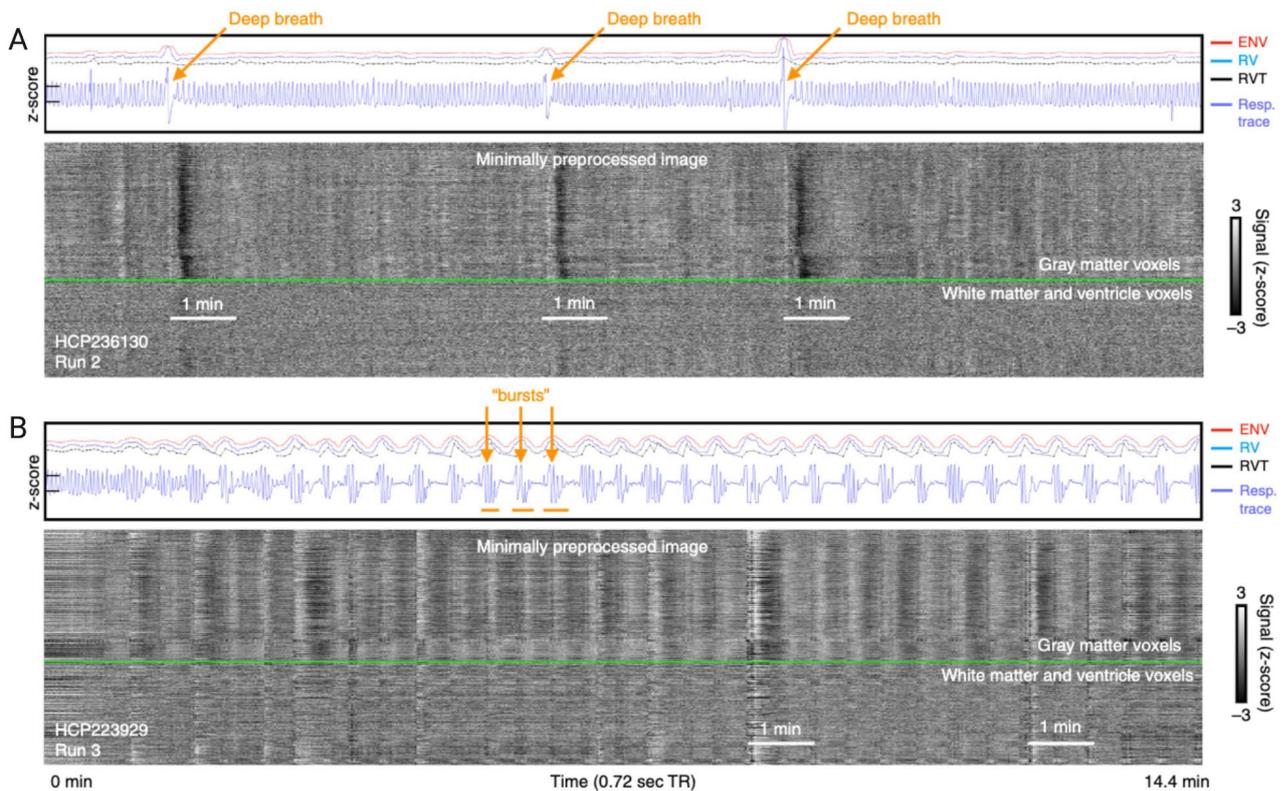


Figure S9. Reproduced from (20). Carpet plots containing (A), deep breaths and (B), bursts. Upper panels show the z-scored respiratory belt traces in blue (y-ticks at left are $z = -1$ and 1), and 3 commonly derived respiratory measures: envelope of the respiratory wave form (ENV), respiratory variation (RV), and respiratory volume over time (RVT), with vertical offsets to enable non-overlapping visualization; scales are identical in all figures. In the grayscale heat maps, all in-brain fMRI signals are shown organized by anatomical compartment (rows are voxel, columns are volumes/time-points), with a green line separating gray matter from white matter and ventricle signals. In (A), three deep breaths are indicated by arrows, with major decreases (vertical black bands) in fMRI signal following each of the breaths. In (B), over two dozen bursts are present (arrows mark several), with accompanying modulations of fMRI signals.

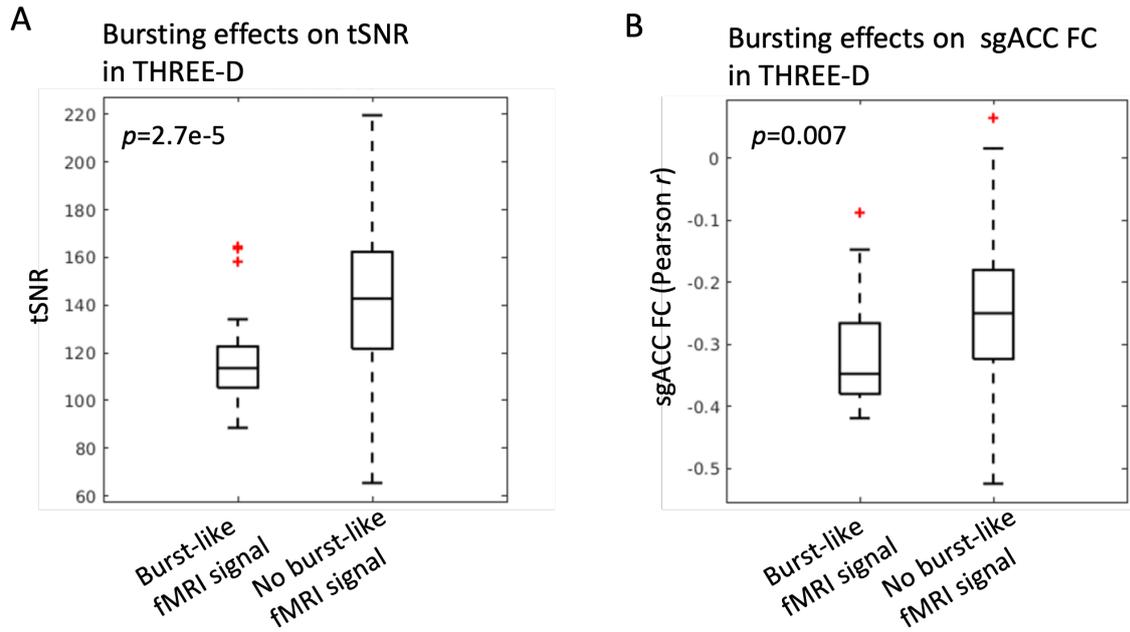


Figure S10. Boxplots showing the difference in (A) temporal signal to noise ratio (tSNR) and (B) the stimulation site's functional connectivity with the subgenual anterior cingulate cortex (sgACC-StimFC) between subjects with bursting-like fMRI signal fluctuations (N = 47) and subjects without (N = 248) in the THREE-D sample. P-value was derived from a two-sample t-test.

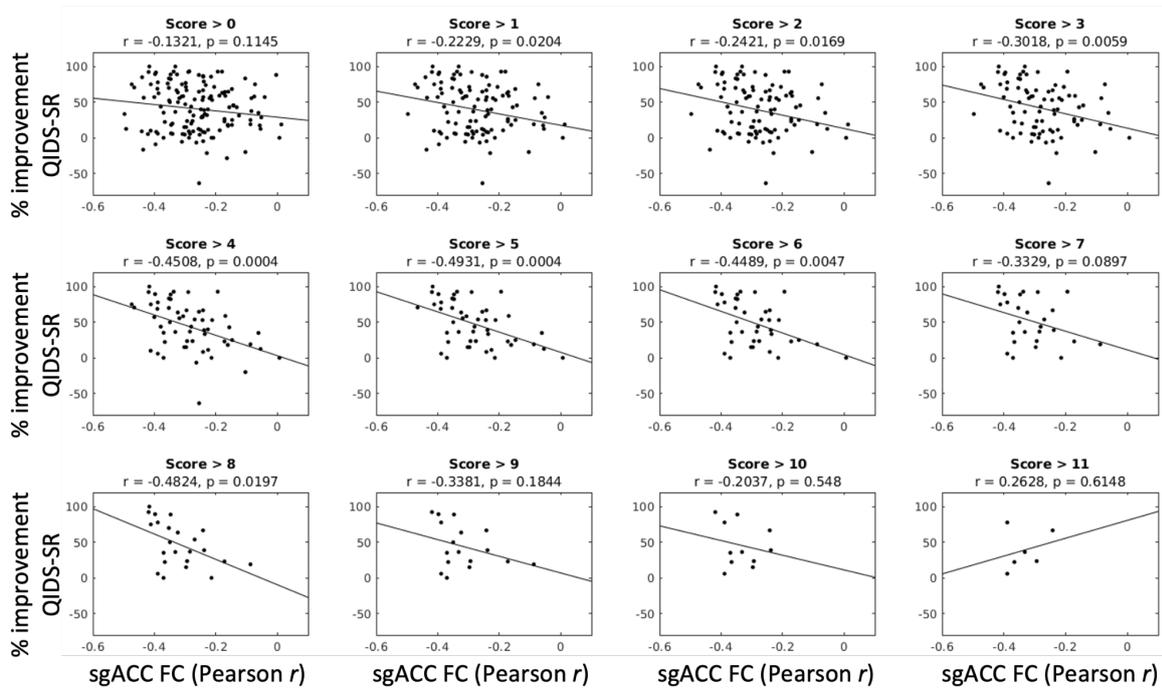


Figure S11. Correlation between the stimulation site's functional connectivity with the subgenual anterior cingulate cortex (sgACC-StimFC) and clinical improvement (QIDS-SR) across the range of thresholds by which 'evidence for bursting' was defined based on cumulative rater decisions. From > 0, at least one rater rated '1' on one scan of subject to >11, all three raters rated '2' on both of a subject's scans. One subject with a QIDS-SR improvement of -150 was omitted for illustrative purposes; the subject was included in all test statistics.

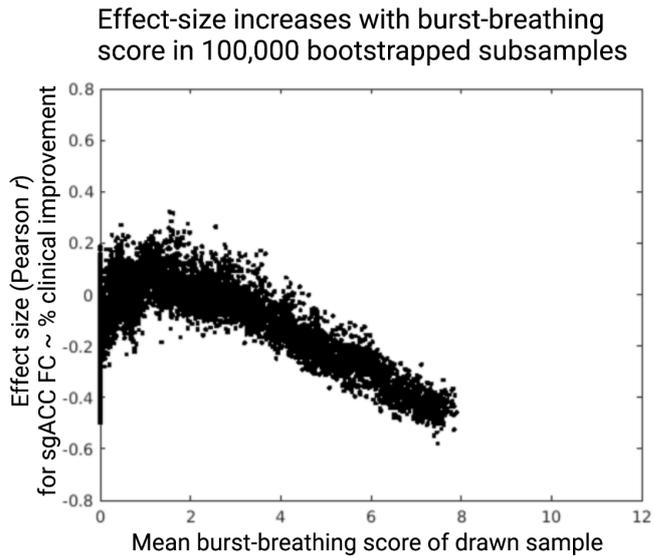


Figure S12. Bootstrapped sub-samples with higher bursting scores show a stronger effect-size (negative correlation between sgACC-StimFC and % improvement on QIDS-SR). For this analysis, we sub-sampled 100,000 samples of 50 subjects across all percentile ranges of burst scores and recorded for each drawn sample the effect size and mean bursting score. Shown is the resulting scatter plot of bursting-scores against effect sizes where each data point represents one of 100,000 samples. This sub-sampling result suggests that the effect size increases linearly with the extent and certainty of bursting present in a sample. Note that the high variability in samples with a mean score of 0, is due to the number of scans with a score of 0 being considerably larger than for other scores (118 subjects). For each iteration, a random bursting-score percentile was set, and a bootstrapped sub-sample of 50 subjects was then drawn without replacement from within a range from the set percentile.

Supplemental Discussion

Possible Contributors to Effect-Size Variance. In addition to the factors discussed in the main text, several other factors may account for the differences in effect size between this and prior reports. Figure 3A clarifies how highly variable results are expected in smaller samples, which may be predisposed to yield larger effects (21, 22), an issue that is magnified when multiple analytical models are employed in a multidimensional variable space (22). Methodological differences may complicate any comparison of the results across studies: a problem that is especially important in studies involving TMS and neuroimaging analysis, both of which feature a large number of free parameters. Relevant stimulation parameters include the stimulation intensity, coil type, scalp-to-cortex distance (23, 24), whether or not E-field modeling was incorporated as a part of the target selection, and the accuracy of the manual co-registration used for neuronavigation (25, 26). The accuracy of the neuronavigation procedure, in particular, might potentially reduce the correlation strength without impacting overall clinical outcomes. The same is true for the choice of stimulation coil, which varied from study to study: prior reports have used the weighted cone heuristic to approximate the E-field distribution. One way in which different coils could thus have introduced variability in these prior reports is by the extent to which their E-field is better or worse approximated with the weighted cone. Only four of the nine prior studies on the matter included details on the used TMS coil (8, 12, 27, 28). Although all of these studies used versions of figure 8 coils (either MagStim D70, MagVenture Cool-B70, or NeuroStar), subtle differences between their ensuing E-field exist that could contribute to result heterogeneity (29).

A second class of free parameters relates to how functional connectivity estimates are derived. Examples include the total duration of the scans; whether participants were instructed to keep their eyes open, closed, or to fixate a cross (which varied across prior reports and could impact vigilance and in turn FC); which type of fMRI sequence was used; whether a simple seed or a weight-map method was used; whether FC was defined based on individual-

level estimates or group estimates; and whether FC strength or distance to an 'optimal' target site was measured, as in two recent reports (6, 30).

A third class of parameters that could generate conflicting results relate to heterogeneity in the clinical population characteristics and the outcome measures used to assess treatment response in a specific patient sample. For example, major depressive disorder is a clinical syndrome, is not a unitary disease entity, and sgACC connectivity could conceivably be more important in predicting treatment outcomes in specific forms of depression but not others. Similarly, some clinical outcome measures could conceivably be more influenced by subjective biases and other sources of noise. Of note, clinical and treatment-related covariates such as age, sex, percent of subjects receiving antidepressant medication, baseline severity, and percent clinical improvement in our sample were all within the range of previously reported studies (6–9, 12, 31), but other unmeasured clinical features of the samples in these various studies may have influenced the results. Our analyses underscore how it is critical to consider methodological differences when comparing results across studies. Because our understanding of how these factors influence fMRI analyses and rTMS outcomes remain rudimentary, we should not necessarily expect to observe the same results in studies employing different recruitment strategies, analytic methods, or rTMS protocols.

Individual differences in network topology may also contribute to differences in treatment outcomes. Recent advances in data acquisition (32) and analysis methods (1) have enabled the investigation of functional brain networks on the level of single individuals. These studies have revealed marked heterogeneity and complexity in the human brain's network architecture (33). For example, the relatively uniform spatial pattern of the default mode network (DMN) as it appears in group-averaged data, contains a far more fine-grained organized structure at the single-subject level (34). Similarly, single-subject parcellation work in densely sampled healthy humans has revealed the complex network organization of structures within the DLPFC, where structures belonging to functionally and topographically distinct networks (dorsal-attention, frontoparietal, salience, cingulo-opercular) border each

other in close proximity (1). If rTMS effects were mediated in part by regulatory effects on the level of functional brain networks, unaccounted interindividual variability in network topology could be one factor contributing to the relatively modest association with clinical outcomes that we observe. Further investigations will be necessary to determine the degree to which rTMS could be optimized when specific networks are precisely targeted based on a patient's individual functional network architecture (35).

References

1. Gordon EM, Laumann TO, Gilmore AW, et al.: Precision Functional Mapping of Individual Human Brains. *Neuron* 2017; 95:791-807.e7
2. Lynch CJ, Power JD, Scult MA, et al.: Rapid Precision Functional Mapping of Individuals Using Multi-Echo fMRI. *Cell Rep* 2020; 33:108540
3. 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
4. Cash RFH, Cocchi L, Lv J, et al.: Personalized connectivity-guided DLPFC-TMS for depression: Advancing computational feasibility, precision and reproducibility. *Hum Brain Mapp* 2021; 42:4155–4172
5. Glasser MF, Coalson TS, Robinson EC, et al.: A multi-modal parcellation of human cerebral cortex. *Nature* 2016; 536:171–178
6. Cash RFH, Cocchi L, Lv J, et al.: Functional Magnetic Resonance Imaging-Guided Personalization of Transcranial Magnetic Stimulation Treatment for Depression. *JAMA Psychiatry* 2020; 1–3
7. Cash RFH, Zalesky A, Thomson RH, et al.: Subgenual Functional Connectivity Predicts Antidepressant Treatment Response to Transcranial Magnetic Stimulation: Independent Validation and Evaluation of Personalization. *Biol Psychiatry* 2019; 86:e5–e7
8. Weigand A, Horn A, Caballero R, et al.: Prospective Validation That Subgenual Connectivity Predicts Antidepressant Efficacy of Transcranial Magnetic Stimulation Sites. *Biol Psychiatry* 2018; 84:28–37
9. Fox MD, Buckner RL, White MP, et al.: Efficacy of transcranial magnetic stimulation targets for depression is related to intrinsic functional connectivity with the subgenual cingulate. *Biol Psychiatry* 2012; 72:595–603
10. Nielsen JD, Madsen KH, Puonti O, et al.: Automatic skull segmentation from MR images for realistic volume conductor models of the head: Assessment of the state-of-the-art. *Neuroimage* 2018; 174:587–598
11. Thielscher A, Antunes A, Saturnino GB: Field modeling for transcranial magnetic stimulation: A useful tool to understand the physiological effects of TMS? *Conf Proc IEEE Eng Med Biol Soc* 2015; 2015:222–225
12. Siddiqi SH, Taylor SF, Cooke D, et al.: Distinct symptom-specific treatment targets for circuit-based neuromodulation. *Am J Psychiatry* 2020; 177:435–446
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; 391:1683–1692
14. Greve DN, Fischl B: Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* 2009; 48:63–72
15. Jenkinson M, Bannister P, Brady M, et al.: Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 2002; 17:825–841

16. Pruim RHR, Mennes M, van Rooij D, et al.: ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *Neuroimage* 2015; 112:267–277
17. Smith SM, Vidaurre D, Beckmann CF, et al.: Functional connectomics from resting-state fMRI. *Trends Cogn Sci* 2013; 17:666–682
18. Ji JL, Spronk M, Kulkarni K, et al.: Mapping the human brain's cortical-subcortical functional network organization. *Neuroimage* 2019; 185:35–57
19. Power JD, Schlaggar BL, Petersen SE: Recent progress and outstanding issues in motion correction in resting state fMRI. *Neuroimage* 2015; 105:536–551
20. Lynch CJ, Silver BM, Dubin MJ, et al.: Prevalent and sex-biased breathing patterns modify functional connectivity MRI in young adults. *Nat Commun* 2020; 11:5290
21. Button KS, Ioannidis JPA, Mokrysz C, et al.: Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 2013; 14:365–376
22. Cremers HR, Wager TD, Yarkoni T: The relation between statistical power and inference in fMRI. *PLoS One* 2017; 12:e0184923
23. Stokes MG, Chambers CD, Gould IC, et al.: Distance-adjusted motor threshold for transcranial magnetic stimulation. *Clin Neurophysiol* 2007; 118:1617–1625
24. Summers PM, Hanlon CA: BrainRuler-a free, open-access tool for calculating scalp to cortex distance. *Brain Stimul* 2017; 10:1009–1010
25. Sparing R, Buelte D, Meister IG, et al.: Transcranial magnetic stimulation and the challenge of coil placement: a comparison of conventional and stereotaxic neuronavigational strategies. *Hum Brain Mapp* 2008; 29:82–96
26. Ruohonen J, Karhu J: Navigated transcranial magnetic stimulation. *Neurophysiol Clin* 2010; 40:7–17
27. Ge R, Downar J, Blumberger DM, et al.: Functional connectivity of the anterior cingulate cortex predicts treatment outcome for rTMS in treatment-resistant depression at 3-month follow-up. *Brain Stimul* 2020; 13:206–214
28. Philip NS, Barredo J, van 't Wout-Frank M, et al.: Network Mechanisms of Clinical Response to Transcranial Magnetic Stimulation in Posttraumatic Stress Disorder and Major Depressive Disorder. *Biol Psychiatry* 2018; 83:263–272
29. Deng Z-D, Lisanby SH, Peterchev AV: Electric field depth-focality tradeoff in transcranial magnetic stimulation: simulation comparison of 50 coil designs. *Brain Stimul* 2013; 6:1–13
30. Siddiqi SH, Weigand A, Pascual-Leone A, et al.: Identification of Personalized Transcranial Magnetic Stimulation Targets Based on Subgenual Cingulate Connectivity: An Independent Replication [Internet]. *Biol Psychiatry* 2021; Available from: <http://dx.doi.org/10.1016/j.biopsych.2021.02.015>
31. Hopman HJ, Chan SMS, Chu WCW, et al.: Personalized prediction of transcranial magnetic stimulation clinical response in patients with treatment-refractory depression using neuroimaging biomarkers and machine learning. *J Affect Disord* 2021; 290:261–271
32. Lynch CJ, Elbau I, Liston C: Improving precision functional mapping routines with multi-

echo fMRI. *Curr Opin Behav Sci* 2021; 40:113–119

33. Gordon EM, Nelson SM: Three types of individual variation in brain networks revealed by single-subject functional connectivity analyses. *Current Opinion in Behavioral Sciences* 2021; 40:79–86
34. Gordon EM, Laumann TO, Marek S, et al.: Default-mode network streams for coupling to language and control systems. *Proc Natl Acad Sci U S A* 2020; 117:17308–17319
35. Lynch CJ, Elbau IG, Ng TH, et al.: Automated optimization of TMS coil placement for personalized functional network engagement. *Neuron* 2022; 110:3263-3277.e4