A neuromarker of sustained attention from whole-brain functional connectivity

Monica D. Rosenberg[1,7], **Emily S. Finn**[2,7], **Dustin Scheinost**[3], **Xenophon Papademetris**[3,4], **Xilin Shen**[3], **R. Todd Constable**[2,3,5], and **Marvin M. Chun**[1,2,6]

[1]Department of Psychology, Yale University, New Haven, CT 06520

[2]Interdepartmental Neuroscience Program, Yale University, New Haven, CT 06520 USA

[3]Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, CT 06520 USA

[4]Department of Biomedical Engineering, Yale University, New Haven, CT 06520 USA

[5]Department of Neurosurgery, Yale School of Medicine, New Haven, CT 06520 USA

[6]Department of Neurobiology, Yale University, New Haven, CT 06520 USA

Abstract

Although attention plays a ubiquitous role in perception and cognition, researchers lack a simple way to measure a person's overall attentional abilities. Because behavioral measures are diverse and difficult to standardize, we pursued a neuromarker of an important aspect of attention, sustained attention, using functional magnetic resonance imaging. To this end, we identified functional brain networks whose strength during a sustained attention task predicted individual differences in performance. Models based on these networks generalized to previously unseen individuals, even predicting performance from resting-state connectivity alone. Furthermore, these same models predicted a clinical measure of attention—symptoms of attention deficit hyperactivity disorder—from resting-state connectivity in an independent sample of children and adolescents. These results demonstrate that whole-brain functional network strength provides a broadly applicable neuromarker of sustained attention.

Introduction

Attention is integral to cognition and perception, underlying performance on almost every task in daily life. However, despite—or maybe because of—attention's pervasiveness,

attention research is increasingly specialized and fragmented, and investigators lack a simple, standardized way to summarize a person's attentional abilities. Although reducing any complex mental process to a single measure risks oversimplification, summary indices are theoretically and practically valuable. For example, intelligence research and education practice depend heavily on the ability to measure *gF*, an index of fluid intelligence[1–3], and working memory research relates numerous behaviors to a fundamental measure of capacity[4–6]. Comparable measures of attention have been elusive because behavioral tasks are diverse and not broadly standardized.

These challenges can be addressed with a brain-based measure of attention, which would summarize global attentional function and help researchers improve comparisons across individuals and track changes in attention longitudinally. As an initial step, we developed a neuromarker of sustained attention, or the ability to maintain focus and performance on a task at hand[7–9]. This neuromarker is based on intrinsic whole-brain functional connectivity, the degree to which brain activity in distinct neural regions is correlated over time. Synchronous fluctuations in the blood oxygenation level-dependent (BOLD) signal, measured with functional magnetic resonance imaging (fMRI), are thought to reflect functional connectivity in that they reveal regions engaging in common or related processing; these can be observed either during task performance or at rest, in the absence of an explicit task. Because sustained attention encompasses a variety of functions, including information selection, enhancement of selected information[10], and inhibition of unselected information[7], it is unsurprising that it involves a wide variety of brain regions, including the frontal and parietal cortices, thalamus, basal ganglia, ventral perceptual areas, and cerebellum[11–14]. Accordingly, whole-brain measures should provide a more holistic measure of attentional abilities than performance on a single task or activity in a single brain region. Practically, an attentional index based on whole-brain networks measured at rest is well-suited to use in both research and clinical contexts given that resting-state data is relatively straightforward to collect and share across acquisition sites and language and cultural barriers.

Here, with a fully cross-validated, data-driven analysis, we demonstrate that the strength of functional brain networks predicts sustained attention in novel individuals. We first model the relationship between connectivity strength and task performance in a subset of individuals as they perform the gradual-onset continuous performance task (gradCPT), a test of sustained attention and inhibition, during fMRI[15–19]. We demonstrate that our network model derived from these data, which we call the Sustained Attention Network model (SAN), predicts the behavioral performance of new individuals from their task-based connectivity. The model also generalizes to the resting state, predicting novel individuals' performance from connectivity observed during rest alone. As a final test of generalizability, we show that the SAN can also predict symptoms of attention deficit hyperactivity disorder (ADHD), which is characterized by deficits in sustained attention and inhibition[20], in children and adolescents collected at an independent research site. These results suggest that whole-brain functional connectivity is a robust neuromarker of sustained attentional abilities.

Results

Network definition

To test whether functional connectivity predicts attentional performance, we scanned 25 individuals as they performed the gradCPT, a test of sustained attention and inhibition that produces a range of behavior across healthy participants[17],[18]. Performance was assessed with sensitivity (d′). Given that head motion confounds analyses of functional connectivity, we confirmed that d′ was not correlated with average frame-to-frame motion during task performance ($r = 0.005$, $p = 0.98$; see online Methods for additional analyses ruling out motion confounds). We also collected resting-state data from each participant.

Network nodes were defined with a 268-node functional brain atlas designed to maximize the similarity of the voxel-wise timeseries within each node[21],[22]. This atlas, which comprises nodes with more coherent timecourses than those defined by the automatic anatomic labeling atlas[22], represents an improvement over anatomical parcellation schemes because anatomical boundaries do not necessarily match functional ones. Whole-brain coverage, including the cerebellum and brainstem, is another advantage of the current atlas. Although defining nodes based on a subset of regions of interest reduces the number of statistical comparisons and thus false positives, it may preclude discovery of informative connections and reduce the network's overall predictive power. Importantly, the problem of false positives can be addressed with cross-validation.

For each participant, a timecourse was calculated for each node by averaging the BOLD signal of all of its constituent voxels at each time point during task performance. Pairwise Pearson correlation coefficients were computed between the timecourses of each possible pair of nodes and were Fisher-normalized. The resulting 268×268 symmetric correlation matrices represented the set of connections or edges in each participant's task-based connectivity profile.

To assess the relevance of functional connections to behavior we performed the following analysis pipeline. First, robust regression between each edge in the connectivity matrices and d′ was performed across subjects. The resulting r-values were statistically thresholded at $p < 0.01$ and separated into a positive tail (edges whose strength indexed higher d′ across subjects) and a negative tail (edges whose strength indexed lower d′ across subjects). Mean r-value was $r = 0.59$ in the positive tail and $r = -0.58$ in the negative tail. When networks were defined on all subjects, the positive tail comprised 1,496 edges and the negative tail, 1,299 edges. Together these represent less than 8 percent of the brain's 35,778 total edges as defined by this atlas.

A single summary statistic, network strength, was used to characterize each participant's degree of connectivity in the positive and negative tails. Positive network strength was calculated by summing the edge strengths (Fisher-normalized r-values) from a participant's connectivity matrix in the edges of the positive tail, and negative network strength was calculated by summing the r-values of the edges in the negative tail. Network strength correlated with d′ across subjects, validating its use as a summary statistic (positive network strength: $r = 0.95$, $p = 1.30e^{-13}$; negative network strength: $r = -0.97$, $p = 2.44e^{-15}$). In

graph-theoretic terms, this statistic is equivalent to a weighted degree measure for each the networks (positive and negative)[23].

To confirm that d′ was more closely related to strength of the whole network than strength of individual edges, we compared the relationship between d′ and network strength to the relationship between d′ and every edge that appeared in the positive or negative tail across subjects. The correlation between d′ and positive network strength ($r = 0.95$) was numerically but not statistically higher than the strongest correlation between d′ and an individual edge in the positive tail ($r = 0.92$), Steiger's z[24] $= 1.25$, $p = 0.2$. It was, however, significantly higher than the second strongest d′-edge correlation in the positive tail ($r = 0.85$), Steiger's $z = 3.29$, $p = 0.001$. The correlation between d′ and negative network strength ($r = -0.97$) was more strongly negative than the strongest correlation between d′ and an individual edge in the negative tail ($r = -0.82$), Steiger's $z = 4.72$, $p = 2.39e^{-6}$. Thus, network strength as a whole better captures individual variability in d′ than any single edge.

Internal validation: Prediction from task connectivity

To determine if network strength predicted task performance in novel individuals, a leave-one-out cross-validation procedure was employed. In each set of $n–1$ participants, predictive networks were defined and used to calculate positive and negative network strengths as described above. Networks ranged in size from 1,099 to 1,540 edges. Next, simple linear models were constructed relating network strength during task performance to d′ in these individuals. Finally, these models were used to predict the left-out individual's d′ based on the strength of his or her positive and negative network during task performance. Pearson correlations between observed and predicted d′ scores were used to assess predictive power. All statistical tests were two-tailed.

Demonstrating that functional connectivity can be used to predict attentional performance in novel individuals, observed and predicted d′ values were significantly correlated (positive network: $r = 0.86$, $p = 3.4e^{-8}$; negative network: $r = 0.87$, $p = 1.6e^{-8}$; Fig. 1). A general linear model (GLM) constructed using strength in both networks also generated significant d′ predictions ($r = 0.84$, $p = 1.3e^{-7}$). However, GLM predictions were not more accurate than the positive (Steiger's $z = 0.51$, $p = 0.61$) and negative (Steiger's $z = 1.78$, $p = 0.08$) networks' predictions, suggesting that these two tails provide some degree of redundant information. Positive and negative networks did not differ in their predictive power (Steiger's $z = 0.45$, $p = 0.65$).

Internal validation: Prediction from rest connectivity

We next demonstrate that predictive networks generalize to resting-state data from novel individuals. To this end, we used the positive and negative network models described in the section labeled above as *Internal validation: Prediction from task connectivity*. However, these models are now applied to data collected at rest to predict the left-out individual's d′; in other words, the summary statistic of network strength is calculated based on the left-out individual's resting-state connectivity matrix rather than their task-based matrix.

Models trained on task data significantly predicted a previously unseen individual's task performance based on his or her resting-state data (correlation between predicted and

observed d′ values, positive network: positive network: $r = 0.49$, $p = 0.014$; negative network: $r = 0.49$, $p = 0.012$; GLM: $r = 0.43$, $p = 0.031$; Fig. 1). The positive and negative network models and the GLM did not differ in their predictive power (Steiger's $|z|$ values < 1.30, p values > 0.19). Although network models did not predict d' scores from resting-state connectivity as well as they did from task-based connectivity (Steiger's z values > 3.34, p values < 0.0009), significant predictions from rest data suggest that attentional abilities are reflected in intrinsic connectivity. This effect cannot be explained by head motion, as average frame-to-frame motion during resting runs was not correlated with d′ ($r = -0.17$, $p = 0.42$).

External validation: ADHD symptom prediction

As an even stronger test of generalizability, we applied these gradCPT network models to a completely independent validation dataset consisting of resting-state fMRI scans from 113 children and adolescents (age range 8 to 16) with and without ADHD diagnoses. These data were collected at Peking University and provided by the ADHD-200 Consortium[25]. In this dataset, attentional ability was assessed using the ADHD Rating Scale IV[26] (ADHD-RS), a clinical measure of ADHD on which a higher score indicates more frequent symptoms and/or a more severe attention deficit. In order to generalize our network model to this new dataset, we defined a *high-attention network* as the set of edges that appeared in the positive network of every iteration of the leave-one-out cross-validation described above in the section titled *Internal validation: Prediction from task connectivity*. A *low-attention network* was defined in an analogous way with edges whose strength was inversely correlated with d′ (Fig. 2). The high-attention network comprised 757 edges, and the low-attention network, 630 edges. In the full gradCPT sample, we constructed linear models relating high- and low-attention network strength (Sustained Attention Network, or SAN, models) to d′.

We then calculated the strength of the high- and low-attention networks during rest in each of the 113 individuals in the Peking University dataset, and submitted these strengths to the SAN models to make predictions about their attentional abilities. The high-attention network model inversely predicted ADHD-RS score ($r = -0.30$, $p = 0.001$; Fig. 3), indicating that individuals with more connectivity in the high-attention network showed less severe symptoms of an attention deficit. The low-attention network model also negatively predicted ADHD-RS score ($r = -0.34$, $p = 2.2e^{-4}$), such that individuals with more connectivity in the low-attention network showed higher symptom severity. Predictions of a GLM defined in the gradCPT dataset were also significantly correlated with ADHD-RS scores ($r = -0.34$, $p = 2.2e^{-4}$). Note that model predictions are inversely correlated with ADHD-RS scores because they were trained to predict d′; thus, higher predictions correspond to better attentional abilities and lower ADHD-RS scores. There was no correlation between average frame-to-frame head motion and observed ADHD-RS score ($r = 0.03$, $p = 0.78$), ruling out this potential confound.

To further confirm that SAN model predictions were specific to attentional abilities, we examined the relationship between predicted ADHD-RS scores and age and IQ, as measured by the Wechsler Intelligence Scale for Chinese Children-Revised[27]. After controlling for age and IQ, SAN model predictions remained significantly correlated with ADHD-RS score.

However, predictions were not correlated with age or IQ after controlling for the other two measures (Table 1). Thus, the model is capturing variance in functional connectivity that is closely related to attention rather than general cognitive ability.

Demonstrating that results were not driven by individual differences in general arousal, SAN model predictions were also anticorrelated with scores on the hyperactivity-impulsivity subscale of the ADHD Rating Scale-IV (positive: $r = -0.26$, $p = 0.006$; negative: $r = -0.32$, $p = 5.75\mathrm{e}^{-4}$; GLM: $r = -0.32$, $p = 4.96\mathrm{e}^{-4}$). That is, if the SAN model were predicting high arousal rather than vigilant attention, it is likely that predictions would correlate positively with hyperactivity scores. Instead, models predicted that hyperactive individuals had worse attention, suggesting that the results are not driven by individual differences in arousal.

With this further validation of the SAN model, we demonstrate that predictive networks not only generalize across cognitive states (task vs. rest), they also generalize across data acquisition site (New Haven vs. Beijing), age group (adults vs. children and adolescents), and—critically, given that we are pursuing a generalizable measure of sustained attention—behavioral measures of attention (gradCPT d′ vs. ADHD symptom scores).

Functional anatomy of attention networks

The high- and low-attention networks spanned numerous cortical, subcortical, and cerebellar nodes. To facilitate characterization of the biological substrates underlying these two networks, we summarized connectivity patterns in two ways. First, we grouped the 268 nodes into macroscale brain regions that were anatomically defined (e.g., cortical lobes) and examined relative numbers of connections between each pair of regions in each network. Second, we grouped nodes into eight canonical networks similar to those previously reported in resting-state literature (e.g., default mode); these networks were defined functionally using the same data used to create the original parcellation[28]. We then examined relative levels of within- and between-network connectivity represented in the high- and low-attention networks. Despite the complexity of the high- and low-attention networks that emerged from our data-driven model construction (see Fig. 2A), several anatomical trends emerged to distinguish them.

In the first analysis, we found that connections between motor cortex, occipital lobes, and the cerebellum were primarily predictors of better sustained attention, whereas connections between temporal and parietal regions, as well as intratemporal and intracerebellar connections, predicted worse attention across subjects (Fig. 2B). The involvement of the cerebellum in both networks provides evidence for a significant role of the cerebellum in attention and cognition[29,30]. In addition, although these findings may be unexpected given the traditional view of ADHD as primarily involving executive control regions and networks, recent work has emphasized the involvement of a variety of brain regions, including motor, occipital and parietal cortex and the cerebellum, in the disorder[31].

In the second analysis based on canonical functional networks, connections within the subcortical-cerebellum network, and connections between the subcortical-cerebellum network and the frontoparietal network appeared more frequently in the low- than the high-attention network (Fig. 2C). Connections between the subcortical-cerebellum network and

the medial frontal, motor, visual I, and visual association networks, on the other hand, appeared more frequently in the high-attention network (Fig. 2C). The involvement of the subcortical-cerebellum and medial frontal networks in the high-attention network mirrors observations of frontal-striatal-cerebellar circuit dysfunction in ADHD[32], and suggests that the connections that are disrupted in ADHD also characterize healthy individuals with poor attentional abilities.

To assess the importance of individual canonical networks to the SAN models, we computationally "lesioned" the high- and low-attention networks to exclude edges from each. That is, in an iterative analysis, we masked connectivity matrices to exclude edges that appeared in one of the eight canonical networks included in Fig. 2C. We then defined network models and predicted attention as described in the manuscript. For example, after excluding edges in the subcortical-cerebellum network, which contained 90 nodes, we submitted 178×178 matrices rather than 268×268 matrices to our analysis pipeline. We found that, in all cases, models missing one of the eight functional networks were still able to predict sustained attention from gradCPT and ADHD data (Supplementary Table 1).

Predictions of a lesioned matrix were significantly worse than predictions of the whole-brain matrix in only one case: when models were trained on a matrix that excluded the frontoparietal network and tested on gradCPT rest data (Steiger's $z = 2.11$, $p = 0.04$), although this did not survive Bonferroni correction for 24 comparisons. There was a trend such that ADHD predictions were worse when models were trained on matrices that excluded the default mode network (Steiger's $z = 1.85$, $p = 0.06$). When models were trained on matrices that excluded the visual I network, predictions from gradCPT task data were more successful than those made by the whole brain (Steiger's $z = 2.63$, $p = 0.01$), but again this did not survive Bonferroni correction. These results further emphasize the fact that models do not rely on strength in a single canonical network, but rather incorporate attention-relevant information from hundreds of diverse within- and between-network connections across the brain.

In addition to these analyses, we measured the importance of individual nodes by ranking them according to their sum of connections in the high- and low-attention networks; the most important nodes are presented in Supplementary Table 2. All of the top 10 most highly connected nodes were located in the cerebellum, temporal or occipital cortices, underscoring the importance of these regions for attentional function. Crucially, though, for most of these nodes the difference between the high- and low-attention networks was not in their overall degree of connectivity, but rather in their specific functional partners (note in Supplementary Table 2 that most of the top nodes had similar numbers of connections in both the high- and low-attention networks). This finding cautions against oversimplifying predictive networks to a handful of regions, instead emphasizing the need to consider specific pairwise connections across the entire brain to best characterize individuals' attentional ability.

ADHD and SAN network overlap

To identify edges that consistently predicted attentional function across datasets, we defined high- and low-ADHD networks in the full Peking University sample. These networks were constructed using the analysis pipeline described in the *Network definition* section above,

except that ADHD-RS score was used as the measure of attention instead of gradCPT d′. In addition, 236 nodes of the original 268 were used due to a lack of whole-brain coverage in some individuals (see online Methods for more information). Strength in the resulting high-ADHD network, containing 595 edges, was correlated with more severe symptoms scores (r = 0.75, p = 2.04e^{-21}); and strength in the low-ADHD network, 477 edges, was correlated with less severe symptoms (r = –0.76, p = 1.20e^{-22}). Note that this analysis is not cross-validated within the Peking University sample; rather, it validates network strength as a summary statistic in this dataset. Demonstrating that ADHD networks generalize to unseen subjects, models based on strength in the high- and low-ADHD networks during task and at rest predicted d′ in the gradCPT sample (Fig. 4); this is the reverse of the analysis described in the *External validation: ADHD symptom prediction* section above, indicating that this method achieves significant predictive power even after exchanging the roles of training and testing datasets.

Networks predicting better or worse sustained attention in both datasets had more common edges than those predicting opposite patterns of attentional function. While the high-attention network and the low-ADHD had 31 edges in common (edges in bold, Fig. 2A), the high-attention and high-ADHD networks had only two. In addition, the low-attention and high-ADHD networks shared 36 edges, while the low-attention and low-ADHD shared none. In a permutation test in which we compared 100 randomly generated positive and negative gradCPT and ADHD networks, overlap did not exceed 10 edges in any case (mean number of overlapping edges = 0.21, standard deviation = 0.53). Thus, the *p*-value associated with 31 and 36 common edges is 1/10,001 (see *Network overlap* in the online Methods for details).

BOLD variance does not predict attention

An important strength of SAN models is that they predict sustained attentional abilities from resting-state data. The use of resting-state data motivated us to use functional connectivity rather than overall activity as a predictor because connectivity can be calculated from data acquired at rest, while overall activity cannot (because there is no absolute measure of activity in resting runs).

To address whether a measure other than connectivity predicted attentional abilities, we tested models defined on BOLD variance. BOLD variance, a measure of the variability in the BOLD signal that can be calculated from resting-state data, is likely influenced by both metabolic function and anatomic factors such as partial volume effects introduced by the gray/white-matter segmentation and/or differing numbers of gray-matter voxels per node due to underlying variation in regional tissue volumes and gyral folding patterns.

BOLD variance models were defined in the same way as functional connectivity models, except that features consisted of 1×268 vectors of BOLD variance (one value per node) rather than 268×268 matrices of functional connections. In a cross-validated analysis analogous to that used to generate SAN models, BOLD variance models were defined on gradCPT task data and used to make predictions from gradCPT task, gradCPT rest, and ADHD data (see online Methods for details). Demonstrating that functional connectivity is a

better predictor of attention than BOLD variance, these models did not successfully predict sustained attentional abilities (Supplementary Table 3).

Discussion

In a group of adults performing a sustained attention task, we identified functional brain networks whose strength predicted individual differences in task success. These whole-brain network models predicted novel individuals' task performance from resting-state data alone, providing evidence for meaningful attention-related signal in patterns of intrinsic connectivity. Demonstrating that models are robust and generalizable, networks defined on sustained-attention task data predicted a clinical measure of ADHD in children and adolescents from a completely independent sample. That is, connections that predicted better task performance in the Yale dataset predicted less severe ADHD symptoms in the Peking University dataset, and connections that predicted worse performance predicted more severe ADHD symptoms. This result—that complex brain network models predict different measures of attention in disparate populations—demonstrates that functional brain networks can serve as a holistic neural index of sustained attention.

The current models, which generalize across two datasets and two measures of sustained attention, make significant progress towards identifying a neuromarker of sustained attention. However, they do not imply that sustained attention is a unitary process. Rather, the overall sustained attention factor measured here likely recruits many cognitive and attentional processes (such as inhibition), which are captured by the data-driven functional connectivity analyses. Future behavioral work with a wider range of tasks is needed to determine whether a single attention factor is feasible[33], such as *g* in intelligence research.

The current result also suggests that models based on functional brain networks are powerful, generalizable predictors of cognitive abilities. Although previous studies have demonstrated that pre-task functional connectivity is correlated with perceptual task performance[34], and that resting-state functional connectivity predicts fluid intelligence within a single data set[28,35], we are not aware of any study to date that has demonstrated the use of functional network models for successful across-dataset prediction of a cognitive ability.

The proposed neuromarker of sustained attention, the Sustained Attention Network (SAN) model, complements existing work on individual differences in attention and offers several advantages. Importantly, SAN models are predictive rather than descriptive in nature and thus contribute to one of the primary goals of human neuroimaging: to identify neuromarkers that can predict a person's educational or health outcomes[36,37]. Here, predictions can be made from resting-state data collected over a short period of time (in the ADHD sample, only eight minutes), which facilitates data sharing and further tests of generalizability. The use of resting-state data is especially advantageous in populations that have difficulty performing tasks, and allows an unbiased way for researchers and clinicians to track and compare attentional function longitudinally across development or training, unconfounded by changes in task performance.

In addition to demonstrating that functional connectivity is a powerful predictor of attentional abilities, our results support recent characterizations of sustained attention as emerging from coordinated activity across wide swaths of cortex as well as subcortical regions and the cerebellum[14,38,39], and demonstrate that attentional mechanisms extend beyond traditional attention regions and networks. For example, although nodes in prefrontal and parietal cortex, which are implicated in numerous tasks requiring the deployment and maintenance of attention[7,12,39], factored into the predictive network models, only 27% of edges in the high-attention and 34% of edges in the low-attention network involved nodes in these regions. Instead, the current results highlight the importance of data-driven analyses that do not constrain features to *a priori* nodes or edges of interest[40].

The fact that network models defined in healthy adults from the Yale-New Haven community predicted ADHD symptoms in children and adolescents from Beijing suggests meaningful overlap between the neural mechanisms important for sustained attention and the neural dysfunction that leads to an ADHD diagnosis. Although valuable research has identified differences in functional connectivity between individuals with ADHD and controls in frontal, parietal, temporal, and occipital cortices as well as in the cerebellum and striatum[41_48], these comparisons do not address whether the connections that go awry in ADHD are disrupted, to a lesser degree, in individuals with subclinical attention problems. The current findings suggest that it may be useful to consider ADHD as a continuum of neural and behavioral dysfunction rather than an all-or-nothing disorder.

Our findings compel a large research program to further validate the proposed Sustained Attention Network model across different attentional operations and tasks. The model presented here is a highly promising starting point, given that it generalizes across acquisition sites and participant populations, relies on a version of the widely used continuous performance task[15_18] and predicts task performance and clinical measures of ADHD. However, while these two measures are both related to attentional abilities, they do not capture the exact same construct, and this is likely why there is significant but not total overlap between edges in the models trained on the two datasets. Stronger claims about the specificity and generalizability of the current model will depend on future work in which models are trained and tested on data from a wide variety of attention tasks; the use of neural data in addition to behavioral measures may even help separate and cluster the many cognitive processes involved in attention. The sharing of resting-state data coupled with behavioral attention task data in public databases such as ADHD-200 will facilitate these efforts. To improve model generalizability and predictive power, researchers can collaborate to identify edges that most consistently predict attention (or another trait or cognitive ability). The analysis pipeline described here can be applied to any dataset that includes fMRI data—ideally, at least some of which is acquired at rest—and a measure of attention, and labs can share the resulting predictive networks. Defining a neuromarker of attention based on edges that appear commonly across tasks may reduce the risk of overfitting and improve generalizability.

In sum, we demonstrate that intrinsic brain connectivity is a powerful predictor of sustained attention. Beyond this finding, the current whole-brain, data-driven functional connectivity

approach can be useful in predicting a wide range of other cognitive abilities and clinical symptoms.

online Methods

gradCPT dataset

Participants—Thirty-one individuals from Yale University and the surrounding community performed a sustained attention task, the gradCPT[17,18], during fMRI data acquisition. Six were excluded for excessive head motion, defined *a priori* as > 2 mm translation or > 3 degrees rotation, in all runs or for a lack of whole-brain coverage, leaving 25 for analysis (13 females, ages 18–32 years, mean age = 22.7 years). All were right handed and had normal or corrected-to-normal vision. Participants gave written informed consent in accordance with the Yale University Human Subjects Committee and were paid for their participation.

A post-hoc power analysis revealed that the statistical power of the gradCPT task analysis (train on *n*–1 participants' task matrices and test on the left-out participant's task matrix) was greater than 0.99. The power of the gradCPT rest analysis (train on *n*–1 participants' task matrices and test on the left-out participant's rest matrix) was greater than 0.62.

Paradigm and stimuli—Participants performed the gradCPT[17,18] during fMRI scanning. Stimuli were grayscale images of city and mountain scenes with a diameter of 256 pixels. Presented at the center of the screen, they subtended a diameter of approximately 7° of visual angle.

On each trial, an image gradually transitioned from one to the next using linear pixel-by-pixel interpolation. Each transition took 800 ms. For 800 ms the current scene transitioned from the previous scene and for the next 800 ms it transitioned to the next. Participants were instructed to respond via button press to city scenes, which occurred randomly 90% of the time, and to withhold response to mountains. Accuracy was emphasized without reference to speed.

Task runs consisted of four 3-min blocks of the gradCPT interleaved with three 30-sec blocks of rest (breaks). Breaks were indicated with a fixation circle in the center of the screen. To warn participants of the upcoming task, a dot replaced the circle for 2 seconds at the end of each break. Eight seconds of fixation, excluded from analyses, were included at the start of each run. During breaks and resting-state runs, participants were instructed to attend to the fixation circle in the center of the screen.

Procedure—Following acquisition of an anatomical magnetization prepared rapid gradient echo (MPRAGE), a 6 min resting scan and three 13:44 min runs of the gradCPT were collected. An additional 6 min resting scan was collected after task runs.

Behavioral analysis—Sensitivity (d′) was used to measure task performance. For each task block, d′ was calculated as z(hit rate)–z(false alarm rate) (in Matlab, norminv(hit rate)–

norminv(false alarm rate)). For each participant, overall d′ values were calculated by averaging d′ across blocks.

Because stimuli were constantly in transition, an iterative algorithm was used to assign key presses to individual trials and determine accuracy[17],[18]. First, the algorithm assigned unambiguous key presses. Unambiguous presses to image *n* were those that occurred after image *n* was 80% cohered and before image *n* + 1 was 40% cohered. Next, any ambiguous presses were assigned to an adjacent trial if one of the two had no response, or to the closest trial if both had no response (unless one was a mountain, in which case participants were given the benefit of the doubt that they had correctly withheld a response). If multiple presses could be assigned to a trial, the fastest response was selected. Slight variations to this algorithm yielded highly similar results.

D′ reliability was calculated with a Spearman-Brown-corrected split-half correlation comparing average performance of odd-numbered task blocks to average performance of even-numbered task blocks. D′ reliability was 0.975, which is considered excellent.

Imaging parameters and preprocessing—FMRI data were collected at the Yale Magnetic Resonance Research Center on a 3T Siemens Trio TIM system equipped with a 32-channel head coil. Functional runs included 824 (task) or 363 (rest) whole-brain volumes acquired using a multiband echo-planar imaging (EPI) sequence with the following parameters: repetition time (TR) = 1000 ms, echo time (TE) = 30 ms, flip angle = 62°, acquisition matrix = 84 × 84, in-plane resolution = 2.5 mm^2, 51 axial-oblique slices parallel to the ac-pc line, slice thickness = 2.5, multiband 3, acceleration factor = 2. MPRAGE parameters were as follows: TR = 2530 ms, TE = 3.32, flip angle = 7°, acquisition matrix = 256 × 256, in-plane resolution = 1.0 mm^2, slice thickness = 1.0 mm, 176 sagittal slices. A 2D T1-weighted image with the same slice prescription as the EPI images was also collected for purposes of registration.

Data were analyzed using BioImage Suite[49] and custom scripts in Matlab (Mathworks). Motion correction was performed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). Linear and quadratic drift, mean signal from cerebrospinal fluid, white matter, and gray matter and a 24-parameter motion model (6 motion parameters, 6 temporal derivatives, and their squares) were also regressed from the data. Finally, data were temporally smoothed with a zero mean unit variance Gaussian filter.

Due to excessive head motion, defined *a priori* as > 2 mm translation or > 3 degrees rotation during a single run, one task run from each of five participants was excluded from analysis, and one resting run was excluded from each of two. Head motion, calculated as mean frame-to-frame displacement, did not correlate with d′ in any of the three task runs (first: $r = 0.08$, $p = 0.71$; second: $r = -0.10$, $p = 0.62$; third: $r = -0.10$, $p = 0.65$). Average d′ across the three task runs was not significantly correlated with average head motion during task runs ($r = 0.005$, $p = 0.98$) or rest runs ($r = -0.17$, $p = 0.42$).

Additional motion controls—As an additional control for motion, we confirmed that predictions of the leave-one-subject-out models described in the *Internal validation:*

Prediction from task connectivity section of the main text did not correlate with mean frame-to-frame head motion. In other words, having established that observed performance was not correlated with head motion, we also verified that predicted performance was not correlated with head motion. Indeed, predictions based on gradCPT subjects' task data were uncorrelated with motion during task (positive network: $r = 0.03$; $p = 0.88$; negative network: $r = 0.04$, $p = 0.84$; GLM: $r = 0.05$, $p = 0.80$), and predictions based on gradCPT subjects' resting-state data were uncorrelated with motion during rest (positive network: $r = -0.05$; $p = 0.80$; negative network: $r = 0.06$, $p = 0.77$; GLM: $r = 0.12$, $p = 0.58$).

We were also unable to predict motion with network models explicitly trained on this variable. That is, "motion prediction models" were defined identically to those described in the *Internal validation: Prediction from task connectivity* section of the main text, except with average frame-to-frame head motion in place of d′. Predictions of these "motion models" based on gradCPT subjects' task data were not significantly correlated with motion during task (positive network: $r = -0.25$; $p = 0.22$; negative network: $r = -0.07$, $p = 0.72$; GLM: $r = -0.18$, $p = 0.40$), and predictions based on gradCPT subjects' resting-state data were uncorrelated with motion during rest (positive network: $r = 0.07$; $p = 0.73$; negative network: $r = -0.28$, $p = 0.18$; GLM: $r = -0.10$, $p = 0.65$). Thus, despite the fact that head motion can pose a confound for functional connectivity analyses of fMRI data, the success of the SAN model appears to be unconfounded by artifacts related to head motion.

Network construction—Network nodes were defined using a groupwise graph-theory-based parcellation algorithm that maximized the similarity of the timeseries of the voxels within each node[21,22]. To obtain the 268-node atlas used in the current study, the parcellation algorithm was applied to resting-state data from an independent sample of 45 healthy adults scanned at the Yale Magnetic Resonance Research Center.

The 268-node atlas was warped from MNI space into single-subject space via concatenation of a series of linear and non-linear registrations between the functional images, 2D and 3D anatomical scans, and the MNI brain. All transformation pairs were calculated independently, combined into a single transform, and inverted, warping the functional atlas into single participant space. This single transformation reduces interpolation error because the functional atlas is warped to an individual with only one transformation. All transformations were estimated using the intensity-based registration algorithms in BioImage Suite.

For each participant, task matrices were calculated using data concatenated across task runs, excluding data collected during the intervening rest breaks (as well as the 6 seconds following them to account for hemodynamic delay). Rest matrices were calculated using data concatenated across rest runs.

ADHD-200 dataset

Participants—Data were provided by the ADHD-200 Consortium[25], a publically available dataset of resting-state fMRI data of children with and without ADHD from eight sites across the globe. The current study includes data from the Peking University site, which had a large number of subjects with relatively low head motion. The Research Ethics Review

Board of Institute of Mental Health, Peking University, approved data collection; informed consent was obtained from each participant's parent, and all children agreed to participate in the study. Detailed descriptions of inclusion criteria and imaging parameters and procedures are available online at fcon_1000.projects.nitrc.org/indi/adhd200.

ADHD diagnosis was established with the Schedule of Affective Disorders and Schizophrenia for Children—Present and Lifetime Version (K-SADS-PL)[50], and the ADHD Rating Scale IV (ADHD-RS)[26] was used to obtain dimensional measures of ADHD symptoms. The ADHD-RS is composed of 18 questions, nine of which assess inattention, or how children attend to tasks or play activities, such as the degree to which a child "fails to give close attention to details" or is "easily distracted by extraneous stimuli." The remaining nine assess hyperactivity and impulsivity levels, such as the degree to which a child "fidgets with hands or feet or squirms in seat" or "interrupts or intrudes on others." Questions are rated on a 4-point Likert scale (0 = rarely or never, 3 = always or very often); higher scores represent more severe and/or more frequent symptoms. Overall ADHD score is calculated as the sum of all responses. Raw scores are converted to percentiles based on each child's age and gender. IQ was assessed with Wechsler Intelligence Scale for Chinese Children-Revised.

The original Peking University dataset consisted of one 8-min run of resting state data collected from 245 subjects (102 patients with an ADHD diagnosis and 143 typically developing controls; 71 females; mean age = 11.7 years; mean ADHD-RS score = 38.3). Data were concatenated across three datasets with slightly different scanning parameters. Subjects were excluded for missing ADHD-RS scores (23 subjects), missing fMRI data in one or more nodes of a 236-node functional atlas (see *Network Construction* below for more details; 7 subjects), or quality control flags provided by the acquisition site (3 subjects). In the remaining 212 subjects, mean frame-to-frame head displacement was correlated with ADHD-RS score, $r = 0.22$, $p = 0.001$. To eliminate this relationship, we incrementally lowered a motion threshold before performing prediction analyses. A threshold of 0.06 mm was selected to minimize the correlation between motion and ADHD-RS score, so 99 subjects with mean frame-to-frame displacement >0.06 mm were excluded from further analysis.

The final set of subjects consisted of 113 individuals (38 patients; 35 females; mean age = 11.8 years; range = 8–16 years; mean ADHD-RS score = 35.5). Of these patients, 25 were medication-naïve; the others' psychostimulant medication was withheld starting at least 48 hours before scanning. All 75 typically developing controls were medication-naïve. There were no correlations between mean frame-to-frame displacement or age and ADHD-RS score (motion and ADHD-RS score: $r = 0.03$, $p = 0.78$; age and ADHD-RS score: $r = -0.06$, $p = 0.56$). In the 112 subjects for whom an IQ score was provided, IQ was inversely related to ADHD-RS ($r = -0.27$, p = 0.004) and age ($r = -0.17$, $p = 0.07$).

Note that individuals in the ADHD-200 dataset were not randomly assigned to groups, but were labeled as "patients" (individuals with ADHD) or "controls" (individuals without ADHD). The information available at http://fcon_1000.projects.nitrc.org/indi/adhd200/ does not specify whether investigators involved in data collection knew each individual's

diagnostic status, but these investigators were not involved in assessing the outcome of the current experiment.

Image preprocessing—Images were slice-time and motion corrected using SPM5 (http://www.fil.ion.ucl.ac.uk) and then iteratively smoothed until the smoothness for any image had a full width half maximum of approximately 6 mm[51]. This iterative smoothing process minimizes motion confounds associated with resting-state fMRI[52]. All further analyses were performed using BioImage Suite[49] unless otherwise specified. Several covariates of no interest were regressed from the data including linear and quadratic drift, six rigid-body motion parameters, mean cerebral-spinal fluid signal, mean white matter signal, and mean global signal. Finally, the data were temporally smoothed with a zero mean unit variance Gaussian filter (cutoff frequency = 0.12 Hz).

Network construction—Network nodes were defined using a subset of nodes of the 268-node functional brain atlas used for the gradCPT network analysis[22]. As some scans did not include full cortex and cerebellum coverage, nodes missing in at least three subjects were removed. This process resulted in the removal of 32 nodes mainly in the inferior portions of the cerebellum, brainstem, temporal poles and orbital frontal cortex (Supplementary Fig. 1). All other steps taken to construct resting-state networks were identical to those described in the gradCPT dataset's *Network construction s*ection above.

Of the 757 edges in the high-attention network, 115 (15.46%) involved nodes that were missing in the ADHD atlas. Of the 630 edges in the low-attention network, 128 (20.32%) involved nodes that were missing in the ADHD atlas. When nodes were sorted by the number of connections they had in the high- and low-attention networks (Supplementary Table 2), none of the top ten were missing in the ADHD atlas. Missing edges (importantly, these were same in all the ADHD-200 subjects analyzed here) were excluded from network strength calculations.

Network overlap

To determine the number of edges that would overlap across datasets by chance, we compared random gradCPT and ADHD networks. First, we shuffled d′ values and defined a positive and negative network exactly as described in the manuscript. That is, for every set of n–1 gradCPT subjects, we selected edges whose strength during task performance correlated with shuffled d′ values at $p < 0.01$. The positive network was defined as edges whose strength was positively related to d′ in each of these 25 iterations, and the negative network was defined as edges whose strength was inversely related to d′ in each of these iterations. By definition, positive and negative networks were mutually exclusive. We repeated this procedure 100 times, resulting in 100 positive networks (mean number of edges = 52.22, standard deviation = 22.60) and 100 random negative networks ($M = 56.59$, $SD = 25.54$).

We also shuffled ADHD-RS scores 100 times and defined 100 random positive and negative networks using ADHD data. As in the manuscript, positive and negative ADHD networks were defined on all 113 subjects rather than on the overlap of leave-one-out networks. Positive ADHD networks contained, on average, 133.54 edges ($SD = 42.18$); negative ADHD networks contained 136.11 edges on average ($SD = 43.93$).

To get overlap statistics, we calculated the number of overlapping edges between every random gradCPT network and every random ADHD network (10,000 comparisons in each of the four possible pairs in the 2×2 design with tail and dataset as factors). Overlap did not exceed 10 edges in any case ($M = 0.21$ edges, $SD = 0.53$ edges). Thus, the p-value associated with obtaining 31 and 36 common edges is 1/10,001.

BOLD variance

To address whether a measure other than functional connectivity predicted attentional abilities in the gradCPT and ADHD datasets, we trained and tested models on BOLD variance, a measure of the variability of the BOLD signal in each node that can be calculated from resting-state data. To this end, we first computed the mean BOLD signal in each frame. This yields an N × 268 matrix of node-wise mean BOLD intensities for each subject for each condition, where N is the number of frames. (This is identical to the first step in calculating connectivity matrices.) For each node, using its N × 1 timecourse vector, we then computed its variance as:

$$variance = sum(mean\ timecourse.\hat{}\ 2)/N$$

This results in a single 1×268 vector of node-wise BOLD variances for task and rest data for each gradCPT subject.

We submitted these vectors to the predictive pipeline described in the manuscript. That is, instead of defining models on 268 × 268 matrices (35,778 features), we defined them on 1 × 268 vectors (268 features). A threshold of $p < 0.20$ (rather than $p < 0.01$, which was used for the functional connectivity models described in the main text) was used for the feature selection step to ensure that at least one node appeared in each predictive model. As with the functional connectivity models, BOLD variance models trained and tested on gradCPT data were trained on n–1 subjects and tested on data from the left-out individual. Models tested on ADHD data were trained on nodes that appeared in all rounds of leave-one-out cross-validation with gradCPT task data.

Models defined on BOLD variance during gradCPT task performance did not predict d′ scores from gradCPT task or rest data (Supplementary Table 3). Although the positive BOLD variance model did predict ADHD-RS scores, predictions were in the unexpected direction (i.e., the model predicted better attentional abilities for subjects with high ADHD-RS scores). This result demonstrates that functional connectivity is a better predictor of attention than BOLD variance, which is likely affected by both metabolic function and anatomic factors.

Prediction range

One thing to note about SAN model predictions is that the range of predicted values is smaller than the range of observed values. That is, models overestimate the abilities of the individuals with the worst attention and underestimate the abilities of the individuals with the best. To investigate whether this effect emerged as a function of the non-Gaussian distribution of observed d′ scores, we used use Spearman's (rank) correlation rather than

robust regression at the edge selection step, and evaluated predictive power using Spearman's correlation between predicted and observed scores. This approach yielded highly significant predictions from gradCPT task and rest data (Supplementary Fig. 2). However, models still overestimated the d' rank of the worst subjects and underestimated the d' rank of the best subjects. Thus, SAN model predictions are best considered relative rather than absolute.

Permutation testing

P values for the leave-one-subject-out analyses were calculated by converting r values using a standard parametric conversion with the assumption that the degrees of freedom was equal to two less than the number of subjects. However, analyses in the leave-one-out folds are not independent, so the number of degrees of freedom is overestimated.

To confirm that our leave-one-subject-out results are still highly significant, we randomly shuffled d' values 1,000 times and ran them through our prediction pipeline, generating null distributions for the analyses presented in Supp. Fig. 2. We used Spearman's correlation at the edge selection step and evaluated predictive power using Spearman's correlation between predicted and observed (randomly shuffled) scores. Based on these null distributions, the P values for leave-one-out predictions from gradCPT task data (the top row of Supp. Fig. 2) are $P < 0.001$. The P values for prediction from gradCPT rest data (the bottom row of Supp. Fig. 2) are $P < 0.006$ (positive network), $P < 0.008$ (negative network), and $P < 0.002$ (GLM). Thus, our results remain highly significant based on non-parametric statistical testing.

We note here that the ADHD analyses presented in Fig. 3 of the main text were not generated with leave-one-out analyses, so P values can be calculated using standard parametric conversion.

Code availability

The 268-node functional parcellation is available online on the BioImage Suite NITRC page (https://www.nitrc.org/frs/?group_id=51). Matlab scripts were written to identify behaviorally relevant edges, model the relationship between edge strength and behavior, and make predictions from novel individual's connectivity matrices. This code is available from the authors upon request.

A supplementary methods checklist is available.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Cattell RB. Intelligence: its structure, growth and action. Advances in psychology. 1987; 35

2. Jaeggi SM, Buschkuehl M, Jonides J, Perrig WJ. Improving fluid intelligence with training on working memory. Proc Natl Acad Sci. 2008; doi: 10.1073/pnas.0801268105

3. Unsworth N, Fukuda K, Awh E, Vogel EK. Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. Cogn Psychol. 2014; 71:1–26. [PubMed: 24531497]

4. Kyllonen PC, Christal RE. Reasoning ability is (little more than) working-memory capacity?! Intelligence. 1990; 14:389–433.

5. Engle RW, Kane MJ, Tuholski SW. Models of working memory: Mechanisms of active maintenance and executive control. 1999; :102–134. DOI: 10.1037/a0021324

6. Luck SJ, Vogel EK. Visual working memory capacity: From psychophysics and neurobiology to individual differences. Trends in Cognitive Sciences. 2013; 17:391–400. [PubMed: 23850263]

7. Chun MM, Golomb JD, Turk-Browne NB. A Taxonomy of External and Internal Attention. Annu Rev Psychol. 2010; 62:73–101. [PubMed: 19575619]

8. Rosenberg MD, Finn ES, Todd Constable R, Chun MM. Predicting moment-to-moment attentional state. Neuroimage. 2015; doi: 10.1016/j.neuroimage.2015.03.032

9. Warm JS, Parasuraman R, Matthews G. Vigilance requires hard mental work and is stressful. Hum Factors. 2008; 50:433–441. [PubMed: 18689050]

10. Desimone R, Duncan J. Neural mechanisms of selective visual attention. Annu Rev Neurosci. 1995; 18:193–222. [PubMed: 7605061]

11. Kastner S, Ungerleider LG. The neural basis of biased competition in human visual cortex. Neuropsychologia. 2001; 39:1263–1276. [PubMed: 11566310]

12. Corbetta M, Shulman GL. Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci. 2002; 3:201–215. [PubMed: 11994752]

13. Posner MI, Rothbart MK. Research on Attention Networks as a Model for the Integration of Psychological Science. Annu Rev Psychol. 2006; 58:1–23. [PubMed: 17029565]

14. deBettencourt MT, Cohen JD, Lee RF, Norman KA, Turk-Browne NB. Closed-loop training of attention with real-time brain imaging. Nat Neurosci. 2015; 18:470–475. [PubMed: 25664913]

15. Rosvold HE, Mirsky AF, Sarason I, Bransome ED, Beck LH. A continuous performance test of brain damage. J Consult Psychol. 1956; 20:343–350. [PubMed: 13367264]

16. Riccio C, Reynolds C, Lowe P. Clinical applications of continuous performance tests: Measuring attention and impulsive responding in children and adults. Arch Clin Neuropsychol. 2001; 20:559–560.

17. Esterman M, Noonan SK, Rosenberg M, Degutis J. In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. Cereb Cortex. 2013; 23:2712–2723. [PubMed: 22941724]

18. Rosenberg M, Noonan S, DeGutis J, Esterman M. Sustaining visual attention in the face of distraction: a novel gradual-onset continuous performance task. Atten Percept Psychophys. 2013; 75:426–439. [PubMed: 23299180]

19. Fortenbaugh FC, et al. Sustained Attention Across the Life Span in a Sample of 10,000 Dissociating Ability and Strategy. Psychol Sci. 2015 0956797615594896.

20. Barkley RA. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. Psychol Bull. 1997; 121:65–94. [PubMed: 9000892]

21. Shen X, Papademetris X, Constable RT. Graph-theory based parcellation of functional subunits in the brain from resting-state fMRI data. Neuroimage. 2010; 50:1027–1035. [PubMed: 20060479]

22. Shen X, Tokoglu F, Papademetris X, Constable RT. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. Neuroimage. 2013; 82:403–415. [PubMed: 23747961]

23. Rubinov M, Sporns O. Complex network measures of brain connectivity: Uses and interpretations. Neuroimage. 2010; 52:1059–1069. [PubMed: 19819337]

24. Steiger JH. Tests for comparing elements of a correlation matrix. Psychological Bulletin. 1980; 87:245–251.

25. Consortium T A.-200. The ADHD-200 Consortium. A Model to Advance the Translational Potential of Neuroimaging in Clinical Neuroscience. Front Syst Neurosci. 2012; 6:62. [PubMed: 22973200]

26. DuPaul, GJ.; Power, TJ.; Anastopoulos, AD.; Reid, R. ADHD Rating Scale-IV: Checklists, norms, and clinical interpretation. Guilford Press; New York: 1998. p. 25

27. Dan L, Yu J, Vandenberg SG, Yuemei Z, CAIHONG T. Report on Shanghai norms for the Chinese translation of the Wechsler Intelligence Scale for Children-Revised. Psychol Rep. 1990; 67:531–541. [PubMed: 2263706]

28. Finn ES, et al. Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. Nat Neurosci. 2015; 18:1664–1671. [PubMed: 26457551]

29. Stoodley CJ. The cerebellum and cognition: evidence from functional imaging studies. Cerebellum. 2012; 11:352–65. [PubMed: 21373864]

30. Buckner RL. The cerebellum and cognitive function: 25 years of insight from anatomy and neuroimaging. Neuron. 2013; 80:807–815. [PubMed: 24183029]

31. Castellanos FX, Proal E. Large-scale brain systems in ADHD: Beyond the prefrontal-striatal model. Trends in Cognitive Sciences. 2012; 16:17–26. [PubMed: 22169776]

32. Krain AL, Castellanos FX. Brain development and ADHD. Clin Psychol Rev. 2006; 26:433–444. [PubMed: 16480802]

33. Huang L, Mo L, Li Y. Measuring the interrelations among multiple paradigms of visual attention: An individual differences approach. Journal of Experimental Psychology: Human Perception and Performance. 2012; 38:414–428. [PubMed: 22250865]

34. Baldassarre A, et al. From the Cover: Individual variability in functional connectivity predicts performance of a perceptual task. Proceedings of the National Academy of Sciences. 2012; 109:3516–3521.

35. Smith SM, et al. Functional connectomics from resting-state fMRI. Trends in Cognitive Sciences. 2013; 17:666–682. [PubMed: 24238796]

36. Gabrieli JDE, Ghosh SS, Whitfield-Gabrieli S. Prediction as a Humanitarian and Pragmatic Contribution from Human Cognitive Neuroscience. Neuron. 2015; 85:11–26. [PubMed: 25569345]

37. Whelan R, et al. Neuropsychosocial profiles of current and future adolescent alcohol misusers. Nature. 2014; 512:185–189. [PubMed: 25043041]

38. Rosenberg MD, Finn ES, Constable RT, Chun MM. Predicting moment-to-moment attentional state.

39. Langner R, Eickhoff SB. Sustaining attention to simple tasks: a meta-analytic review of the neural mechanisms of vigilant attention. Psychol Bull. 2013; 139:870–900. [PubMed: 23163491]

40. Turk-Browne NB. Functional Interactions as Big Data in the Human Brain. Science (80-). 2013; 342:580–584.

41. Cao Q, et al. Abnormal neural activity in children with attention deficit hyperactivity disorder: a resting-state functional magnetic resonance imaging study. Neuroreport. 2006; 17:1033–1036. [PubMed: 16791098]

42. Tian L, et al. Altered resting-state functional connectivity patterns of anterior cingulate cortex in adolescents with attention deficit hyperactivity disorder. Neurosci Lett. 2006; 400:39–43. [PubMed: 16510242]

43. Uddin LQ, et al. Network homogeneity reveals decreased integrity of default-mode network in ADHD. J Neurosci Methods. 2008; 169:249–54. [PubMed: 18190970]

44. Wang L, et al. Altered small-world brain functional networks in children with attention-deficit/hyperactivity disorder. Hum Brain Mapp. 2009; 30:638–649. [PubMed: 18219621]

45. Fair DA, et al. Atypical default network connectivity in youth with attention-deficit/hyperactivity disorder. Biol Psychiatry. 2010; 68:1084–91. [PubMed: 20728873]

46. Qiu M, et al. Changes of Brain Structure and Function in ADHD Children. Brain Topogr. 2011; 24:243–252. [PubMed: 21191807]

47. Tomasi D, Volkow ND. Abnormal functional connectivity in children with attention-deficit/hyperactivity disorder. Biol Psychiatry. 2012; 71:443–50. [PubMed: 22153589]

48. Cocchi L, et al. Altered Functional Brain Connectivity in a Non-Clinical Sample of Young Adults with Attention-Deficit/Hyperactivity Disorder. J Neurosci. 2012; 32:17753–17761. [PubMed: 23223295]

49. Joshi A, et al. Unified Framework for Development, Deployment and Robust Testing of Neuroimaging Algorithms. Neuroinformatics. 2011; 9:69–84. [PubMed: 21249532]

50. Kaufman J, et al. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. J Am Acad Child Adolesc Psychiatry. 1997; 36:980–988. [PubMed: 9204677]

51. Friedman L, Glover GH. The FBIRN Consortium. Reducing interscanner variability of activation in a multicenter fMRI study: Controlling for signal-to-fluctuation-noise-ratio (SFNR) differences. Neuroimage. 2006; 33:471–481. [PubMed: 16952468]

52. Scheinost D, Papademetris X, Constable RT. The impact of image smoothness on intrinsic functional connectivity and head motion confounds. Neuroimage. 2014; 95:13–21. [PubMed: 24657356]
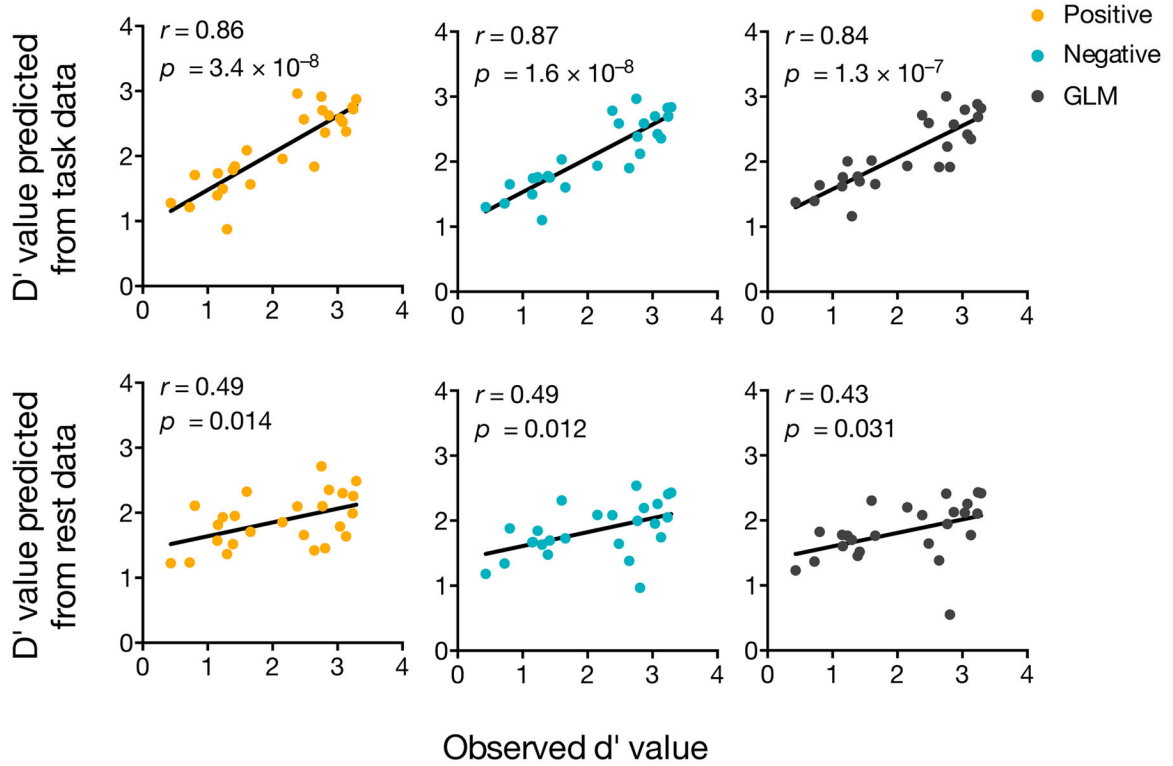
Fig. 1.
Functional connectivity models predict sustained attention performance. Scatter plots show correlations between observed gradCPT d′ values and predictions by positive and negative networks and general linear models (GLM) that take into account positive and negative network strength. Network models were iteratively trained on task data from $n-1$ subjects in the gradCPT data set and tested on task data (top row) and resting-state data (bottom row) from the left-out individual.
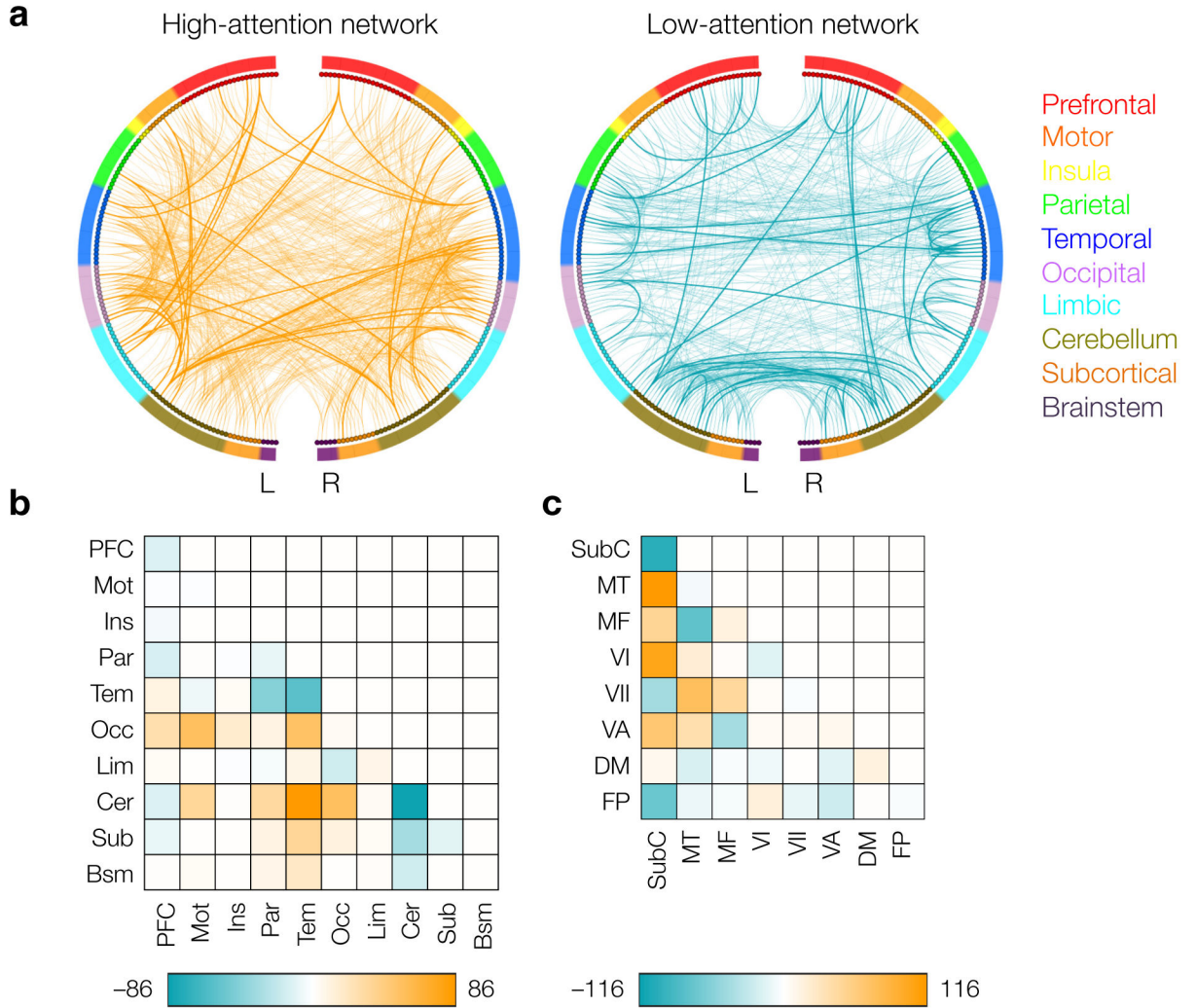
Fig. 2.

Functional connections predicting gradCPT performance and ADHD-RS scores. (A) The 757 edges in the high-attention network (predicting higher d′ values in the gradCPT sample and lower ADHD-RS scores in the ADHD-200 sample) are visualized in orange. The 630 edges in the low-attention network (predicting lower d′ values in the gradCPT sample and higher ADHD-RS scores in the ADHD-200 sample) are visualized in blue. Edges that appear in both the gradCPT and ADHD networks appear in bold. Macroscale regions include prefrontal cortex (PFC), motor cortex (Mot), insula (Ins), parietal (Par), temporal (Tem), occipital (Occ), limbic (including the cingulate cortex, amygdala and hippocampus; Lim), cerebellum (Cer), subcortical (thalamus and striatum; Sub), brainstem (Bsm). (B) Differences in the number of edges between each pair of macroscale regions, calculated by subtracting the number of edges in the low-attention network from the number in the high-attention network. (C) Differences in the number of edges between each pair of canonical networks, calculated by subtracting the number of edges in the low-attention network from the number in the high-attention network. Canonical networks[28] include the subcortical-

cerebellum (SubC), motor (MT), medial frontal (MF), visual I (VI), visual II (VII), visual association (VA), default mode (DM), and frontoparietal (FP).
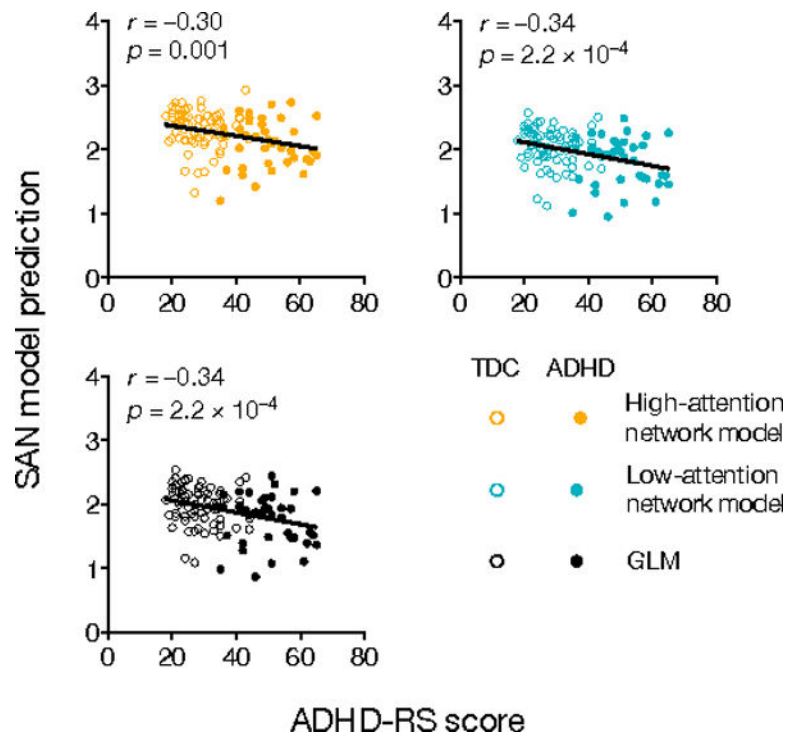
Fig. 3.
Sustained Attention Network (SAN) models, defined with gradCPT subjects, significantly predict scores on the ADHD-Rating Scale (ADHD-RS) in an independent sample of children and adolescents from the ADHD-200 dataset. Predictions are negatively correlated with ADHD-RS scores because models were trained to predict d′; thus, higher predictions correspond to better attentional abilities and lower ADHD-RS scores. These individuals were diagnosed with ADHD (solid dots) or as typically developing controls (TDC, hollow dots).
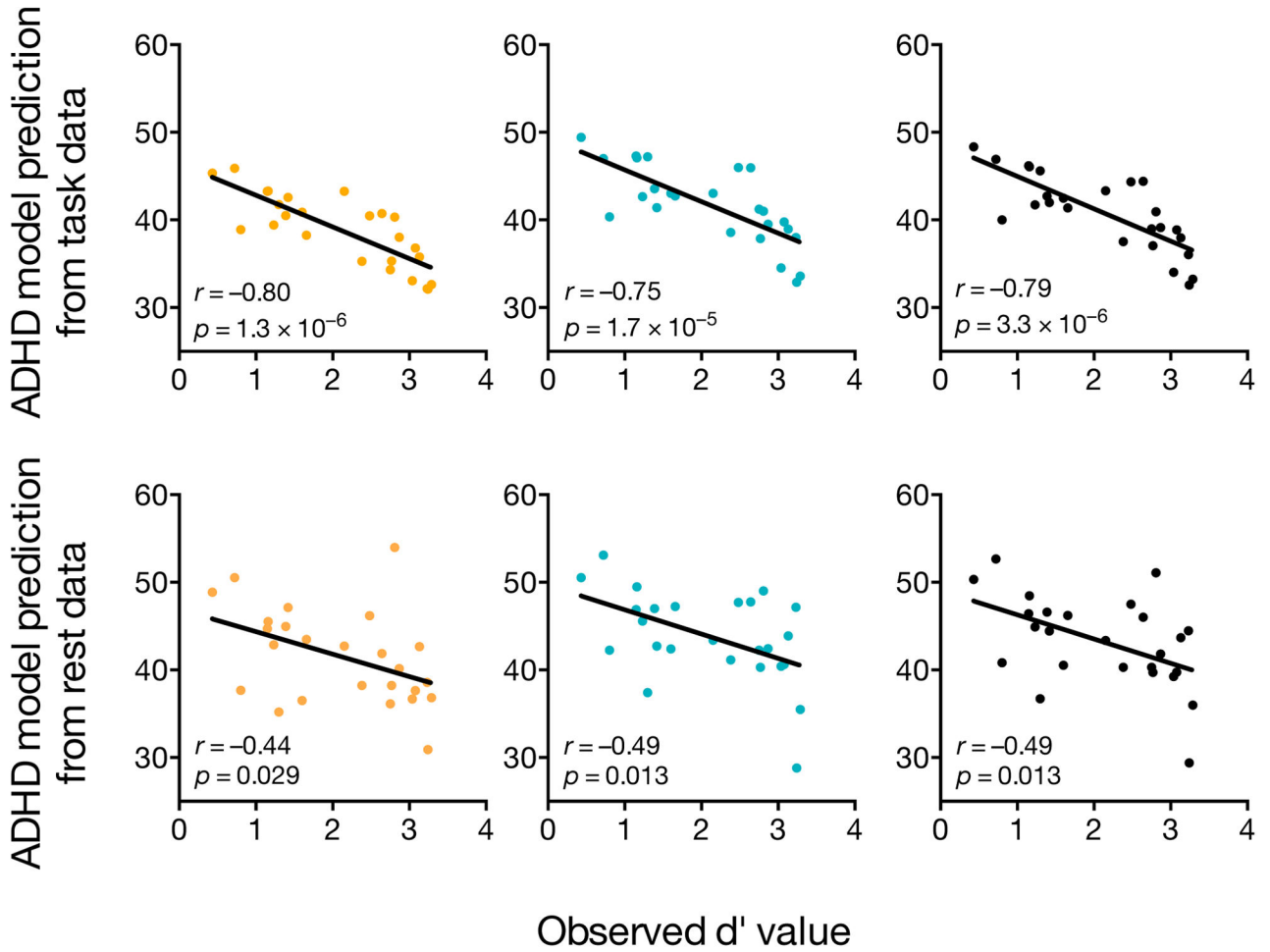
Fig. 4.
Connectivity models defined on ADHD-200 data predict gradCPT performance in an
independent group of participants. Scatter plots show predictions of models defined using
edges negatively (orange) and positively (blue) related to ADHD-RS scores in ADHD-200
resting state data. Predictions of a GLM, which incorporates low- and high-ADHD network
strength, are shown in black. These models were applied to gradCPT task (top) and resting-
state data (bottom).

Table 1

Partial correlations between SAN model predictions and ADHD-RS scores, IQ, and age. One subject did not have an IQ score, so these correlations were performed on 112 individuals. Correlations between ADHD-RS scores, IQ, and age are provided in the *ADHD-200 dataset: Participants* section of the online Methods.

		ADHD-RS score		IQ		Age	
		r-value	*p*-value	*r*-value	*p*-value	*r*-value	*p*-value
Predictive network	**Positive**	−0.27	0.0047	0.11	0.25	−0.09	0.33
	Negative	−0.30	0.0015	0.13	0.17	0.07	0.45
	GLM	−0.30	0.0016	0.13	0.16	0.10	0.29