Transdiagnostic connectome-based predictive modeling reveals where circuits related to self-reported clinical symptoms impinge upon brain networks supporting cognition

Alexander J Simon[1], Anja Samardzija[2], Santino Iannone[1], Flor Parra Rodriguez[2], Saloni Mehta[3], Fuyuze Tokoglu[3], Maolin Qiu[3], Jagriti Arora[3], Kira Y Tang[3], Anna Q Flanagan[3], Rachel Katz[4], Gerard Sanacora[4], Scott W. Woods[4], Vinod H. Srihari[4], Xilin Shen[3], 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 Biomedical Engineering, Yale University, New Haven, CT, USA

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

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

Abstract

A key assumption of the NIMH's RDoC framework is that disordered circuits in the brain should manifest in observable behaviors, including psychiatric symptomatology and cognitive deficits. However, how disordered circuitry impacts multiple behaviors remains poorly understood. Connectome-based predictive modeling (CPM) applied to functional MRI connectivity data can identify networks associated with specific behavioral measures across individuals. Prediction strength reflects how closely a measure relates to network connectivity, while derived networks provide evidence of where an individual's disordered circuits are located. Using CPM, we predicted a broad range of self-reported clinical and objective cognitive measures in a large, transdiagnostic sample with extensive fMRI data (n = 317). Prediction performance varied substantially across instruments, with objective cognitive tests yielding stronger models than self-reported clinical measures ($p < 0.001$). To test whether circuits underlying cognitive deficits related to symptomatology reside in regions where networks overlap, we examined the prediction strength of these sparsely shared circuits. Their connectivity strongly predicted cognitive performance and were primarily localized within the frontoparietal network and between the frontoparietal and default mode networks. These findings reveal how much various behavioral measures reflect brain networks and how circuits within the shared network space contribute to cognitive deficits associated with symptomatology.

Introduction

Identifying neurobiological correlates of real-world behavioral measures can teach us about how neural dysfunction manifests 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 psychiatric symptoms and cognitive deficits stem from deviations from normal neural functioning that span diagnostic categories led to the conceptualization of the NIMH's RDoC initiative[1,2]. In accordance with this framework, 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 inferred from fMRI connectivity data in a predictive modeling framework[3,10]. The cross-validation employed in building the predictive models assures that the identified brain-behavior associations are generalizable[11,12]. 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[13], craving[14], PTSD[15], and social function[16]. While numerous studies have shown that CPM can identify functional connectivity correlates for individual measures (often within specific diagnostic categories), only a handful of studies have examined relationships between functional connectivity and multiple measures in large samples[17-19]. However, these studies used methods different from CPM in their predictive modeling. They also all utilized data from the Adolescent Brain Cognitive Development (ABCD) study. While the ABCD study contains rich fMRI and phenotypic data, brain networks undergo substantial reorganization from childhood to adulthood[20], potentially impacting the

59 derived brain-behavior models. In addition to ABCD data, one of these studies utilized data from
60 the Human Connectome Project (HCP)[18] to predict a variety of measures. The HCP also contains
61 high quality fMRI and phenotypic data, but the sample only consists of healthy young adults. Thus,
62 network correlates of clinically-relevant metrics were unable to be identified. The Yale
63 NeuroConnect dataset, described in detail here[21], provides an opportunity to assess, in a clinically
64 diverse transdiagnostic dataset with extensive fMRI testing (>50 min rest + task), the extent to
65 which a wide range of diagnostically related self-reported clinical symptomatology measures
66 (referred to throughout as 'clinical measures' for brevity) and objective cognitive testing
67 instruments have clearly identifiable network correlates and which do not.
68
69 The RDoC framework suggests that disordered circuits that affect diagnostically-related
70 subjective symptoms can also impact cognitive abilities. This is corroborated by a large body of
71 evidence indicating that cognitive deficits often co-occur with psychiatric symptoms[22]. However,
72 disordered circuits vary substantially across individuals[23] with similar symptoms/conditions and
73 may not completely impinge on a specific network related to a particular symptom or cognitive
74 process. Rather, an individual's disordered circuits are more likely to partially affect multiple
75 networks linked to various symptom and cognitive measures. This phenomenon may underlie the
76 highly variable cognitive deficits seen in psychiatric populations[22,24]. According to this hypothesis,
77 individuals with disordered circuits impinging on networks related to both symptoms and cognition
78 should experience cognitive deficits that co-occur with symptomatology, while individuals with
79 disordered circuits impinging on networks only related to either symptoms should have psychiatric
80 symptoms without cognitive deficits (or vice versa). While similar work has shown that networks
81 associated with mental health measures and cognitive functioning show little overlap[17], the extent
82 to which connectivity within these overlapping regions contribute to the high amount of variance
83 seen in relationships between subjective measures indexing mental health and objective cognitive

assessments remains unknown. To address this question, refinements to computational approaches used for brain-behavior modeling are warranted. Here, we show that CPM can be leveraged to identify circuits that putatively link inconsistencies in cognitive perturbations[22,24] with high inter-individual variability in network dysfunctions seen in psychiatric populations[23].

This study has two primary goals. First, using a comprehensive testing battery administered outside the magnet, we determine which of the many clinical and cognitive measures yield CPM models with high prediction strength. This allows us to characterize the extent to which each measure is linked to identifiable brain network correlates while quantifying prediction strength. Furthermore, predicting an array of measures enabled us to address the extent to which measurement noise and other distribution characteristics influence prediction strength, which are important considerations when interpreting brain-based predictive models[25]. Second, we examine the extent to which connectivity in circuits shared between networks associated with both clinical and cognitive measures can predict cognitive performance. With relevance to precision medicine, we show that modeling how disordered circuits related to an individual's symptomatology impinge upon networks supporting cognition can help explain the variability in cognitive deficits that co-fluctuate with symptomatology. This also demonstrates an advancement in predictive modeling approaches by demonstrating that constraining CPM to networks shared across measures in training data can enhance the precision of functional network localization and identify networks supporting relationships between multiple phenotypes.

Results

In the present study, we utilized a demographically and clinically diverse transdiagnostic population[21] (*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)[26], the Behavior Rating Inventory of Executive Function (BRIEF)[27], the Brief Symptom Inventory (BSI)[28], the Interpersonal Reactivity Index (IRI)[29], the Positive and Negative Affect Schedule (PANAS)[30], the Pittsburgh Sleep Quality Index (PSQI)[31], and the Perceived Stress Scale (PSS)[32]. 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)[33], the letter-number sequencing, symbol search, and cancellation subtests from the Wechsler Adult Intelligence Scales (WAIS)[34], the vocabulary and matrix reasoning subtests of the Wechsler Abbreviated Scale of Intelligence 2nd Edition (WASI)[35], the finger windows and list learning subtests of the Wide Range Assessment of Memory and Learning 2nd Edition (WRAML)[36], the reading subtest of the Wide Range Achievement Test 5th Edition (WRAT)[37], and the Boston Naming Test 2nd edition (BNT)[38]. The fMRI data was preprocessed with standard preprocessing methods (described in greater detail in **Methods** and Samardzija et al.[21]). The Shen 268-node atlas[39] 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 1000 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.

134

135 **Prediction strength of clinical measures**

136 To determine whether a measure had sufficient prediction strength to yield reliable brain network
137 correlates, measures must have had $pFDR < 0.05$ and prediction strength > 0.1. In neuroimaging,
138 prediction strengths of 0.1 are generally considered weak, but are accurate enough if also
139 statistically significant to yield meaningful insights about the neurobiological features important to
140 these models[40]. Of 63 clinical measures, 38 (60.32%) met our criteria for being predicted strongly
141 enough to have identifiable whole brain functional connectivity networks (**Fig 1**). The other 25
142 measures failed to yield sufficient prediction strength to continue with further analyses. When
143 examining predictive model performance across scales and measures, several notable patterns
144 emerged. Of the 7 clinical assessment scales, models built with measures related to sleep quality
145 tended to have the highest prediction strength, with 5 of the top 10 strongest predictions belonging
146 to the PSQI. The mean prediction strength across all PSQI measures was $\rho = 0.170$. Measures
147 pertaining to positive affect from the PANAS also yielded strong predictive models, with 5 out of
148 6 of these measures providing reliable network correlates (mean $\rho = 0.151$). PANAS measures
149 related to negative affect tended to yield slightly lower prediction strength, with 4 out of 7 providing
150 reliable network correlates (mean $\rho = 0.120$). $\rho\rho\rho\rho\rho\rho$Prediction strength values and associated
151 prediction *p*-values can be found in **Supplemental Table 2**, and distributions of clinical measures
152 can be found in **Supplemental Fig 1**.

153

154 Interestingly, there were 4 separate instances where the same measures quantified by different
155 assessment scales yielded widely different predictions. The prediction strength for 'fear' was
156 significantly higher when measured by the ATQ than when measured by the PANAS ($\rho = 0.239$
157 vs $\rho = 0.105$; $p < 0.001$), as was 'sadness' ($\rho = 0.165$ vs $\rho = 0.099$; $p < 0.001$). PANAS 'hostility'

was better predicted than BSI 'hostility' (ρ = 0.181 vs ρ = 0.123; $p < 0.001$). 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 (ρ = 0.164 vs ρ = 0.122; $p < 0.001$).

Additional models were run covarying for years of education and income in addition to sex and age. In general, the added covariates did not have a strong effect on prediction strength, as the predictions were strongly correlated with the original models (**Supplemental Fig 2**; r = 0.946, $p <$ 0.001). The networks resolved were also highly similar (**Supplemental Fig 3**; mean Dice = 0.808 +/- 0.058 stdev). In the interest of examining the effects of zero-inflation on the predictions, models were also re-fit with zero-responders removed. Again, the prediction strengths were similar to the original models (**Supplemental Fig 4**; r = 0.794, $p < 0.001$), as were the network derivations (**Supplemental Fig 5**; mean Dice = 0.636 +/- 0.189 stdev). However, removing zero-responses did influence the predictability and thus the networks derived for a few measures (specifically, subjective sleep quality, sleep duration, and use of sleep medication had network similarities of Dice < 0.3).

Prediction strength of cognitive measures

Nearly all cognitive measures yielded excellent 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 (ρ = 0.428). Predictions of measures from the D-KEFS were highly variable, though all produced (range: ρ = 0.159 to ρ = 0.374). In general, measures from the twenty-questions (mean ρ = 0.288) and trail-making tests (mean ρ = 0.320) were the most strongly predicted measures from D-KEFS. Performance measures from the WAIS (mean ρ = 0.319), the WRAML (mean ρ = 0.263), and the

183 WRAT (ρ = 0.302) were generally predicted strongly. Measures from the BNT (mean ρ = 0.224)
184 and the D-KEFS verbal fluency test (mean ρ = 0.177) were predicted less strongly. The two
185 cognitive measures that CPM did not generate reliable models for (WRAML verbal learning
186 intrusion and BNT number correct following a stimulus cue) had the highest percentage of
187 individuals performing at ceiling (see **Supplemental Fig 6**; 53.16% and 77.19%, respectively).
188 These results highlight that predictive models linking external trait measures to brain connectivity
189 patterns via CPM can be built across a wide range of cognitive metrics, with higher prediction
190 strengths representing greater confidence in the derived network correlates. Prediction strength
191 values, associated prediction *p*-values, and brief descriptions of the cognitive abilities theoretically
192 tested by each measure can be found in **Supplemental Table 3**.
193
194 Additional models were run covarying for years of education and income in addition to sex and
195 age. The added covariates did not have a strong effect on prediction strength, as the predictions
196 were strongly correlated with the original models (**Supplemental Fig 7**; $r = 0.967$, $p < 0.001$) nor
197 did they substantially change the networks resolved (**Supplemental Fig 8**; mean Dice = 0.719
198 +/- 0.067 stdev). In the interest of examining the effects of zero-inflation on the predictions, models
199 were also re-fit with zero-responders removed. Similarly, the prediction strengths were highly
200 similar to the original models (**Supplemental Fig 9**; $r = 0.928$, $p < 0.001$), as were the network
201 derivations (**Supplemental Fig 10**; mean Dice = 0.849 +/- 0.105 stdev).
202
203 **What drives prediction strength?**
204 In addition to CPM significantly predicting a higher percentage of cognitive than clinical measures,
205 the prediction strength of the cognitive measures that were successfully modeled were higher
206 than the clinical measures' (**Supplemental Fig 11**; cognitive measures' mean prediction strength:
207 $\rho = 0.289$; clinical measures' mean prediction strength: $\rho = 0.168$; $p < 0.001$). This highlights that

208 CPM yields more accurate and robust functional network correlates for cognitive than for clinical
209 measures.
210
211 We next sought to understand the factors that influence model performance leading to better
212 prediction strength for some measures compared to others, noting that the functional connectivity
213 data remained constant across all models. Due to the differences in prediction strength between
214 clinical and cognitive assessments, we examined how prediction strength varied across
215 instruments and subdomains tested in clinical and cognitive assessments separately. One-way
216 ANOVAs revealed that prediction strength strongly varied as a function of instrument/subdomain
217 for clinical (**Fig 3a**; $F(13,62986) = 3059.98$, *Cohen's F* = 0.795, $p < 0.001$) and cognitive measures
218 (**Fig 3b**; $F(13,27986) = 1194.96$, *Cohen's F* = 0.745, $p < 0.001$).
219
220 Recent work questioned the extent to which brain-based predictive modeling is influenced by
221 measurement noise[25]. To address this, we tested whether the predictive accuracy on these
222 measures was related to their intraclass coefficients (ICCs) reported by the manuals and seminal
223 papers describing each of the testing instruments. We also examined whether model performance
224 was related to other characteristics of the measures' data (skewness, variance, proportion
225 responding at floor/ceiling, and the number of unique responses). Clinical prediction strength was
226 not correlated with ICC in the measures with reported ICC values ($\rho = 0.091$, *pFDR* = 0.725), nor
227 was it related to the measures' skewness ($\rho = 0.188$, *pFDR* = 0.725), the variance of the
228 responses ($\rho = 0.150$, *pFDR* = 0.725), the proportion responding at floor ($\rho = 0.109$, *pFDR* =
229 0.725) and ceiling ($\rho = 0.045$, *pFDR* = 0.848), and the number of unique responses ($\rho = -0.025$,
230 *pFDR* = 0.848) (**Fig 3c**). Similarly, cognitive prediction strength was not correlated with ICC in the
231 measures with reported ICC values ($r = -0.122$, *pFDR* = 0.895), nor was it correlated with
232 response variance ($\rho = 0.342$, *pFDR* = 0.150), ceiling effects ($\rho = 0.028$, *pFDR* = 0.895), or the

number of unique responses ($\rho = 0.305$, $pFDR = 0.172$). However, prediction strength of cognitive measures was significantly correlated with floor effects ($\rho = -0.541$, $pFDR = 0.017$*) and was trending towards being significantly correlated with the distribution's skewness ($\rho = -0.444$, $pFDR = 0.053$) (**Fig 3d**).

Network similarity across measures

Since disordered circuits could impinge upon both clinical and cognitive networks, we examined the degree of overlap between predictive networks for all clinical and cognitive measures (**Fig 4**). This was achieved by sorting the predictive networks based on similarity using an agglomerative hierarchical clustering algorithm. As expected, the similarities between the measures' networks mirrored the correlational structure of the measures themselves since the networks are derived from the measures. There was a nearly identical correspondence between these networks and the networks' Dice similarities when using an unthresholded edge selection approach ($r = 0.99$), indicating that the edge selection threshold that was applied likely did not impact overall network similarity structure.

In general, the clinical measures were most related to other clinical measures, while the cognitive measures were most related to other cognitive measures. The only exception was high intensity pleasure (ATQ), which was more closely related to cognition than to other clinical measures. The Dice similarity between clinical-cognitive networks was generally much lower than between clinical-clinical (unpaired t-test: $t(788.66) = -31.68$, $p < 0.001$) and cognitive-cognitive networks (unpaired t-test: $t(389.11) = -23.39$, $p < 0.001$). Additionally, after adjusting for multiple comparisons using an FDR correction, only 101 pairs of clinical and cognitive measures were significantly correlated ($pFDR < 0.05$), and their effect sizes were typically weak (maximum $\rho = 0.234$, minimum $\rho = -0.227$). The significantly correlated pairs of clinical and cognitive measures

258 are highlighted in **Supplemental Fig 12**. Together, this suggests that brain networks that vary as
259 a function of cognitive performance are largely distinct from those networks that vary
260 systematically with clinically relevant symptomatology.
261
262 We then grouped the brain networks predicting clinical and cognitive measures into 9 clusters
263 based on the hierarchical clustering results. This number was determined using the elbow
264 method[41] (**Supplemental Fig 13**), which is a commonly used approach for determining an
265 appropriate number of clusters to group data into. These clusters were selected to visualize how
266 the measures grouped together based on network similarity, and to compare the regions
267 important in driving CPM predictions across clusters, rather than to make a claim about how many
268 clusters exist in this dataset. The patterns that the CPM networks loaded onto each of the 9
269 clusters corresponded to what the measures theoretically assess. Of these 9 clusters, 5 were
270 composed of clinical measures and 4 were composed of cognitive measures. For discussion
271 purposes, each of the clusters were given a name that generally reflects the broad measures
272 composing them. The cluster labels are displayed underneath the dendrogram in **Fig 4**, which
273 align with the sorted measures displayed in the matrix below. It is important to note that these
274 names do not encompass all the behavioral phenotypes present in each cluster. Additionally,
275 since they were not orthogonal, important features were often shared across clusters. For
276 example, cognitive control abilities were likely important to performing all cognitive tests, not just
277 the tests in the 'cognitive control' cluster. This phenomenon is reflected by the substantial overlap
278 in the network features between clusters.
279
280 **Localizations of networks important to predictions**
281 We examined the contributions of specific networks to prediction strength of clinical measures
282 using the canonical Yeo17 networks as a reference[42] (**Fig 5a**). For identifying the nodes and

edges important to predictions across clusters, we compared positively predicting network features from the positive affect cluster to negatively predicting network features from the 4 other clinical clusters (and vice versa) because these cluster's measures were anticorrelated. Doing so allowed us to effectively put the brain correlates of measures where higher scores indicated better mental health on the same scale as measures where higher scores indicated worse mental health. 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

307 indicative of better performance on the task, and not just a higher score, since sometimes a higher
308 score was undesirable. The nodes contributing to predicting cognitive measures displayed
309 substantial overlap across the cognitive measures' clusters for both positively (mean pairwise
310 Jaccard similarity = 0.197 +/- 0.092 SD, one-tailed $p < 0.001$) and negatively predicting nodes
311 (mean pairwise Jaccard similarity = 0.210 +/- 0.118 SD, one-tailed $p < 0.001$). Despite this,
312 examining the top 10% highest degree nodes in each cluster's predictive networks showed that
313 only nodes within the visual A and somatomotor B networks were positively associated with
314 cognitive performance and only one node in the temporal parietal network was negatively
315 associated with cognitive performance across all clusters. A full list of the top 10% highest degree
316 positive and negative predicting nodes for each cognitive measure cluster can be found in
317 **Supplemental Tables 6 and 7**. There was significant overlap between the nodes important to
318 positively and negatively predicting cognitive measures (Jaccard similarity = 0.294, one-tailed p
319 < 0.001), and these nodes often were in the cerebellum. However, this overlap was less
320 pronounced than it was in the clinical measures (Jaccard similarity = 0.294 vs 0.500). The nodes
321 that were only important to positive predictions were typically located toward the bottom of the
322 functional gradient in lower-level unimodal cortex, while nodes that were only important in
323 negative predictions were frequently localized near the top of the functional gradient in higher-
324 level transmodal cortical regions[43] (see **Supplemental Fig 14**). Examining the top 10% of
325 canonical network pairs where the predictive edges were most concentrated showed that dorsal
326 attention A to dorsal attention B, default mode A to cerebellum, limbic A to limbic B, somatomotor
327 A/B to cerebellum, and visual A to somatomotor B network connectivity was important to positive
328 predictions across all cognitive measures' clusters. Furthermore, dorsal attention B to default
329 mode A, default mode B to temporal parietal, somatomotor A/B to dorsal attention B, somatomotor

330 B to default mode A, and within somatomotor A network connectivity were important for driving
331 negative predictions across all cognitive measures' clusters.
332
333 **Predicting cognition from impinging circuits within clinical networks**
334 In **Fig 4** we highlighted that the correlations were relatively weak, and network overlaps were
335 sparse between clinical and cognitive measures. This finding suggests that cognitive alterations
336 that co-occur with clinical symptomatology are highly variable between individuals. Given that
337 dysfunctional circuits are heterogeneously distributed in individuals with similar
338 symptoms/conditions[23], we hypothesized that the precise localization of an individual's disordered
339 circuitry can be inferred from assessing how their clinical networks impinge upon networks
340 supporting cognitive performance. Therefore, we sought to determine the extent to which circuits
341 that span both systems explain variance in cognitive deficits that co-occur with clinical measures.
342 This was achieved by identifying the edges that were related to both clinical and cognitive
343 measures for all possible clinical-cognitive measure pairings within the training folds of the CPM
344 procedure and then predicting cognitive measures individually from these constrained sets of
345 overlapping edges (**Fig 6a**). Importantly, these models were only trained on edges that predicted
346 poor symptomatology and poor cognition (as opposed to shared edges predictive of poor
347 symptomatology and good cognition; more details in **Methods**).
348
349 Connectivity in the overlapping edges significantly predicted ($pFDR < 0.05$) cognition 777 times
350 out of 988 total pairs of clinical-cognitive network overlaps (38 clinical x 26 cognitive measures;
351 **Fig 6b**). The cognitive measures were plotted in descending order according to how strongly they
352 were predicted by impinging clinical networks. The 5 cognitive measures that were predicted the
353 strongest were matrix reasoning, finger windows, verbal learning delayed recall, verbal learning,

354 and coding. The 5 impinging clinical networks that predicted these measures most strongly were
355 the paranoid ideation (BSI), hostility (PANAS), sleep duration (PSQI), hostility (BSI), and the
356 global severity index (BSI) networks. As an additional control analysis, models were trained on
357 edges from clinical measures' networks that did not overlap with cognitive networks and were
358 tested to predict cognitive measures. Importantly, the number of training edges selected for the
359 control models were equal to the number of edges that overlapped between clinical and cognitive
360 networks. Of these 777 significant predictions from overlapping edges, 762 were significantly
361 stronger than cognitive predictions from an equal number of non-overlapping edges within
362 symptom networks ($pFDR < 0.05$). The only instances where the overlapping edge predictions
363 were not significantly stronger than the non-overlapping edge predictions occurred in the some of
364 the weakest overlapping edge predictions (**Supplemental Fig 15**).
365
366 The number of overlapping edges was strongly correlated with the prediction strength (**Fig 6c**; r
367 $= 0.590$, $p < 0.001$), indicating that the more widespread the impingements of clinical networks
368 onto cognitive, the more likely that connectivity in these circuits predicted cognition. However,
369 there were 153 instances where cognitive measures were strongly predicted ($\rho > 0.3$) by
370 overlapping networks that contained <100 edges. Furthermore, the prediction strengths yielded
371 by these constrained sets of overlapping edges tended to be much stronger than the correlations
372 between the clinical and cognitive measures' values (**Fig 6d**; $t(987) = -59.18$, $p < 0.001$),
373 highlighting that shared circuits between clinical and cognitive measures can predict cognition
374 even when the measures themselves are uncorrelated. Taken together, these results suggest
375 that disordered circuits that impinge upon networks supporting cognition account for a significant
376 amount of the variance in cognitive deficits co-occurring with symptomatology.
377

Finally, we examined where the location of network overlaps important to explaining the variability in clinical-cognitive relationships were. To achieve this, the number of times that each edge was predictive across the 777 overlapping network pairs was identified by summing the binarized predictive network masks. Most predictive edges only overlapped between clinical and cognitive networks a small proportion of times (441 out of 777 overlapping predictive networks shared <50 edges). This was to be expected given the diversity of clinical and cognitive measures assessed. Nevertheless, a subset of edges that were repeatedly predictive across numerous overlapping clinical-cognitive network pairs. Specifically, 291 edges were predictive >50 times, and 96 were predictive >100 times (**Fig 7a**). To examine how these predictive overlapping edges were distributed across the brain, the number of times that overlapping edges were predictive was averaged across all edges within each set of canonical network-network pairs (**Fig 7b**). This way, predictive edges that overlapped more frequently were weighed more heavily.

The overlapping edges that most frequently positively predicted cognitive deficits were distributed between the frontoparietal control C to default mode B, visual A to frontoparietal control C, within brainstem, limbic A to frontoparietal control C, and within the limbic A network. The overlapping edges that most frequently negatively predicted cognitive deficits were mostly distributed within the frontoparietal control C network. On average, edges within each of these network-network pairs appeared in overlapping predictive models >7 times. Examining the highest degree nodes within these overlapping networks revealed that the 5 nodes most important in the positive predictive model (edge strength increased with deficit score) for cognitive deficits were in the left dorsolateral prefrontal cortex, right ventral occipitotemporal cortex, bilateral posterior inferior temporal gyrus, and right anterior inferior temporal gyrus. The 5 nodes most important in the negative predictive model (edge strength decreased with increasing deficit score) for cognitive deficits were in the left anterior inferior temporal gyrus, bilateral basal ganglia, right orbitofrontal

403 cortex, and left posterior inferior temporal gyrus. Lists of these nodes' indices, their centroids' MNI
404 coordinates, and their canonical network assignments can be found in **Supplemental Table 8**.
405 Overall, these results highlight that performing CPM on sets of edges shared between measures
406 in the training data can yield much more confined solutions for mapping networks related to
407 cognition. Furthermore, this approach revealed that circuits distributed within and between higher
408 order networks are related to cognitive deficits that co-occur with subjective clinical
409 symptomatology measures.
410

411 **Discussion**

412 This study had two primary goals. First, we aimed to comprehensively evaluate which clinical and
413 cognitive measures have clear brain correlates as determined by prediction strength in CPM. The
414 findings showed that, in general, there is a tighter coupling between brain networks and external
415 instruments for cognitive tests compared to clinical measures. We also show that this relationship
416 is not uniform across measures, as prediction strength varied substantially across measures in
417 both classes and was not driven by differences in measurement stability. Our second goal was to
418 test a hypothesis upon which the RDoC framework is based: that disordered circuits that affect
419 clinical symptomatology should also impact cognition. Our results indicated that the networks
420 implicated in clinical symptomatology impinge upon networks that support cognitive functioning,
421 and that these overlapping networks explain a significant amount of variance in the relationships
422 between clinically relevant symptoms and cognitive performance metrics. As might be expected,
423 the networks associated with symptoms do not completely align with any specific cognitive
424 network. This work establishes a framework for how to use brain-behavior modeling to locate
425 functional networks that support and facilitate relationships between multiple behavioral domains.
426

427 **Clinical predictions and associated networks**

428 The broad testing of clinical measures that we employed in this study allowed us to systematically
429 evaluate the extent to which a variety of clinical assessment instruments can be linked to
430 functional brain networks. We identified reliable functional connectivity correlates of many clinical
431 measures collected in this dataset (38 out of 63 predicted at $\rho > 0.1$ and $pFDR < 0.05$). Varying
432 prediction strength between instruments highlights that some clinical assessment tools are better
433 suited to capture brain network function than others. Among these, the PSQI, the BRIEF
434 (behavioral regulation measures), PANAS, ATQ (non-aggressive negative affect measures) and
435 BSI all tended to be well modeled in the brain. Conversely, measures from the IRI, BRIEF
436 (metacognition measures), ATQ, and the PSS did not consistently yield reliable brain network
437 models. We also observed that some network models of sleep measures were strongly influenced
438 by zero-responses. This suggests that there were a mix of circuits in the original network models
439 that were driven by asymptomatic patients and circuits linked to sleep pathology. Future studies
440 should seek to further understand the relevance of these differing circuits in clinical populations.
441

442 Each of these clinical assessment instruments utilized self-reporting. The subjective nature of
443 self-reports introduces several confounds that could interfere with CPM prediction strength, such
444 as response style differences related to introspection abilities[44], social[45] and cultural
445 backgrounds[46], or differences in how respondents and assessment administrators interpret the
446 questions[47]. Measures that are less prone to these influences are thus more likely to have clear
447 network correlates. Understanding how to refine self-report scales to yield better behavioral
448 predictions will improve the extent to which they reflect network functioning, thereby enhancing
449 their utility in identifying neural generators of clinical dysfunction. Better test instruments combined
450 with digital phenotyping with HER and EMA data in its many different forms will help to improve

451 the efficacy of precision medicine[48-50]. Brain models such as CPM can inform the further
452 development of these assessment tools.
453
454 The considerable overlap in the networks important to predicting all clinical measures suggests
455 that CPM was capturing a low-dimensional general factor of psychopathology (p-factor) posited
456 by the HiTOP framework and elsewhere[51,52]. Each cluster's unique deviations from this general
457 p-factor network, therefore, track network dysfunctions specific to the clusters' phenotypic
458 makeups. Our results strongly implicate the thalamus and cerebellum as key nodes in this p-factor
459 network. These brain structures and their clinical relevance are commonly overlooked, and
460 researchers have argued that this omission is stifling progress in neuroscience[53]. However, there
461 is evidence suggesting the functional connections involving the cerebellum and thalamus contain
462 clinically relevant information[54,55]. Although not involved in predicting clinical measures across all
463 clusters, nodes within the somatomotor network were also frequently involved in predicting clinical
464 measures. Like the cerebellum and deeper brain structures such as the thalamus, the
465 somatomotor network has also been relatively overlooked in psychiatric research. An emerging
466 body of research has linked somatomotor network dysfunction to clinical pathology[56-59]. Our
467 results add to this mounting evidence implicating the thalamus, cerebellum, and SM network in
468 symptomatology.
469
470 **Cognitive predictions and associated networks**
471 We also evaluated a comprehensive battery of cognitive tests to identify the extent to which they
472 reliably reflect brain network function. Nearly all cognitive measures yielded models with good
473 prediction strength (26 out of 28 predicted at $\rho > 0.1$ and $pFDR < 0.05$). More than 50% of
474 participants performed at ceiling for the two measures that were not modeled well (verbal learning

475 intrusion (WRAML) and number correct following a stimulus cue (BNT)). These were also the two
476 cognitive measures with the lowest range of scores (see **Supplemental Fig 6**). These results
477 highlight the importance of utilizing cognitive tests with a sufficiently wide distribution of scores to
478 identify reliable functional network associations of cognition. However, considering that there was
479 not a significant correlation between measurement score variance and prediction strength across
480 all cognitive measures (see **Fig 3**), the distribution of the scores does not account for all of the
481 variance in prediction performance.
482
483 Since the cognitive tests utilized objective response time- and accuracy-based performance
484 metrics, they were not susceptible to the biases mentioned previously to the same degree as
485 subjective self-report measures. This was likely a major factor contributing to their prediction
486 strength being substantially higher than the clinical measures' (clinical measures' mean $\rho = 0.168$;
487 cognitive measures' mean $\rho = 0.289$). There are several other noteworthy potential factors
488 influencing this difference in prediction strength. It is possible that the networks supporting
489 cognition are less variable than the ones that support psychiatric symptoms. The instruments
490 used to model symptoms could also simply reflect brain function to a lesser extent than cognitive
491 tests. Finally, the cognitive tasks we utilized could have engaged networks that support cognition
492 to a greater extent than networks related to clinical measures. The clinical measures' predictions
493 may therefore be boosted from tasks that probe states more closely related to what they are
494 assessing. Any of these possibilities could contribute to this difference in prediction strength.
495 Regardless, the higher prediction strength for cognitive tests reported here is accompanied by
496 greater confidence in the derived brain network correlates.
497
498 The networks that predicted cognitive measures also showed considerable overlap with each
499 other, suggesting that these common networks track with the g-factor commonly reported in the

500 literature[60]. Given that all the tests probe a spectrum of cognitive processes, this g-factor network
501 likely reflects a mix of the various cognitive components that interact with each other to facilitate
502 test performance. The nodes important for positively predicting cognitive performance often layed
503 at the opposite ends of the principal cortical functional gradient[43] from negatively predicting nodes,
504 highlighting a departure from the nature of clinical networks, where the same nodes were often
505 important to both the positive and negative predictive models. Interestingly, this pattern did not
506 apply to cerebellar nodes, as areas of the cerebellum that were involved in the positive predictive
507 networks for cognition were also commonly involved in the negative predictive networks.
508 Furthermore, our findings also strongly implicate the SM network in supporting cognitive function,
509 in line with the recent report that systems supporting higher-order cognition are interdigitated
510 within motor cortex[61].
511
512 **Other factors contributing to prediction strength**
513 Recent work has suggested that measurement stability, as reflected by ICC, is a major component
514 underlying predictive modeling performance[25]. We directly tested this by correlating the prediction
515 strength of clinical and cognitive measures with the ICCs reported by the authors of those scales.
516 We observed effects of instrument/subdomain and response variability on predictions, while an
517 effect of ICC was absent across all measures, including those that failed to model well (see **Fig**
518 **3**). This does not rule out the possibility that measurement stability is still a factor contributing to
519 brain-based predictive modeling success. Rather, it suggests that if a measurement is sufficiently
520 stable, we can generate a valid brain-behavior model. In most of the measures we used, the ICC
521 values were high (>0.6). Thus, the influence of this on prediction performance should be minimal,
522 and instead the wide range of responses for any given phenotype should dominate.
523

Additionally, CPM can directly test how the development of the measures can lead to measurable differences in the extent to which they engage clear brain correlates[62]. As demonstrated here, there were 4 instances where different instruments provided measurements with the same label that purportedly measured the same phenotype but yielded widely different predictions (see **Fig 1**). These findings highlight how the framing of the questions probing symptoms or how index scores composed of multiple related measures can improve the confidence of the network correlates derived from predictive modeling.

Relationships between clinical and cognitive measures

While a large body of work has demonstrated cognitive deficits in psychiatric populations, the findings here indicate that the relationships between symptom and cognitive measures were weak, likely reflecting the highly variable cognitive alterations within clinical populations[22,24,63,64]. When examining the clinical-cognitive relationships more closely, we observed that significant correlations between clinical and cognitive measures were evenly distributed across clusters, suggesting that these measures might be related through more general p- and g-factor correlations. Although we were interested in identifying brain networks associated with both poor symptomatology and impaired cognition, not all clinical measures were associated with poorer cognitive performance. Better performance on cognitive tests that contained a language component (e.g., the BNT) were correlated with negative affect and depression symptoms. One potential interpretation of this finding could be that enhanced verbal abilities feed into more ruminative thoughts and higher levels of depression symptoms[65]. Interestingly, depression symptoms were also negatively associated with performance on WAIS digit symbol coding, a test that assesses cognitive control. This dissociation suggests further interrogation is needed to more thoroughly understand the various cognitive alterations that co-occur with depression symptoms.

549 **Shared networks explain variance in clinical and cognitive measure relationships**

550 Examining the intersection of the networks for the clinical and cognitive measures revealed that
551 overlap was sparse. These results extend upon similar recent findings that predictive network
552 features are distinct across mental health and cognitive measures[17] by revealing that connectivity
553 within these shared networks explains a significant amount of the variance in cognitive deficits
554 that co-occur with self-reported clinical symptomatology. This suggests that connectivity in
555 regions where networks overlap could account for much of the heterogeny in cognitive dysfunction
556 seen in psychiatric populations[22,24,63,64]. This also aligns with the RDoC hypothesis that disordered
557 circuits within an individual likely do not impinge directly on a single network, but instead partially
558 affect multiple networks spanning both symptoms and cognitive measures. It also supports the
559 notion that the precise location and distribution of the disordered circuits relate to an individual's
560 multivariate symptom profile. Thus, identifying overlapping networks can potentially help inform
561 cognitive intervention personalization in clinical populations, which often assume that
562 symptomatology stems from cognitive deficits[22] and that cognitive remediation can improve
563 symptoms[66].

564

565 To investigate these overlapping networks, we refined CPM such that models were only trained
566 on edges that were related to both cognitive measures and clinical measures. This step improved
567 the specificity of CPM for network mapping. Many of the cognitive measures utilized here reflect
568 a wide range of abilities that are not necessarily a pure reflection of subjective clinical
569 symptomatology. Some measures, such as matrix reasoning, index abilities like general fluid
570 intelligence (a key component of IQ[67]) that are not inherently reflective of pathology. Although
571 clinical symptomatology has been associated with reductions in fluid intelligence, these
572 associations are generally small in magnitude, indicating that substantial variability in fluid
573 intelligence exists independent of clinical symptomatology[68]. Therefore, if general fluid intelligence

574 (or other aspects of cognition) is affected by psychiatric dysfunction, this likely occurs through a
575 partial perturbation of the relevant circuits, rather than impacting entire systems. Our approach
576 enabled us to confine the cognitive measures' predictive networks to just the edges that were
577 impinged upon by the networks defined for the clinical measures, allowing us to better understand
578 where in the brain circuits supporting cognition are impacted by symptomatology. This may also
579 be useful in other multivariate modeling frameworks, such as examining how brain networks
580 related to anxiety and depression are impacted by poor sleep quality.
581
582 Given that patients who experience cognitive impairment may respond differently to various forms
583 of treatment than unimpaired patients[63,64,69] and that psychiatric patients often consider cognitive
584 impairments among their primary priorities for treatment[70,71], identifying the loci of these
585 interactions is important and a potential strategy for developing novel diagnostic biomarkers and
586 precision treatment targets. Examining overlapping networks localization revealed that much of
587 the overlap associated with cognitive dysfunction was distributed between task-positive and task-
588 negative networks (e.g., frontoparietal control to default mode, and frontoparietal control to limbic),
589 while network associated with cognitive performance was mostly limited to the task-positive
590 networks (e.g., frontoparietal control). Other research has shown that a breakdown in the
591 anticorrelation structure between task-positive and task-negative networks is a common feature
592 across several dimensions of psychopathology spanning multiple diagnostic categories[72]. Such
593 findings also have been observed in case-control studies of specific diagnoses. Abnormal
594 connectivity withing the frontoparietal control network and breakdowns in the segregation
595 between the default mode and frontoparietal network have been observed in patients with major
596 depressive disorder[73,74], which is thought to underlie these patient's deficits in cognitive control[74].
597 Similar alterations in the connectivity of these networks have also been observed in obsessive-
598 compulsive disorder[75], bipolar disorder[76], and schizophrenia[77,78]. The results we present here

599 corroborate these findings and suggest a mechanistic explanation of how disordered circuitry can
600 impact cognitive deficits related to symptomatology transdiagnostically. Additionally, the left
601 dLPFC was among the key hubs in these overlapping networks. Given that recent findings
602 suggest that individuals with treatment resistant depression who also have cognitive impairments
603 respond better to TMS applied to the left dLPFC than patients with intact cognition[64], this indicates
604 that this region may be a promising target in precision medicine settings. Future work should
605 therefore further examine the utility of these overlapping circuits in biotyping and guiding
606 personalized interventions.
607
608 **Limitations**
609 There are some noteworthy limitations to the current study. First, without an external sample to
610 validate our results, we cannot be certain that the network correlates of various behavioral
611 measures will generalize to unseen data. This is one of the first studies to utilize a large
612 transdiagnostic sample, with both extensive fMRI data (>50 min) and comprehensive phenotyping
613 conducted outside the magnet[21]. Similar data to evaluate whether these results generalize is
614 lacking. Second, some of the studies that we obtained ICC values from were inconsistent in the
615 time intervals between test and re-test, sample sizes, and age groups utilized. These differences
616 could have biased ICC values, potentially obscuring any effects on prediction accuracy. Third, we
617 acknowledge that there are many different methods to determine "optimal" clustering solutions
618 that each typically yield widely different results, and that there is often no clear determination of
619 how many clusters exist[79]. The clustering performed here was solely intended for reducing the
620 dimensionality of the data and showing how similar the measures were to each other based on
621 their network correlates. Their labels were therefore only used to help describe the types of
622 measures composing each cluster. Fourth, the biases introduced via self-reporting reduce the
623 prediction strength of clinical measures and add noise to our network estimates for those

624 measures. However, most instruments available for assessing psychiatric symptoms in clinical
625 contexts rely on self-reporting, and objective assessments are lacking. Fifth, networks associated
626 with clinical measures may be more heterogeneously affected than networks associated with
627 cognitive functioning, especially in a transdiagnostic population that includes many healthy
628 individuals. The asymmetry may underlie the differences in prediction strength between clinical
629 and cognitive measures and should be further explored in populations with less expected
630 heterogeneity in networks associated with clinical measures. Sixth, we have loosely used the
631 terms cognitive and clinical or symptom measures. Some of these objective cognitive tests were
632 designed to ascertain clinically-relevant information regarding cognitive dysfunction, and some of
633 these clinical measures were designed to assess subjective, real-world cognitive functioning.
634 None of these measures are purely cognitive nor purely clinical and the categorical distinction
635 here is not binary. As the data shows, these are not independent systems. Nevertheless, it helps
636 to frame the discussion in this context which is commonly used in the literature. Finally, our results
637 beg a deeper investigation into the relationship between symptoms and cognition. We have
638 assumed here that we are modeling fixed traits, but it is well known that brain-state can strongly
639 influence findings in network neuroscience[80]. Symptoms and cognition wax and wane in both
640 healthy and clinical populations. Understanding how these models may vary, and how the
641 relationship between symptoms and cognitive variables interact over time could be very
642 informative in clinical contexts[80].
643

644 **Methods**

645 **Participants**

646 The sample reported in this study includes data from 317 demographically diverse participants
647 with at least 50 minutes of high-quality functional MRI data and extensive phenotyping outside of
648 the magnet. Participants were recruited from advertisements broadly distributed throughout the

New Haven community and referrals from Yale clinics. This transdiagnostic population experienced a wide range of symptoms and symptom severities, often with multiple psychiatric diagnoses. All participants provided written informed consent in accordance with a protocol approved by the Yale IRB. A more thorough description of this dataset can be found elsewhere[21].

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[81,82]. Motion correction was done using SPM12[83]. Nonlinear

registration of the MPRAGE to the MNI template was performed using BioImage Suite[84], and linear registration of the functional to the structural images was done through a combination of FSL[85] 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 in the study if they completed all eight fMRI scan runs and had a grand mean frame-to-frame displacement of less than 0.15 mm and a maximum mean frame-to-frame displacement of less than 0.2 mm. Two participants were scanned with a slightly shorter scanning time (25 seconds shorter) and were included in the dataset.

The Shen 268 node atlas[39] was applied to the preprocessed data, parcellating it into 268 functionally coherent nodes. Pearson correlations of the time series between all node pairs were computed and subsequently z-transformed to generate 8 functional connectivity matrices for each participant. The averages of all functional connectivity matrices were computed for each run/participant. Since individuals' functional connectomes are more similar to themselves, regardless of the task, than they are to other individuals[3], the average connectome will still likely contain relevant information about an individual's trait characteristics. By including connectivity data collected during multiple different cognitive tasks in the averaged connectomes, patterns of task-evoked brain activity that may differentially enhance behavioral prediction were accounted for[6]. Analyses and visualizations were conducted using both MATLAB and python.

Clinical assessments

697 A battery of commonly used self-report questionnaires probing a wide range of clinically relevant
698 measures were collected immediately following the MRI scans (same day). They were selected
699 to index aspects numerous constructs from each of the six major domains of basic human
700 neurobehavioral functioning specified by the RDoC framework[1,2] (negative valence systems,
701 positive valence systems, cognitive systems, social processes, arousal and regulatory systems,
702 and sensorimotor systems). The assessments consisted of the Adult Temperament
703 Questionnaire (ATQ)[26], the Behavior Rating Inventory of Executive Function (BRIEF)[27], the Brief
704 Symptom Inventory (BSI)[28], the Interpersonal Reactivity Index (IRI)[29], the Positive and Negative
705 Affect Schedule (PANAS)[30], the Pittsburgh Sleep Quality Index (PSQI)[31], and the Perceived
706 Stress Scale (PSS)[32]. Two of the questionnaires (BSI and BRIEF) provided both raw and age and
707 sex normalized scores. In this study, we used the raw scores for neuro-behavioral modeling to
708 maintain consistency across clinical measures. Some participants did not complete all clinical
709 assessment scales. They were included in all analyses for which they had data. A detailed
710 breakdown of the number of subjects (*n*s) for each measure can be found in **Supplemental Table**
711 **9**.
712

713 **Cognitive assessments**

714 A series of neuropsychological evaluations was administered by a trained evaluator to objectively
715 test a wide array of cognitive functions related to the cognitive constructs specified by the RDoC
716 framework[1,2] (attention, perception, cognitive control, language, declarative memory, and working
717 memory). These were conducted in the same post-imaging session as the clinical assessments.
718 In total, we administered the verbal fluency, trail making, color-word interference, and twenty
719 questions subtests from the Delis-Kapan Executive Function Scale (D-KEFS)[33], the letter-number
720 sequencing, symbol search, and cancellation subtests from the Wechsler Adult Intelligence
721 Scales (WAIS)[34], the vocabulary and matrix reasoning subtests of the Wechsler Abbreviated

722 Scale of Intelligence 2nd Edition (WASI)[35], the finger windows and list learning subtests of the
723 Wide Range Assessment of Memory and Learning 2nd Edition (WRAML)[36], the reading subtest of
724 the Wide Range Achievement Test 5th Edition (WRAT)[37], and the Boston Naming Test 2nd edition
725 (BNT)[38]. A few of the cognitive tests provided both raw and age and sex normalized scores. In
726 this study, we used the raw scores for predictive modeling to maintain consistency across all
727 measures. Some participants did not complete all cognitive tests. They were included in all
728 analyses for which they had data. A detailed breakdown of the number of subjects (*n*s) for each
729 measure can be found in **Supplemental Table 10**.
730
731 **Predictive modeling**
732 We used CPM[3,10] to identify functional brain networks that were related to each clinical and
733 cognitive measure separately. Each predictive model was trained and tested in a 10-fold cross-
734 validation, which was repeated 1000 times to obtain a distribution of predictions. Within each
735 training fold, edges were selected if they were significantly correlated ($p < 0.05$) with behavior
736 after controlling for age and sex using partial correlations. Given that cognitive abilities are often
737 correlated with years of education[86], and many clinical symptoms are often influenced by
738 income[87], these demographic variables were not controlled for in the main analyses (though they
739 were in a supplemental analysis). Prediction strength was determined by comparing the observed
740 clinical/cognitive measures to the predicted measures using Spearman correlations. Permutation
741 tests were utilized for significance testing. Specifically, the number of null prediction values that
742 were greater than the median prediction for each measure was divided by the number of
743 permutations to derive the *p*-values. The *p*-values were FDR-corrected for 91 comparisons (63
744 clinical + 28 cognitive measures). Prediction strengths between the same measures assessed by
745 different instruments were formally compared by comparing the difference between the two

746 median prediction strengths to the difference of median prediction strengths from the two
747 distributions predictions after shuffling, bootstrapped 1000 times. The number of instances when
748 the median difference from the shuffled predictions exceeded the true median prediction strength
749 differences was divided by 1000 (the number of bootstraps) to derive *p*-values. Networks for each
750 measure were derived by identifying the edges that were predictive in >50% of permutations. The
751 resulting networks were binarized, with a '1' or '0' indicating an edge's network membership. In
752 the interest of addressing the extent to which this 50% threshold influenced the overall network
753 mapping structure, networks that were weighted based on the number of times an edge was
754 predictive across permutations were also computed.
755
756 **Assessing the effects of instruments and subdomains on prediction strength**
757 We utilized one-way ANOVAs to assess the effect of measurement instruments and subdomains
758 on prediction strength for clinical and cognitive measures separately (**Fig 3a/b**). To do so,
759 prediction strength values from measures within the same clinical assessment subdomain or
760 cognitive testing battery were grouped together and compared to other subdomains/tests. The
761 ATQ, BRIEF, and PANAS inventories explicitly specified that the individual measures belonged
762 to separate subdomains (e.g., PANAS measures were split into 'positive' and 'negative' affect
763 domains). These separate subdomains were treated as separate categories for this analysis. The
764 BSI, IRI, PSQI, and PSS inventories did not specify how measures could be grouped into separate
765 subdomains. Therefore, all measures within these instruments were categorized as belonging to
766 the same subdomain for this analysis. Several of the cognitive testing batteries were composed
767 of sets of multiple separate cognitive tests. The D-KEFS consisted of the twenty-questions test,
768 the color-word interference test, the verbal fluency test, and the trail making test. Additionally, the
769 WASI was composed of a vocabulary test and a matrix reasoning test, the WRAML was
770 composed of a verbal learning test and a finger windows test, and the WAIS consisted of

771 cancellation, letter-number sequencing, digit-symbol coding, and symbol search tests. These
772 were all categorized as belonging to different subdomains for this analysis, since they were all
773 different cognitive tests. Conversely, all the metrics from the BNT captured various aspects of
774 BNT performance and were thus considered as belonging to the same subdomain.
775
776 **Obtaining measurement reliability data**
777 Test-retest reliability (ICC) measures were obtained from the original papers and/or technical and
778 user manuals published for each clinical and cognitive inventory. The BRIEF provided separate
779 ICC data from a normative sample and a mixed clinical/healthy adult population. We used the
780 ICC data from the mixed clinical/healthy adult sample, since it more closely matched our study's
781 participant population. The original paper validating the ATQ provided ICC data collected from 2
782 separate studies, one conducted in undergraduate students and the other in adults between 26
783 and 91 years old (median age = 57 years). We utilized the data from the study conducted in
784 undergraduate students because it provided ICC data for all the ATQ measures that we used
785 here, while the other study did not. Additionally, our sample's median age (31.17 years) is closer
786 to the median age of typical undergraduate populations. The IRI reported ICC data from males
787 and females separately. We averaged those results for use in the analysis here. There were no
788 ICC data reported for the 'total questions asked' measure from the twenty-questions subtest in
789 the D-KEFS, the verbal learning intrusion measure in the WRAML, or for the performance
790 measures on the BNT (except for 'total score'). Therefore, these measures were omitted from
791 analyses examining the relationship between ICC and prediction performance.
792
793 **Calculating distribution characteristics**

For each clinical and cognitive measure, skewness, variance, proportion performing at floor and ceiling, and the number of unique responses were calculated and compared to prediction strength. Skewness was calculated using python's scipy library. Variance was calculated by first performing min-max scaling to normalize and rescale each variable to a range from 0 to 1. The proportion performing at floor and ceiling were calculated by identifying the number of people with the min/max scores and dividing that number by the number of respondents.

Identifying relationships between measures and calculating network similarity

Spearman correlations were calculated for each pair of measures that CPM significantly predicted, and the derived *p*-values were FDR corrected for multiple comparisons. Network similarity between measures was assessed using Dice similarity coefficients. If there was a positive relationship between two measures, then the Dice coefficient was calculated between the positively predicting network for measure A and the positively predicting network for measure B, and then calculated once more between the negatively predicting network for measure A and the negatively predicting network for measure B. If there was a negative relationship between two measures, then the Dice coefficient was calculated between the positively predicting network for measure A and the negatively predicting network for measure B, and then calculated once more between the negatively predicting network for measure A and the positively predicting network for measure B. A weighted average of these two Dice coefficients was taken and used to examine overall network similarity between measures. This same approach was performed when comparing the unthresholded networks, except weighted Dice coefficients were utilized.

Comparing network similarity across measures

We sorted the measures based on network similarity in the interest of identifying meaningful commonalities and differences between predictive networks across measures. To achieve this,

819 we input the inverse of this full Dice similarity matrix (i.e., Dice distances) into an agglomerative
820 hierarchical clustering algorithm using Ward's method. This yielded a dendrogram that sorted the
821 measures based on their predictive networks' similarities. The measures were grouped into
822 clusters based on the hierarchical clustering results. The number of clusters chosen was
823 determined using the elbow method[41], which evaluates where the incremental decrease in within-
824 cluster sum of squares (WCSS) begins to slow across increasing values of *k*. The point of
825 maximum curvature (i.e., the "elbow") was objectively identified using the KneeLocator
826 algorithm[88]. This is a commonly used approach for determining an appropriate number of clusters
827 in a dataset. However, we performed clustering with the intention of reducing dimensionality,
828 showing how the measures grouped based on network similarity, and to compare important
829 features driving CPM predictions across measures, rather than to make any claims about how
830 many clusters exist in this dataset.
831
832 To identify the extent to which the same nodes were important to positively and negatively
833 predicting clinical and cognitive measures, we calculated the Jaccard similarity between vectors
834 representing the top 10% highest degree nodes in the positive and negative predictive networks
835 for each cluster. Overlap was summarized as the mean of the pairwise Jaccard values, and one-
836 sided permutation tests (10,000 iterations) were utilized for significance testing. The *p*-values
837 were calculated as the proportion of null replicates that exceeded the observed mean Jaccard
838 similarity.
839
840 **Identifying localizations of predictive networks**
841 The binarized positively and negatively predicting networks for measures within a cluster were
842 summed separately and Z-scored. This yielded aggregate positively and negatively predicting
843 networks for each cluster, with edges most important to predictions across measures weighted

844 most heavily. Before summing the networks, we ensured that the directionality of all the measures
845 within a cluster was similar so that networks positively related to symptoms or cognition would not
846 be mistakenly combined with networks negatively related to symptoms or cognitions. For
847 example, if there was one measure in a cluster where a higher score was desirable while lower
848 scores were desirable for the remaining measures, then the positive network for the first measure
849 would be summed with the negative networks for the other measures (and vice versa). The
850 distribution of edges within and between canonical networks was evaluated by summing the
851 edges within each canonical network pair from the aggregated cluster networks. We utilized the
852 Yeo17 networks[42] as network definitions, plus the cerebellum, brainstem, and subcortical regions.
853 The nodes important to predictions were examined by calculating degree centrality for each node.
854 The degree values were then projected onto the FreeSurfer template brain's ('fsaverage') pial
855 surface, along with cerebellar and brainstem surface reconstructions rendered in Bioimage Suite.
856 These degree plots were created using Python's nilearn package[89].
857
858 **Predicting cognition from network overlaps**
859 Prior to using clinical and cognitive network overlaps to predict cognitive performance, measures
860 in which a higher score was more desirable were inverted in their direction so that for all
861 measures, a higher score was less desirable. This was done because we were interested in how
862 poor clinical symptomatology related to cognitive deficits. Inverting the scores facilitated the ease
863 by which shared edges that were positively/negatively associated with both clinical and cognitive
864 measures in the training data could be identified. CPM models were subsequently run on all
865 measures within the same training/testing cross validation folds. Within each training fold, for all
866 possible pairwise combinations of clinical and cognitive measures, edges that were positively
867 correlated ($p < 0.05$) with both clinical and cognitive measures (after controlling for age and sex
868 using partial correlations) were summed together with the negative sum of the edges that were

negatively associated with both clinical and cognitive measures (also after controlling for age and sex using partial correlations). A linear model describing the relationship between the cognitive measure and the summed edges associated with both the cognitive and clinical measures were constructed. Then, that model was applied to the held-out testing data to predict the cognitive measure. This was performed for all 38 clinical measures and 26 cognitive measures that were previously determined to be modelable by CPM, yielding a total of 988 different cognitive measure predictions from sets of overlapping edges. These 10-fold cross-validation models were repeated 1000 times to obtain a distribution of predictions. Prediction strength was determined by comparing the observed cognitive measures to the predicted measures using Spearman correlations. Permutation tests were utilized for significance testing. Specifically, the number of null prediction values that were greater than the median prediction for each measure was divided by the number of permutations to derive the *p*-values. The *p*-values were FDR-corrected for 988 comparisons. Predictive overlapping networks were derived by identifying the edges that were predictive in >50% of permutations. The resulting networks were binarized, with a '1' or '0' indicating an edge's network membership.

To demonstrate that the specificity of the overlapping edges to predicting cognitive measures, an additional control analysis was performed in which an equal number of edges to the number of overlapping edges were randomly selected from the clinical networks in regions that did not overlap with the cognitive networks. This was similarly performed in the 10-fold cross validation procedure described above. Prediction strengths from these control analyses were compared to the prediction strengths from the overlapping network predictions by comparing the difference between median prediction strengths of the two sets of predictions to the difference of median prediction strengths from the two sets of predictions after shuffling, bootstrapped 1000 times. The

number of instances when the median difference from the shuffled predictions exceeded the true median prediction strength differences was divided by 1000 (the number of bootstraps) to derive *p*-values. The *p*-values were FDR-corrected for 988 comparisons.

Localizing overlapping edges that predicted cognition

Localization of overlapping edges that were predictive of cognition was achieved by summing the binarized overlapping network masks. By doing this, the frequency that predictive edges overlapped were indexed and weighed more heavily in the localization analyses. These overlapping edge frequency indices were averaged for all edges within each set of canonical network pairs to summarize the how the edges were distributed across networks. Again, the Yeo17 networks[42] were utilized as network definitions, plus the cerebellum, brainstem, and subcortical regions. In addition to edge distributions across networks, the localizations of the nodes commonly implicated in overlapping edge predictions were examined. To do so, the matrices indexing the frequencies of positively and negatively predicting overlapping edges were summed separately. This yielded two 268-element vectors corresponding to the nodal degree for each node in the Shen268 atlas. The degree centralities of the summed overlapping positively and negatively predicting networks were visualized on the brain surface the same way as before using Python's nilearn package[89].

Acknowledgements

This work was supported by funding from the National Institutes of Health (MH121095 to D.S. and R.T.C. and MH138347 and EB034720 to R.T.C.). AJS is also supported by the National Science Foundation's Graduate Research Fellowship Program. We would like to thank Xenophon Papademetris for help with data visualization and Sydney Smith for helpful conversations about data interpretation.

918

919 **Code availability**

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

924

925 **Competing interests**

926 G.S. has served as a consultant or scientific advisory board member to Axsome Therapeutics,
927 Biogen, Biohaven Pharmaceuticals, Boehringer Ingelheim International, Bristol-Myers Squibb,
928 Clexio, Cowen, Denovo Biopharma, ECR1, EMA Wellness, Engrail Therapeutics, Gilgamesh,
929 Janssen, Levo, Lundbeck, Merck, Navitor Pharmaceuticals, Neurocrine, Novartis, Noven
930 Pharmaceuticals, Perception Neuroscience, Praxis Therapeutics, Sage Pharmaceuticals, Seelos
931 Pharmaceuticals, Vistagen Therapeutics and XW Labs; and received research contracts from
932 Johnson & Johnson (Janssen), Merck and Usona. G.S. holds equity in Biohaven Pharmaceuticals
933 and is a co-inventor on a US patent (8,778,979) held by Yale University and a co-inventor on US
934 provisional patent application no. 047162-7177P1 (00754), filed on 20 August 2018 by Yale
935 University Office of Cooperative Research. Yale University has a financial relationship with
936 Janssen Pharmaceuticals and may receive financial benefits from this relationship. The University
937 has put multiple measures in place to mitigate this institutional conflict of interest. Questions about
938 the details of these measures should be directed to Yale University's Conflict of Interest office.
939 V.H.S. has served as a scientific advisory board member to Takeda and Janssen. The remaining
940 authors declare no competing interests.

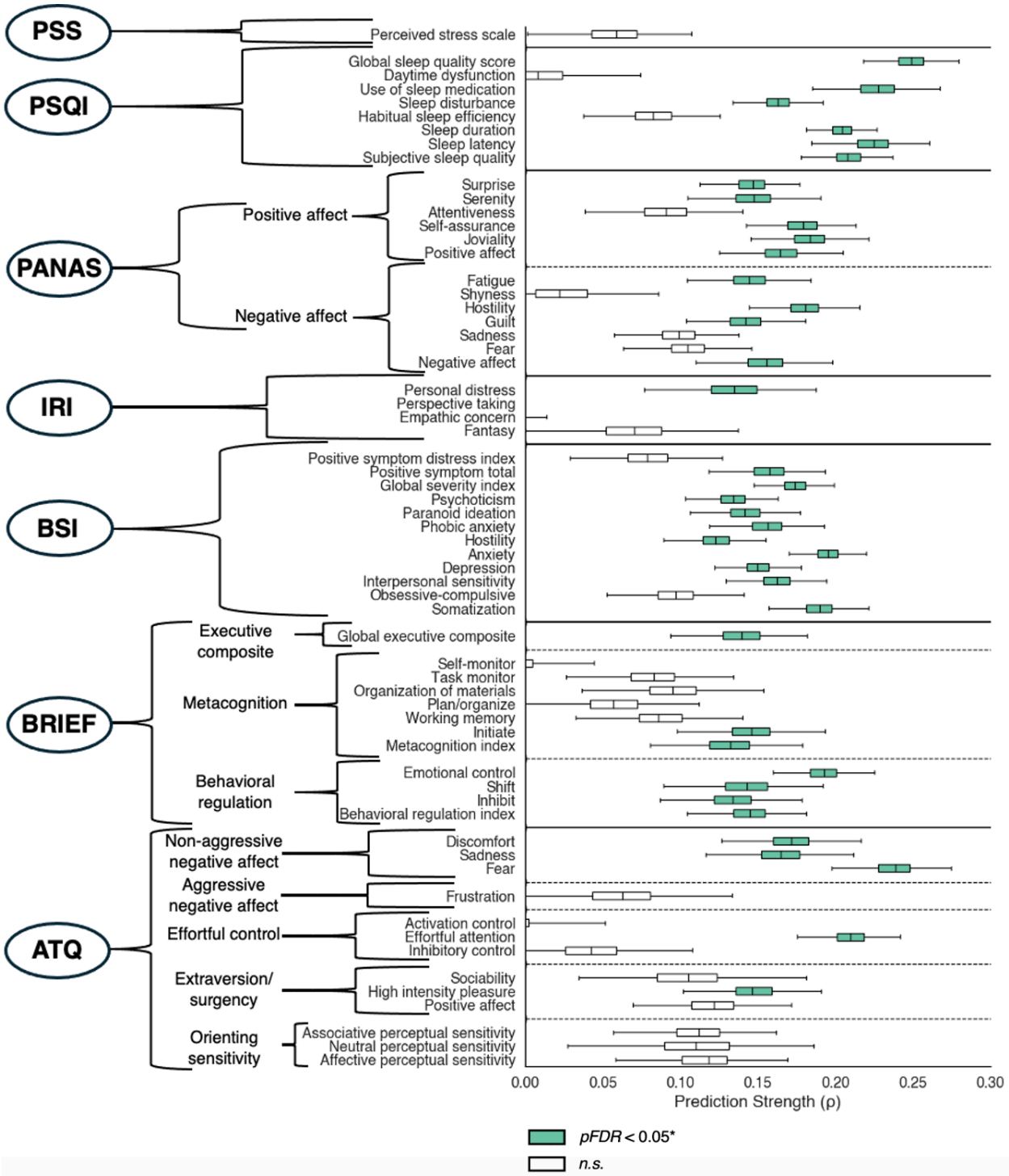

Figure 1. Prediction strength of clinical measures. The boxplots on the right show the prediction strength of each measure and are colored green if significantly predicted by CPM (*pFDR* < 0.05). Measures are grouped according to the instrument they belong to, as indicated by the encircled bolded acronym 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.

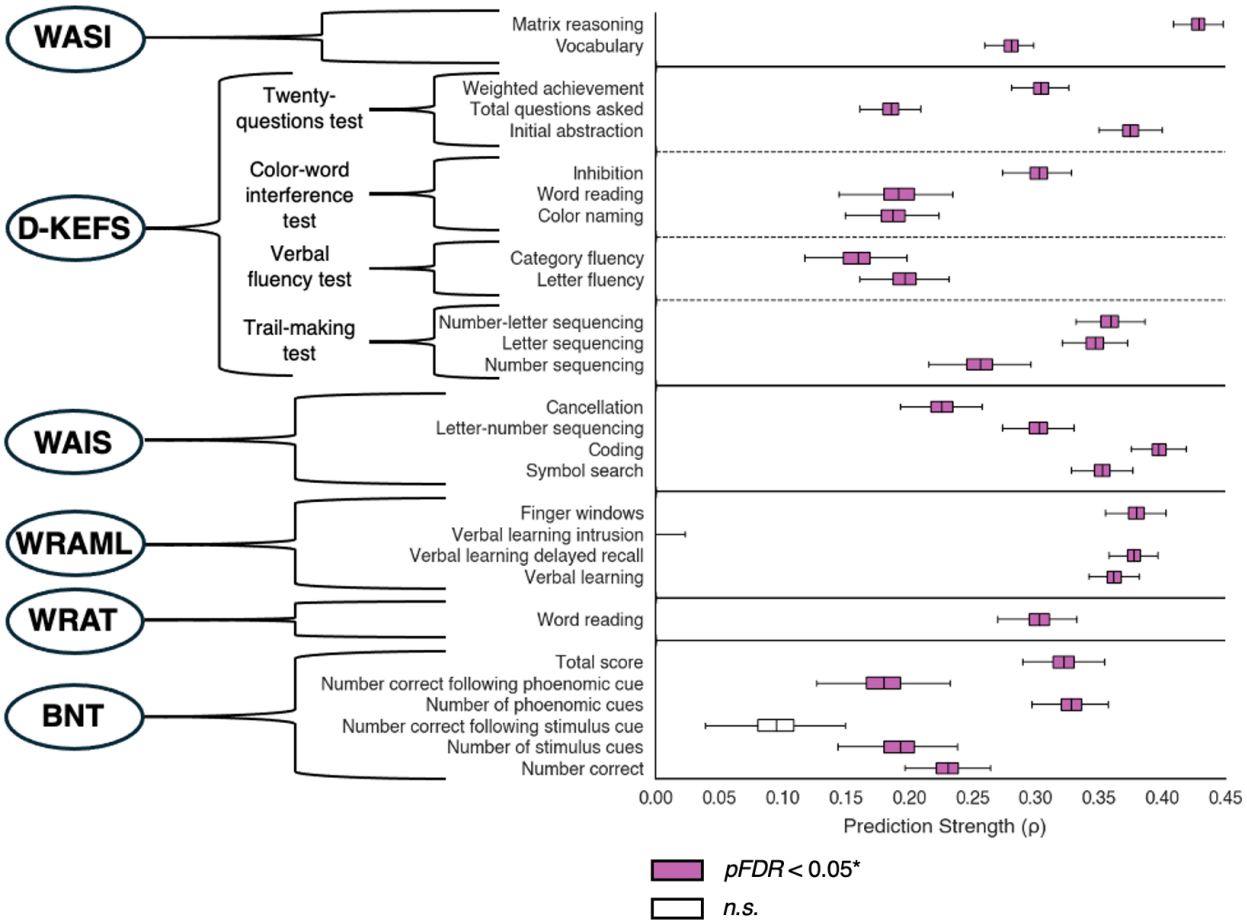

Figure 2. Prediction strength of cognitive measures. The boxplots on the right show the prediction strength of each measure and are colored purple if significantly predicted by CPM (*pFDR* < 0.05). Measures are grouped according to the testing battery they belong to, as indicated by the encircled bolded 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.

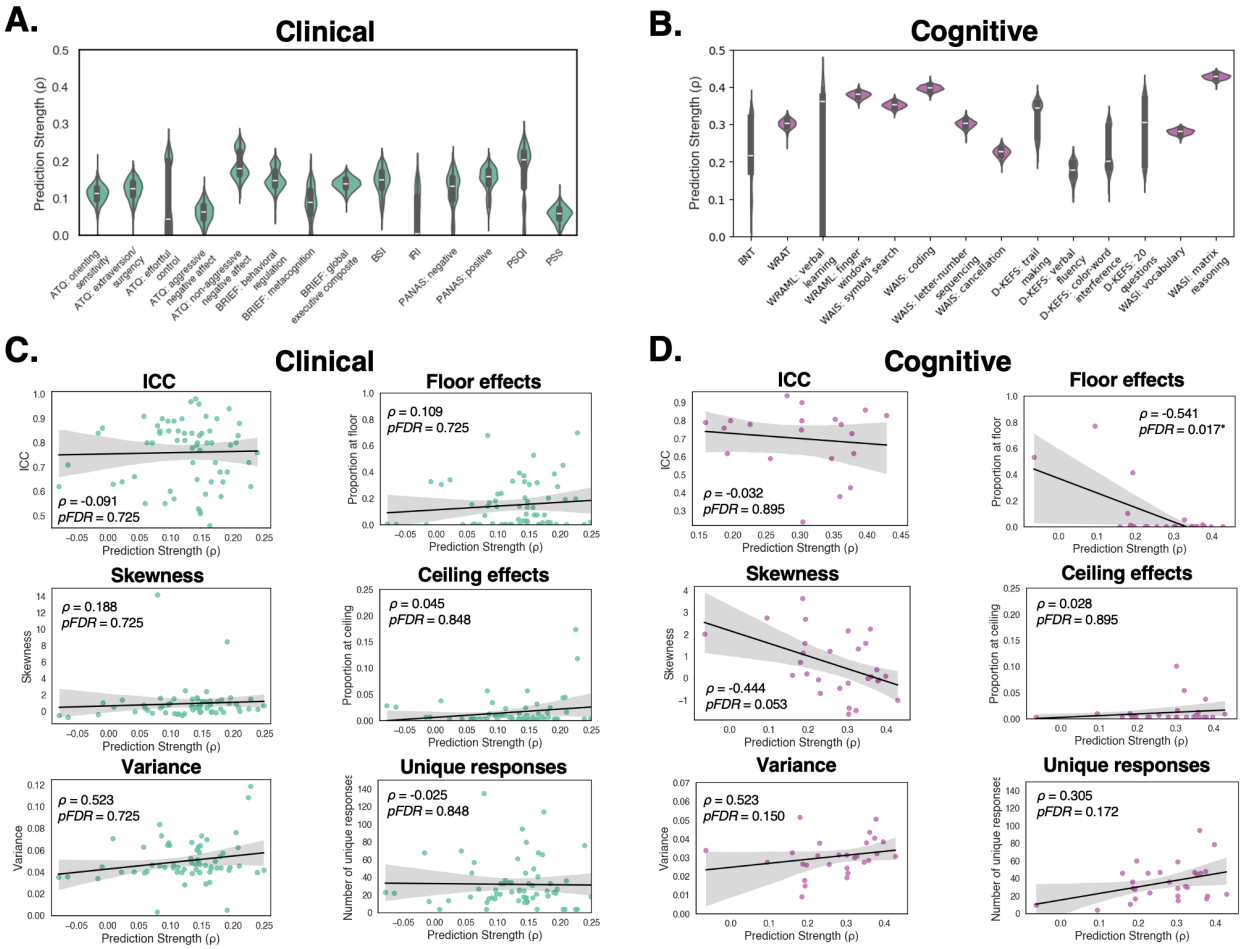

Figure 3. **Prediction strength was driven by phenotype measured. A)** Clinical predictions varied significantly by instrument and subdomain (one way ANOVA: $p < 0.001$). **B)** Cognitive predictions varied significantly by tests and subtests (one way ANOVA: $p < 0.001$). **C)** In clinical measures, prediction strength was unrelated to test-retest reliability (ICC) (top left), skewness (middle left), normalized response variance (bottom left), floor and ceiling effects (top and middle right), and the number of unique responses (bottom right). **D)** In cognitive measures, prediction strength was significantly related to floor effects (top right) and was trending towards being significantly related to skewness (middle left). Prediction strength was unrelated to test-retest reliability (ICC) (top left), the normalized response variance (bottom left), ceiling effects (middle right), and the number of unique responses (bottom right).

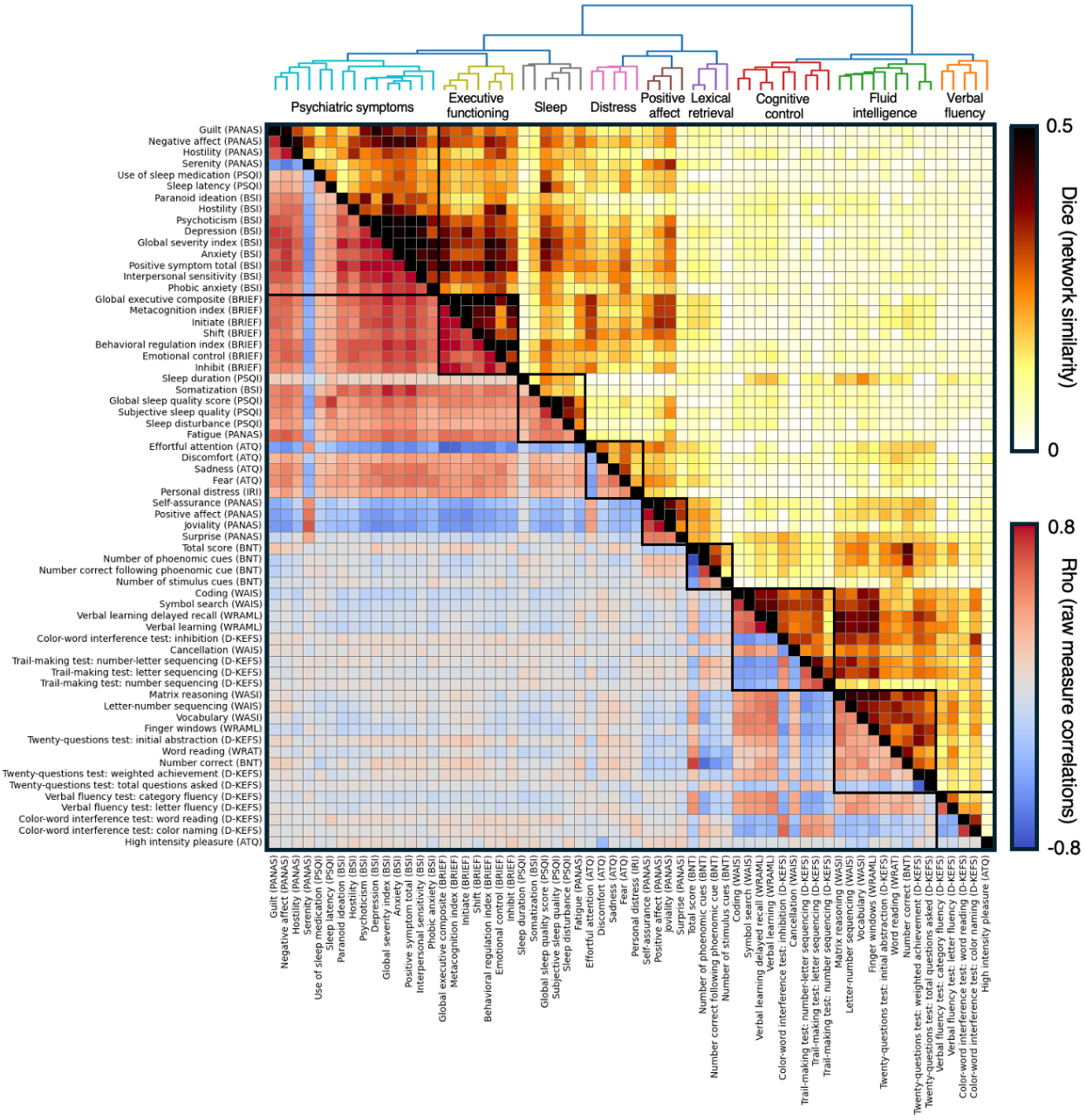

Figure 4. Relationships between measures. The measures' predictive networks were sorted according to their Dice similarities by an agglomerative hierarchical clustering algorithm (upper triangle). A dendrogram displayed above the matrix highlights how the measures grouped together based on their network 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).

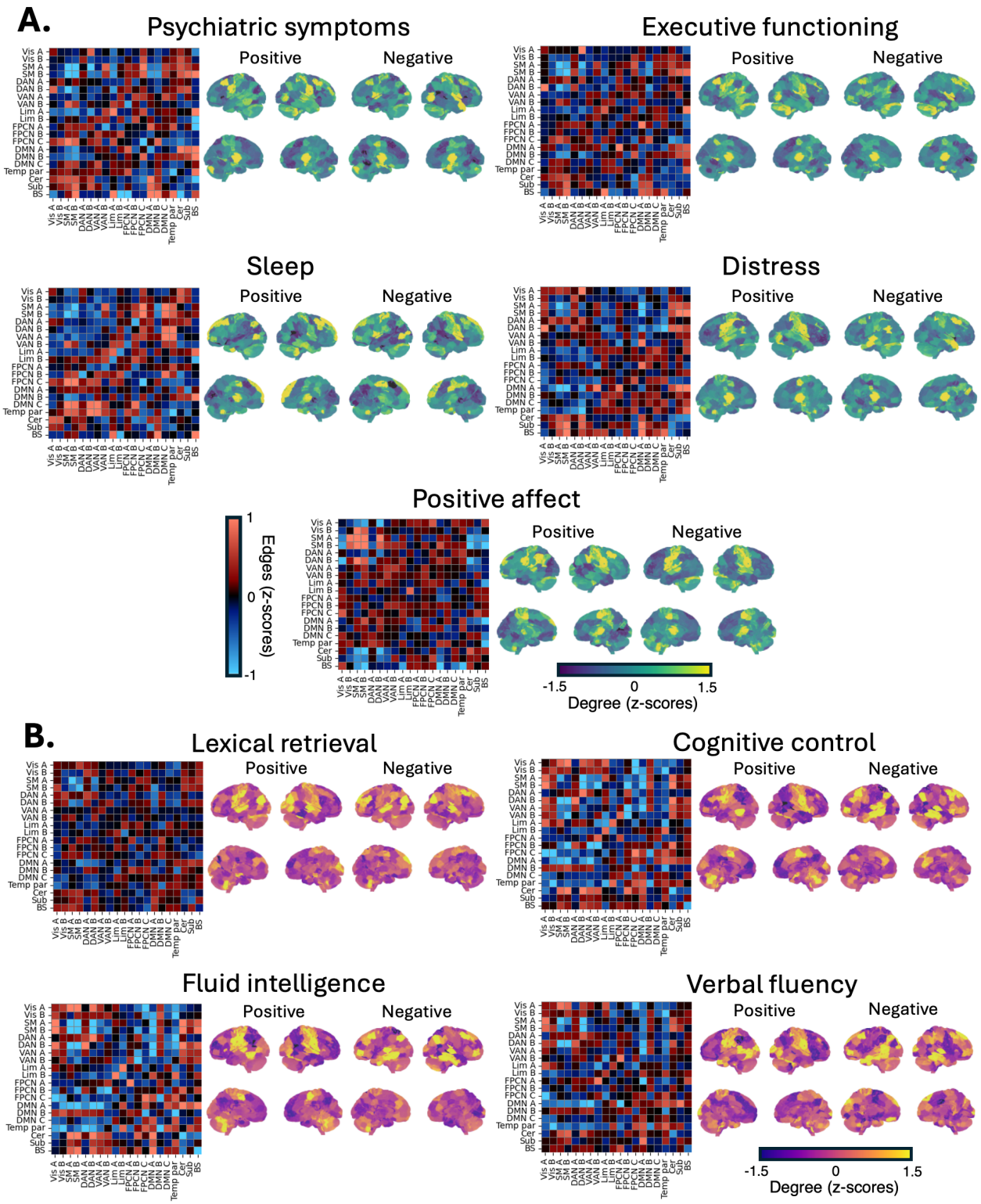

Figure 5. Brain networks driving predictions. A) The clusters summarizing networks important to predicting clinical measures are displayed with the descriptive labels 'psychiatric symptoms', 'executive functioning', 'sleep', 'distress', and 'positive affect'. **B)** The clusters summarizing networks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

988 ‘lexical retrieval’, ‘cognitive control’, ‘fluid intelligence, and ‘verbal fluency’. For each cluster,
989 degree plots highlight the important nodes for predictions across all measures, and matrices
990 highlight the distribution of edges connecting canonical networks. Networks are labeled according
991 to the Yeo 17 network definition, plus cerebellar, subcortical, and brainstem regions.
992

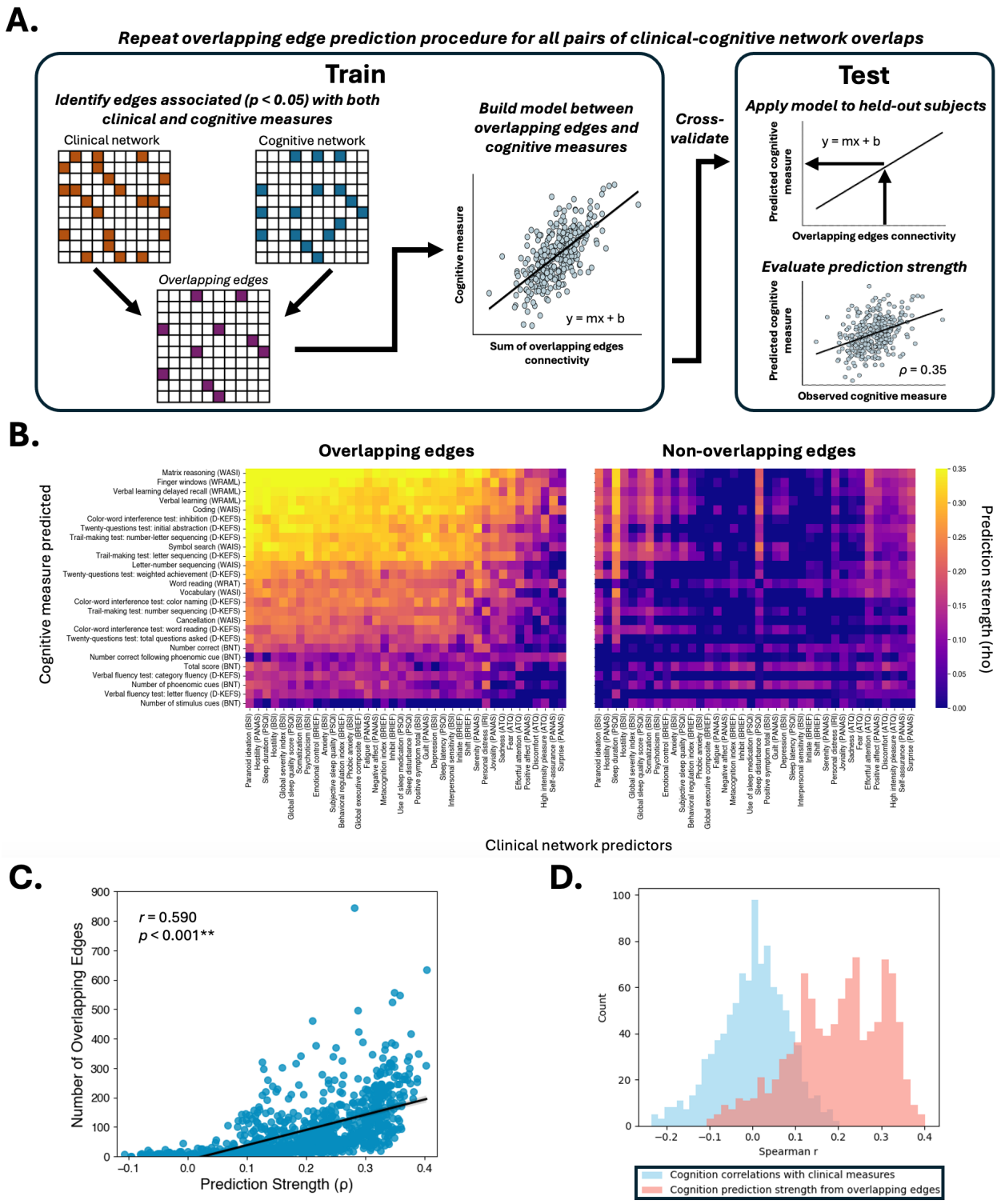

Figure 6. Predicting cognitive deficits from clinical-cognitive network overlaps. A) An overview of how the shared networks were identified and how predictive models were generated. Within each fold of a 10-fold cross validation, edges were identified that were significantly correlated ($p < 0.05$) in the training data with each behavioral measure for all 38 clinical measures and 26 cognitive measures. For every possible pairwise combination of clinical and cognitive

1001 measures (38 x 26 = 988 in total), the edges that were related to both measures in the training
1002 data were selected for constructing brain-behavior models. These models were then applied to
1003 the held-out testing data to predict the cognitive measures. This approach confined the cognitive
1004 measures' predictive networks to just the edges that were impinged upon by the clinical measures'
1005 networks. B) Prediction strength from overlapping edges. The cognitive measures were predicted
1006 (y-axis) from the edges in the clinical measures' networks (x-axis) that their networks overlapped
1007 with in the training data (left). Predictions that were not significant ($pFDR > 0.05$) were artificially
1008 set to 0 for visualization purposes. To demonstrate the specificity of this effect, cognitive
1009 measures were also predicted from edges in the clinical measures' networks that did not impinge
1010 upon cognitive networks in the training data (right). C) There was a significant relationship
1011 between the number of overlapping edges and prediction strength ($r = 0.590$, $p < 0.001$**). D) The
1012 prediction strength from the overlapping networks tended to be much stronger than the
1013 correlations between the clinical and cognitive measures' values (paired t-test: $p < 0.001$**).
1014

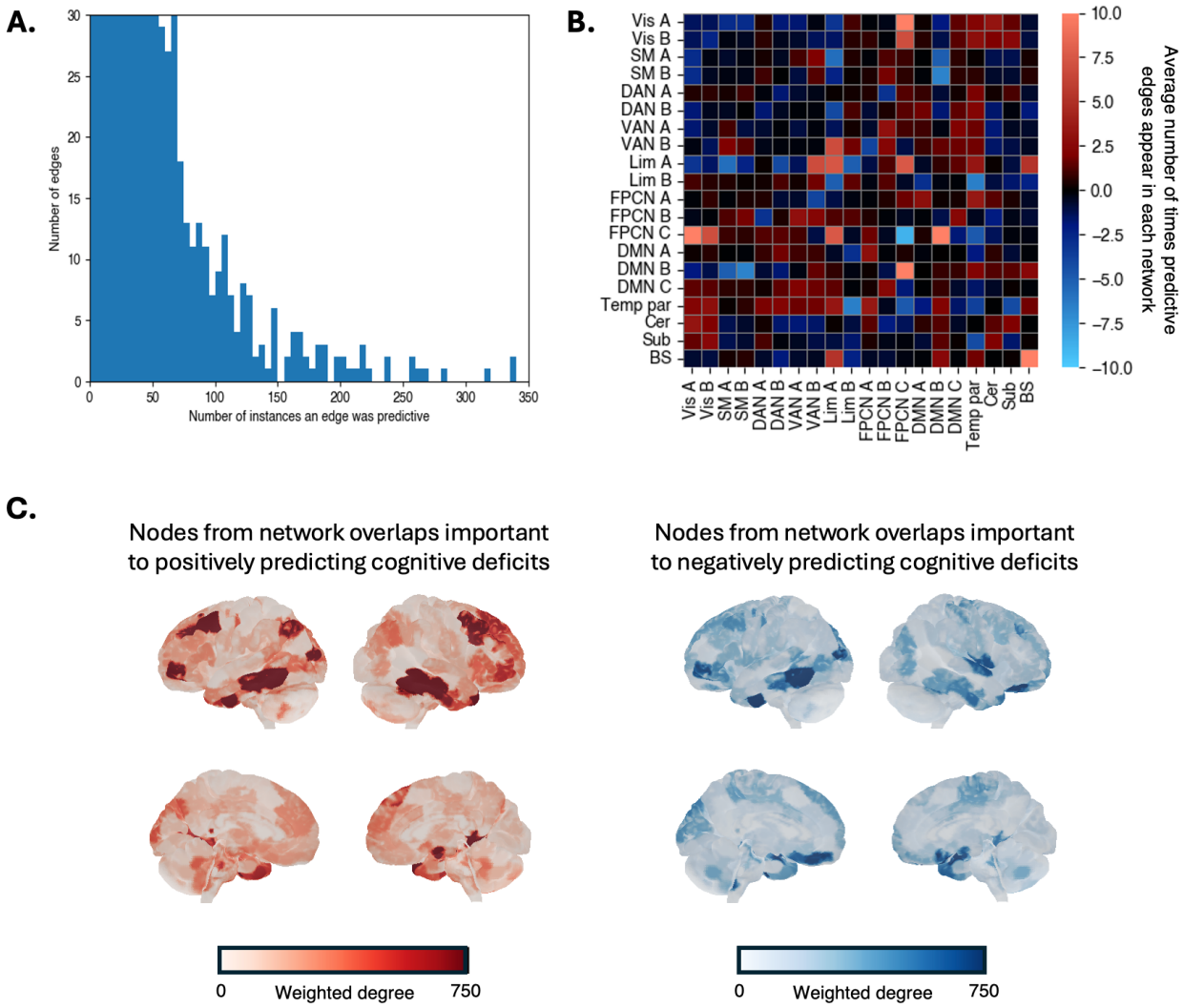

Figure 7. Localizations of overlapping edges that are frequently predictive of cognitive deficits. A) Histogram of how often overlapping edges are predictive of cognitive deficits. Most predictive edges only overlapped between clinical and cognitive networks a small proportion of times (441 out of 777 overlapping predictive networks shared <50 edges). However, there were edges that were frequently implicated in network overlaps. B) A summary of the canonical networks that the overlapping edges were distributed across. The frequency that overlapping edges were predictive was averaged across all edges within each set of canonical network-network pairs. Therefore, edges that often overlapped were given greater importance. Networks were labeled according to the Yeo 17 network definition, plus cerebellar, subcortical, and brainstem regions. The overlapping edges that most frequently positively predicted cognitive deficits were distributed between the frontoparietal control C to default mode B, visual A to frontoparietal control C, within brainstem, limbic A to frontoparietal control C, and within the limbic A network. The overlapping edges that most frequently negatively predicted cognitive deficits were distributed within the frontoparietal control C network. Negatively predicting edges were given negative values for plotting purposes. C) Surface plots displaying the overlapping edges' nodal degree (weighted by the number of times a predictive edge overlapped). The 5 nodes most important to positively predicting cognitive deficits were in the left dlPFC, right ventral

occipitotemporal cortex, bilateral posterior inferior temporal gyrus, and right anterior inferior temporal gyrus. The 5 nodes most important to negatively predicting cognitive deficits were in the left anterior inferior temporal gyrus, bilateral basal ganglia, right orbitofrontal cortex, and left posterior inferior temporal gyrus.

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 Spisak, T., Bingel, U. & Wager, T. D. Multivariate BWAS can be replicable with moderate sample sizes. *Nature* **615**, E4-E7 (2023). https://doi.org:10.1038/s41586-023-05745-x

13 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

14 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

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

16 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

17 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

18 Ooi, L. Q. R. *et al.* Comparison of individualized behavioral predictions across anatomical, diffusion and functional connectivity MRI. *Neuroimage* **263**, 119636 (2022). https://doi.org:10.1016/j.neuroimage.2022.119636

19 Zhi, D. *et al.* Triple Interactions Between the Environment, Brain, and Behavior in Children: An ABCD Study. *Biol Psychiatry* **95**, 828-838 (2024). https://doi.org:10.1016/j.biopsych.2023.12.019

20 Grayson, D. S. & Fair, D. A. Development of large-scale functional networks from birth to adulthood: A guide to the neuroimaging literature. *Neuroimage* **160**, 15-31 (2017). https://doi.org:10.1016/j.neuroimage.2017.01.079

21 Samardzija, A. *et al.* A Transdiagnostic fMRI Dataset With 300+ Deeply Phenotyped Subjects Across Resting and Task States. *medRxiv*, 2025.2007. 2015.25331595 (2025).

22 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

23 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

24 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

25 Gell, M. *et al.* How measurement noise limits the accuracy of brain-behaviour predictions. *Nature Communications* **15**, 1-12 (2024).

26 Evans, D. E. & Rothbart, M. K. Developing a model for adult temperament. *Journal of research in personality* **41**, 868-888 (2007).

27 Roth, R. M., Isquith, P. K. & Gioia, G. A. Behavior Rating Inventory of Executive Function®--Adult Version. *Archives of Clinical Neuropsychology* (2005).

28 Derogatis, L. R. & Melisaratos, N. The Brief Symptom Inventory: an introductory report. *Psychol Med* **13**, 595-605 (1983).

29 Davis, M. H. A multidimensional approach to individual differences in empathy. (1980).

30 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

31 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

32 Cohen, S., Kamarck, T. & Mermelstein, R. A global measure of perceived stress. *Journal of health and social behavior*, 385-396 (1983).

33 Delis, D. C., Kaplan, E. & Kramer, J. H. Delis-Kaplan executive function system. *Assessment* (2001).

34 Lichtenberger, E. O. & Kaufman, A. S. *Essentials of WAIS-IV assessment*. (John Wiley & Sons, 2012).

35 Wechsler, D. Wechsler abbreviated scale of intelligence. (1999).

36 Adams, W. Wide-Range Assessment of Memory and Learning. *The Corsini Encyclopedia of Psychology*, 1-2 (2010).

37 Wilkinson, G. S. & Robertson, G. J. Wide range achievement test 4. *Journal of Clinical and Experimental Neuropsychology* (1993).

38 Kaplan, E., Goodglass, H. & Weintraub, S. Boston naming test. *The Clinical Neuropsychologist* (1983).

39 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

40 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

41 Cui, M. Introduction to the k-means clustering algorithm based on the elbow method. *Accounting, Auditing and Finance* **1**, 5-8 (2020).

42 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

43 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

44 Nisbett, R. E. & Wilson, T. D. Telling more than we can know: Verbal reports on mental processes. *Psychological review* **84**, 231 (1977).

45 Lira, B. *et al.* Large studies reveal how reference bias limits policy applications of self-report measures. *Scientific reports* **12**, 19189 (2022).

46 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).

47 Schober, M. F. & Conrad, F. G. Does conversational interviewing reduce survey measurement error? *Public opinion quarterly*, 576-602 (1997).

48 Pendergrass, S. & Crawford, D. Using Electronic Health Records To Generate Phenotypes For Research. *Current Protocols in Human Genetics* **100** (2018). https://doi.org:10.1002/cphg.80

49 Russell, M. & Gajos, J. Annual Research Review: Ecological momentary assessment studies in child psychology and psychiatry. *Journal of child psychology and psychiatry, and allied disciplines* (2020). https://doi.org:10.1111/jcpp.13204

50 Balaskas, A., Schueller, S., Cox, A. & Doherty, G. Ecological momentary interventions for mental health: A scoping review. *PloS one* **16** (2021). https://doi.org:10.1371/journal.pone.0248152

51 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

52 Lahey, B. B. *et al.* Is there a general factor of prevalent psychopathology during adulthood? *Journal of abnormal psychology* **121**, 971 (2012).

53 Wang, B., LeBel, A. & D'Mello, A. M. Ignoring the cerebellum is hindering progress in neuroscience. *Trends in Cognitive Sciences* (2025).

54 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

55 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

56 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

57 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

58 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

59 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).

60 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

61 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

62 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

63 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

64 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

65 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).

66 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).

67 Cattell, R. B. Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology* **54**, 1-22 (1963). https://doi.org:10.1037/h0046743

68 Keyes, K. M., Platt, J., Kaufman, A. S. & McLaughlin, K. A. Association of Fluid Intelligence and Psychiatric Disorders in a Population-Representative Sample of US Adolescents. *JAMA Psychiatry* **74**, 179-188 (2017). https://doi.org:10.1001/jamapsychiatry.2016.3723

69 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

70 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

71 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

72 Xia, C. H. *et al.* Linked dimensions of psychopathology and connectivity in functional brain networks. *Nat Commun* **9**, 3003 (2018). https://doi.org:10.1038/s41467-018-05317-y

73 Liston, C. *et al.* Default mode network mechanisms of transcranial magnetic stimulation in depression. *Biol Psychiatry* **76**, 517-526 (2014). https://doi.org:10.1016/j.biopsych.2014.01.023

74 Kaiser, R. H., Andrews-Hanna, J. R., Wager, T. D. & Pizzagalli, D. A. Large-Scale Network Dysfunction in Major Depressive Disorder: A Meta-analysis of Resting-State Functional Connectivity. *JAMA Psychiatry* **72**, 603-611 (2015). https://doi.org:10.1001/jamapsychiatry.2015.0071

75 Stern, E. R., Fitzgerald, K. D., Welsh, R. C., Abelson, J. L. & Taylor, S. F. Resting-State Functional Connectivity between Fronto-Parietal and Default Mode Networks in Obsessive-Compulsive Disorder. *PLOS ONE* **7**, e36356 (2012). https://doi.org:10.1371/journal.pone.0036356

76 Chou, T., Dougherty, D. D., Nierenberg, A. A. & Deckersbach, T. Restoration of default mode network and task positive network anti-correlation associated with mindfulness-based cognitive therapy for bipolar disorder. *Psychiatry Res Neuroimaging* **319**, 111419 (2022). https://doi.org:10.1016/j.pscychresns.2021.111419

77 Manoliu, A. *et al.* Aberrant dependence of default mode/central executive network interactions on anterior insular salience network activity in schizophrenia. *Schizophr Bull* **40**, 428-437 (2014). https://doi.org:10.1093/schbul/sbt037

78 Williamson, P. Are anticorrelated networks in the brain relevant to schizophrenia? *Schizophr Bull* **33**, 994-1003 (2007). https://doi.org:10.1093/schbul/sbm043

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

80 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).

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

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

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

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

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

86 Lovden, M., Fratiglioni, L., Glymour, M. M., Lindenberger, U. & Tucker-Drob, E. M. Education and Cognitive Functioning Across the Life Span. *Psychol Sci Public Interest* **21**, 6-41 (2020). https://doi.org:10.1177/1529100620920576

87 Sareen, J., Afifi, T. O., McMillan, K. A. & Asmundson, G. J. G. Relationship Between Household Income and Mental Disorders: Findings From a Population-Based Longitudinal Study. *Archives of General Psychiatry* **68**, 419-427 (2011). https://doi.org:10.1001/archgenpsychiatry.2011.15

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

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