Transdiagnostic connectome-based predictive modeling of many behavioral phenotypes reveals brain network mediators of clinical-cognitive relationships

Alexander J Simon[1], Anja Samardzija[3], Santino Iannone[1], Flor Parra Rodriguez[3], Saloni Mehta[2], Jagriti Arora[2], Fuyuze Tokoglu[2], Maolin Qiu[2], Kira Y Tang[2], Anna Q Flanagan[2], Rachel Katz[4], Gerard Sanacora[4], Scott W. Woods[4], Vinod H. Srihari[4], Xilin Shen[2], Dustin Scheinost[1,2,3], R. Todd Constable[1,2,3]

Corresponding authors: Alexander Simon (aj.simon@yale.edu) & Todd Constable (todd.constable@yale.edu)

1 Interdepartmental Neuroscience Program, Yale School of Medicine, New Haven, CT, USA

2 Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, CT, USA

3 Department of Biomedical Engineering, Yale University, New Haven, CT, USA

4 Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA

Abstract

Connectome-based predictive modeling (CPM) applied to functional MRI connectivity data can identify brain networks that vary with behavioral measures across subjects. The prediction strength also provides an index of how closely an external instrument relates to specific brain networks, potentially impacting their clinical interpretation. Here we use a deeply phenotyped transdiagnostic population ($n = 317$) to evaluate CPM performance across a variety of clinical and cognitive measures. The findings revealed a wide range of predictive performance for external instruments, with cognitive tests generally predicting better than self-reported clinical measures (unpaired *t*-test, $p < 0.001$). Testing the hypothesis that networks supporting cognition should be apparent in networks related to symptomatology, we examined the networks' overlap. The overlap was sparse, but primarily identified the thalamus, cerebellum, somatomotor networks, and the dorsolateral prefrontal cortex as key hubs in mediating relationships between clinical and cognitive measures. The findings reveal the extent to which external measures reflect underlying brain networks and highlight that examining network overlap can identify networks specific to clinically relevant cognitive dysfunction.

Introduction

Identifying neurobiological correlates of real-world behavioral measures can teach us about how certain neural dysfunctions manifest as clinical symptoms and cognitive deficits. Such information can improve precision psychiatry by increasing our understanding of the etiology of disorders, refining diagnostic and treatment strategies, and guiding novel therapeutic development. The recognition that many psychiatric symptoms and cognitive deficits exist on a transdiagnostic spectrum and stem from deviations from normal neural functioning led to the conceptualization of the NIMH's RDoC initiative[1,2]. In accordance with this framework, many studies have utilized functional MRI (fMRI) to link high-dimensional patterns of functional connectivity distributed across the brain to a wide range of continuous behavioral measures[3-9].

Connectome-based predictive modeling (CPM) is a well-established approach for linking behavioral measures to underlying brain networks measured by fMRI connectivity data[3,10] in a predictive modeling framework. The cross-validation employed in building the predictive models assures that the identified brain-behavior associations are generalizable[11]. A large body of work has demonstrated CPM's ability to relate functional networks to numerous cognitive and clinically relevant measures, such as sustained attention[5], memory[12], craving[13], PTSD[14], and social function[15]. A key assumption of the RDoC formalism is that a clinical symptom (like anxiety) should involve networks that vary with severity of the symptom and that the same networks generate the symptom independent of the patients' diagnostic category[2,16]. While numerous studies have shown that CPM can identify functional connectivity correlates for specific measures, few studies combine a large transdiagnostic sample with both extensive fMRI data and outside-the-magnet clinical and cognitive phenotyping. The data reported here provides an opportunity to assess the extent to which a wide range of diagnostically related clinical symptomatology measures (referred

56 to throughout as ‘clinical measures’) and cognitive instruments have clearly identifiable network
57 correlates and which do not in a clinically diverse transdiagnostic dataset[17]. Furthermore,
58 predicting an array of measures enabled us to address the extent to which measurement noise
59 influences prediction strength, which is an important consideration when interpreting brain-based
60 predictive models[18].
61
62 The human brain consists of many complex networks that support a diverse array of traits. CPM
63 can be used to map the functional network architecture underlying a variety of clinical and
64 cognitive measures thus providing a means to assess network commonalities. The RDoC
65 framework suggests that disordered circuits can affect both diagnostically-related symptoms and
66 cognitive abilities. This is corroborated by a large body of evidence indicating that cognitive
67 deficits often co-occur with psychiatric symptoms[19]. However, disordered circuits vary
68 substantially across individuals[20], and may not completely impinge on a network related to a
69 particular symptom or cognitive process. Rather, an individual’s disordered circuits are more likely
70 to partially affect multiple networks linked to various symptom and cognitive measures. This
71 phenomenon may underlie the highly variable cognitive deficits seen in psychiatric
72 populations[16,19]. According to this hypothesis, individuals with disordered circuits impinging on
73 networks related to both symptoms and cognition should experience cognitive deficits that co-
74 occur with symptomatology, while individuals with disordered circuits impinging on networks only
75 related to either symptoms should have psychiatric symptoms without cognitive deficits (or vice
76 versa). While it is unlikely that any disordered circuits would overlap completely with a specific
77 symptom or cognitive network, brain-behavior modeling provides a means for assessing the
78 effects of a partial overlap. Therefore, we sought to identify where the networks impacted by
79 clinically relevant symptomatology overlapped with networks supporting cognition. Given that
80 patients who experience cognitive impairment may respond differently to various forms of

treatment than unimpaired patients[21-23] and that psychiatric patients often consider cognitive impairments among their primary priorities for treatment[24,25], identifying the loci of these interactions is an important research focus that could lead to novel diagnostic biomarkers and precision treatment targets. Furthermore, this would provide a putative link between inconsistencies in cognitive perturbations[16,19] and high inter-individual variability in network dysfunctions seen in psychiatric populations[20].

This study has two primary goals. First, we determine which of the many clinical and cognitive measures obtained using a comprehensive testing battery administered outside the magnet yield CPM models with high prediction strength and which do not. This allows us to characterize the degree that each measure reflects the functioning of identifiable brain networks and to compare prediction strength across measures (i.e., the correlation between observed and predicted values). Second, we utilize the networks revealed by CPM in the first aim to examine where the networks implicated in clinical symptomatology overlap with networks that support cognitive functioning. By doing so, we demonstrate how CPM can precisely locate functional networks that mediate relationships between symptoms and cognition. This approach improves the specificity of mapping networks to behavior by highlighting subsets of edges within networks associated with clinical symptomatology that co-fluctuate with cognitive dysfunction.

Results

In the present study, we utilized a demographically and clinically diverse transdiagnostic population[26] (n=317; demographic data is presented in **Supplemental Table 1**). The dataset consists of fMRI data from 2 resting-state and 6 task runs, as well as questionnaires collected outside the magnet consisting of 7 real-world self-report clinical assessment scales comprised of a total of 63 distinct measures, and 6 neuropsychological tests yielding 28 distinct cognitive

performance measures. The clinical assessment instruments were the Adult Temperament Questionnaire (ATQ)[27], the Behavior Rating Inventory of Executive Function (BRIEF)[28], the Brief Symptom Inventory (BSI)[29], the Interpersonal Reactivity Index (IRI)[30], the Positive and Negative Affect Schedule (PANAS)[31], the Pittsburgh Sleep Quality Index (PSQI)[32], and the Perceived Stress Scale (PSS)[33]. The cognitive testing instruments were the verbal fluency, trail making, color-word interference, and twenty questions subtests from the Delis-Kapan Executive Function Scale (D-KEFS)[34], the letter-number sequencing, symbol search, and cancellation subtests from the Wechsler Adult Intelligence Scales (WAIS)[35], the vocabulary and matrix reasoning subtests of the Wechsler Abbreviated Scale of Intelligence 2nd Edition (WASI)[36], the finger windows and list learning subtests of the Wide Range Assessment of Memory and Learning 2nd Edition (WRAML)[37], the reading subtest of the Wide Range Achievement Test 5th Edition (WRAT)[38], and the Boston Naming Test 2nd edition (BNT)[39]. The fMRI data was preprocessed with standard preprocessing methods (described in greater detail in **Methods** and Ref 26). The Shen 268-node atlas[40] was applied to the preprocessed data, parcellating it into 268 functionally coherent nodes, and functional connectivity matrices were derived for each participant/run. The averages of all 8 functional connectivity matrices were computed for each participant and used as input in building the CPMs. CPM was run for all clinical and cognitive measures in a 10-fold cross validation. Prediction strength was quantified as the Spearman correlation (ρ) between the observed measurements and the predicted values. Significance was determined by comparing the prediction strengths from 1000 CPM permutations to a null distribution generated from running CPM on randomly permuted observations. The proportion of null predictions that exceeded the median prediction strength represented the predictive models' *p*-values. All *p*-values were FDR corrected for multiple comparisons.

Prediction strength of clinical measures

To determine whether a measure had sufficient prediction strength to yield reliable brain network correlates, measures must have had $pFDR < 0.05$ and prediction strength > 0.1. In neuroimaging, prediction strengths of 0.1 are generally considered weak, but are accurate enough if also statistically significant to yield meaningful insights about the neurobiological features important to these models[41]. Of 63 clinical measures, 38 (60.32%) met our criteria for being predicted strongly enough to have identifiable whole brain functional connectivity networks (**Fig 1**). The other 25 measures failed to yield sufficient prediction strength to continue with further analyses. When examining predictive model performance across scales and measures, several notable patterns emerged. Of the 7 clinical assessment scales, models built with measures related to sleep quality tended to have the highest prediction strength, with 5 of the top 10 strongest predictions belonging to the PSQI. The mean prediction strength across all PSQI measures was $\rho = 0.170$. Measures pertaining to positive affect from the PANAS also yielded strong predictive models, with 5 out of 6 of these measures providing reliable network correlates (mean $\rho = 0.151$). PANAS measures related to negative affect tended to yield slightly lower prediction strength, with 4 out of 7 providing reliable network correlates (mean $\rho = 0.120$). $\rho\rho\rho\rho\rho\rho$Prediction strength values and associated prediction *p*-values can be found in **Supplemental Table 2**, and distributions of clinical measures can be found in **Supplemental Fig 1**.

Interestingly, there were 4 separate instances where the same measures quantified by different assessment scales yielded widely different predictions. The prediction strength for 'fear' was significantly higher when measured by the ATQ than when measured by the PANAS ($\rho = 0.239$ vs $\rho = 0.105$; paired t-test: $t(999) = 192.12$, $p < 0.001$), as was 'sadness' ($\rho = 0.165$ vs $\rho = 0.099$; paired t-test: $t(999) = 88.63$, $p < 0.001$). PANAS 'hostility' was better predicted than BSI 'hostility' ($\rho = 0.181$ vs $\rho = 0.123$; paired t-test: $t(999) = 98.23$, $p < 0.001$). Furthermore, the prediction

strength for 'positive affect' was higher when measured as an index of all measures related to positive affect by the PANAS than when assessed as a single measure by the ATQ ($\rho = 0.164$ vs $\rho = 0.122$; paired t-test: $t(999) = 54.26$, $p < 0.001$).

Prediction strength of cognitive measures

Nearly all cognitive measures yielded strong prediction strength by CPM. Successful models were built for 26 of 28 (92.86%) measures from the cognitive panel (**Fig 2**). The matrix reasoning test from the WASI provided the model with the highest prediction strength ($\rho = 0.428$). Predictions of measures from the D-KEFS were highly variable, though all produced (range: $\rho = 0.159$ to $\rho = 0.374$). In general, measures from the twenty-questions (mean $\rho = 0.288$) and trail-making tests (mean $\rho = 0.320$) were the most strongly predicted measures from D-KEFS. Performance measures from the WAIS (mean $\rho = 0.319$), the WRAML (mean $\rho = 0.263$), and the WRAT ($\rho = 0.302$) were generally predicted strongly. Measures from the BNT (mean $\rho = 0.224$) and the D-KEFS verbal fluency test (mean $\rho = 0.177$) were predicted less strongly. The two cognitive measures that CPM did not generate reliable models for (WRAML verbal learning intrusion and BNT number correct following a stimulus cue) had the highest percentage of individuals performing at ceiling (see **Supplemental Fig 2**; 53.16% and 77.19%, respectively). These results highlight that predictive models linking external trait measures to brain connectivity patterns via CPM can be built across a wide range of cognitive metrics, with higher prediction strengths representing greater confidence in the derived network correlates. Prediction strength values, associated prediction *p*-values, and brief descriptions of the cognitive abilities theoretically tested by each measure can be found in **Supplemental Table 3**.

What drives prediction strength?

179 In addition to CPM significantly predicting a higher percentage of cognitive than clinical measures,
180 the prediction strength of the cognitive measures that were successfully modeled were higher
181 than the clinical measures' (**Supplemental Fig 3**; cognitive measures' mean prediction strength:
182 $\rho = 0.289$; clinical measures' mean prediction strength: $\rho = 0.168$; $t(62) = 8.39$, $p < 0.001$). This
183 highlights that CPM yields more accurate and robust functional network correlates for cognitive
184 than for clinical measures.
185
186 We next sought to understand the factors that influence model performance leading to better
187 prediction strength for some measures compared to others, noting that the functional connectivity
188 data remained constant across all models. Due to the differences in prediction strength between
189 clinical and cognitive assessments, we examined how prediction strength varied across
190 instruments and subdomains tested in clinical and cognitive assessments separately. One-way
191 ANOVAs revealed that prediction strength strongly varied as a function of instrument/subdomain
192 for clinical (**Fig 3a**; $F(13,62986) = 3059.98$, *Cohen's F* = 0.795, $p < 0.001$) and cognitive measures
193 (**Fig 3b**; $F(13,27986) = 1194.96$, *Cohen's F* = 0.745, $p < 0.001$).
194
195 Recent work questioned the extent to which brain-based predictive modeling is influenced by
196 measurement noise[18]. To address this, we tested whether the predictive accuracy on these
197 measures was related to their intraclass coefficients (ICCs) reported by the manuals and seminal
198 papers describing each of the testing instruments. We also examined whether model performance
199 was related to other characteristics of the measures' data (distribution skewness and variance).
200 Clinical prediction strength was not correlated with ICC in the measures with reported ICC values
201 ($r = 0.026$, *pFDR* = 0.802), nor was it related to the measures' skewness ($r = 0.072$, *pFDR* =
202 0.802) or variance ($r = 0.032$, *pFDR* = 0.802) (**Fig 3c**). Similarly, cognitive prediction strength was
203 not correlated with ICC in the measures with reported ICC values ($r = -0.122$, *pFDR* = 0.802), nor

204 was it correlated with skewness (r = -0.456, $pFDR$ = 0.108) or variance (r = 0.262, $pFDR$ = 0.534)
205 (**Fig 3d**). Together, these results suggest that CPM performance depends on how well the scale
206 captures aspects of cognition or clinical symptoms and how tightly coupled these measures are
207 to brain organization. The differential results across tests do not appear to be significantly driven
208 by measurement stability or distribution of scores across subjects.
209
210 **Network similarity across measures**
211 Since disordered circuits could impinge upon both clinical and cognitive networks, we examined
212 the degree of overlap between predictive networks for all clinical and cognitive measures (**Fig 4**).
213 This was achieved by sorting the predictive networks based on similarity using an agglomerative
214 hierarchical clustering algorithm. As expected, the similarities between the measures' networks
215 mirrored the correlational structure of the measures themselves since the networks are derived
216 from the measures. In general, the clinical measures were most related to other clinical measures,
217 while the cognitive measures were most related to other cognitive measures. The only exception
218 was high intensity pleasure (ATQ), which was more closely related to cognition than to other
219 clinical measures. The Dice similarity between clinical-cognitive networks was generally much
220 lower than between clinical-clinical (unpaired t-test: $t(788.66) = -31.68$, $p < 0.001$) and cognitive-
221 cognitive networks (unpaired t-test: $t(389.11) = -23.39$, $p < 0.001$). Additionally, after adjusting for
222 multiple comparisons using an FDR correction, only 101 pairs of clinical-cognitive measures were
223 significantly correlated ($pFDR < 0.05$). Together, this suggests that brain networks that vary as a
224 function of cognitive performance are largely distinct from those networks that vary systematically
225 with clinically relevant symptomatology.
226
227 We then grouped the brain networks predicting clinical and cognitive measures into 9 clusters
228 based on the hierarchical clustering results. This number was determined using the elbow

229 method[42] (**Supplemental Fig 4**), which is a commonly used approach for determining an
230 appropriate number of clusters to group data into. These clusters were selected to visualize how
231 the measures grouped together based on network similarity, and to compare the regions
232 important in driving CPM predictions across clusters, rather than to make a claim about how many
233 clusters exist in this dataset. The patterns that the CPM networks loaded onto each of the 9
234 clusters corresponded to what the measures theoretically assess. Of these 9 clusters, 5 were
235 composed of clinical measures and 4 were composed of cognitive measures. For discussion
236 purposes, each of the clusters were given a name that generally reflects the broad measures
237 composing them. The cluster labels are displayed underneath the dendrogram in **Fig 4**, which
238 align with the sorted measures displayed in the matrix below. It is important to note that these
239 names do not encompass all the behavioral phenotypes present in each cluster. Additionally,
240 since they were not orthogonal, important features were often shared across clusters. For
241 example, cognitive control abilities were likely important to performing all cognitive tests, not just
242 the tests in the 'cognitive control' cluster. This phenomenon is reflected by the substantial overlap
243 in the network features between clusters.
244
245 **Localizations of networks important to predictions**
246 We examined the contributions of specific networks to prediction strength of clinical measures
247 using the canonical Yeo17 networks as a reference[43] (**Fig 5a**). For identifying the nodes and
248 edges important to predictions across clusters, we compared positively predicting network
249 features from the positive affect cluster to negatively predicting network features from the 4 other
250 clinical clusters (and vice versa) because these cluster's measures were anticorrelated. Doing so
251 allowed us to effectively put the brain correlates of measures where higher scores indicated better
252 mental health on the same scale as measures where higher scores indicated worse mental health.
253 The nodes driving the prediction strength for the clinical measures displayed substantial overlap

across the clinical measures' clusters for both positively (mean pairwise Jaccard similarity = 0.252 +/- 0.088 SD, one-tailed $p < 0.001$) and negatively predicting nodes (mean pairwise Jaccard similarity = 0.188 +/- 0.068 SD, one-tailed $p < 0.001$). Examining the top 10% highest degree nodes in each cluster's predictive networks revealed that the thalamus was the only region important to positively predicting every cluster of clinical measures, and only the cerebellum contained nodes that were important to negatively predicting each cluster of clinical measures. A full list of the top 10% highest degree positive and negative predicting nodes for each clinical measure cluster can be found in **Supplemental Tables 4 and 5**. Many of the same nodes that were important in positive predictions were also implicated in negative predictions (Jaccard similarity = 0.500, one-tailed $p < 0.001$). Furthermore, examining the top 10% of canonical network pairs where the predictive edges were most concentrated showed that limbic A to brainstem connectivity was important to all positive predictions, while within somatomotor A/B and somatomotor A to dorsal attention B were important to negative predictions across all clinical measures' clusters.

Similarly, the networks supporting prediction strength in the cognitive measures were assessed (**Fig 5b**). As with the clinical measures, we flipped the sign of the cognitive measures' predictive networks before comparing across measures to ensure that all positively predicting networks were indicative of better performance on the task, and not just a higher score, since sometimes a higher score was undesirable. The nodes contributing to predicting cognitive measures displayed substantial overlap across the cognitive measures' clusters for both positively (mean pairwise Jaccard similarity = 0.197 +/- 0.092 SD, one-tailed $p < 0.001$) and negatively predicting nodes (mean pairwise Jaccard similarity = 0.210 +/- 0.118 SD, one-tailed $p < 0.001$). Despite this, examining the top 10% highest degree nodes in each cluster's predictive networks showed that

only nodes within the visual A and somatomotor B networks were positively associated with cognitive performance and only one node in the temporal parietal network was negatively associated with cognitive performance across all clusters. A full list of the top 10% highest degree positive and negative predicting nodes for each cognitive measure cluster can be found in **Supplemental Tables 6 and 7**. There was significant overlap between the nodes important to positively and negatively predicting cognitive measures (Jaccard similarity = 0.294, one-tailed $p < 0.001$), and these nodes often were in the cerebellum. However, this overlap was less pronounced than it was in the clinical measures (Jaccard similarity = 0.294 vs 0.500). The nodes that were only important to positive predictions were typically located toward the bottom of the functional gradient in lower-level unimodal cortex, while nodes that were only important in negative predictions were frequently localized near the top of the functional gradient in higher-level transmodal cortical regions[44] (see **Supplemental Fig 5**). Examining the top 10% of canonical network pairs where the predictive edges were most concentrated showed that dorsal attention A to dorsal attention B, default mode A to cerebellum, limbic A to limbic B, somatomotor A/B to cerebellum, and visual A to somatomotor B network connectivity was important to positive predictions across all cognitive measures' clusters. Furthermore, dorsal attention B to default mode A, default mode B to temporal parietal, somatomotor A/B to dorsal attention B, somatomotor B to default mode A, and within somatomotor A network connectivity were important for driving negative predictions across all cognitive measures' clusters.

Networks mediating clinical-cognitive relationships

In **Fig 4** we highlighted that the correlations were relatively weak, and network overlaps were sparse between clinical and cognitive measures. However, given that cognitive alterations co-occuring with psychiatric symptomatology are hypothesized to emerge from disordered circuits

302 that span both systems, we sought to determine where these network overlaps are located, and
303 if their connectivity mediates relationships between significantly correlated pairs of clinical and
304 cognitive measures (p*FDR* < 0.05). This was achieved using the sum the overlapping edges'
305 connectivity as the mediator variable in mediation analyses for each of the 101 significantly
306 correlated clinical-cognitive measure pairs (**Fig 6a**). We did not have an *a priori* hypotheses
307 regarding the directionality of these mediations, so we conducted two mediation analyses for each
308 pair of correlated clinical-cognitive measures. In one, the clinical measures were the dependent
309 variables, and the cognitive measures were the independent variables, and in the other, the
310 dependent and independent variables were switched. The mediation analyses were significant (*p*
311 < 0.01) bidirectionally for all 101 pairs of clinical-cognitive relationships examined (**Fig 6b**). For
312 each clinical-cognitive relationship pair, we reported the larger of the two indirect mediation
313 coefficients.
314
315 We then examined the localization of these mediating networks by grouping them by the clusters
316 that each pair of related measures belonged to (**Fig 6c**). Even within the same pairs of clusters,
317 the edges in the shared clinical-cognitive networks showed little overlap. However, the nodes
318 implicated in the shared networks were highly consistent across clinical-cognitive measure pairs.
319 The nodes and important canonical network-network connections involved in generally mediating
320 clinical-cognitive relationships are highlighted in **Fig 7**. Of the top 10% highest degree nodes
321 important to mediating clinical-cognitive relationships, 8 were in the cerebellum, 6 were in the
322 somatomotor networks, 2 covered the thalamus (bilaterally), 2 were in frontoparietal control C
323 (bilateral dLPFC), 2 were in brainstem, and 1 was in each the ventral attention B, dorsal attention
324 B, limbic A, and default mode B. A full list of these nodes and the MNI coordinates of their
325 centroids can be found in **Supplemental Table 8**. Interestingly, the first and third highest degree
326 nodes were the left and right thalamus (respectively), while 6 out of the 10 highest degree nodes

were in the cerebellum. Examining the top 10% canonical network pairs where the mediating edges were concentrated corroborated these results. Most of these mediating edges were concentrated between the cerebellum and subcortical nodes (e.g., the thalamus), between cerebellum and somatomotor networks, within somatomotor networks, between somatomotor and subcortical regions (thalamus), and between somatomotor and higher order networks (e.g., frontoparietal control C (dLPFC), default mode B, dorsal attention B, ventral attention B, limbic A). These results highlight the importance of the cerebellum, the thalamus, somatomotor cortices, and bilateral dLPFC in generally mediating clinical-cognitive relationships.

Discussion

This study had two primary goals. First, we aimed to comprehensively evaluate which clinical and cognitive measures reflect clear brain correlates as determined by prediction strength in CPM. The findings showed that, in general, there is a tighter coupling between brain networks and external instruments for cognitive tests compared to clinical measures. We also show that this relationship is not uniform across measures, as prediction strength varied substantially across measures in both classes and was not driven by differences in measurement stability. Our second goal was to test a hypothesis upon which the RDoC framework is based: that evidence for disordered circuits should be apparent in both cognitive and clinical measures. Our results indicated that the networks implicated in clinical symptomatology impinge upon networks that support cognitive functioning, and that specific networks mediate relationships between clinically relevant symptoms and cognitive performance metrics. As might be expected, the networks associated with symptoms do not cleanly overlap with any specific cognitive network. This work further establishes a framework for how to use brain-behavior modeling to locate functional networks that support and facilitate relationships between multiple behavioral domains.

351

352 **Clinical predictions and associated networks**

353 The broad testing of clinical measures that we employed in this study allowed us to systematically
354 evaluate the extent to which a variety of clinical assessment instruments reflect brain network
355 functioning. We identified reliable functional connectivity correlates of many clinical measures
356 collected in this dataset (38 out of 63 predicted at $\rho > 0.1$ and $pFDR < 0.05$). Varying prediction
357 strength between instruments highlights that some clinical assessment tools are better suited to
358 capture brain network functioning than others. Among these, the PSQI, the BRIEF (behavioral
359 regulation measures), PANAS, ATQ (non-aggressive negative affect measures) and BSI all
360 tended to be well modeled in the brain. Conversely, measures from the IRI, BRIEF (metacognition
361 measures), ATQ, and the PSS did not consistently yield reliable brain network models.

362

363 Each of these clinical assessment instruments utilized self-reporting. The subjective nature of
364 self-reports introduces several confounds that could interfere with CPM prediction strength, such
365 as response style differences related to introspection abilities[45], social[46] and cultural
366 backgrounds[47], or differences in how respondents and assessment administrators interpret the
367 questions[48]. Measures that are less prone to these influences are thus more likely to have clear
368 network correlates. Understanding how to refine self-report scales to yield better behavioral
369 predictions will improve the extent to which they reflect network functioning, thereby enhancing
370 their utility in identifying neural generators of clinical dysfunction.

371

372 The considerable overlap in the networks important to predicting all clinical measures suggests
373 that CPM was capturing a low-dimensional general factor of psychopathology (p-factor) posited
374 by the HiTOP framework and elsewhere[49,50]. Each cluster's unique deviations from this general

375 p-factor network, therefore, track network dysfunctions specific to the clusters' phenotypic
376 makeups. Our results strongly implicate the thalamus and cerebellum as key nodes in this p-factor
377 network. These brain structures and their clinical relevance are commonly overlooked, and
378 researchers have argued that this omission is stifling progress in neuroscience[51]. However, there
379 is evidence suggesting the functional connections involving the cerebellum and thalamus contain
380 clinically relevant information[52,53]. Although not involved in predicting clinical measures across all
381 clusters, nodes within the somatomotor network were also frequently involved in predicting clinical
382 measures. Like the cerebellum and deeper brain structures such as the thalamus, the
383 somatomotor network has also been relatively overlooked in psychiatric research. An emerging
384 body of research has linked somatomotor network dysfunction to clinical pathology[54-57]. Our
385 results add to this mounting evidence implicating the thalamus, cerebellum, and SM network in
386 symptomatology.
387
388 **Cognitive predictions and associated networks**
389 We also evaluated a comprehensive battery of cognitive tests to identify the extent to which they
390 reliably reflect brain network functioning. Nearly all cognitive measures yielded models with good
391 prediction strength (26 out of 28 predicted at $\rho > 0.1$ and $pFDR < 0.05$). More than 50% of
392 participants performed at ceiling for the two measures that were not modeled well (verbal learning
393 intrusion (WRAML) and number correct following a stimulus cue (BNT)). These were also the two
394 cognitive measures with the lowest range of scores (see **Supplemental Fig 2**). These results
395 highlight the importance of utilizing cognitive tests with a sufficiently wide distribution of scores to
396 identify reliable functional network associations of cognition. However, considering that there was
397 not a significant correlation between measurement score variance and prediction strength across

398 all cognitive measures (see **Fig 3**), the distribution of the scores likely only affects prediction
399 performance up to a certain extent.
400
401 Since the cognitive tests utilized objective response time- and accuracy-based performance
402 metrics, they were not susceptible to the biases mentioned previously to the same degree as
403 subjective self-report measures. This was likely a major factor contributing to their prediction
404 strength being substantially higher than the clinical measures' (clinical measures' mean $\rho = 0.168$;
405 cognitive measures' mean $\rho = 0.289$). It is also possible that the networks that support cognition
406 are less variable than the ones that support psychiatric symptoms, or that the instruments used
407 to model symptoms are more flawed and reflect brain functioning to a lesser extent than cognitive
408 tests. Both possibilities could contribute to this difference in prediction strength. Regardless, the
409 higher prediction strength for cognitive tests is accompanied by greater confidence in derived
410 brain network correlates.
411
412 The networks that predicted cognitive measures also showed considerable overlap with each
413 other, suggesting that these common networks track with the g-factor commonly reported in the
414 literature[58]. Given that all the tests probe numerous cognitive processes, this g-factor network
415 likely reflects a mix of the various cognitive components that interact with each other to facilitate
416 test performance. The nodes important for positively predicting cognitive performance often lied
417 at the opposite end of the principal cortical functional gradient[44] from negatively predicting nodes,
418 highlighting a departure from the nature of clinical networks, where the same nodes were often
419 important to positively and negatively predicting measures. Interestingly, this pattern did not apply
420 to cerebellar nodes, as areas of the cerebellum that were involved in positively predicting
421 cognition were also commonly involved in negatively predicting cognition. Furthermore, our

422 findings also strongly implicate the SM network in supporting cognitive functioning, in line with the
423 recent report that systems supporting higher-order cognition are interdigitated within motor
424 cortex[59].
425
426 **Other factors contributing to prediction strength**
427 Recent work has suggested that measurement stability, as reflected by ICC, is a major component
428 underlying predictive modeling performance[18]. We directly tested this by correlating the prediction
429 strength of clinical and cognitive measures with the ICCs reported by the authors of those scales.
430 We observed a very strong effect of instrument/subdomain on predictions, while an effect of ICC
431 was absent across all measures, including those that failed to model well (see **Fig 3**). This does
432 not rule out the possibility that measurement stability is an important factor contributing to brain-
433 based predictive modeling success. Rather, it suggests that if a measurement is sufficiently
434 stable, then we can generate a valid brain-behavior model. Most of the ICC values in the
435 measures we used were high (>0.6). Thus, the effect that this could have had on prediction
436 performance would be minimal, and the effect of the phenotype dominates.
437
438 Additionally, CPM can directly test how the development of the measures can lead to measurable
439 differences in the extent to which they engage clear brain correlates[60]. As demonstrated here,
440 there were 4 instances where different instruments provided measurements with the same label
441 that purportedly measured the same phenotype but yielded widely different predictions (see **Fig**
442 **1**). These findings highlight how the framing of the questions probing symptoms or how index
443 scores composed of multiple related measures can improve the confidence of the network
444 correlates derived from predictive modeling.
445

446 **Relationships between clinical and cognitive measures**

447 While a large body of work has demonstrated cognitive deficits in psychiatric populations, the
448 findings here indicate that the relationships between symptom and cognitive measures were weak
449 (maximum $\rho = 0.234$, minimum $\rho = -0.227$). Only 10.22% of clinical-cognitive measure pairs were
450 significant after correcting for multiple comparisons. Furthermore, the effect sizes of the
451 relationships that were significant were small. Given that the ICC values for these measures were
452 all greater than 0.6, this finding is probably not a result of measurement noise, but rather likely
453 reflects the highly variable cognitive alterations within clinical populations[16,19,21,22].

454

455 When examining the clinical-cognitive relationships more closely, some interesting patterns
456 arose. Significant correlations between clinical and cognitive measures were evenly distributed
457 across clusters, suggesting that these measures might be related through more general p- and
458 g-factor correlations. However, not all clinical measures were associated with poorer cognitive
459 performance. In fact, better performance on cognitive tests that contained a language component
460 (e.g., the BNT) were typically correlated with worse affect. This positive association with the BNT
461 extended to depression symptoms. One potential interpretation of this finding could be that
462 enhanced verbal abilities feed into more ruminative thoughts and higher levels of depression
463 symptoms[61]. Interestingly, depression symptoms were also negatively associated with
464 performance on WAIS digit symbol coding, a test that assesses cognitive control. This
465 dissociation between how depression symptoms were positively correlated with lexical retrieval
466 (as assessed by BNT) but negatively correlated with cognitive control could provide useful
467 information for studies investigating ways to identify depression biotypes. The implications of how
468 this can inform precision medicine frameworks should be further explored, especially in light of
469 recent evidence that cognitively impaired depression patients respond differently to
470 antidepressants than patients without cognitive impairment[21].

471

472 **Shared networks mediate clinical-cognitive relationships**

473 Mirroring the correlations between clinical and cognitive measures, the network overlaps were
474 also sparse. This finding is consistent with a recent report that predictive network features are
475 distinct across mental health and cognitive measures[62]. Our results extend these findings by
476 revealing that these sparsely overlapping networks mediate clinical-cognitive relationships. These
477 findings suggest that varying degrees of dysfunction within overlapping networks could explain
478 the heterogeneous cognitive deficits seen in psychiatric populations[16,19,21,22]. This aligns with the
479 RDoC hypothesis stating that disordered circuits within an individual likely do not impinge directly
480 on a single network, but instead, they partially affect multiple networks linked to various symptoms
481 and cognitive measures. Identifying these overlapping networks can guide cognitive intervention
482 personalization in clinical populations, which often assume that symptomatology stems from
483 cognitive deficits[19] and that cognitive remediation can improve symptoms[63].

484

485 Even though significant clinical-cognitive relationships were mostly evenly spread across cluster-
486 cluster pairs, nodes localized to the thalamus, cerebellum, and somatomotor network were most
487 important in mediating these relationships. Bilateral dLPFC was also involved, corroborating
488 recent findings that individuals with treatment resistant depression who also have cognitive
489 impairments respond better to TMS applied to dLPFC than patients with intact cognition[22].
490 Additionally, evidence has emerged recently suggesting that the thalamus, cerebellum, and
491 somatomotor networks may be promising treatment targets for a variety of symptoms across
492 several populations, such as depression, psychosis, and obsessive-compulsive disorder[64-69].
493 Interestingly, although the thalamus and cerebellum were not necessarily the most important
494 nodes to predicting cognitive performance measures, our results suggest they were the most
495 important hubs for generally maintaining clinically-relevant cognitive functioning. Future work

496 should further examine whether neuromodulation or selective pharmacological manipulation
497 targeting neurons in these areas can effectively improve cognitive deficits in psychiatric
498 populations. The patterns by which individual pairs of overlapping clinical-cognitive networks
499 deviated from this common mediating network may provide additional translational insights, such
500 as which networks underlie the opposing relationships between depression symptoms and lexical
501 retrieval versus cognitive control.
502
503 This study highlights how this approach can be useful for refining networks identified by brain-
504 behavior modeling to pinpoint the specific networks that track with cognitive deficits co-occurring
505 with clinical symptomatology. Our findings corroborate a growing body of literature highlighting
506 that the thalamus, cerebellum, somatomotor network, and dLPFC are important to both clinical
507 symptomatology and cognition. Studies probing the causal nature of these relationships and their
508 implications in precision medicine are warranted.
509
510 **Limitations**
511 There are some noteworthy limitations to the current study. First, without an external sample to
512 validate our results, we cannot be certain that the network correlates of various behavioral
513 measures will generalize to unseen data. This is one of the first studies to have a large
514 transdiagnostic sample, extensive fMRI data (>48 min), and comprehensive phenotyping
515 conducted outside the magnet. Therefore, similar data to evaluate whether these results
516 generalize is lacking. Second, some of the studies that we obtained ICC values from were
517 inconsistent in the time intervals between test and re-test, sample sizes, and age groups utilized.
518 These differences could have biased ICC values, potentially obscuring any effects on prediction
519 accuracy. Third, we acknowledge that there are many different methods to determine “optimal”
520 clustering solutions that each typically yield widely different results, and that there is often no clear

521 determination of how many clusters exist[70]. The clustering performed here was solely intended
522 for reducing the dimensionality of the data and showing how similar the measures were to each
523 other based on their network correlates. Their labels were therefore only used to help describe
524 the types of measures composing each cluster. Fourth, the biases introduced via self-reporting
525 reduce the prediction strength of clinical measures and add noise to our network estimates for
526 those measures. However, most instruments available for assessing psychiatric symptoms in
527 clinical contexts rely on self-reporting, and objective assessments are lacking. Finally, our results
528 beg a deeper investigation into the relationship between symptoms and cognition. We have
529 assumed here that these measures reflect fixed traits, but it is well known that symptoms and
530 cognition wax and wane in clinical populations. Understanding how these models may vary, and
531 how the relationship between symptoms and cognitive variables interact over time could be very
532 informative in clinical contexts[71].
533
534 **Methods**
535 **Participants**
536 The sample reported in this study includes data from 317 demographically diverse participants
537 with at least 48 minutes of high-quality functional MRI data and extensive phenotyping outside of
538 the magnet. Participants were recruited from advertisements broadly distributed throughout the
539 New Haven community and referrals from Yale clinics. This transdiagnostic population
540 experienced a wide range of symptoms and symptom severities, often with multiple psychiatric
541 diagnoses. All participants provided written informed consent in accordance with a protocol
542 approved by the Yale IRB. A more thorough description of this dataset can be found elsewhere[26].
543
544 **Imaging acquisition protocol**

The imaging data was collected at Yale on a 3T Siemens Prisma scanner with a 64-channel head coil. A high-resolution T1-weighted MPRAGE (TR=2,400 ms, TE=1.22 ms, flip angle=8°, voxel size 1 mm^3) was acquired for each participant. The functional data were obtained using a multiband EPI sequence (TR=1,000 ms, TE=30 ms, flip angle=55°, slice thickness=2 mm, multiband factor=5). A total of 6-task and 2-resting-state runs were acquired, each lasting 6 minutes and 49 seconds. The first and last functional scans were resting-state runs. The six task runs were all designed as event-related designs and comprised a working memory (n-back), an inhibition (stop signal task), a decision making (card guessing), an emotional perception (eyes), a continuous performance task (the gradCPT), and a more naturalistic movie watching task. The task order was randomized and counterbalanced across participants. There was no block structure to the tasks and they were designed as state-manipulation tasks for functional connectivity analyses unlike a typical task paradigm designed for conventional GLM analyses. We ensured that participants understood task instructions and received practice. Responses were recorded on a button box. Analyses were restricted to participants who completed all fMRI scans runs.

Functional MRI data processing

Skull stripping of the structural scans was done using an optimized version of the FMRIB's Software Library (FSL) pipeline[72,73]. Motion correction was done using SPM12[74]. Nonlinear registration of the MPRAGE to the MNI template was performed using BioImage Suite[75], and linear registration of the functional to the structural images was done through a combination of FSL[76] and BioImage Suite. The remaining preprocessing steps were performed in BioImage Suite, including global signal regression, high-pass and low-pass filtering, and regression of motion parameters. All registered data were visually examined to ensure whole-brain coverage, adequate registration, and the absence of artifacts or other quality issues. Subjects were included

570 in the study if they completed all eight fMRI scan runs and had a grand mean frame-to-frame
571 displacement of less than 0.15 mm and a maximum mean frame-to-frame displacement of less
572 than 0.2 mm. Two participants were scanned with a slightly shorter scanning time (25 seconds
573 shorter) and were included in the dataset.
574
575 The Shen 268 node atlas[40] was applied to the preprocessed data, parcellating it into 268
576 functionally coherent nodes. Pearson correlations of the time series between all node pairs were
577 computed and subsequently z-transformed to generate 8 functional connectivity matrices for each
578 participant. The averages of all functional connectivity matrices were computed for each
579 run/participant. Since individuals' functional connectomes are more similar to themselves,
580 regardless of the task, than they are to other individuals[3], the average connectome will still likely
581 contain relevant information about an individual's trait characteristics. Analyses and visualizations
582 were conducted using both MATLAB and python.
583
584 **Clinical assessments**
585 A battery of commonly used self-report questionnaires probing a wide range of clinically relevant
586 measures were collected immediately following the MRI scans (same day). These assessments
587 include the Adult Temperament Questionnaire (ATQ)[27], the Behavior Rating Inventory of
588 Executive Function (BRIEF)[28], the Brief Symptom Inventory (BSI)[29], the Interpersonal Reactivity
589 Index (IRI)[30], the Positive and Negative Affect Schedule (PANAS)[31], the Pittsburgh Sleep Quality
590 Index (PSQI)[32], and the Perceived Stress Scale (PSS)[33]. Two of the questionnaires (BSI and
591 BRIEF) provided both raw and age and sex normalized scores. In this study, we used the raw
592 scores for neuro-behavioral modeling to maintain consistency across clinical measures. Some
593 participants did not complete all clinical assessment scales. They were included in all analyses

594 for which they had data. A detailed breakdown of the number of subjects (*n*s) for each measure
595 can be found in **Supplemental Table 9**.
596
597 **Cognitive assessments**
598 A series of neuropsychological evaluations was administered by a trained evaluator to test a wide
599 array of cognitive functions. These were conducted in the same post-imaging session as the
600 clinical assessments. In total, we administered the verbal fluency, trail making, color-word
601 interference, and twenty questions subtests from the Delis-Kapan Executive Function Scale (D-
602 KEFS)[34], the letter-number sequencing, symbol search, and cancellation subtests from the
603 Wechsler Adult Intelligence Scales (WAIS)[35], the vocabulary and matrix reasoning subtests of the
604 Wechsler Abbreviated Scale of Intelligence 2nd Edition (WASI)[36], the finger windows and list
605 learning subtests of the Wide Range Assessment of Memory and Learning 2nd Edition
606 (WRAML)[37], the reading subtest of the Wide Range Achievement Test 5th Edition (WRAT)[38], and
607 the Boston Naming Test 2nd edition (BNT)[39]. A few of the cognitive tests provided both raw and
608 age and sex normalized scores. In this study, we used the raw scores for predictive modeling to
609 maintain consistency across all measures. Some participants did not complete all cognitive tests.
610 They were included in all analyses for which they had data. A detailed breakdown of the number
611 of subjects (*n*s) for each measure can be found in **Supplemental Table 10**.
612
613 **Predictive modeling**
614 We used CPM[3,10] to identify functional brain networks that were related to each clinical and
615 cognitive measure separately. Each predictive model was trained and tested in a 10-fold cross-
616 validation, which was repeated 1000 times to obtain a distribution of predictions. Within each
617 training fold, edges were selected if they were significantly correlated ($p < 0.05$) with behavior

618 after controlling for age and sex using partial correlations. Prediction strength was determined by
619 comparing the observed clinical/cognitive measures to the predicted measures using Spearman
620 correlations. Permutation tests were utilized for significance testing. Specifically, the number of
621 null prediction values that were greater than the median prediction for each measure was divided
622 by the number of permutations to derive the *p*-values. The *p*-values were FDR-corrected for 91
623 comparisons (63 clinical + 28 cognitive measures). Networks for each measure were derived by
624 identifying the edges that were predictive in >50% of permutations. The resulting networks were
625 binarized, with a '1' or '0' indicating an edge's network membership.
626
627 **Assessing the effects of instruments and subdomains on prediction strength**
628 We utilized one-way ANOVAs to assess the effect of measurement instruments and subdomains
629 on prediction strength for clinical and cognitive measures separately (**Fig 3a/b**). To do so,
630 prediction strength values from measures within the same clinical assessment subdomain or
631 cognitive testing battery were grouped together and compared to other subdomains/tests. The
632 ATQ, BRIEF, and PANAS inventories explicitly specified that the individual measures belonged
633 to separate subdomains (e.g., PANAS measures were split into 'positive' and 'negative' affect
634 domains). These separate subdomains were treated as separate categories for this analysis. The
635 BSI, IRI, PSQI, and PSS inventories did not specify how measures could be grouped into separate
636 subdomains. Therefore, all measures within these instruments were categorized as belonging to
637 the same subdomain for this analysis. Several of the cognitive testing batteries were composed
638 of sets of multiple separate cognitive tests. The D-KEFS consisted of the twenty-questions test,
639 the color-word interference test, the verbal fluency test, and the trail making test. Additionally, the
640 WASI was composed of a vocabulary test and a matrix reasoning test, the WRAML was
641 composed of a verbal learning test and a finger windows test, and the WAIS consisted of
642 cancellation, letter-number sequencing, digit-symbol coding, and symbol search tests. These

643 were all categorized as belonging to different subdomains for this analysis, since they were all
644 different cognitive tests. Conversely, all the metrics from the BNT captured various aspects of
645 BNT performance and were thus considered as belonging to the same subdomain.
646
647 **Obtaining measurement reliability data**
648 Test-retest reliability (ICC) measures were obtained from the original papers and/or technical and
649 user manuals published for each clinical and cognitive inventory. The BRIEF provided separate
650 ICC data from a normative sample and a mixed clinical/healthy adult population. We used the
651 ICC data from the mixed clinical/healthy adult sample, since it more closely matched our study's
652 participant population. The original paper validating the ATQ provided ICC data collected from 2
653 separate studies, one conducted in undergraduate students and the other in adults between 26
654 and 91 years old (median age = 57 years). We utilized the data from the study conducted in
655 undergraduate students because it provided ICC data for all the ATQ measures that we used
656 here, while the other study did not. Additionally, our sample's median age (31.17 years) is closer
657 to the median age of typical undergraduate populations. The IRI reported ICC data from males
658 and females separately. We averaged those results for use in the analysis here. There were no
659 ICC data reported for the 'total questions asked' measure from the twenty-questions subtest in
660 the D-KEFS, the verbal learning intrusion measure in the WRAML, or for the performance
661 measures on the BNT (except for 'total score'). Therefore, these measures were omitted from
662 analyses examining the relationship between ICC and prediction performance.
663
664 **Identifying relationships between measures and calculating network similarity**
665 Spearman correlations were calculated for each pair of measures that CPM significantly
666 predicted, and the derived *p*-values were FDR corrected for multiple comparisons. Network

667 similarity between measures was assessed using Dice similarity coefficients. If there was a
668 positive relationship between two measures, then the Dice coefficient was calculated between
669 the positively predicting network for measure A and the positively predicting network for measure
670 B, and then calculated once more between the negatively predicting network for measure A and
671 the negatively predicting network for measure B. If there was a negative relationship between two
672 measures, then the Dice coefficient was calculated between the positively predicting network for
673 measure A and the negatively predicting network for measure B, and then calculated once more
674 between the negatively predicting network for measure A and the positively predicting network for
675 measure B. A weighted average of these two Dice coefficients was taken and used to examine
676 overall network similarity between measures.
677
678 **Comparing network similarity across measures**
679 We sorted the measures based on network similarity in the interest of identifying meaningful
680 commonalities and differences between predictive networks across measures. To achieve this,
681 we input the inverse of this full Dice similarity matrix (i.e., Dice distances) into an agglomerative
682 hierarchical clustering algorithm using Ward's method. This yielded a dendrogram that sorted the
683 measures based on their predictive networks' similarities. The measures were grouped into
684 clusters based on the hierarchical clustering results. The number of clusters chosen was
685 determined using the elbow method[42], which evaluates where the incremental decrease in within-
686 cluster sum of squares (WCSS) begins to slow across increasing values of *k*. The point of
687 maximum curvature (i.e., the "elbow") was objectively identified using the KneeLocator
688 algorithm[77]. This is a commonly used approach for determining an appropriate number of clusters
689 in a dataset. However, we performed clustering with the intention of reducing dimensionality,
690 showing how the measures grouped based on network similarity, and to compare important

features driving CPM predictions across measures, rather than to make any claims about how many clusters exist in this dataset.

To identify the extent to which the same nodes were important to positively and negatively predicting clinical and cognitive measures, we calculated the Jaccard similarity between vectors representing the top 10% highest degree nodes in the positive and negative predictive networks for each cluster. Overlap was summarized as the mean of the pairwise Jaccard values, and one-sided permutation tests (10,000 iterations) were utilized for significance testing. The *p*-values were calculated as the proportion of null replicates that exceeded the observed mean Jaccard similarity.

Identifying localizations of predictive networks

The binarized positively and negatively predicting networks for measures within a cluster were summed separately and Z-scored. This yielded aggregate positively and negatively predicting networks for each cluster, with edges most important to predictions across measures weighted most heavily. Before summing the networks, we ensured that the directionality of all the measures within a cluster was similar so that networks positively related to symptoms or cognition would not be mistakenly combined with networks negatively related to symptoms or cognitions. For example, if there was one measure in a cluster where a higher score was desirable while lower scores were desirable for the remaining measures, then the positive network for the first measure would be summed with the negative networks for the other measures (and vice versa). The distribution of edges within and between canonical networks was evaluated by summing the edges within each canonical network pair from the aggregated cluster networks. We utilized the Yeo17 networks[43] as network definitions, plus the cerebellum, brainstem, and subcortical regions. The nodes important to predictions were examined by calculating degree centrality for each node.

716 The degree values were then projected onto the FreeSurfer template brain's ('fsaverage') pial
717 surface, along with cerebellar and brainstem surface reconstructions rendered in Bioimage Suite.
718 These degree plots were created using Python's nilearn package[78].
719
720 **Network mediators of relationships between clinical and cognitive measures**
721 For each pair of significantly correlated clinical-cognitive measures, we computed the shared
722 network connectivity by identifying and summing the edges that were predictive for both
723 measures. If the correlation between a pair of clinical-cognitive measures was positive, then the
724 shared network connectivity was defined as the sum of the positively predicting edges shared by
725 both measures minus the sum of negatively predicting edges shared by both measures.
726 Conversely, if the correlation between the measures was negative, then the shared network
727 connectivity was defined as the sum of edges that were positively predicting for one measure but
728 negatively predicting for the other minus the sum of edges that were negative predicting for the
729 first measure but positively predicting for the other. Causal mediation analyses were conducted
730 using Python's pingouin package[79] to test whether the shared network connectivity mediated
731 relationships between clinical and cognitive measures. Path coefficients were estimated with
732 ordinary least squares regression, and the indirect effect was evaluated using 1,000 bootstrap
733 resamples with a two-tailed $\alpha = 0.01$. Significance was determined by whether the 99% bootstrap
734 confidence interval excluded zero. The mediation analyses were conducted bidirectionally for all
735 pairs of significantly correlated clinical-cognitive relationships, and since we did not have any
736 hypotheses on the directionality of the mediations, we reported the larger of the two indirect
737 mediation coefficients.
738
739 The nodes important to mediating these relationships were examined by 1) deriving binarized
740 matrices for each pair of significantly correlated clinical-cognitive measure pairs that indicated

741 whether each edge was predictive in both measures, 2) summing these matrices across all
742 significantly correlated clinical-cognitive measure pairs within each clinical-cognitive cluster, 3) Z-
743 scoring this summed matrix, and 4) calculating the degree centrality for each node. The degree
744 centrality of the summed shared networks for all significantly correlated clinical-cognitive measure
745 pairs within each clinical-cognitive cluster was visualized on the brain surface the same way as
746 before using Python's nilearn package[78].
747

748 **Acknowledgements**

749 This work was supported by funding from the National Institutes of Health (MH121095 to D.S. and
750 R.T.C). AJS is also supported by the National Science Foundation's Graduate Research
751 Fellowship Program. We would like to thank Xenophon Papademetris for help with data
752 visualization and Sydney Smith for helpful conversations about data interpretation.
753

754 **Code availability**

755 The code used in all of the main analyses conducted in this study can be found at:
756 https://github.com/YaleMRRC/TDS_measure_predictions_and_mediations.git. All dependency
757 files required to run this code are also provided in this GitHub repository (e.g., cortical surface
758 files), as is the relevant input data.
759

760 **Competing interests**

761 G.S. has served as a consultant or scientific advisory board member to Axsome Therapeutics,
762 Biogen, Biohaven Pharmaceuticals, Boehringer Ingelheim International, Bristol-Myers Squibb,
763 Clexio, Cowen, Denovo Biopharma, ECR1, EMA Wellness, Engrail Therapeutics, Gilgamesh,
764 Janssen, Levo, Lundbeck, Merck, Navitor Pharmaceuticals, Neurocrine, Novartis, Noven
765 Pharmaceuticals, Perception Neuroscience, Praxis Therapeutics, Sage Pharmaceuticals, Seelos

766 Pharmaceuticals, Vistagen Therapeutics and XW Labs; and received research contracts from
767 Johnson & Johnson (Janssen), Merck and Usona. G.S. holds equity in Biohaven Pharmaceuticals
768 and is a co-inventor on a US patent (8,778,979) held by Yale University and a co-inventor on US
769 provisional patent application no. 047162-7177P1 (00754), filed on 20 August 2018 by Yale
770 University Office of Cooperative Research. Yale University has a financial relationship with
771 Janssen Pharmaceuticals and may receive financial benefits from this relationship. The University
772 has put multiple measures in place to mitigate this institutional conflict of interest. Questions about
773 the details of these measures should be directed to Yale University's Conflict of Interest office.
774 V.H.S. has served as a scientific advisory board member to Takeda and Janssen. The remaining
775 authors declare no competing interests.

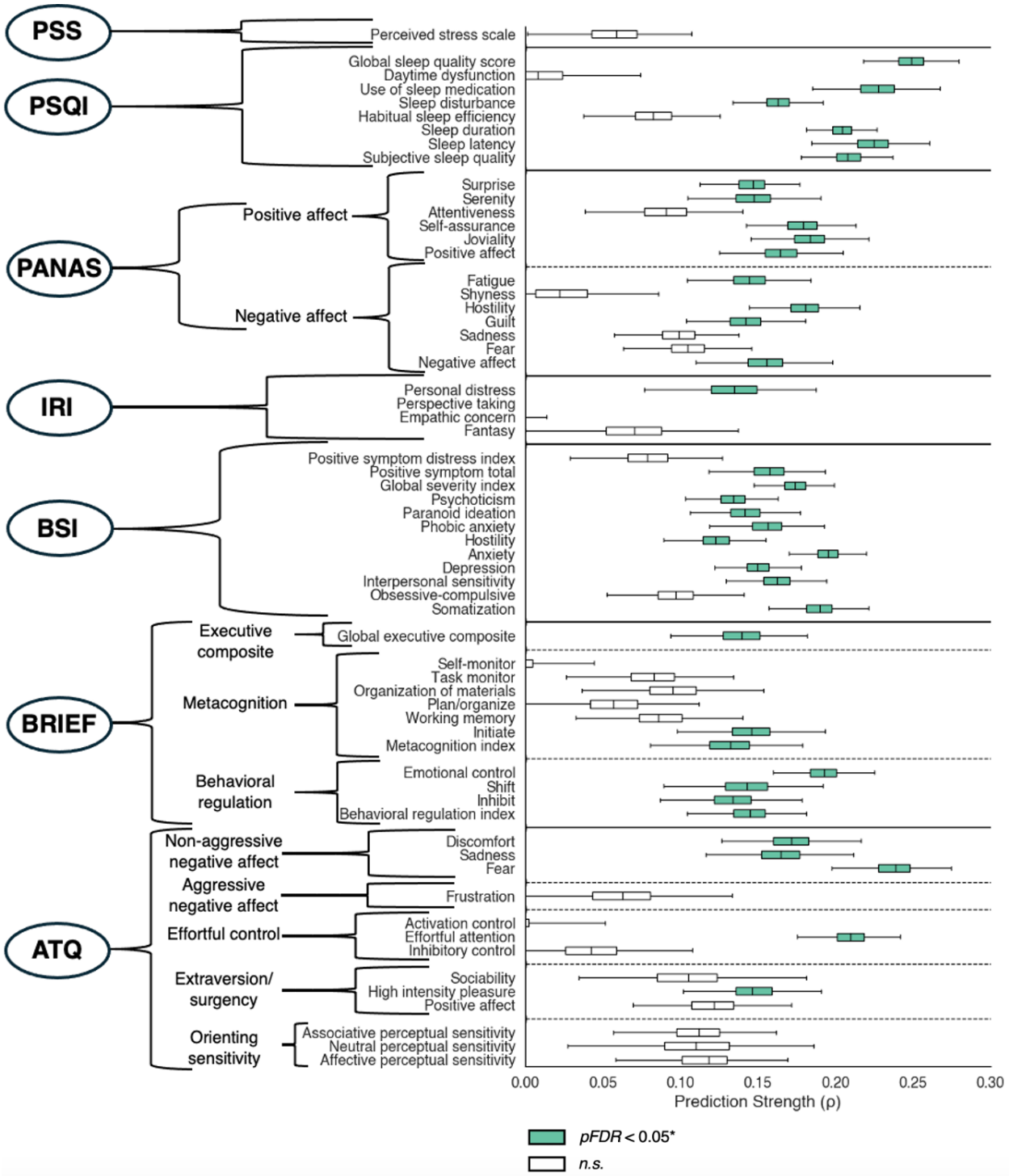


Figure 1. Prediction strength of clinical measures. The boxplots on the right show the prediction strength of each measure and are colored green if significantly predicted by CPM (*pFDR* < 0.05). Measures are grouped according to the instrument they belong to, as indicated by the encircled bolded acronym to the left of the outermost brackets, and the solid horizontal black lines between the boxplots. Measures within instruments were grouped into subdomains if applicable, which are indicated by the middle-level of brackets and the dashed black lines between the boxplots.

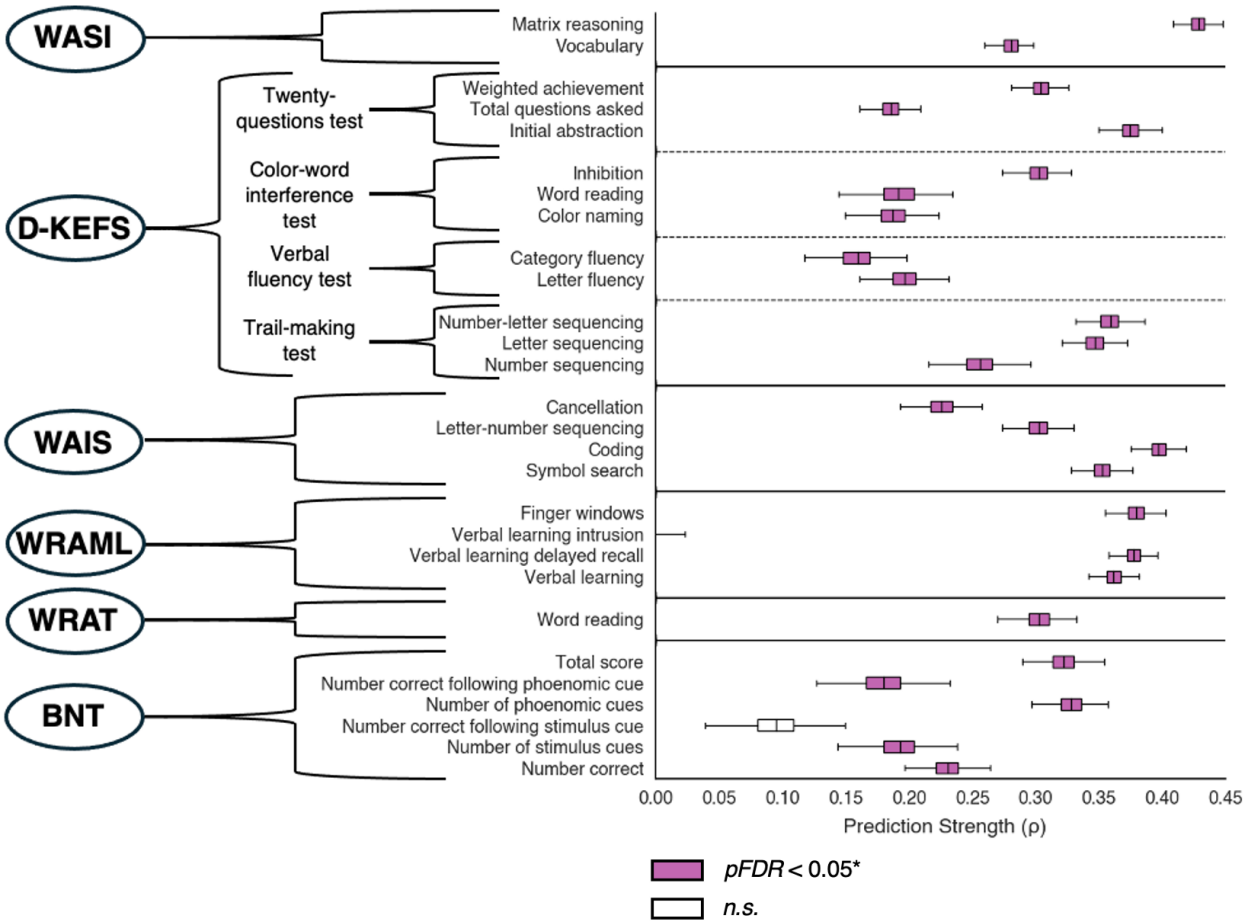


Figure 2. Prediction strength of cognitive measures. The boxplots on the right show the prediction strength of each measure and are colored purple if significantly predicted by CPM (*pFDR* < 0.05). Measures are grouped according to the testing battery they belong to, as indicated by the encircled bolded instrument acronym to the left of the outermost brackets, and the solid horizontal black lines between the boxplots. Measures within the D-KEFS were further subdivided because this battery was composed of subtests that each yielded multiple performance metrics. This is indicated by the middle-level of brackets and the dashed black lines between the boxplots.

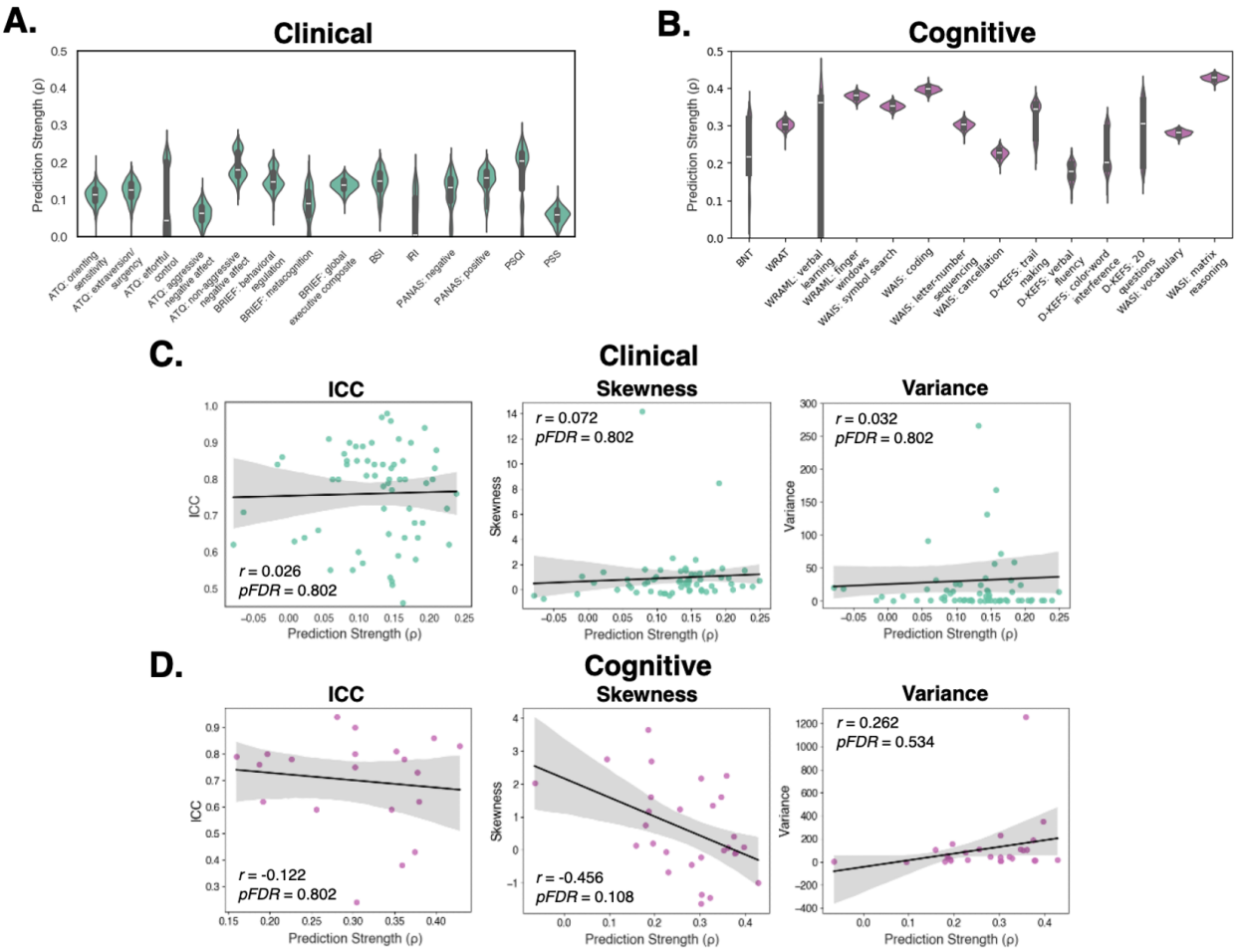


Figure 3. **Prediction strength was driven by phenotype measured. A)** Clinical predictions varied significantly by instrument and subdomain (one way ANOVA: $p < 0.001$). **B)** Cognitive predictions varied significantly by tests and subtests (one way ANOVA: $p < 0.001$). Prediction strength was unrelated to other measurement characteristics, such as test-retest reliability (ICC) (top), skewness (middle), and variance (top) for clinical (panel **C**) and cognitive measures (panel **D**).

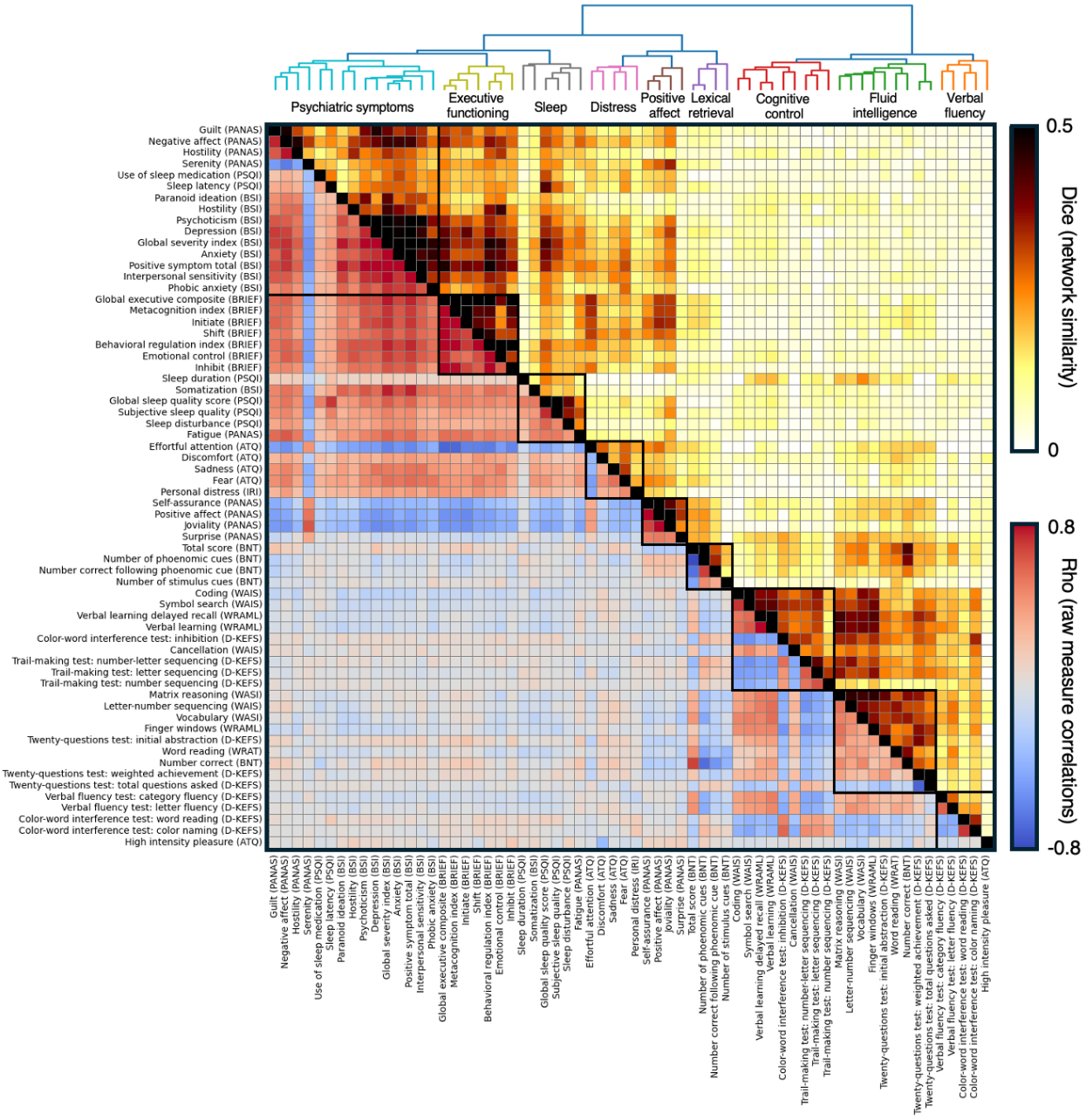


Figure 4. Relationships between measures. The measures' predictive networks were sorted according to their Dice similarities by an agglomerative hierarchical clustering algorithm (upper triangle). A dendrogram displayed above the matrix highlights how the measures grouped together based on their network similarity. Descriptive labels were given to the 9 clusters and displayed underneath the dendrogram. The measures belonging to each cluster were surrounded black borders on the matrix and colored differently on the dendrogram. The clinical measures were displayed on the top/left and the cognitive measures were displayed on the bottom/right. The network similarity was mirrored by the correlational structure between the raw measures (lower triangle).

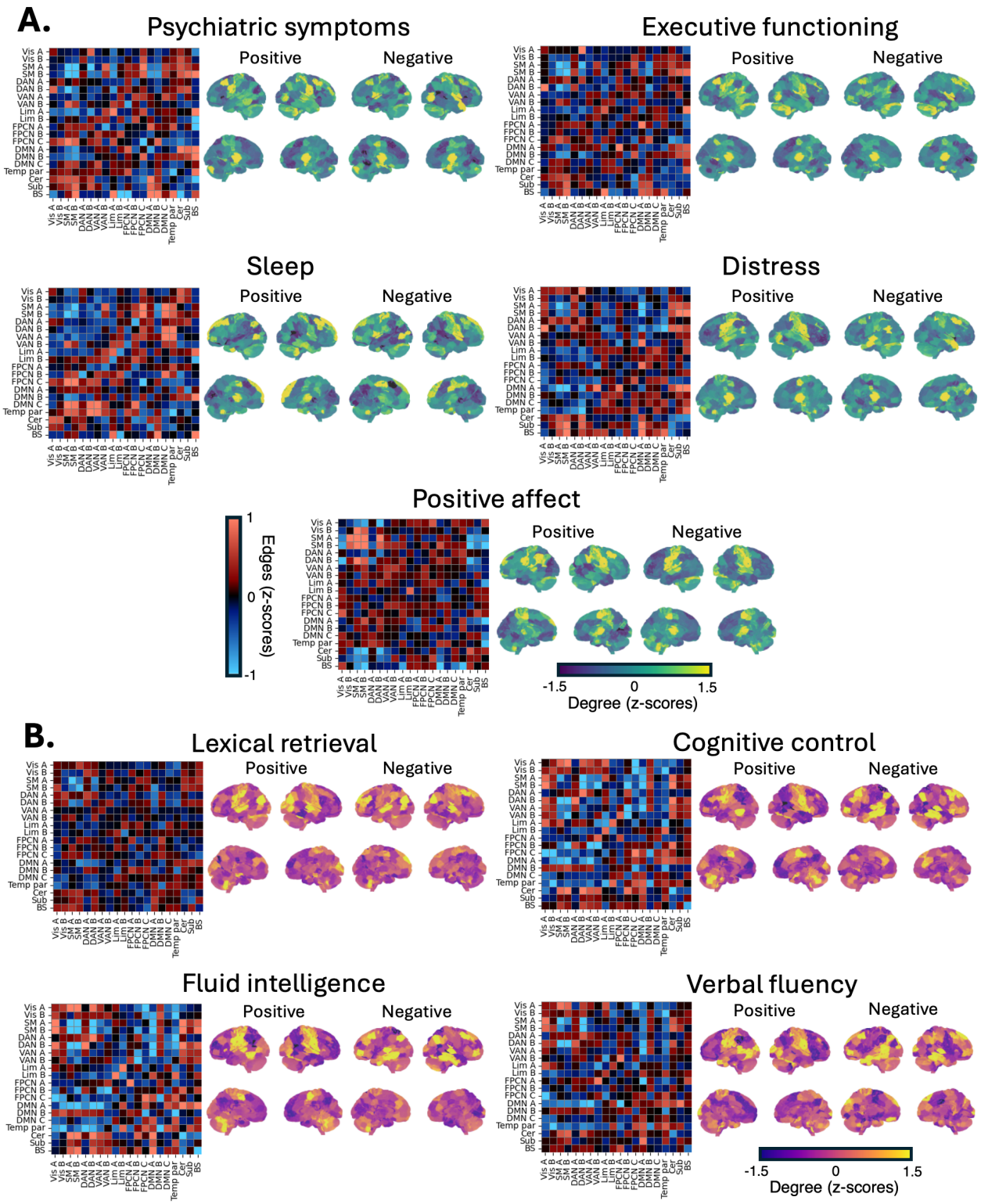
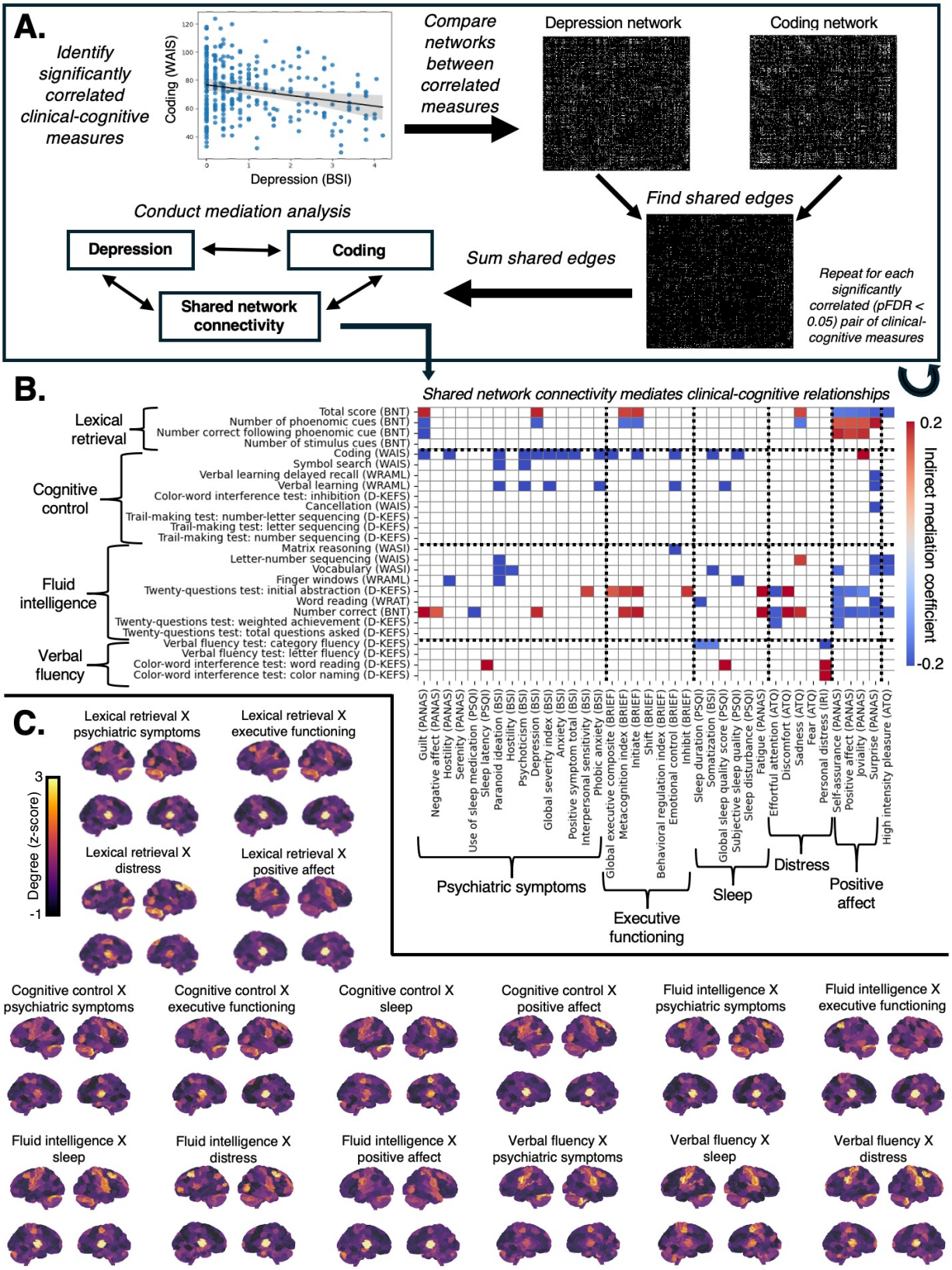


Figure 5. Brain networks driving predictions. A) The clusters summarizing networks important to predicting clinical measures are displayed with the descriptive labels 'psychiatric symptoms', 'executive functioning', 'sleep', 'distress', and 'positive affect'. **B)** The clusters summarizing networks important to predicting cognitive measures are displayed with the descriptive labels

‘lexical retrieval’, ‘cognitive control’, ‘fluid intelligence, and ‘verbal fluency’. For each cluster, degree plots highlight the important nodes for predictions across all measures, and matrices highlight the distribution of edges connecting canonical networks. Networks are labeled according to the Yeo 17 network definition, plus cerebellar, subcortical, and brainstem regions.


A.
Identify significantly correlated clinical-cognitive measures
Coding (WAIS)
Depression (BSI)
Compare networks between correlated measures
Depression network
Coding network
Find shared edges
Sum shared edges
Repeat for each significantly correlated (pFDR < 0.05) pair of clinical-cognitive measures
Conduct mediation analysis
Depression
Coding
Shared network connectivity
B.
Shared network connectivity mediates clinical-cognitive relationships
Lexical retrieval
Cognitive control
Fluid intelligence
Verbal fluency
Total score (BNT)
Number of phoenomic cues (BNT)
Number correct following phoenomic cue (BNT)
Number of stimulus cues (BNT)
Coding (WAIS)
Symbol search (WAIS)
Verbal learning delayed recall (WRAML)
Verbal learning (WRAML)
Color-word interference test: inhibition (D-KEFS)
Cancellation (WAIS)
Trail-making test: number-letter sequencing (D-KEFS)
Trail-making test: letter sequencing (D-KEFS)
Trail-making test: number sequencing (D-KEFS)
Matrix reasoning (WASI)
Letter-number sequencing (WAIS)
Vocabulary (WASI)
Finger windows (WRAML)
Twenty-questions test: initial abstraction (D-KEFS)
Word reading (WRAT)
Number correct (BNT)
Twenty-questions test: weighted achievement (D-KEFS)
Twenty-questions test: total questions asked (D-KEFS)
Verbal fluency test: category fluency (D-KEFS)
Verbal fluency test: letter fluency (D-KEFS)
Color-word interference test: word reading (D-KEFS)
Color-word interference test: color naming (D-KEFS)
Guilt (PANAS)
Negative affect (PANAS)
Hostility (PANAS)
Serenity (PANAS)
Use of sleep medication (PSQI)
Sleep latency (PSQI)
Paranoid ideation (BSI)
Hostility (BSI)
Psychoticism (BSI)
Depression (BSI)
Global severity index (BSI)
Anxiety (BSI)
Positive symptom total (BSI)
Interpersonal sensitivity (BSI)
Phobic anxiety (BSI)
Global executive composite (BRIEF)
Metacognition index (BRIEF)
Initiate (BRIEF)
Shift (BRIEF)
Behavioral regulation index (BRIEF)
Emotional control (BRIEF)
Inhibit (BRIEF)
Sleep duration (PSQI)
Somatization (BSI)
Global sleep quality score (PSQI)
Subjective sleep quality (PSQI)
Sleep disturbance (PSQI)
Fatigue (PANAS)
Effortful attention (ATQ)
Discomfort (ATQ)
Sadness (ATQ)
Fear (ATQ)
Personal distress (IRI)
Self-assurance (PANAS)
Positive affect (PANAS)
Joviality (PANAS)
Surprise (PANAS)
High intensity pleasure (ATQ)
Psychiatric symptoms
Executive functioning
Sleep
Distress
Positive affect
0.2
-0.2
Indirect mediation coefficient
C.
Lexical retrieval X psychiatric symptoms
Lexical retrieval X executive functioning
Lexical retrieval X distress
Lexical retrieval X positive affect
Degree (z-score)
3
-1
Cognitive control X psychiatric symptoms
Cognitive control X executive functioning
Cognitive control X sleep
Cognitive control X positive affect
Fluid intelligence X psychiatric symptoms
Fluid intelligence X executive functioning
Fluid intelligence X sleep
Fluid intelligence X distress
Fluid intelligence X positive affect
Verbal fluency X psychiatric symptoms
Verbal fluency X sleep
Verbal fluency X distress

825 **Figure 6. Shared networks mediate clinical-cognitive relationships. A)** An overview of how
826 the shared networks were identified and how mediation analyses were conducted for significantly
827 correlated pairs of clinical and cognitive measures. Predictive networks for each measure in a
828 correlated pair were compared, and edges that were predictive for both measures were summed.
829 This sum was used as a mediating variable to test whether the shared network connectivity
830 mediated the relationships between the correlated clinical and cognitive measure pairs. This was
831 repeated for every pair of significantly correlated pair of clinical and cognitive measures ($pFDR <$
832 0.05). **B)** This matrix displays the indirect mediation coefficients for all the significantly correlated
833 clinical-cognitive measure pairs. The shared network connectivity significantly mediated every
834 pair of significant clinical-cognitive correlations. Measures were grouped by their cluster
835 assignment in the visualization. **C)** Average degree centrality for the mediating networks across
836 all clinical-cognitive measures within all possible pairs of clusters were computed and projected
837 onto cortical surfaces to highlight the regions important for mediating clinical-cognitive
838 relationships within cluster pairs.

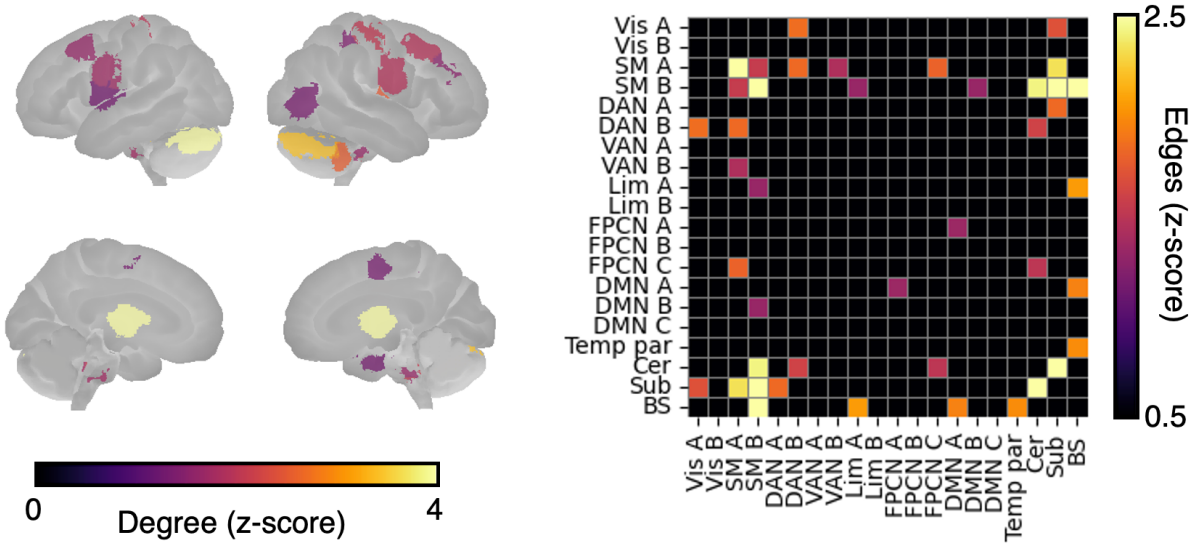


Figure 7. Networks important to generally mediating clinical-cognitive relationships. Average degree centrality for all mediating networks was computed and projected onto a cortical surface to highlight the regions important for generally mediating clinical-cognitive relationships (left). The surface plot was thresholded to display the top 10% most important nodes for mediations. The matrix on the right highlights the distribution of edges connecting canonical networks. The matrix was thresholded to only display the top 10% most important canonical network-network connections for mediations. Networks were labeled according to the Yeo 17 network definition, plus cerebellar, subcortical, and brainstem regions.

References

1 Insel, T. *et al.* Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry* **167**, 748-751 (2010). https://doi.org:10.1176/appi.ajp.2010.09091379

2 Cuthbert, B. N. & Insel, T. R. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med* **11**, 126 (2013). https://doi.org:10.1186/1741-7015-11-126

3 Finn, E. S. *et al.* Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat Neurosci* **18**, 1664-1671 (2015). https://doi.org:10.1038/nn.4135

4 Dubois, J., Galdi, P., Paul, L. K. & Adolphs, R. A distributed brain network predicts general intelligence from resting-state human neuroimaging data. *Philos Trans R Soc Lond B Biol Sci* **373** (2018). https://doi.org:10.1098/rstb.2017.0284

5 Rosenberg, M. D. *et al.* A neuromarker of sustained attention from whole-brain functional connectivity. *Nat Neurosci* **19**, 165-171 (2016). https://doi.org:10.1038/nn.4179

6 Greene, A. S., Gao, S., Scheinost, D. & Constable, R. T. Task-induced brain state manipulation improves prediction of individual traits. *Nat Commun* **9**, 2807 (2018). https://doi.org:10.1038/s41467-018-04920-3

7 Kong, R. *et al.* Spatial Topography of Individual-Specific Cortical Networks Predicts Human Cognition, Personality, and Emotion. *Cereb Cortex* **29**, 2533-2551 (2019). https://doi.org:10.1093/cercor/bhy123

8 Sripada, C. *et al.* Prediction of neurocognition in youth from resting state fMRI. *Mol Psychiatry* **25**, 3413-3421 (2020). https://doi.org:10.1038/s41380-019-0481-6

9 Beaty, R. E. *et al.* Robust prediction of individual creative ability from brain functional connectivity. *Proc Natl Acad Sci U S A* **115**, 1087-1092 (2018). https://doi.org:10.1073/pnas.1713532115

10 Shen, X. *et al.* Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nat Protoc* **12**, 506-518 (2017). https://doi.org:10.1038/nprot.2016.178

11 Rosenberg, M. D. & Finn, E. S. How to establish robust brain–behavior relationships without thousands of individuals. *Nature neuroscience* (2022). https://doi.org:10.1038/s41593-022-01110-9

12 Barron, D. S. *et al.* Transdiagnostic, Connectome-Based Prediction of Memory Constructs Across Psychiatric Disorders. *Cereb Cortex* **31**, 2523-2533 (2021). https://doi.org:10.1093/cercor/bhaa371

13 Garrison, K. A. *et al.* Transdiagnostic Connectome-Based Prediction of Craving. *Am J Psychiatry* **180**, 445-453 (2023). https://doi.org:10.1176/appi.ajp.21121207

14 Ben-Zion, Z. *et al.* Connectome-Based Predictive Modeling of PTSD Development Among Recent Trauma Survivors. *JAMA network open* **8**, e250331-e250331 (2025).

15 Lake, E. M. R. *et al.* The Functional Brain Organization of an Individual Allows Prediction of Measures of Social Abilities Transdiagnostically in Autism and Attention-Deficit/Hyperactivity Disorder. *Biol Psychiatry* **86**, 315-326 (2019). https://doi.org:10.1016/j.biopsych.2019.02.019

16 Feczko, E. *et al.* The Heterogeneity Problem: Approaches to Identify Psychiatric Subtypes. *Trends Cogn Sci* **23**, 584-601 (2019). https://doi.org:10.1016/j.tics.2019.03.009

17 Tiego, J. *et al.* Precision behavioral phenotyping as a strategy for uncovering the biological correlates of psychopathology. *Nature Mental Health* **1**, 304-315 (2023).
18 Gell, M. *et al.* How measurement noise limits the accuracy of brain-behaviour predictions. *Nature Communications* **15**, 1-12 (2024).
19 Rock, P. L., Roiser, J. P., Riedel, W. J. & Blackwell, A. D. Cognitive impairment in depression: a systematic review and meta-analysis. *Psychol Med* **44**, 2029-2040 (2014). https://doi.org:10.1017/S0033291713002535
20 Segal, A. *et al.* Regional, circuit and network heterogeneity of brain abnormalities in psychiatric disorders. *Nat Neurosci* **26**, 1613-1629 (2023). https://doi.org:10.1038/s41593-023-01404-6
21 Hack, L. M. *et al.* A Cognitive Biotype of Depression and Symptoms, Behavior Measures, Neural Circuits, and Differential Treatment Outcomes: A Prespecified Secondary Analysis of a Randomized Clinical Trial. *JAMA Netw Open* **6**, e2318411 (2023). https://doi.org:10.1001/jamanetworkopen.2023.18411
22 Tozzi, L. *et al.* A cognitive neural circuit biotype of depression showing functional and behavioral improvement after transcranial magnetic stimulation in the B-SMART-fMRI trial. *Nat Ment Health* **2**, 987-998 (2024). https://doi.org:10.1038/s44220-024-00271-9
23 Tozzi, L. *et al.* Personalized brain circuit scores identify clinically distinct biotypes in depression and anxiety. *Nat Med* **30**, 2076-2087 (2024). https://doi.org:10.1038/s41591-024-03057-9
24 Fitzgerald, H. M. *et al.* Treatment Goals in Schizophrenia: A Real-World Survey of Patients, Psychiatrists, and Caregivers in the United States, with an Analysis of Current Treatment (Long-Acting Injectable vs Oral Antipsychotics) and Goal Selection. *Neuropsychiatr Dis Treat* **17**, 3215-3228 (2021). https://doi.org:10.2147/NDT.S330936
25 McNaughton, E. C. *et al.* Patient attitudes toward and goals for MDD treatment: a survey study. *Patient Prefer Adherence* **13**, 959-967 (2019). https://doi.org:10.2147/PPA.S204198
26 Samardzija, A. *et al.* A Transdiagnostic fMRI Dataset With 300+ Deeply Phenotyped Subjects Across Resting and Task States. *medRxiv*, 2025.2007. 2015.25331595 (2025).
27 Evans, D. E. & Rothbart, M. K. Developing a model for adult temperament. *Journal of research in personality* **41**, 868-888 (2007).
28 Roth, R. M., Isquith, P. K. & Gioia, G. A. Behavior Rating Inventory of Executive Function®--Adult Version. *Archives of Clinical Neuropsychology* (2005).
29 Derogatis, L. R. & Melisaratos, N. The Brief Symptom Inventory: an introductory report. *Psychol Med* **13**, 595-605 (1983).
30 Davis, M. H. A multidimensional approach to individual differences in empathy. (1980).
31 Watson, D., Clark, L. A. & Tellegen, A. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Pers Soc Psychol* **54**, 1063-1070 (1988). https://doi.org:10.1037//0022-3514.54.6.1063
32 Buysse, D. J., Reynolds, C. F., 3rd, Monk, T. H., Berman, S. R. & Kupfer, D. J. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res* **28**, 193-213 (1989). https://doi.org:10.1016/0165-1781(89)90047-4
33 Cohen, S., Kamarck, T. & Mermelstein, R. A global measure of perceived stress. *Journal of health and social behavior*, 385-396 (1983).
34 Delis, D. C., Kaplan, E. & Kramer, J. H. Delis-Kaplan executive function system. *Assessment* (2001).

35 Lichtenberger, E. O. & Kaufman, A. S. *Essentials of WAIS-IV assessment*. (John Wiley & Sons, 2012).
36 Wechsler, D. Wechsler abbreviated scale of intelligence. (1999).
37 Adams, W. Wide-Range Assessment of Memory and Learning. *The Corsini Encyclopedia of Psychology*, 1-2 (2010).
38 Wilkinson, G. S. & Robertson, G. J. Wide range achievement test 4. *Journal of Clinical and Experimental Neuropsychology* (1993).
39 Kaplan, E., Goodglass, H. & Weintraub, S. Boston naming test. *The Clinical Neuropsychologist* (1983).
40 Shen, X., Tokoglu, F., Papademetris, X. & Constable, R. T. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *Neuroimage* **82**, 403-415 (2013). https://doi.org:10.1016/j.neuroimage.2013.05.081
41 Tian, Y. & Zalesky, A. Machine learning prediction of cognition from functional connectivity: Are feature weights reliable? *Neuroimage* **245**, 118648 (2021). https://doi.org:10.1016/j.neuroimage.2021.118648
42 Cui, M. Introduction to the k-means clustering algorithm based on the elbow method. *Accounting, Auditing and Finance* **1**, 5-8 (2020).
43 Yeo, B. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol* **106**, 1125-1165 (2011). https://doi.org:10.1152/jn.00338.2011
44 Margulies, D. S. *et al.* Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 12574-12579 (2016). https://doi.org:10.1073/pnas.1608282113
45 Nisbett, R. E. & Wilson, T. D. Telling more than we can know: Verbal reports on mental processes. *Psychological review* **84**, 231 (1977).
46 Lira, B. *et al.* Large studies reveal how reference bias limits policy applications of self-report measures. *Scientific reports* **12**, 19189 (2022).
47 Heine, S. J., Lehman, D. R., Peng, K. & Greenholtz, J. What's wrong with cross-cultural comparisons of subjective Likert scales?: The reference-group effect. *Journal of personality and social psychology* **82**, 903 (2002).
48 Schober, M. F. & Conrad, F. G. Does conversational interviewing reduce survey measurement error? *Public opinion quarterly*, 576-602 (1997).
49 Kotov, R. *et al.* The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *J Abnorm Psychol* **126**, 454-477 (2017). https://doi.org:10.1037/abn0000258
50 Lahey, B. B. *et al.* Is there a general factor of prevalent psychopathology during adulthood? *Journal of abnormal psychology* **121**, 971 (2012).
51 Wang, B., LeBel, A. & D'Mello, A. M. Ignoring the cerebellum is hindering progress in neuroscience. *Trends in Cognitive Sciences* (2025).
52 Hwang, W. J. *et al.* Thalamic Connectivity System Across Psychiatric Disorders: Current Status and Clinical Implications. *Biol Psychiatry Glob Open Sci* **2**, 332-340 (2022). https://doi.org:10.1016/j.bpsgos.2021.09.008
53 Phillips, J. R., Hewedi, D. H., Eissa, A. M. & Moustafa, A. A. The cerebellum and psychiatric disorders. *Front Public Health* **3**, 66 (2015). https://doi.org:10.3389/fpubh.2015.00066

54 Kebets, V. *et al.* Somatosensory-Motor Dysconnectivity Spans Multiple Transdiagnostic Dimensions of Psychopathology. *Biol Psychiatry* **86**, 779-791 (2019). https://doi.org:10.1016/j.biopsych.2019.06.013

55 Keane, B. P. *et al.* Functional dysconnectivity of visual and somatomotor networks yields a simple and robust biomarker for psychosis. *Mol Psychiatry* **30**, 1539-1547 (2025). https://doi.org:10.1038/s41380-024-02767-3

56 Long, Y. *et al.* Altered resting-state dynamic functional brain networks in major depressive disorder: Findings from the REST-meta-MDD consortium. *Neuroimage Clin* **26**, 102163 (2020). https://doi.org:10.1016/j.nicl.2020.102163

57 Ray, D., Bezmaternykh, D., Mel'nikov, M., Friston, K. J. & Das, M. Altered effective connectivity in sensorimotor cortices is a signature of severity and clinical course in depression. *Proceedings of the National Academy of Sciences* **118**, e2105730118 (2021).

58 Colom, R., Jung, R. E. & Haier, R. J. Distributed brain sites for the g-factor of intelligence. *Neuroimage* **31**, 1359-1365 (2006). https://doi.org:10.1016/j.neuroimage.2006.01.006

59 Gordon, E. M. *et al.* A somato-cognitive action network alternates with effector regions in motor cortex. *Nature* **617**, 351-359 (2023). https://doi.org:10.1038/s41586-023-05964-2

60 Samardzija, A. *et al.* Testing the Tests: Using Connectome-Based Predictive Models to Reveal the Systems Standardized Tests and Clinical Symptoms are Reflecting. *bioRxiv*, 2024.2010.2024.619737 (2024). https://doi.org:10.1101/2024.10.24.619737

61 Karpinski, R. I., Kolb, A. M. K., Tetreault, N. A. & Borowski, T. B. High intelligence: A risk factor for psychological and physiological overexcitabilities. *Intelligence* **66**, 8-23 (2018).

62 Chen, J. *et al.* Shared and unique brain network features predict cognitive, personality, and mental health scores in the ABCD study. *Nat Commun* **13**, 2217 (2022). https://doi.org:10.1038/s41467-022-29766-8

63 Anguera, J. A., Gunning, F. M. & Areán, P. A. Improving late life depression and cognitive control through the use of therapeutic video game technology: A proof-of-concept randomized trial. *Depression and anxiety* **34**, 508-517 (2017).

64 Hua, J. P. Y., Abram, S. V. & Ford, J. M. Cerebellar stimulation in schizophrenia: A systematic review of the evidence and an overview of the methods. *Front Psychiatry* **13**, 1069488 (2022). https://doi.org:10.3389/fpsyt.2022.1069488

65 Yang, C. *et al.* The thalamus is the causal hub of intervention in patients with major depressive disorder: Evidence from the Granger causality analysis. *NeuroImage: Clinical* **37**, 103295 (2023).

66 Onofrj, M. *et al.* The central role of the Thalamus in psychosis, lessons from neurodegenerative diseases and psychedelics. *Transl Psychiatry* **13**, 384 (2023). https://doi.org:10.1038/s41398-023-02691-0

67 Hawken, E. R. *et al.* Transcranial Magnetic Stimulation of the Supplementary Motor Area in the Treatment of Obsessive-Compulsive Disorder: A Multi-Site Study. *Int J Mol Sci* **17**, 420 (2016). https://doi.org:10.3390/ijms17030420

68 Hu, Y. T. *et al.* Motor cortex repetitive transcranial magnetic stimulation in major depressive disorder - A preliminary randomized controlled clinical trial. *J Affect Disord* **344**, 169-175 (2024). https://doi.org:10.1016/j.jad.2023.10.058

69 Garg, S., Sinha, V. K., Tikka, S. K., Mishra, P. & Goyal, N. The efficacy of cerebellar vermal deep high frequency (theta range) repetitive transcranial magnetic stimulation

(rTMS) in schizophrenia: A randomized rater blind-sham controlled study. *Psychiatry Res* **243**, 413-420 (2016). https://doi.org:10.1016/j.psychres.2016.07.023

70 Hennig, C. What are the true clusters? *Pattern Recognition Letters* **64**, 53-62 (2015).

71 Greene, A. S., Horien, C., Barson, D., Scheinost, D. & Constable, R. T. Why is everyone talking about brain state? *Trends in Neurosciences* **46**, 508-524 (2023).

72 Smith, S. M. *et al.* Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23**, S208-S219 (2004).

73 Lutkenhoff, E. S. *et al.* Optimized brain extraction for pathological brains (optiBET). *PloS one* **9**, e115551 (2014).

74 Frackowiak, R. S., Friston, K. J., Frith, C. D., Dolan, R. J. & Mazziotta, J. Human brain function. (2004).

75 Joshi, A. *et al.* Unified framework for development, deployment and robust testing of neuroimaging algorithms. *Neuroinformatics* **9**, 69-84 (2011).

76 Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W. & Smith, S. M. Fsl. *Neuroimage* **62**, 782-790 (2012).

77 Satopaa, V., Albrecht, J., Irwin, D. & Raghavan, B. in *2011 31st international conference on distributed computing systems workshops.* 166-171 (IEEE).

78 Abraham, A. *et al.* Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics* **8**, 14 (2014).

79 Vallat, R. Pingouin: statistics in Python. *J. Open Source Softw.* **3**, 1026 (2018).