Brain functional connectome defines a transdiagnostic dimension associated with cognition and psychopathology in youth

Authors:

Xiang Xiao[1], Christopher Hammond[2], Betty Jo Salmeron[1], Hong Gu[1], Tianye Zhai[1], Hieu Nguyen[1], Hanbing Lu[1], Thomas Ross[1], Yihong Yang[1*]

Affiliations:

1. Neuroimaging Research Branch, National Institute on Drug Abuse, National Institutes of Health, Baltimore, Maryland, USA.

2. Department of Psychiatry, Johns Hopkins University, Baltimore, Maryland, USA.

* To whom correspondence may be addressed: Email: yihongyang@intra.nida.nih.gov.

Yihong Yang, Ph.D.

Neuroimaging Research Branch, National Institute on Drug Abuse, National Institutes of Health, Baltimore, Maryland, USA;

ABSTRACT

The search for psychiatric biomarkers has remained elusive, in part, due to high comorbidity, low specificity, and poor concordance between neurobiological abnormalities and existing diagnostic categories. Developmental shifts in symptom expression and brain function across the lifespan further complicate biomarker identification. Recent studies suggest that focusing on cognition may be a pathway forward: Cognitive dysfunction is a common feature across psychiatric disorders. Individual differences in cognition may reflect variability in the connectivity of underlying neurocognitive brain networks, and predict psychopathology at different developmental periods. In the present study we identified brain-based dimensions of general cognitive capacity and psychopathology using sparse canonical correlation analysis (sCCA) in a sample of 7,383 preadolescents from the Adolescent Brian Cognitive Development (ABCD) study. This analysis revealed correlated patterns of functional connectivity with cognitive control capacity and psychopathology. Specifically, the results identified a single connectome-based latent brain variate that was positively correlated with performance on cognitive measures across domains and negatively correlated with parent-reported psychopathology across diagnoses and domains. Functional connectivity loadings for the brain variate were across distributed cortical and subcortical brain networks and a dose-dependent relationship with psychopathology based upon the cumulative number of psychiatric diagnoses was observed. These findings provide preliminary evidence for a connectome-based biomarker that indexes individual differences in cognitive control and predicts transdiagnostic psychopathology in a dose-dependent fashion.

.

INTRODUCTION

Over the past fifty years since the inception of the diagnostic and statistical manual (DSM) and the standardization of psychiatric nomenclature in the 1970s, psychiatry has focused on establishing diagnostic categories based upon clinical symptoms. The absence of established biomarkers to aid in diagnosis and treatment selection is a medical diagnostic quandary unique to psychiatry and has limited progress in the field (Singh & Rose, 2009). Further complicating psychiatric nosology are issues of comorbidity and specificity (Pincus, Tew, & First, 2004). There is a high degree of comorbidity across psychiatric disorders and many symptoms are present across diagnostic categories (van Loo & Romeijn, 2015). Furthermore, emerging biological research points to shared genetic risk and overlapping structural and functional abnormalities across psychiatric disorders, that could, in part, explain some of the comorbidity across conditions (Goodkind et al., 2015; Lee et al., 2013; McTeague, Goodkind, & Etkin, 2016; Sharma et al., 2017). This disconnection between current psychiatric nosology and biological findings, highlights the need to examine neurobiological substrates and clinical symptoms that are shared across diagnoses.

On top of this complexity, developmental factors play a large role in brain and behavioral expression related to psychiatric disorders impacting biomarker identification. Three quarters of all psychiatric disorders emerge before the age of 21 years with 35% emerging before the age of 14 years (Solmi et al., 2021). Few studies have examined transdiagnostic features and neurobiological correlates of psychiatric conditions in youth samples. The presentation of psychiatric symptoms and disorders and brain structure and function change across the lifespan with marked shifts occurring during adolescence (Gogtay et al., 2004; Paus, Keshavan, & Giedd, 2008). As such, biological markers of psychopathology risk may differ as a function of age and developmental stage. Prior research suggests that neural and genetic correlates of psychiatric disorders including psychosis, mood, and addictive disorders vary across age periods (Thapar & Riglin, 2020). These findings highlight the importance of taking age- and developmental-stage into account in biomarker identification. Additionally, they suggest that focusing on

the prepubertal-to-pubertal transition age period as this represents a critical period of vulnerability during which approximately 40% of all psychiatric conditions emerge (Solmi et al., 2021).

Given the problems identifying biomarkers of specific psychiatric disorders, the field has shifted towards transdiagnostic and dimensional investigations, spearheaded by the NIMH's Research Domain Criteria (RDoC) (Mcteague, 2016). RDoC is a biologically grounded framework for studying psychiatric conditions that takes a dimensional approach and conceptualizes individual differences in symptoms as emerging from mixed-dimensional abnormalities of specific brain circuits (Cuthbert & Insel, 2013). Recent imaging techniques that characterize low frequency fluctuations in blood-oxygen-level-dependent (BOLD) fMRI signal between disparate brain regions at rest (i.e., resting-state fMRI), network analytic approaches, and multivariate prediction model techniques now enable the investigation of changes in spatial and temporal variance in connectivity between large-scale canonical brain networks as they relate to disease processes (Bassett & Bullmore, 2009; Menon, 2011; Sui, Jiang, Bustillo, & Calhoun, 2020)

One approach to dissect biological heterogeneity and improve understanding of comorbidity across psychiatric disorders is to focus on neuropsychological features or symptom clusters that are present across diagnostic categories. Recent studies suggest that focusing on cognition may be a pathway forward. Cognitive dysfunction is a common feature across psychiatric disorders (Gualtieri & Morgan, 2008; Millan et al., 2012). Studies of cognition suggest a hierarchical framework with both higher order processes (e.g., executive function [EF] and planning),and lower order processes (e.g., motor skills, perception, memory), and an underlying, largely heritable, latent factor reflecting general cognition (g-factor) has been described (Barbey, 2018). Measures of general cognition and higher order cognitive processes have been shown to predict socio-occupational stability, academic success, and quality of life (Barbey, 2018; Snyder, Miyake, & Hankin, 2015). Additionally, recent research has identified genetic, epigenetic, developmental and environmental factors that collectively disrupt brain networks integrating and modulating cognition (Friedman & Miyake, 2017; Friedman et al., 2016; Sha, Wager, Mechelli, & He, 2019). Individual differences in cognition may reflect variability in the connectivity of underlying

neurocognitive brain networks, and predict psychopathology (Reineberg, Andrews-Hanna, Depue, Friedman, & Banich, 2015; Shanmugan et al., 2016). Relationships between cognition and psychopathology are complex and may be bidirectional (Romer & Pizzagalli, 2021). Problematically, to date, much of the research in this area has focused on cognitive functioning within specific psychiatric disorder categories (e.g., schizophrenia) (Lesh, Niendam, Minzenberg, & Carter, 2011) without examining relative differences in cognition in other comorbid psychiatric disorders. Still, a growing number of studies have shown evidence of cognitive deficits in various domains to be a common presentation in multiple psychiatric disorders (Millan et al., 2012).

In parallel with general cognition, psychopathology also appears to exhibit a hierarchical framework with most disorders falling into larger externalizing (e.g., attention deficit hyperactivity disorder (ADHD), conduct disorder) and internalizing (e.g., depression and anxiety) domains (Kim & Eaton, 2015). Evidence from epidemiological studies have also shown evidence of a dimensional general psychopathology factor (p-factor), parallel to the g-factor, that cuts across disorder boundaries and is predictive of lifespan functional impairment and prospective psychopathology beyond current symptom-based prediction (Caspi et al., 2014). In adults and children, higher p-factor scores predict worse performance on higher order cognitive tasks related to working memory, planning and is associated with lower academic achievement and lower IQ (Caspi et al., 2014; Caspi & Moffitt, 2018). In fact, general cognition and psychopathology scores are commonly anticorrelated (Caspi et al., 2014; Romer & Pizzagalli, 2021). Much work remains to be done in this area. For example, whether these cognition-psychopathology relationships emerge as the product of common underlying neurocognitive deficits across psychiatric disorders and whether unique disorder and domain-specific relationships are present remain unanswered questions in the field (Kim & Eaton, 2015). Given these findings, the question of whether there are functional connectivity alterations within the brain networks subserving cognition that contribute to psychopathology is the main area of focus for our present study.

In the present study, we seek to delineate functional network abnormalities associated with higher-order and domain specific cognitive processes and broad array of psychopathology in preadolescence. Using data from the ABCD study (Bjork, Straub, Provost, & Neale, 2017), which includes fine-grained analyses of psychopathology (measured both dimensionally and categorically) and cognitive control capacity (general and domain specific), we applied sparse canonical correlation analysis (sCCA), a recently developed machine learning technique. The sCCA is a multivariate method capable of discovering complex linear relationships between two high dimensional datasets that simultaneously identifies a series of orthogonal latent dimensions in the two data sets, while ensuring a maximized correlation between the two dimensions paired by their orders (Witten et al., 2009). This approach has been successfully applied in prior works in neurodegenerative diseases and normal brain-behavior relationships. In our case we applied this technique to the 'brain data' indexed with the resting-state functional connectome (RSFC) and 'behavioral data' which includes diverse cognitive tasks and assessments of a broad range of different forms of psychopathology. Prior works discovering biotypes within categories of psychopathology were mostly based solely on imaging features such as rsFC, gray matter volume (GMV), and/or gray matter density (GMD) (Clementz et al., 2016; Drysdale et al., 2017). In contrast, with the approach of the current study, we seek to link a broad range cognitive domains and psychiatric symptoms to individual differences in functional brain networks.

RESULTS

Generalizable brain-behavior associated dimensions in the youth population

To identify the intrinsic association between the functional connectome of youth brain and multi-dimensional behavioral assessments, we conducted a sparse Canonical Correlation Analysis (sCCA) on the two datasets. SCCA (Witten, Tibshirani, & Hastie, 2009) is a data-driven approach seeking to find latent dimensions shared by two high-dimensional datasets. Particularly, sCCA simultaneously identifies a series of orthogonal latent dimensions in the two data sets, while ensuring a maximized correlation between the two dimensions paired by their orders.

The analysis was performed on 7,383 pre-adolescent youth (3,714 females, aged 119.5 ± 7.4 months at scanning) from the baseline cohort of Adolescent Brian Cognitive Development (ABCD) study (Volkow et al., 2018). The criteria for participant inclusion/exclusion are detailed in the methods section. The ABCD study is the largest long-term study of brain development and child health in the United States (Volkow et al., 2018). In the baseline cohort, 11,875 children aged 9–10 years were recruited through school- and community-settings from 21 research sites across the country, covering a wide range of sociodemographic variations (Garavan et al., 2018). This population-based design has distinct advantages for studying brain-behavior associations, allowing for assessment of the full continuum of brain and behavioral outcomes and their relative variability as expressed in the population, and optimizing generalizability of the findings (Falk et al., 2013; Paus, 2010).

One input for the sCCA was the dataset of resting-state functional connectome (RSFC). RSFC was derived from resting-state functional MRI data for each participant. The image data were pre-processed using a validated pipeline (Fair et al., 2020) to ensure the inter-subject spatial alignment and to minizine the artificial signals involved during the acquisition of MR images. Individual RSFC was constructed using the 352-area parcellation scheme of (Gordon et al., 2016), optimized for homogeneity of within-region BOLD activity. The 352 areas cover both cortical and sub-cortical regions, and are a priori

assigned to 12 large-scale communities. Parcel-wise Pearson's correlations were calculated and transformed to Z scores for individual brains, with the 61,776 upper-triangular values used to describe that individual's functional connectome. Due to its high association with individual differences, the functional connectome has also been conceptualized as the human brain 'functional fingerprint' (Finn et al., 2015).

Before submitting the brain features into the sCCA, two preprocessing steps were conducted at the group level. First, head motion and scanner type were regressed out from the brain data set as covariates. Then we applied a dimension-reduction procedure based on Principle Component Analysis (PCA) on these brain features. Following previous practices (Smith et al., 2015; Xia et al., 2018), 51 PCs (consistent with the dimension of the behavior dataset) explaining the greatest variance were kept for further analysis. Note in the cross validation, PCA was only performed on the training dataset, and the resultant coefficients were applied to the testing dataset.

The other input to the sCCA was the behavioral measures. The behavioral assessment dataset included 51 dimensional scales broadly measuring individual differences across two major domains: cognition and psychological functioning (i.e., psychopathology). Cognitive measures included 20 scales derived from participants' performance on 15 neuro-cognitive tests (Luciana et al., 2018). Psychopathology measures included 31 scales derived from participants' baseline mental health battery (Barch et al., 2018). Outcome variables from these assessments indexed general cognition and psychopathology as well as specific neurocognitive and psychological symptom domains. This allowed us to conduct fine-grained analyses to characterize whether general versus domain-specific cognitive and psychological symptom domains were associated with distinct or overlapping RSFC brain features. Specific measures included in the analysis are detailed in the methods section. Like the brain features, PCA was performed on the behavioral feature dataset, with all 51 PCs kept for further analysis. In this way, 100% of the variance in the dataset was kept, and therefore the PCA was considered as a procedure of de-collinearity for the behavioral dataset. In the hold-out validation, PCA was only performed on the

training dataset, with the resulting coefficients applied to the test set. The procedure of generating the sCCA model is schematically illustrated in Figure 1.

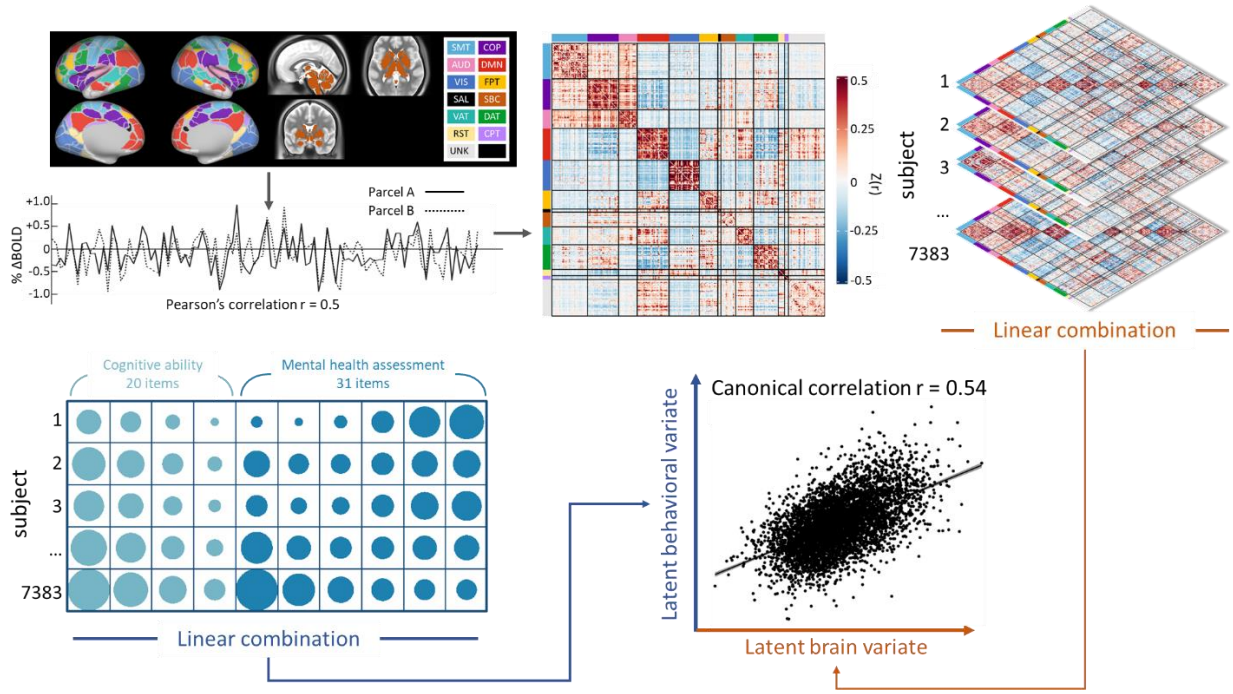

Figure 1 Scheme of sparse canonical correlation analysis (sCCA). (A) Construction of functional connectome. Preprocessed time-series were extracted from 352 parcels defined by the Gordon atlas (Gordon et al., 2016). Then Pearson's correlation of the time series was calculated for each pair of the brain parcels. The 352 by 352 correlation matrix forms the functional connectome of individual brains. In Gordon's atlas, 352 parcels are assigned into functional modules: sensory motor network (SMT), auditory network (AUD), cingulo-opercular (COP), default mode network (DMN), visual network (VIS), frontal-parietal network (FPT), salience network (SAL), subcortical network (SBC), ventral attention network (VAT), dorsal attention network (DAT), retrosplenial temporal network (RST), cingulo-parietal network (CPT), unknown network (UNK). **(B) Connectome dataset.** Functional connectome data 7383 young participants (including 5906 as the training set and 1477 as the test set) were included as the brain feature set. **(C) Behavioral-assessment dataset.** The behavioral assessments consisted of 51 behavioral assessments, including 20 scales for cognitive ability and 31 scales for mental-health related constructs. **(D) Youth population in the aligned latent space identified by sCCA.** SCCA simultaneously identifies pairs of linear subspaces in the two data sets. By projecting the data set to the linear subspaces, one can get a pair of brain and behavior latent variates, and correlation between these pair of variates are maximized. These two aligned latent variates themselves form a new space, which provides a joint perspective to view the individual differences across the population.

To validate the generalizability of the identified brain-behavior association, the whole sCCA analysis was conducted in a hold-out framework (Mihalik et al., 2020). The 7,383 participants were randomly split into a training set of 5906 subjects (80% of the total samples) and a test set of 1477 subjects (20% of the total samples). Within the training set, we tuned the sCCA regularization parameters by performing a grid-searching procedure based upon a nested 10-fold cross validation. SCCA was

performed on the training set, and the resultant canonical correlations along the 51 canonical modes are plotted in Figure 2A. The first 5 modes out of the 51 in total were statistically significant ($p < 0.01$, Bonferroni corrected) according to a permutation test. Then we applied the trained sCCA model to the test set; the resultant canonical correlations along the 51 modes are plotted in Figure 2B. From the permutation test on the test set, only the first CC mode was statistically significant ($p < 0.01$, Bonferroni corrected). Therefore, the first CC mode ($r = 0.51$ in training set, and $r = 0.49$ in the test set, shown in Figure 2B and Figure 2D respectively) was identified as the generalizable brain-behavior association in the youth population.

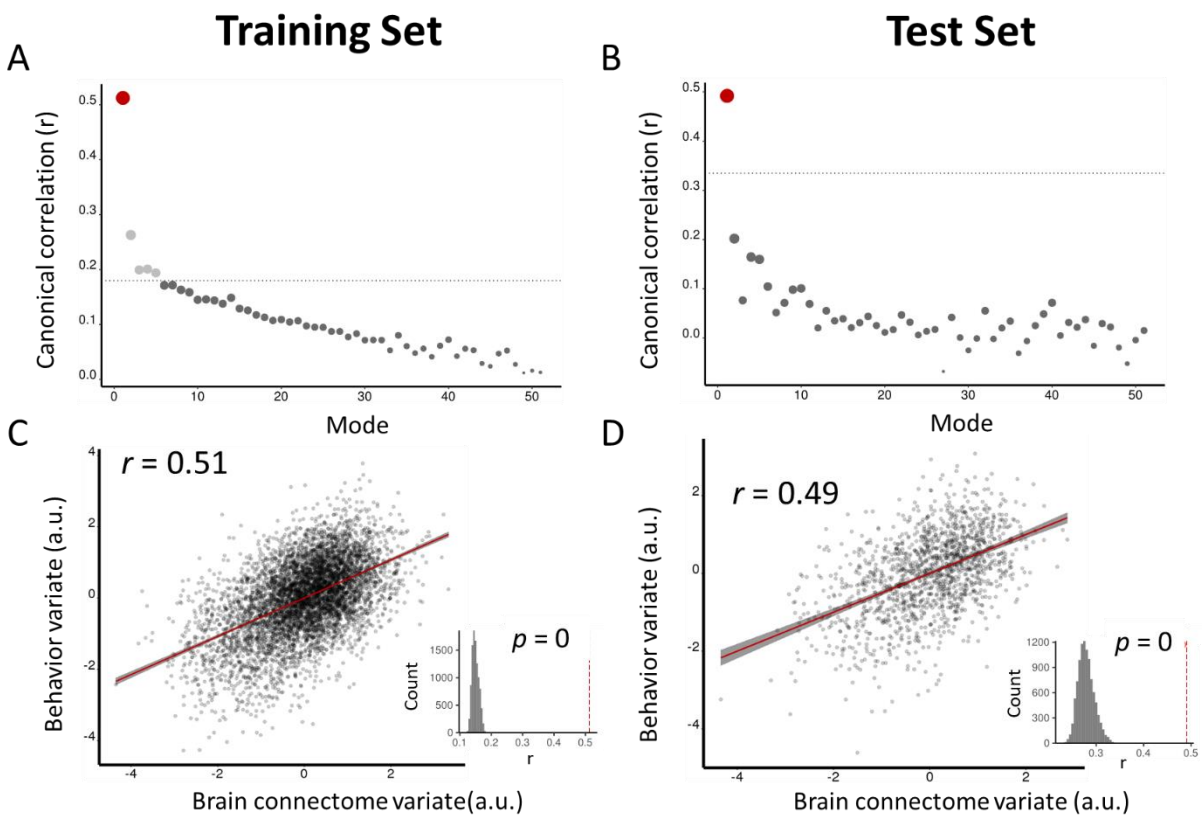

Figure 2 Canonical correlations (CCs) between the brain and behavioral datasets. (A) Canonical correlations of canonical modes in the training set. Dots in the plot **i**ndicate CCs of the 51 canonical modes from the training set. Canonical modes are sorted in descending order of related eigen values. Dash line in the plot represents the 99% CI of the null distribution of CC generated from permutated datasets. In the training set, top five (the dot in red and four dots of light gray) canonical modes show significant correlation in comparison to the null distribution**. (B) Canonical correlations of canonical modes in the test set.** Correlation modes in the test set were obtained from the hold-out validation, i.e., by applying the sCCA model trained from training set on the data in the test set. Dots in the plot **i**ndicate CCs of the 51 canonical modes from the training set. Dash line in the plot represents the 99% CI of the null distribution of CC generated from permutated datasets. In the training set, only the

first canonical mode shows a significant CC in comparison to the null distribution**. (C) and (D) Scatter plots of the associated brain and behavior variate scores, for training and test set, respectively.** Dots in the plots represent individuals in the space formed by brain and behavior variates. Correlation between these two variates shows significant and consistent correlations in both the training set (Pearson's correlation $r = 0.51$, $p < 0.01$, $N = 5906$) and the test set (Pearson's correlation $r = 0.49$, $p < 0.01$, $N = 1477$). Inset plots represent the result of permutation test conducted in training and testing data set; the actual correlations for mode 1 are significantly higher (indicated with red dash in the plots) than the null in each data set.

Individual differences in the latent brain-behavior space

After we identified the pair of associated latent dimensions from the brain and behavior datasets, the two aligned dimensions formed a new subspace (Mihalik et al., 2020) that provides a joint perspective to view the individual differences across the population.

We first examined the loadings of each behavioral assessment. As shown in Figure3A, nearly all assessments related to cognitive ability show positive loadings on the latent behavioral dimension, whereas most of the psychopathology related constructs show negative loadings. Furthermore, within the psychopathology related constructs, assessments belonging to the spectrum of externalizing problems show higher loadings (-0.36 for Rule Breaking Problem, and -0.29 for Aggressive Behavior), while those in the internalizing spectrum (-0.15 for Anxious Depression, -0.14 for Withdrawn Depressive Problem and -0.14 for CBCL Somatic Problem) show lower loadings.

Along the behavioral loading, population-level distribution of behavior assessments manifested divergent patterns in the latent space formed by the two variates of brain and behavior. Figure.3B shows three representative distributions. First, for the composite score of total intelligence, which locates in the positive end of the behavioral loading axis, its distribution shows a negative-to-positive gradient in the diagonal direction in the joint space. Second, the score of externalizing problem, which carries a high negative behavioral loading, also showed a gradient in the diagonal direction but in an opposite trend to the total intelligence. Third, the internalizing problem, which carries a behavioral loading near zero, shows an even distribution across the embedded population in comparison with the other two assessments.

Accordingly, to view from brain-centered perspective, we projected each the embedded distributions to the latent dimension of brain connectome variate. As plotted in Figure 3C, assessments of cognitive ability show ascending trends along the latent brain dimension, while assessments of mental disorder related constructs show descending trends. This indicates that the single multivariate pattern of functional connectome serves as a common source of population-level variance in behaviors spanning domains of cognition and psychopathology.

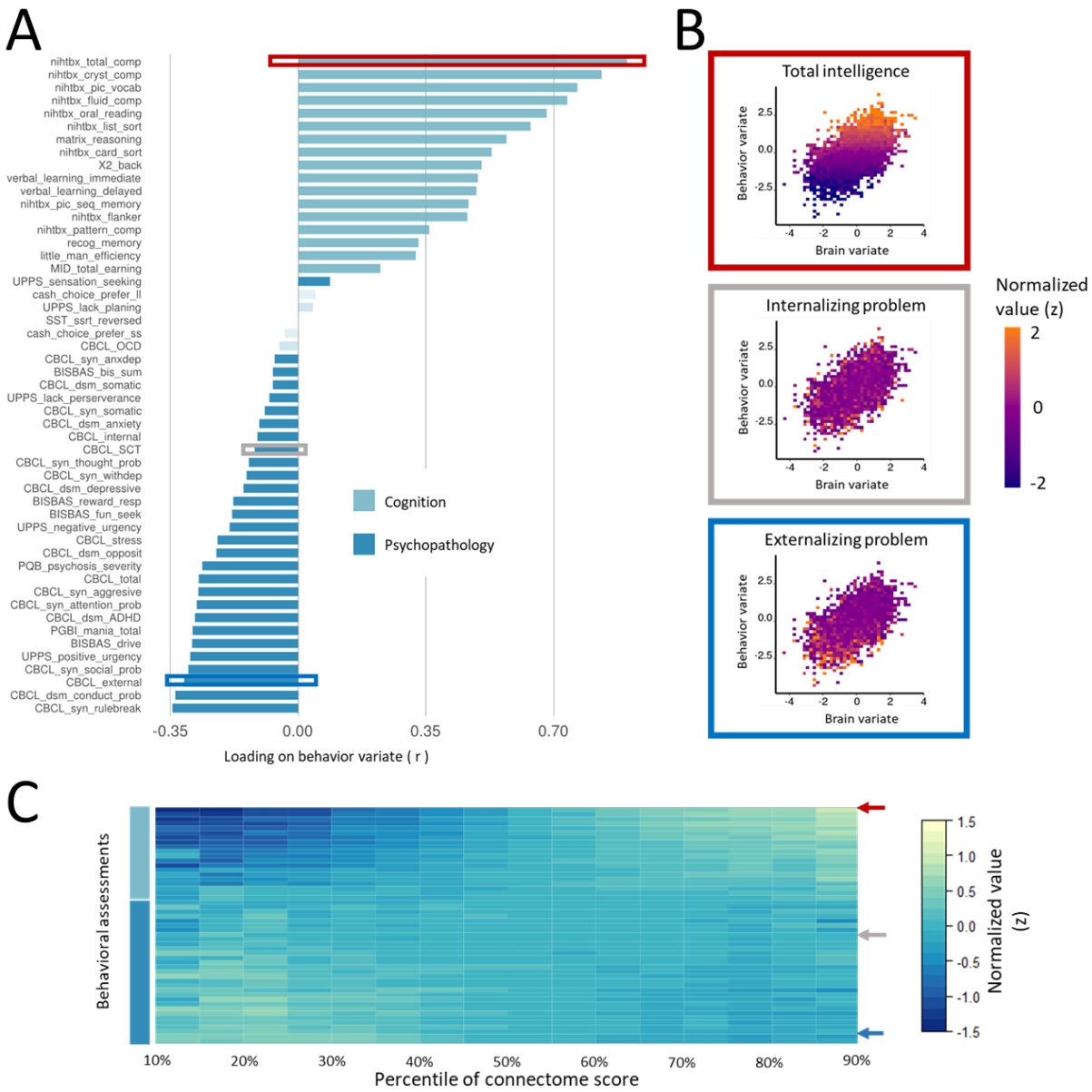

Figure 3 (A) Loadings of behavioral assessments on the latent behavioral dimension. Behavioral loadings are plotted in a descending order. Assessments in the domain of cognitive ability (light blue) show positive loadings, while the mental disorder related constructs (dark blue) tend to show negative loadings. **(B) Exemplar distribution of behavioral assessment scores represented in the aligned latent space.** The rectangular plot shows distribution of behavioral scores across the embedded population. Individuals in the training set are plotted in the aligned latent space detected by sCCA, and z-scores of each assessment are averaged in each 2D rectangular bin. Among the three behavioral scales, Total Intelligence score (red) and Externalizing Problem score (blue) show gradients in the embedded population in opposite directions, while Internalizing Problem score (gray) is distributed evenly across the population. **(C) Distribution of behavioral assessment scores along the connectome variate score.** Individuals in the training set are grouped according to the populational percentile of connectome score, then the score of each assessment are averaged for each of the percentile bins. Each row of the heatmap indicate the distribution of one behavioral assessment. And the behavioral assessments are ordered by their behavioral loadings. Colored arrows indicate rows related to the three exemplar assessments.

Prediction for a broad band of cognitive and psychopathological assessments using connectome of individual brains

The above hypothesis of common neural source was tested within the framework of connectome-based prediction (Finn et al., 2015; Rosenberg et al., 2015; Shen et al., 2017) in two folds. First, the connectome-derived variate should be able to predict the univariate assessments of cognitive ability and psychopathology in unseen population, i.e., cross-subject generalizability. Second, the connectome variate estimated from behavioral assessments in one domain should be able to predict behavioral assessments in the other domain, i.e., the cross-domain generalizability.

For cross-subject generalizability, we first trained the sCCA model with the connectome data and full behavioral assessments in the training set. Then the estimated canonical projection of the connectome variate was adopted as the prediction model. Finally, the canonical projection was applied on the individual's connectome in the test set, and the resultant connectome variate was used to predict each of their behavioral assessments. The result of the multi-behavioral prediction is shown in the Figure 4. Specifically, all assessments in the cognitive ability domain, except for the cash choice task and the stop-signal reaction time, are significantly predicted (FDR-adjusted $p < 0.05$, $N = 1477$). The predictive mental-disorder related constructs (FDR-adjusted $p < 0.05$, $N = 1477$) spans psychopathological scales mostly in the externalizing spectrum, e.g., CBCL scales of Rule-breaking Behavior, and Conduct Problems; other CBCL syndrome of Social Problem, and Oppositional Defiant Problem; impulsivity related scales of Positive Urgency, Negative Urgency, Reward Reponses and Fun Seeking; and the scale of Mania.

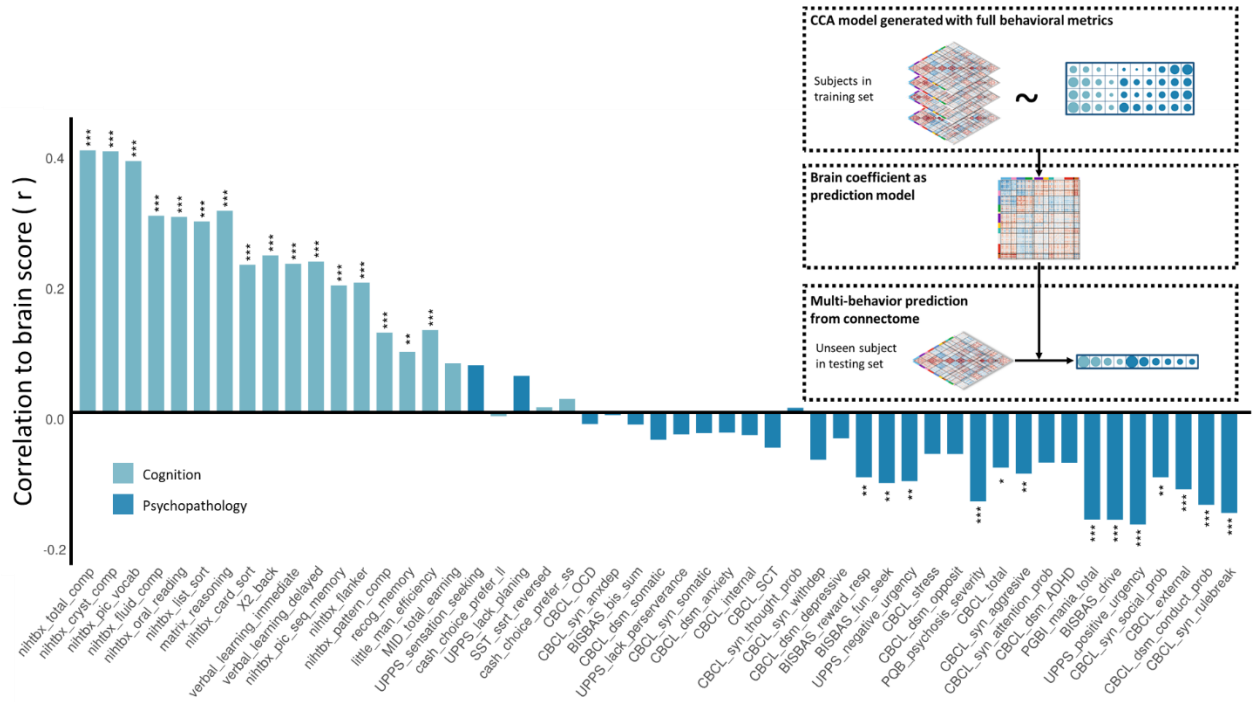

Figure 4 Behavioral assessment prediction from the connectome variate score. The bar plot shows the performance of using sCCA estimated connectome model to predict a behavior assessment scores in an unseen population. Behavioral assessments are ordered by the rank of their behavioral loadings. Inset plot shows the procedure of the cross-population prediction. Behavioral assessment predictive ability are assessed with Pearson's correlation, and the resultant *p* values are corrected for multiple comparison using the FDR procedure.

For the trans-domain generalizability, we trained the sCCA model with the connectome data and only partial set of the behavioral assessments. Specifically, we trained two sCCA models, one with only the cognitive ability assessments, and the other with only the mental disorder assessments. These two models were then applied to predict behaviors in the hold-out test set, analogous to the previous analysis. As shown in Figures 5A-B, the correlation of each behavioral assessment is generally unchanged in comparison to the original model, when using the connectome model trained with either behavioral subset. In particular, the item-wise correlations of the prediction scores are $r = 0.99$ ($p < 0.01$, $N = 51$) between the cognitive ability model vs. the full model, and $r = 0.99$ ($p < 0.01$, $N = 51$) between the mental disorder related construct model vs. the full model.

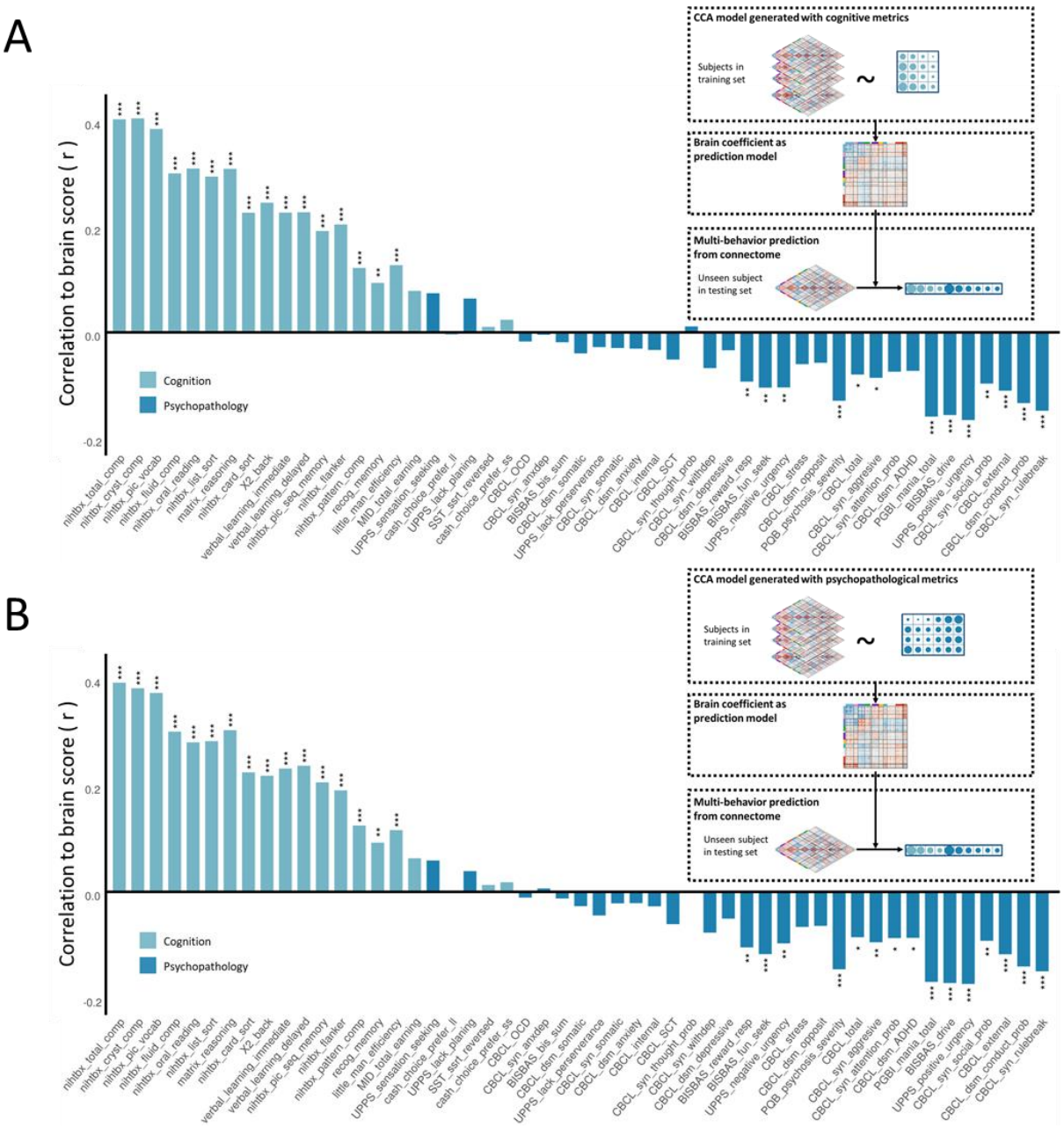

Figure 5 Predicting behavioral assessments with connectome variate scores trained from partial datasets. (A) Predictive ability of the sCCA model trained only with cognitive ability measures (B) Predictive ability of the sCCA model trained only with mental disorder related constructs. The bar plots show the performance of using sCCA estimated connectome model to predict a broad band of behavior assessment scores in an unseen population. Behavioral assessments are ordered by the rank of their behavioral loadings in the full model. Insets show the procedures of the cross-population prediction. Behavioral assessment predictive ability is measured with Pearson's correlation, and the resultant *p* values are corrected for multiple comparison using the FDR procedure.

Low-scored connectomes are prevalent across clinical diagnostic categories

The above finding that the connectome variate is capable of predicting a broad range of psychopathology suggests its clinical potential as a transdiagnostic neural marker. To validate this, we examined the association between the connectome variate to clinical diagnoses of Kiddie Schedule for Affective Disorders and Schizophrenia for DSM-5 (KSADS-5). Here KSADS-5 scales are independent from behavioral metrics used in generating the sCCA model above.

We assigned participants of the whole data set into groups according to their current diagnosis from the baseline interview of KSADS-5. Children who were diagnosed without any disorder were assigned as the control group. Connectome variate scores were compared between each diseased group and the control group, using Welch's two-sample t-tests. Groups with an insufficient sample size (<20) were excluded. As shown in the Figure 6 and Table S1, all groups of children with a psychiatric disorder diagnosis showed a lower score of the connectome variate, with Conduct Disorder, Obsessive Compulsive Disorder, Delusion, Specific Phobia, Separation Anxiety and Homicidal Ideation reaching statistical significance (all FDR-adjusted $p < 0.05$).

To investigate whether low-scored connectomes associated to comorbid syndromes, we further categorized the participants of the whole data set into Groups (0, 1, 2 and ≥3) according to the number of KSADS diagnosis they met. As shown in Figure S1, the connectome variate score decreased with increasing number of co-occurring syndromes. A main effect of group was observed (One-way ANOVA $F(3,7378) = 4.62$, $p = 0.003$), and post-hoc tests revealed that children with 2 and ≥3 diagnosis showed significantly lower connectome variate scores compared to the control group (2 vs. 0: $t(378) = 2.68$, $p = 0.04$; ≥3 vs. 0 $t(259) = 5.21$, $p < 0.0001$), and children with ≥3 syndromes showed a significant lower connectome variate score compared to the other two diseased groups (≥3 vs. 2: $t(474) = 2.65$, $p = 0.04$, ; ≥3 vs. 1: $t(474) = 3.83$, $p = 0.0008$).

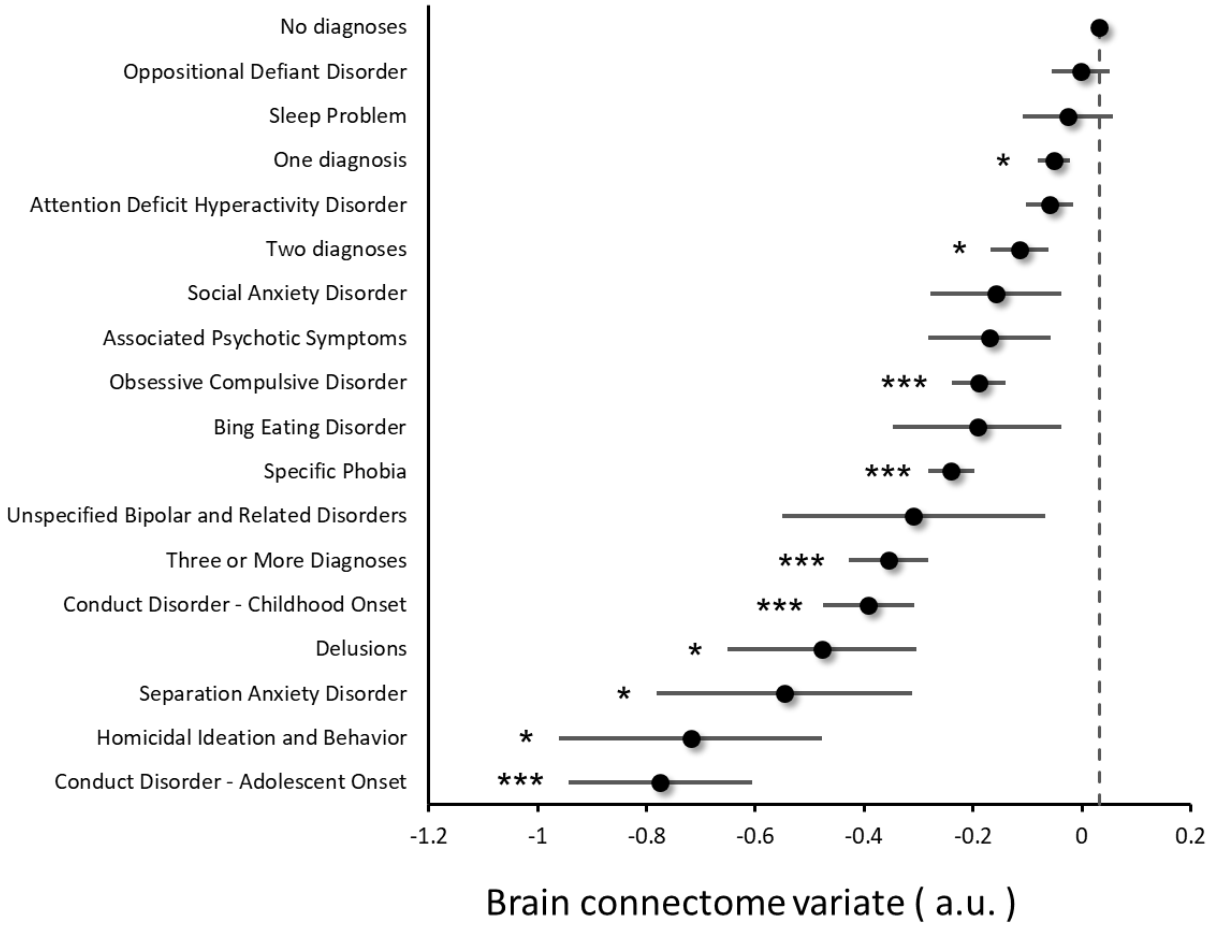

Figure 6 Group difference of connectome variate score between children with and without psychiatric disorder diagnoses. Dots and lines in the plot indicate mean and standard error of the connectome variate score in each group of KSADS diagnosis. Group without any diagnosis is set as the baseline. Welch's *t*-tests were used to compare the difference to the baseline group. P-values were FDR-adjusted for multiple comparisons. * FDR-adjusted $p < 0.05$, ** FDR-adjusted $p < 0.01$, *** FDR-adjusted $p < 0.001$

Profile of functional connectivity (FC) contributing to the connectome variate

As shown in above results the sCCA based model identified a common connectome variate shared by individual differences in both measures of cognition and mental disorders. Above, we explored the cognitive/behavioral aspects of this multi-dimensional connectome variate. We next address how the individual FC patterns across distributed brain areas contribute to this latent variate.

To address this, we first calculated the FC loading of each ROI pair as the Pearson's correlation between the FC and the connectome variate score across participants. The edgewise loadings (or loading

matrix) of the training population are shown in Figure 7A. A positive loading indicates the FC contributes positively to the brain score, while a negative loading indicates the opposite. The profile of edgewise loadings (Figure 7A) was further confirmed to be consistent with the difference of rsFC matrices between groups of highest 20% total intelligence score and lowest 20% total intelligence score (Pearson's $r = 0.87$, $p < 10^{-4}$, $N = 61776$; Figure S2A), and between groups with 0 KSADS diagnosis and 3+ KSADS diagnosis (Pearson's $r = 0.50$, $p < 10^{-4}$, $N = 61776$; Figure S3A).

Second, we investigated whether the FC loadings showed dominance in specific nodes or were distributed equally across the brain. For each of brain node, we separately summed the positive and negative loadings of the adjacent edges, respectively. Inequality of the summed loadings across brain nodes was quantified by the Gini coefficient(Cowell, 2000; Gini, 1912). Significant inequality across the brain areas was observed for both the positive (Gini = 0.203, $p = 0$, permutation test) and negative loadings (Gini = 0.207, $p = 0$ permutation test), indicating a non-uniform distribution of the FC loadings across the brain nodes (Figure S4). As shown in Figure 7B and Figure S5, positive loadings were predominantly in regions encompassing cortical areas of the anterior insula, anterior cingulate, dorsal lateral frontal, dorsal medial frontal, anterior temporal and temporal-parietal junction regions, as well as subcortical regions of the caudate, putamen, accumbens, thalamus and brain stem. Negative loadings were predominantly in sensory cortical regions encompassing the parietal and occipital cortex as well as the motor and premotor areas in the frontal cortex. As shown in Figure 7C and 7D , this pattern aligned (Spearman's *rho* = 0.36, $p < 1\times10^{-4}$, $N = 352$) with the sensory-to-association axis of human brain measured by the 'principal gradient of functional connectivity' (Margulies et al., 2016). The profile of nodewise loadings (Figure 7B) was further confirmed to be consistent with the nodewise difference of rsFC between groups of highest 20% total intelligence score and lowest 20% total intelligence score (Pearson's $r = 0.83$, $p < 10^{-4}$, $N = 352$; Figure S2B), and between groups with 0 KSADS diagnosis and 3+ KSADS diagnosis (Pearson's $r = 0.44$, $p < 10^{-4}$, $N = 352$; Figure S3B).

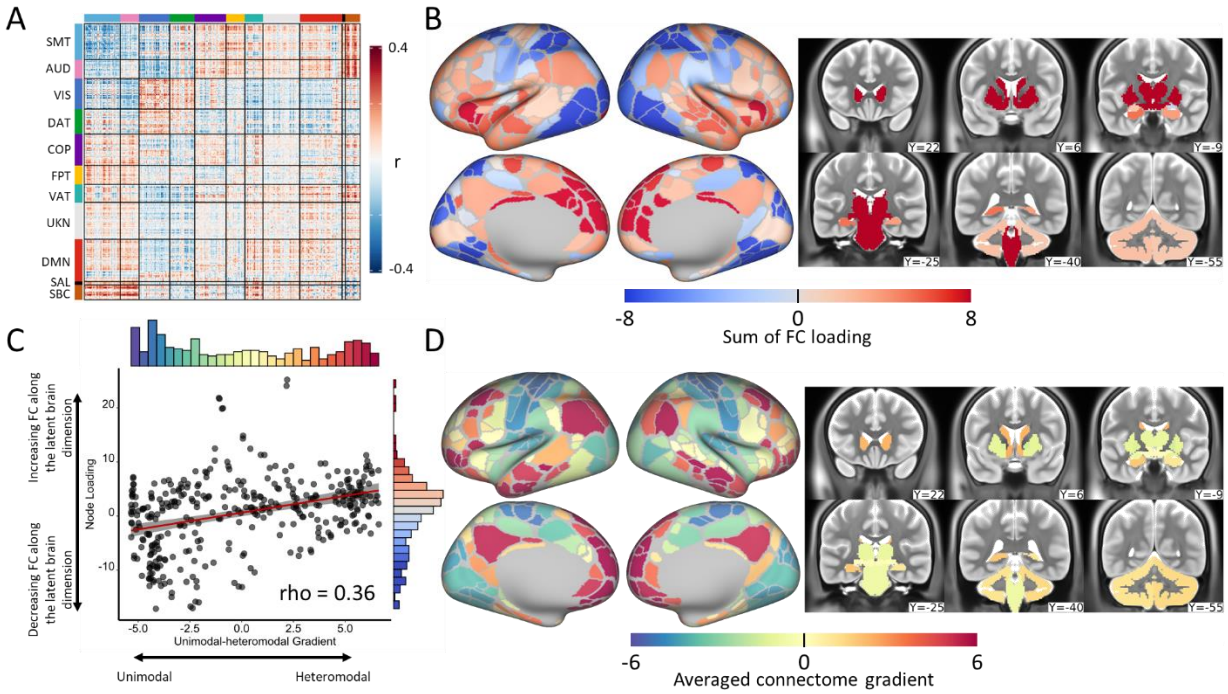

Figure 7 (A) Loading of functional connectivity (FC) of each pair of brain ROIs to the brain connectome variate. (B) Map of FC loading summed for each brain ROI. (C) Spatial correlation between nodal sum of FC loading and the value of principal functional connectome gradient. (D) The value of principal functional connectome gradient in each ROI. The image data was adopted from the result of (Margulies et al., 2016)**.** Colormap of the marginal histograms in C are shared with B and D in separate.

DISCUSSION

Leveraging a large population-based dataset of community dwelling preadolescents that included comprehensive measurements of RSFC, cognition, and psychopathology and a multi-view machine learning approach, we identified a multivariate pattern of functional connectivity that was positively correlated with performance on cognitive measures across domains and negatively correlated with parent reported psychopathology across diagnoses and domains. This connectome-based latent brain variate was capable of predicting individual differences in broadband behavioral metrics spanning domains of cognition and psychopathology and showed a dose-dependent relationship with categorical measures of psychopathology based upon the cumulative number of psychiatric diagnoses present. Moreover, the brain variate showed within sample replicability and stability following sensitivity testing. For example, when training the sCCA model with only one of the two behavioral domains (e.g., psychopathology measures only), the identified brain variate remained stable, and succeeded in predicting behavioral metrics in the other domain (e.g., cognition). Functional connectivity loadings on the brain variate were observed across a distributed network of cortical and subcortical brain regions, with our post-hoc analyses identifying positive and negative hubs in cognitive-control systems that contributed dominantly to the transdiagnostic neurocognitive dimension. Together these results indicate that specific patterns of functional connectivity during development are integral to cognitive control capacity and adaptive functioning, and are disrupted in preadolescents with broad-spectrum psychopathology across dimensions and diagnoses.

The identified biomarker shared by cognitive deficits and psychopathology suggests a new validated transdiagnostic dimension for RDoC. Identification of psychiatric biomarkers is essential for both understanding the ontology of and developing precise treatments for psychiatric disorders. In the past decade, efforts in the field of psychiatry has been made in searching for brain circuits liable to specific disorders, and research has been typically focused on comparing groups with single diagnosis to the healthy control. However, the high rate of co-morbidity of psychiatric disorders challenged the

traditional nosology system as it may not map the complex structure of human experience and biological deficit manifestation as mental disorders, or failed to ‘carve the nature at its joint’(Caspi & Moffitt, 2018; Cuthbert & Insel, 2013; Goldberg, 2015). This inspired the search for core brain-behavior association that crossed traditional diagnostic boundaries (Barch, 2017, 2020), as initiated by the RDoC (Cuthbert & Insel, 2013; Insel et al., 2010). In addressing this question, progresses have been made to identify neural correlates of cognitive function deficits commonly found in population with divergence of psychiatric diagnoses, represented by work focusing on executive function deficit. In a meta-analysis (Goodkind et al., 2015), brain morphological findings from traditional case-control studies were synthesized across a range of psychiatric diseases, and loss of gray matter was identified at regions encompassing bilateral anterior insula and anterior cingulate cortex, which is overlapped with the cingulo-opercular network that responds to tasks involving cognitive functions. A more recent meta-analysis (McTeague et al., 2017) moved this one step forward by identifying the pervasive hypoactivation of this circuit during cognitive tasks, comparing population with various disorders vs. the healthy control. These finding provided valuable insights into common circuit abnormalities that, in part, explain the presence of cognitive deficits across a variety of psychiatric diseases. However, a gap between a common deficit across diagnoses to a transdiagnostic neural dimension lies in whether this identified neural marker is associate with the individual difference in both the targeted behavior (e.g. working memory deficit) and the severity of psychiatric conditions. This gap limited the value of the identified neural marker in clinical applications. In parallel, the general factor of psychopathology (i.e., the p-factor) proposed from theoretical psychiatry was suggested to play a complementary role. Other than a common factor among mental disorders like the cognitive deficits, p-factor captures positive correlations with all the psychiatric symptoms (Caspi et al., 2014; Caspi & Moffitt, 2018), thus satisfies the concept of transdiagnostic dimension. However, unlike cognitive deficits, p-factor is derived from the symptoms defined by traditional psychiatric diagnoses; despite critics on this circularity, p-factor itself provides limited information about specific neurocognitive processing underlying. Hypothesis has been proposed that p-factor and cognitive deficits may mirror the same maladaptation of human brain (Caspi & Moffitt, 2018), based on the anticorrelation

between cognitive functions and p-factor observed in epidemiologic research in community samples (Caspi & Moffitt, 2018; Keyes, Platt, Kaufman, & McLaughlin, 2017). Our current finding provided unique evidence in supporting this hypothesis. First, on the identified CCA mode, metrics of cognition and psychopathology loaded on the positive and negative end of the same latent dimension of brain connectome, rather than orthogonal dimensions separately. Second and even more importantly, sCCA analysis conducted separately with each behavioral dataset identified exact the same brain connectome variate. These two together indicate that the two sets of behavioral measurements traditionally treated in separate domains are covarying with the same pattern of functional brain connectome in youth population. Furthermore, the connectome variate was validated as a transdiagnostic dimension in its liability to the number of co-occurring syndromes in individuals.

The current finding provides a brain-centered view to understand the latent structure among behaviors in different domains. Though the equation between brain and behavior is the key concept in identifying RDoC dimensions, selection of the targeted behavioral measurements is challenging. This challenge is even severer when latent structures like p-factor and g-factor are of interest, due to the problem of 'task impurity' (Miyake & Friedman, 2012; Watts, Lane, Bonifay, Steinley, & Meyer, 2020).

Taking g-factor as an example, this factor is conceptualized to reflect the core executive function that supports individual's performance in all the cognitive tasks. No single cognitive task can measure the pure executive function directly. Instead, the general executive function can only be observed when embedded within the context of specific cognitive tasks, but the performance in each task is mixed with the pure executive function and task specific components with unknown weightings. Decomposition the pooled behavior data has been long used in the field. However, the dependency on decomposition model may bear ambiguous conclusions. Joint decomposition of these multidimensional behavioral data with biological metrics is a promising approach to deal with this problem, and help to identify latent dimensions capturing covariance in the behavioral data while manifest biological significance (Kaczkurkin et al., 2020; H.-T. Wang et al., 2018; Xia et al., 2018). As revealed in our finding, the total

composite score of cognitive functions loaded with the largest extend on the connectome variate. This confirmed that this measurement captured the variance of the identified connectome pattern with best sensitivity in comparing to the other measurements. The measurement with negative while large extend loadings such as symptoms in the external spectra and impulsivity traits may capture impairment in cognitive control that is liable to these symptoms. Meanwhile, symptoms in the internalizing spectra loaded less or even insignificantly on the connectome variate. This dissociation may suggest a neuro-biological boundary between syndromes from these high-order spectra.

The current finding indicates a pattern of system-level functional configuration at rest of human brain liable to cognitive deficit and psychopathology. As shown in the result, functional connectivity among nodes across the brain loaded on the connectome variate in a non-random pattern. Specifically, functional connectivity loadings were highest in brain systems underlying cognitive control (Corbetta & Shulman, 2002; Dosenbach et al., 2006; Power & Petersen, 2013) and could be grouped into two clusters according to their loading on the connectome variate. The first cluster, with dominant positive loading, included brain areas of dACC, AI, dLPFC and inferior parietal lobule, and subcortical regions of striatal nuclei, thalamus and brain stem. The second cluster, with dominate negative loading, encompasses brain areas in frontoparietal system such as superior parietal lobule and premotor and FEF region, as well as primary regions of visual and sensory-motor areas. Though both clusters are recruited in cognitively demanding tasks (Dosenbach et al., 2006) they are thought to support distinct processes of cognitive control. For example, the first cluster supports goal-directed orienting of attention at start of task while the second cluster may play a larger role in conflict monitoring during maintaining of task (Corbetta & Shulman, 2002; Dosenbach et al., 2006). The positively loaded cluster showed heavy loading of cortical-striatal-thalamus loop circuits of the salience network (Fettes, Schulze, & Downar, 2017; Seeley et al., 2007), which has been found to play a critical role in cognitive control as a communication pathway integrating interoceptive and visceromotor information, and dysfunction of the network has been found in multiple psychiatric disorders (Menon, 2011; Uddin, 2015). Previous brain morphologic studies have

shown that loss of gray matter in key cortical nodes of this network, dACC and AI, correlated with cognitive deficits (McTeague et al., 2016) and psychopathology (Romer et al., 2021) at individual levels. Our results confirmed the critical role of these circuits in function and dysfunction of cognitive control, and provide new insight on how they support the individual difference from perspective of functional integration (Mesulam, 1998).

The current finding of biological dimension from ABCD study may open a neurocognitive perspective in understanding typical/atypical development. Cognitive capacity and risk of psychopathology per se are critical to brain development, due to their association to general life quality in the coming adolescent and later period of children's lives. The longitudinal and age-controlled population design of ABCD study provides an opportunity to understand the causal relationship between development of psychiatric disorders and reconfiguration of brain networks, and how the processes are affected by genetic, epigenetic and environmental factors. The identified neuro-cognitive biomarker should be valuable in both tracking the developmental trajectory of youth brain and discovering a causal neural target for early prevention or interference of negative events emerging in adolescent that would lead to psychiatric disorders and/or drug addiction.

In conclusion, in this study we discovered a single multivariate pattern of whole-brain functional connectivity that is shared by cognitive capacity and psychopathology in a large community sample of pre-adolescent youth. This connectome-based latent brain variate was stable, cross-validated within the ABCD sample, and explained individual differences across a wide range of cognitive and psychopathology domains and psychiatric disorder categories. Two subnetworks of the cognitive control system were found to contribute significantly to the connectome variate in distinct connectivity patterns. Consistent with other studies using case control or univariate analytic approaches, our findings revealed a transdiagnostic circuit-level abnormality linked to cognitive control capacity that predicted multiple psychiatric symptoms across domains and disorders. Further testing of this brain dimension for out of sample validation in an independent dataset and within the longitudinal design of ABCD study, will

determine it’s potential as a transdiagnostic biomarker for tracking typical/atypical developmental trajectory of youth brain and determine it’s viability as a neural target for prevention or early intervention of psychiatric disorders and/or drug addiction in young people.

METHODS

Participants

Neuroimaging data and behavioral assessments of 11,875 children aged at 9- to 10-years were obtained from the Adolescent Brian Cognitive Development (ABCD) study (Volkow et al., 2018). The large and long-term ongoing project aiming to characterize psychological and neurobiological development from pre-adolescence to young adulthood across 21 research sites in the US; description of the recruitment is detailed in (Garavan et al., 2018). Participants and their families are recruited through school- and community-settings in 21 centers across the US, following locally and centrally approved Institutional Review Board procedures.

ABCD study data includes yearly behavioral assessments and biennial MRI scans https://abcdstudy.org/scientists/protocols/ . In the current study, we only analyzed the baseline data. Demographic and behavioral assessments data were accessed from Curated Annual Release 3.0 (https://nda.nih.gov/abcd/query/abcd-curated-annual-release-3.0.html). Preprocessed data were obtained from the collection 3165 provided by Developmental Cognition and Neuroimaging (DCAN) Labs https://collection3165.readthedocs.io/en/stable/release_notes.html. Description of the processing pipeline is described in the section of image preprocessing as below.

As shown in Figure.S6, participants were excluded due to missing information of the scanning, neurological condition, data availability, failure to pass the quality control of Freesurfer, or inadequate time points of the fMRI data after motion censoring. As a result, data of 7383 participants (3714 females, aged 119.5 ± 7.4 months at scanning) were included in the final analysis. Demographic information of the included participant is summarized in Table S2.

Assessments of youth cognitive ability

Cognitive ability of the participants were quantified by 20 scores derived from their performance in 15 available neuro-cognitive tests (Luciana et al., 2018) and 3 neuroimaging tasks (Casey et al., 2018).

The NIH Toolbox cognition battery includes seven tasks measuring multiple aspects of cognition, which generates the following scores: Picture Vocabulary Test Score (nihtbx_pic_vocab); Flanker Inhibitory Control and Attention Test Score (nihtbx_flanker); List Sorting Working Memory Score (nihtbx_list_sort); Dimensional Change Card Sort Test Score (nihtbx_card_sort); Pattern Comparison Processing Speed Test Score (nihtbx_pattern_comp); Picture Sequence Memory Test Score (nihtbx_pic_seq_memory); Oral Reading Recognition Test Score (nihtbx_oral_reading). From to the direct scores of these tests, the NIH Toolbox further derivates 3 higher-order composited scores: Cognition Fluid Composite Score (nihtbx_fluidcomp); Crystallized Composite Score (nihtbx_cryst_comp); and Cognition Total Composite Score (nihtbx_total_comp). For the above measures, uncorrected raw scores were used due to the undergoing revision of the NIH Toolbox (Luciana et al., 2018; Rosenberg et al., 2020).

The Matrix Reasoning Test (matrix_reasoning) is a subtest from the Wechsler Intelligence Scale for Children Fifth Edition (WISC-V) and is a reliable measure for perceptual reasoning. Test performance in the test is measured by the total number of correct trials, and transformed into a standard score that is transformed into a standard score (Luciana et al., 2018; Rosenberg et al., 2020).

The Rey Auditory Verbal Learning Test is a test of auditory learning, memory, and recognition. The test requires the participant to listen to and recall a list of unrelated words over the learning trials. After the initial learning trials, participants were required to hear and recall another list of words as a distractor. Recall of the initial list was tested immediately and 30-min later after the distractor task. The performance of verbal learning in the immediate (verbal_learning_immediate) and the delayed

(verbal_learning_delayed) tests were measured by the number of words correctly recalled.(Rosenberg et al., 2020)

The Intertemporal Cash Choice Task (Wulfert, Block, Santa Ana, Rodriguez, & Colsman, 2002) is the test of the participant's delay of gratification. Participants were required to choose between a smaller-sooner ($75 in 3 days) vs. a larger-later ($115 in 3 months) hypothetical reward. In the current analysis, the choices were coded in two dummy variables consisting of preference for smaller-sooner (cash_choice_prefer_ss) and preference for larger-later (cash_choice_prefer_ll)

The Little Man Task (Acker & Acker, 1982) involves mental rotation and visual-spatial processing. The task includes trials that require the participant to indicate whether the cartoon of a little man is holding a briefcase with his left or right hand, while the little man varies his position and facing directions (Luciana et al., 2018). Performance is measured with efficiency (little_man_efficiency) (Rosenberg et al., 2020), which is defined as the ratio of the percent accuracy to mean response time of correct trials.

The Emotional N-back Task (EN-back) is an in-scanner block-design task in the ABCD study which engages short-term memory and emotion regulation (Casey et al., 2018). The task includes 2-back conditions and 0-back conditions. In the 0-back condition, participants were required to indicate whether the picture represented in the current trial is the same as the target picture shown at the begin of the task. In the 2-back condition, participants were required to indicate whether the picture in the current trial is the same to the target picture shown in 2 trials earlier. The task includes four types of stimuli: photographs of faces with happy, fearful and neutral expressions and photographs of places. In the current analysis, the percent accuracy of 2-back pooled across all types of stimuli was used as the measurement of performance (2_back).

The Recognition Memory Task is the post-scan test of the in-scanner EN-back task(Casey et al., 2018). During the task, participants were shown photographs presented in the EN-back task as well as

equal number of novel photographs. They were asked to indicate whether the presented photograph was 'new' or 'old' in each trial. The performance (recog_memory) in the task was measured with the sensitivity index (*d'*) averaged across stimulus types (Rosenberg et al., 2020).

The Stop-signal Task (SST) is an in-scanner task in the ABCD study. It measures impulsivity and impulse control (Casey et al., 2018). During the "Go" trials (83.33% of the total trials), participants were instructed to indicate the direction of the presented arrow (the "Go Signal" pointing towards left/right) "as quickly and accurately as possible." During the "Stop" trials (16.67% of the total trials), an unpredictable up-right arrow appeared as a "Stop Signal" after the "Go Signal". In response to the "Stop Signal," the participant was required to withhold their response to the arrow. The time delay between go and stop signal, defined as the stop-signal delay, was changed in a staircase fashion in an attempt to ensure each participant reached an accuracy of ~50%. Performance of the task (SST_ssrt_reversed) was derived from the stop-signal reaction time (SSRT), which was defined by subtracting the mean stop-signal delay across stop trials from the mean reaction time across correct go trials. The SSRT were sign-reversed following (Rosenberg et al., 2020), so that the higher score relates to better impulse inhibition.

The Monetary Incentive Delay (MID) task is an in-scanner task in the ABCD study. It engages reward processing, including anticipation and receipt of reward and losses, trial-by-trial motivation in speeded responses to win or avoid loss (Casey et al., 2018). Each trial of the task begins with an incentive cue indicating the type of the trial from 5 possibilities (Win $.20, Win $5, Lose $.20, Lose $5, $0-no money at stake). After a jittered anticipation period (1500–4000 ms) , a target stimulus with varying duration (150-500 ms) appeared and the participant was required to respond to a target stimulus to win the reward or to avoid the loss. The duration of the target stimulus was changed in a staircase fashion in an attempt to ensure each participant reached an accuracy of ~60%. Performance of the task (MID_total_earning) was measured with the total amount earned by each child (Rosenberg et al., 2020). Measurements and abbreviations for cognitive ability are listed in Table S3.

Assessments of participant mental health

Participant mental health and related individual traits were assessed by the baseline mental health battery (Barch et al., 2018). The battery incorporates reliable and validated scales covering a broad range of mental health related constructs, and provides the feasibility of tracking a large-scale population of 9- or 10-year-old children into their adulthood. In the ABCD project, both categorical and dimensional approaches were included. We utilized these two types of assessments for different purposes. Specifically, the dimensional assessments were used for estimating the brain-behavior association using the sCCA model,

due to their advantage in capturing the full range of variance from the diseased to the normal (Cuthbert & Insel, 2013). In contrast, the categorical assessments were used for external validation, due to their compatibility with psychiatric diagnosis used in clinical practice.

Dimensional assessments of mental health

In the current study, 30 dimensional scores of mental health in youth derived from baseline assessments were used to estimate the sCCA based brain-behavior association. The measurements include the parent/caregiver reported Child Behavioral Checklist (CBCL), scale of mania, youth reported scales of impulsivity and psychosis risk.

The parent/caregiver-reported CBCL (Achenbach & Verhulst, 2010) was used to assess the emotional and behavioral problems of the participants. It includes 119 problem items rated with 3-level scales: 0 = Not True, 1 = Somewhat or Sometimes True, and 2 = Very True or Often True. From the rating on these items, the CBCL further derives composite scales including: 8 syndrome scales of Anxious/Depressive symptom (CBCL_syn_anxdep), Withdrawn/Depressed symptom (CBCL_syn_withdep), Somatic Complaints (CBCL_syn_somatic), Social Problems (CBCL_syn_social_prob), Thought Problems (CBCL_syn_thought_prob), Attention Problems (CBCL_syn_attention_prob), Rule-Breaking Behaviors (CBCL_syn_rulebreak) and Aggressive Behavior (CBCL_syn_aggresive); 3 higher-order scales of Internalizing Problems (CBCL_internal), Externalizing

Problems (CBCL_external) and Total Problems (CBCL_total); 6 DSM oriented scales of Depressive Problems (CBCL_dsm_depressive), Anxiety Problems (CBCL_dsm_anxiety), Somatic Problems (CBCL_dsm_somatic), Attention Deficits (CBCL_dsm_ADHD), Oppositional Defiant Problems (CBCL_dsm_opposit), Conduct Problems (CBCL_dsm_conduct_prob) and 3-scale scores of Sluggish Cognitive Tempo (CBCL_SCT), Obsessive-Compulsive Problems (CBCL_OCD) and Stress Problems (CBCL_stress). For each of the 20 scales, raw scores were used in the current analysis.

The parent/caregiver-reported ten-item Mania Scale (Youngstrom, Frazier, Demeter, Calabrese, & Findling, 2008) derived from the 73-item Parent General Behavior Inventory (PGBI) for Children and Adolescents (Youngstrom, Findling, Danielson, & Calabrese, 2001) was used to assess dimensional mania symptoms of the youth. The sum score (PGBI_mania_total) was included in the current analysis.

A youth-report version of UPPS-P (Urgency, Perseverance, Premeditation and Sensation seeking) scale was used to assess the dimensional impulsivity of children (Barch et al., 2018; Zapolski, Stairs, Settles, Combs, & Smith, 2010). Five dimensional scales of Lack of Planning (UPPS_lack_planing), Sensation Seeking (UPPS_sensation_seeking), Positive Urgency (UPPS_positive_urgency), Negative Urgency (UPPS_negative_urgency) and Lack of Perserverance (UPPS_lack_perserverance) were included in the analysis.

A modified version (Pagliaccio et al., 2016) of Behavioral Inhibition & Behavioral Activation Scales (BIS/BAS) (Carver & White, 1994) reported by the participants was used to assess the psychobiological models of two broad motivational systems, the behavioral activation system (BAS) and the behavioral inhibition system (BIS). The BIS/BAS scale has a dimensional scale to assess Behavioral Inhibition (BISBAS_bis_sum). and three scales to assess Behavioral Activation: Drive (BISBAS_drive), Reward Responsiveness (BISBAS_reward_resp), and Fun Seeking (BISBAS_fun_seek). All the 4 scales were included in the current analysis.

Youth-report Prodromal Questionnaire Brief Version (PQ-B) (Loewy, Pearson, Vinogradov, Bearden, & Cannon, 2011) was used to assess the psychosis risk symptoms. The severity score of prodromal psychosis (PQB_psychosis_severity) was included in the current analysis.

Measurements and abbreviations of dimensional mental health assessment are listed in Table S4.

Categorical assessments of mental health

The categorical diagnostic assessment is measured with the computerized Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS) for DSM-5 (KSADS-5) (Barch et al., 2018; Kobak, Kratochvil, Stanger, & Kaufman, 2013). In the baseline assessment, modules covering externalizing, psychosis, and eating disorders were reported by parents, while modules of mood disorders, separation anxiety, social anxiety, generalized anxiety, sleep and suicidality were reported by the participants with support from trained research staff. The rationale has been discussed in (Barch et al., 2018). In the ABCD dataset, KSADS-5 includes Current Diagnoses (CD) and Past Diagnoses (PD). For CD of a disorder, symptoms are currently at their worst during the episode, while PD indicates that the symptoms of the episode have been resolved for at least two months (Kaufman et al., 1997).

Image acquisition and processing

In the ABCD data study, MRI data of the participants are collected from 21 sites across the US. Structural and functional images are acquired using three 3 tesla scanner platforms: Siemens Prisma and Prisma_fit, General Electric Discovery MR750 and SIGNA Creator, and Phillips Medical System Ingenia and Achieva dStream, all equipped with multi-channel coils in adult-size that are capable of multiband echo planar imaging (EPI) acquisitions. Structural MRI images are collected with T1-weighted magnetization-prepared rapid acquisition gradient echo and T2-weighted fast spin echo with variable flip angle scans. Resting-state and task functional images are collected with high spatial and temporal

resolution simultaneous multi-slice /multiband EPI. The resting-state fMRI acquisition includes four 5-min scans, during which participants were instructed to keep their eyes open and passively watch a cross hair. Parameters and the protocol of acquisition are detailed elsewhere (Casey et al., 2018) .

Structural and functional MR images were preprocessed and housed in the ABCD-BIDS Community Collection (ABCC) from the Developmental Cognition and Neuroimaging (DCAN) Labs. The preprocessing abcd-hcp-pipeline mainly consists of two stages. The first stage is a modified version of Human Connectome Project (HCP)'s minimal preprocessing pipeline (Glasser et al., 2013) to map fMRI volumes from the native space of each individual into a standard grayordinate spatial coordinate system, a combined cortical surface and subcortical volume coordinate system. The second stage used the DCAN BOLD Processing (DBP) software (Fair et al., 2020) to remove nuisance signals including artifacts induced by in-scanner head motion and participant respiration.

The modified HCP’s minimal preprocessing pipeline includes processing of both structural and functional MR images. For the structural MRIs, the processing includes denoising and N4 bias field correction, segmentation of brain tissue, volume-based spatial normalization of the sub-cortical brain and folding-based surface spatial normalization of the cortex. For fMRI, the processing includes correction of the gradient-nonlinearity-induced distortions, frame alignment, calculation of the motion parameters, non-linear registration to the MNI standard space, and volume-to-surface mapping to the grayordinate space.

The DBP is modified from the procedure of (Power et al., 2014), with correction of respiratory effects from motion estimates within multi-band data (Fair et al., 2020). The processing includes de-meaning and de-trending; nuisance-signal regression including: mean time series from white matter and CSF; the global signal (GS); and the 24-parameter Volterra expansion from motion measures estimated by re-alignment (Friston, Williams, Howard, Frackowiak, & Turner, 1996; Yan et al., 2013); band-pass filtering between 0.008 and 0.09 Hz using a 2nd order Butterworth filter; and motion censoring. In the DBP the framewise displacement (FD) used for the motion censoring is corrected by filtering out the respiratory frequencies (0.31 Hz and 0.43 Hz). In the current analysis, frames of fMRI data with corrected

FD > 0.2 mm were censored (Fair et al., 2020; Power et al., 2014) before entering the construction of functional connectome. To ensure a reliable estimation of the functional connectivity, 1645 participants with remaining time series < 6 min after censoring were excluded for further analysis (Birn et al., 2013).

Construction of functional connectome

Functional connectome was constructed from inter-regional BOLD synchrony of 352 areas across an individual brain. For the cortex, the Gordon atlas (Gordon et al., 2016) was used to divide the brain into 333 areas. The Gordon parcellation uses resting-state functional connectivity (RSFC)-boundary mapping to parcellate the cortex of human brain, ensuring that vertices within the same parcel show a homogenous RSFC patterns. The 333 cortical parcels were further assigned in to 12 functional communities or networks using an Infomap based procedure (Laumann et al., 2015). The 12 functional networks include Sensory Motor Network (SMT), Auditory Network (AUD), Visual Network (VIS), Default Network (DMN), Frontal-parietal Network (FPT), Cingulo-opercular Network (CON), Salience Network (SAL), Dorsal Attention Network (DAT), Ventral Attention Network (VAT), Cingulo Parietal Network (COP), Retrosplenial Temporal Network (RST) and Limbic Network (LIMB). The parcellation atlas also includes 19 subcortical nuclei of L/R_Cerebellum, L/R_Pallidum, L/R_Caudate, L/R_Putamen, L/R_Accumbens, L/R_Thalamus, L/R_Amygdala, L/R_Hippocampus, L/R_Ventral_Diencephalon, Brain_Stem, and all the subcortical regions are assigned into the one Subcortical Network (SUB). For each brain, the preprocessed BOLD signals within each parcel were extracted and averaged. The functional connectivity between every pair of parcels was measured as the z-transformed Pearson's correlation between the two averaged timeseries. A 352 by 352 correlation matrix for each participant can then be obtained, which represents the brain's profile of functional connectome.

Pre-CCA processing

For each participant, the connectome features consist of 61,776 edges, and the behavioral features consist of 51 metrics. Before submitting them for sCCA analysis, we performed two pre-sCCA processing on the data sets. First, to minimize effects of head motion on CCA results and to control different scanner platforms, we regressed head motion measured by the mean framewise displacement (FD) of the scrubbed data, and the MRI scanner type from the brain data set across participants. Second, Principle Component Analyses (PCA) were performed separately on the two feature sets to transform them into orthogonal feature spaces (principal components, PCs) while keeping a major portion of keeping variance of the original feature sets. For each of the transformed feature sets, the top 51 PCs with highest variance were retained for the sCCA analysis. Above procedure produces 51 orthogonal components for the connectome feature set, with 24% of the total variance retained, and 51 orthogonal components for the behavioral feature set, with 100% of the total variance retained.

Sparse canonical correlation analysis

The sparse canonical correlation analysis procedure introduced by (Witten et al., 2009) was performed to exam the brain-behavioral association (H. T. Wang et al., 2020; Xia et al., 2018). The Witten method is based on penalized matrix decomposition. In this framework, CCA is expressed as a problem of constrained optimization:

$$\arg\max_{\mathbf{u},\mathbf{v}} \mathbf{u}^T\mathbf{X}^T\mathbf{Y}\mathbf{v}, \textit{subject to } \|\mathbf{u}\|_2^2 = 1, \|\mathbf{v}\|_2^2 = 1, \|\mathbf{u}\|_1 = C_1 \textit{ and } \|\mathbf{v}\|_2 = C_2$$

Here, $\mathbf{X}$, $\mathbf{Y}$ are the multi-dimensional dataset of brain and behavior, $\mathbf{u}$ and $\mathbf{v}$ are coefficient matrices projecting each of the data sets onto the latent variates. C_1 and C_2 are regularization parameters that constrain the sparsity of $\mathbf{u}$ and $\mathbf{v}$. Stronger constraints (i.e., smaller value of C_1 or C_2) shrink the model and will usually result a reduced performance for fitting the training data but better generalizability to unseen data. The final regularization parameters used in the sCCA model were determined from a grid-searching procedure detailed below.

Hold-out validation

The generalizability of the sCCA model was validated within a hold-out framework (Figure.S8). The whole data set of 7383 subjects was randomly split into a training set of 5906 subjects (80% of the total samples) and a test set of 1477 subjects (20% of the total samples). We tuned the regularization parameters and generated the sCCA model (i.e., coefficients of **u** and **v**) only on the training set, and then applied the sCCA model to the test set for validation.

Grid-search for regularization parameters

Regularization parameters C_1 and C_2 are optimized via a grid-searching procedure and nested cross-validation within the training set.

For each regularization parameter the grid covered the full ranged [0, 1] with a step of 0.11. This resulted in 100 cells in the grid, where each cell represents a combination of the two parameters.

For a given cell, a 10-fold nested cross-validation was performed within the training set. In each fold, the training set was randomly split into a nested training set (90% of the samples) and a nested test set (10% of the samples). A sCCA model was first estimated from the nested training set, with regularization parameters corresponding to the given grid cell. The estimated model was then applied to the nested test set and the primary canonical correlation was calculated as the validation score. Finally, the parameter combination was scored by the averaged validation scores across the 10-folds. Regularization parameters C_1 and C_2 were optimized to be 0.67 and 0.56 for the brain and behavior feature sets respectively.

The regularization parameters were optimized by selecting the combination with highest score from all the 100 cells. When there is a tie, parameters with smaller values were preferred, maximizing the simplicity and generalizability of the model.

The grid search described above was performed only in the training set, with the optimal regularization parameters applied to the test set.

Permutation testing for sCCA

Permutations tests were conducted to assess the significance of the canonical correlations, i.e., the correlation between each pair of the brain and behavior variates. The rationale of permutation is to break the sample-wise pairing between brain and behavior features. To generate a permutated dataset, we shuffled the order of behavior features while keeping the brain feature set in the original order for each permutation An sCCA model was estimated from the permutated dataset with same regularization parameters of the true model, and the value of the primary canonical correlation was recorded as r_i for the ith permutation. Repeating the permutations 10,000 times, we can obtain a null distribution of the canonical correlation coefficient r_i. The null distribution represents the result sCCA can provide when no actual association exists between brain and behavior. For a canonical correlation r in the true sCCA model the corresponding p value was estimated by the proportion of r_i exceed the true r.

$$p = \frac{\sum \begin{cases} 1, if\ r_i > r \\ 0, if\ r_i \leq r \end{cases}}{10000}$$

The level of significance was set at $p = 0.05$ after Bonferroni correction for multiple comparisons.

Cross-domain behavior prediction with the latent brain variate

To assess the association between the detected latent brain connectome variate to the individual difference in each of the behavioral assessments, we conducted an analysis of cross-domain behavior prediction. First, an sCCA model was estimated from the training set following procedures described above. Second, the first column vector u_1 from the coefficient matrix **u** were used as the brain model.

Third, u_1 was applied to the connectome features in the test set to estimate the brain variate in the unseen population. Finally, in the test set, Pearson's correlation tests were performed between the brain variate estimate and each of the 51 behavioral assessments. The level of significance was at $p = 0.05$, multiple comparisons adjusted using the Benjamini and Hochberg procedure (Benjamini & Hochberg, 1995) to control the false discovery rate (FDR) . The Pearson's correlation test and FDR adjustment were performed using the R package 'stats' v3.6.2

Transdiagnostic liability of connectome mode to psychiatric disorders

To assess the the association between of the detected latent brain variate to the clinical diagnosis, we examined the brain variate among participants who meet the current and lifetime psychiatric diagnosis in comparison to the group of no-disorder. The clinical diagnosis was assessed via the KSADS-5. The lifetime psychiatric diagnosis was generated by pooling both current and past diagnosis. In addition to the KSADS-5 disorders, we categorized participants of the whole data set into having 0, 1, 2 or 3+ disorders. Disorders with insufficient sample size (<20) were excluded. Due to the unbalanced sample sizes between the diseased group and the no-disorder group, we used the Welch two-sample *t*-test (Delacre, Lakens, & Leys, 2017) to compare differences between the two groups. The level of significance was at $p = 0.05$, and the *p*-values were FDR-adjusted (Benjamini & Hochberg, 1995) for multiple comparisons. The Welch two-sample *t*-test and FDR adjustment was performed using the R package 'stats' v3.6.2.

FC loadings on the composite brain score

Contribution of each FC to the composite brain score was measured by the FC loading defined as the Pearson's correlation between the FC and the brain connectome variate across the training population. As only the brain connectome variate in the first canonical correlation mode is of interest, we only investigated FC loading to this variate. A positive or a negative loading indicates the given FC either increases or decreases along the variate.

Permutation test for the inequality of nodal FC loadings

Inequality of nodal FC loadings was tested for the positive and negative loading matrices separately. For each of the two loading matrices, absolute FC loadings were summed at each node. Here absolute value was used for negative loadings because Gini coefficient only applies for non-negative values. The Gini coefficient (Gini, 1912) was calculated to quantify the inequality across the 352 nodal sums. We then permuted the elements in the loading matrix 1000 times and calculated the Gini coefficient from each of the permuted matrices forming the null distribution. The *p*-value of the test was calculated by the proportion of permuted Gini coefficients exceeding the true one.

The Gini coefficient ranges between 0 to 1. A Gini of 0 indicates all the observed measurements are perfectly equal, while a Gini approaching to 1 indicates all the measurements but one has a zero value, a totally unequal distribution. The calculation of the Gini coefficient and the related Lorenz curve (Lorenz, 1905) was using the R package 'DescTools' version 0.99. (https://andrisignorell.github.io/DescTools/); see (Cowell, 2000) for further details of the Gini coefficient and Lorenz curve.

Permutation test for nodal FC loadings

To identify dominant nodes with positive/negative loadings, we calculated the nodal FC loadings from the full loading matrix. We then permuted the elements in the loading matrix 1000 times. For each of the 352 nodes, nodal FC loading derived from the 1000 permuted matrices formed the null distribution of its nodal FC loading. The *p*-value of the test was calculated by the proportion of permuted nodal FC loading exceeded the true one. To deal with multiple comparisons across the 352 nodes, the p-values were FDR adjusted and thresholded at 0.05.

REFERENCES

Achenbach, T. M., & Verhulst, F. (2010). *Achenbach system of empirically based assessment (ASEBA)*. Burlington, Vermont.

Acker, W., & Acker, C. (1982). Bexley Maudsley automated processing screening and Bexley Maudsley category sorting test manual. *Windsor, England: NFER-Nelson*.

Barbey, A. K. (2018). Network Neuroscience Theory of Human Intelligence. *Trends in Cognitive Sciences*, *22*(1), 8–20. https://doi.org/10.1016/j.tics.2017.10.001

Barch, D. M. (2017). The neural correlates of transdiagnostic dimensions of psychopathology. *American Journal of Psychiatry*, *174*(7), 613–615. https://doi.org/10.1176/appi.ajp.2017.17030289

Barch, D. M. (2020). What Does It Mean to Be Transdiagnostic and How Would We Know? *American Journal of Psychiatry*, *177*(5), 370–372. https://doi.org/10.1176/appi.ajp.2020.20030243

Barch, D. M., Albaugh, M. D., Avenevoli, S., Chang, L., Clark, D. B., Glantz, M. D., … Sher, K. J. (2018). Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: Rationale and description. *Developmental Cognitive Neuroscience*, *32*(September 2017), 55–66. https://doi.org/10.1016/j.dcn.2017.10.010

Bassett, D. S., & Bullmore, E. T. (2009). Human brain networks in health and disease. *Current Opinion in Neurology*, *22*(4), 340–347. https://doi.org/10.1097/WCO.0b013e32832d93dd

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Birn, R. M., Molloy, E. K., Patriat, R., Parker, T., Meier, T. B., Kirk, G. R., … Prabhakaran, V. (2013). The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *NeuroImage*, *83*, 550–558. https://doi.org/10.1016/j.neuroimage.2013.05.099

Bjork, J. M., Straub, L. K., Provost, R. G., & Neale, M. C. (2017). The ABCD study of neurodevelopment: Identifying neurocircuit targets for prevention and treatment of adolescent substance abuse. *Current Treatment Options in Psychiatry*, *4*(2), 196–209. https://doi.org/10.1007/s40501-017-0108-y

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology*, *67*(2), 319–333. https://doi.org/10.1037/0022-3514.67.2.319

Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., … Dale, A. M. (2018). The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, *32*(May 2017), 43–54. https://doi.org/10.1016/j.dcn.2018.03.001

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S. J., Harrington, H., Israel, S., … Moffitt, T. E. (2014). The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*, *2*(2), 119–137. https://doi.org/10.1177/2167702613497473

Caspi, A., & Moffitt, T. E. (2018). All for one and one for all: Mental disorders in one dimension. *American Journal of Psychiatry*, *175*(9), 831–844. https://doi.org/10.1176/appi.ajp.2018.17121383

Clementz, B. A., Sweeney, J. A., Hamm, J. P., Ivleva, E. I., Ethridge, L. E., Pearlson, G. D., … Tamminga, C. A. (2016). Identification of Distinct Psychosis Biotypes Using Brain-Based Biomarkers. *The American Journal of Psychiatry*, *173*(4), 373–384. https://doi.org/10.1176/appi.ajp.2015.14091200

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215. https://doi.org/10.1038/nrn755

Cowell, F. A. (2000). Measurement of inequality. In *Handbook of income distribution* (Vol. 1, pp. 87–166). Elsevier. https://doi.org/10.1016/S1574-0056(00)80005-6

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine*, *11*(1), 126. https://doi.org/10.1186/1741-7015-11-126

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use welch's t-Test instead of student's t-Test. *International Review of Social Psychology*, *30*(1), 92–101. https://doi.org/10.5334/irsp.82

Dosenbach, N. U. F., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., … Petersen, S. E. (2006). A Core System for the Implementation of Task Sets. *Neuron*, *50*(5), 799–812. https://doi.org/10.1016/j.neuron.2006.04.031

Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., … Liston, C. (2017). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine*, *23*(1), 28–38. https://doi.org/10.1038/nm.4246

Fair, D. A., Miranda-Dominguez, O., Snyder, A. Z., Perrone, A., Earl, E. A., Van, A. N., … Dosenbach, N. U. F. (2020). Correction of respiratory artifacts in MRI head motion estimates. *NeuroImage*, *208*(October 2019), 116400. https://doi.org/10.1016/j.neuroimage.2019.116400

Falk, E. B., Hyde, L. W., Mitchell, C., Faul, J., Gonzalez, R., Heitzeg, M. M., … Schulenberg, J. (2013). What is a representative brain? Neuroscience meets population science. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(44), 17615–17622. https://doi.org/10.1073/pnas.1310134110

Fettes, P., Schulze, L., & Downar, J. (2017). Cortico-striatal-thalamic loop circuits of the orbitofrontal cortex: Promising therapeutic targets in psychiatric illness. *Frontiers in Systems Neuroscience*, *11*(April), 1–23. https://doi.org/10.3389/fnsys.2017.00025

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., … Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, *18*(11), 1664–1671. https://doi.org/10.1038/nn.4135

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *86*, 186–204. https://doi.org/10.1016/j.cortex.2016.04.023

Friedman, N. P., Miyake, A., Altamirano, L. J., Corley, R. P., Young, S. E., Rhea, S. A., & Hewitt, J. K. (2016). Stability and change in executive function abilities from late adolescence to early adulthood: A longitudinal twin study. *Developmental Psychology*, *52*(2), 326–340. https://doi.org/10.1037/dev0000075

Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J., & Turner, R. (1996). Movement-related effects in fMRI time-series. *Magnetic Resonance in Medicine*, *35*(3), 346–355. https://doi.org/10.1002/mrm.1910350312

Garavan, H., Bartsch, H., Conway, K., Decastro, A., Goldstein, R. Z., Heeringa, S., … Zahs, D. (2018). Recruiting the ABCD sample: Design considerations and procedures. *Developmental Cognitive Neuroscience*, *32*(April), 16–22. https://doi.org/10.1016/j.dcn.2018.04.004

Gini, C. (1912). Variabilità e mutabilità. *Reprinted in Memorie Di Metodologica Statistica (Ed. Pizetti E*.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., … Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, *80*, 105–124. https://doi.org/10.1016/j.neuroimage.2013.04.127

Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., … Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America*,

101(21), 8174–8179. https://doi.org/10.1073/pnas.0402680101

Goldberg, D. (2015). Psychopathology and classification in psychiatry. *Social Psychiatry and Psychiatric Epidemiology*, *50*(1), 1–5. https://doi.org/10.1007/s00127-014-0924-y

Goodkind, M., Eickhoff, S. B., Oathes, D. J., Jiang, Y., Chang, A., Jones-Hagata, L. B., … Etkin, A. (2015). Identification of a common neurobiological substrate for mental Illness. *JAMA Psychiatry*, *72*(4), 305–315. https://doi.org/10.1001/jamapsychiatry.2014.2206

Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M., & Petersen, S. E. (2016). Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cerebral Cortex*, *26*(1), 288–303. https://doi.org/10.1093/cercor/bhu239

Gualtieri, C. T., & Morgan, D. W. (2008). The frequency of cognitive impairment in patients with anxiety, depression, and bipolar disorder: an unaccounted source of variance in clinical trials. *The Journal of Clinical Psychiatry*, *69*(7), 1122–1130. https://doi.org/10.4088/jcp.v69n0712

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D., Quinn, K., … Wang, P. (2010). Research Domain Criteria (RDoC): Toward a. *American Journal of Psychiatry Online*, (July), 748–751.

Kaczkurkin, A. N., Moore, T. M., Sotiras, A., Xia, C. H., Shinohara, R. T., & Satterthwaite, T. D. (2020). Approaches to Defining Common and Dissociable Neurobiological Deficits Associated With Psychopathology in Youth. *Biological Psychiatry*, *88*(1), 51–62. https://doi.org/10.1016/j.biopsych.2019.12.015

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., … Ryan, N. (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(7), 980–988. https://doi.org/10.1097/00004583-199707000-00021

Keyes, K. M., Platt, J., Kaufman, A. S., & McLaughlin, K. A. (2017). Association of fluid intelligence

and psychiatric disorders in a population-representative sample of US adolescents. *JAMA Psychiatry*, *74*(2), 179–188. https://doi.org/10.1001/jamapsychiatry.2016.3723

Kim, H., & Eaton, N. R. (2015). The hierarchical structure of common mental disorders: Connecting multiple levels of comorbidity, bifactor models, and predictive validity. *Journal of Abnormal Psychology*, *124*(4), 1064–1078. https://doi.org/10.1037/abn0000113

Kobak, K. A., Kratochvil, C. J., Stanger, C., & Kaufman, J. (2013). Computerized screening of comorbidity in adolescents with substance or psychiatric disorders. *Anxiety Disorders and Depression.(La Jolaa, CA)*.

Laumann, T. O., Gordon, E. M., Adeyemo, B., Snyder, A. Z., Joo, S. J., Chen, M. Y., … Petersen, S. E. (2015). Functional System and Areal Organization of a Highly Sampled Individual Human Brain. *Neuron*, *87*(3), 657–670. https://doi.org/10.1016/j.neuron.2015.06.037

Lee, S. H., Ripke, S., Neale, B. M., Faraone, S. V, Purcell, S. M., Perlis, R. H., … Wray, N. R. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics*, *45*(9), 984–994. https://doi.org/10.1038/ng.2711

Lesh, T. A., Niendam, T. A., Minzenberg, M. J., & Carter, C. S. (2011). Cognitive Control Deficits in Schizophrenia: Mechanisms and Meaning. *Neuropsychopharmacology*, *36*(1), 316–338. https://doi.org/10.1038/npp.2010.156

Loewy, R. L., Pearson, R., Vinogradov, S., Bearden, C. E., & Cannon, T. D. (2011). Psychosis risk screening with the Prodromal Questionnaire - Brief Version (PQ-B). *Schizophrenia Research*, *129*(1), 42–46. https://doi.org/10.1016/j.schres.2011.03.029

Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, *9*(70), 209–219.

Luciana, M., Bjork, J. M., Nagel, B. J., Barch, D. M., Gonzalez, R., Nixon, S. J., & Banich, M. T. (2018).

Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery. *Developmental Cognitive Neuroscience*, *32*(July 2017), 67–79. https://doi.org/10.1016/j.dcn.2018.02.006

Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., … Smallwood, J. (2016). 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*(44), 12574–12579. https://doi.org/10.1073/pnas.1608282113

Mcteague, L. M. (2016). Reconciling RDoC and DSM approaches in clinical psychophysiology and neuroscience. *Psychophysiology*, *53*(3), 323–327. https://doi.org/10.1111/psyp.12602

McTeague, L. M., Goodkind, M. S., & Etkin, A. (2016). Transdiagnostic impairment of cognitive control in mental illness. *Journal of Psychiatric Research*, *83*, 37–46. https://doi.org/10.1016/j.jpsychires.2016.08.001

McTeague, L. M., Huemer, J., Carreon, D. M., Jiang, Y., Eickhoff, S. B., & Etkin, A. (2017). Identification of common neural circuit disruptions in cognitive control across psychiatric disorders. *American Journal of Psychiatry*, *174*(7), 676–685. https://doi.org/10.1176/appi.ajp.2017.16040400

Menon, V. (2011). Large-scale brain networks and psychopathology: A unifying triple network model. *Trends in Cognitive Sciences*, *15*(10), 483–506. https://doi.org/10.1016/j.tics.2011.08.003

Mesulam, M. M. (1998). From sensation to cognition. *Brain*, *121*(6), 1013–1052. https://doi.org/10.1093/brain/121.6.1013

Mihalik, A., Ferreira, F. S., Moutoussis, M., Ziegler, G., Adams, R. A., Rosa, M. J., … Mourão-Miranda, J. (2020). Multiple Holdouts With Stability: Improving the Generalizability of Machine Learning Analyses of Brain–Behavior Relationships. *Biological Psychiatry*, *87*(4), 368–376. https://doi.org/10.1016/j.biopsych.2019.12.001

Millan, M. J., Agid, Y., Brüne, M., Bullmore, E. T., Carter, C. S., Clayton, N. S., … Young, L. J. (2012). Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. *Nature Reviews. Drug Discovery*, *11*(2), 141–168. https://doi.org/10.1038/nrd3628

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Pagliaccio, D., Luking, K. R., Anokhin, A. P., Gotlib, I. H., Hayden, E. P., Olino, T. M., … Barch, D. M. (2016). Revising the BIS/BAS Scale to study development: Measurement invariance and normative effects of age and sex from childhood through adulthood. *Psychological Assessment*, *28*(4), 429–442. https://doi.org/10.1037/pas0000186

Paus, T. (2010). Population neuroscience: Why and how. *Human Brain Mapping*, *31*(6), 891–903. https://doi.org/10.1002/hbm.21069

Paus, T., Keshavan, M., & Giedd, J. N. (2008). Why do many psychiatric disorders emerge during adolescence? *Nature Reviews Neuroscience*, *9*(12), 947–957. https://doi.org/10.1038/nrn2513

Pincus, H. A., Tew, J. D., & First, M. B. (2004). Psychiatric comorbidity: is more less? *World Psychiatry : Official Journal of the World Psychiatric Association (WPA)*, *3*(1), 18–23. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16633444%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1414654

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*, 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048

Power, J. D., & Petersen, S. E. (2013). Control-related systems in the human brain. *Current Opinion in*

Neurobiology, *23*(2), 223–228. https://doi.org/10.1016/j.conb.2012.12.009

Reineberg, A. E., Andrews-Hanna, J. R., Depue, B. E., Friedman, N. P., & Banich, M. T. (2015). Resting-state networks predict individual differences in common and specific aspects of executive function. *NeuroImage*, *104*, 69–78. https://doi.org/10.1016/j.neuroimage.2014.09.045

Romer, A. L., Elliott, M. L., Knodt, A. R., Sison, M. L., Ireland, D., Houts, R., … Hariri, A. R. (2021). Pervasively Thinner Neocortex as a Transdiagnostic Feature of General Psychopathology. *American Journal of Psychiatry*, *178*(2), 174–182. https://doi.org/10.1176/appi.ajp.2020.19090934

Romer, A. L., & Pizzagalli, D. A. (2021). Is executive dysfunction a risk marker or consequence of psychopathology? A test of executive function as a prospective predictor and outcome of general psychopathology in the adolescent brain cognitive development study®. *Developmental Cognitive Neuroscience*, *51*, 100994. https://doi.org/10.1016/j.dcn.2021.100994

Rosenberg, M. D., Finn, E. S., Scheinost, D., Papademetris, X., Shen, X., Constable, R. T., & Chun, M. M. (2015). A neuromarker of sustained attention from whole-brain functional connectivity. *Nature Neuroscience*, *19*(1), 165–171. https://doi.org/10.1038/nn.4179

Rosenberg, M. D., Martinez, S. A., Rapuano, K. M., Conley, M. I., Cohen, A. O., Cornejo, M. D., … Casey, B. J. (2020). Behavioral and Neural Signatures of Working Memory in Childhood. *The Journal of Neuroscience*, *40*(26), 5090–5104. https://doi.org/10.1523/JNEUROSCI.2841-19.2020

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., … Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience*, *27*(9), 2349–2356. https://doi.org/10.1523/JNEUROSCI.5587-06.2007

Sha, Z., Wager, T. D., Mechelli, A., & He, Y. (2019). Common Dysfunction of Large-Scale Neurocognitive Networks Across Psychiatric Disorders. *Biological Psychiatry*, *85*(5), 379–388. https://doi.org/https://doi.org/10.1016/j.biopsych.2018.11.011

Shanmugan, S., Wolf, D. H., Calkins, M. E., Moore, T. M., Ruparel, K., Hopson, R. D., … Satterthwaite, T. D. (2016). Common and dissociable mechanisms of executive system dysfunction across psychiatricdisorders in youth. *American Journal of Psychiatry*, *173*(5), 517–526. https://doi.org/10.1176/appi.ajp.2015.15060725

Sharma, A., Wolf, D. H., Ciric, R., Kable, J. W., Moore, T. M., Vandekar, S. N., … Satterthwaite, T. D. (2017). Common Dimensional Reward Deficits Across Mood and Psychotic Disorders: A Connectome-Wide Association Study. *The American Journal of Psychiatry*, *174*(7), 657–666. https://doi.org/10.1176/appi.ajp.2016.16070774

Shen, X., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., & Constable, R. T. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols*, *12*(3), 506–518. https://doi.org/10.1038/nprot.2016.178

Singh, I., & Rose, N. (2009). Biomarkers in psychiatry. *Nature*, *460*(7252), 202–207. https://doi.org/10.1038/460202a

Smith, S. M., Nichols, T. E., Vidaurre, D., Winkler, A. M., Behrens, T. E. J., Glasser, M. F., … Miller, K. L. (2015). A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature Neuroscience*, *18*(11), 1565–1567. https://doi.org/10.1038/nn.4125

Snyder, H. R., Miyake, A., & Hankin, B. L. (2015). Advancing understanding of executive function impairments and psychopathology: Bridging the gap between clinical and cognitive approaches. *Frontiers in Psychology*, *6*(MAR), 328. https://doi.org/10.3389/fpsyg.2015.00328

Solmi, M., Radua, J., Olivola, M., Croce, E., Soardo, L., Salazar de Pablo, G., … Fusar-Poli, P. (2021). Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. *Molecular Psychiatry*. https://doi.org/10.1038/s41380-021-01161-7

Sui, J., Jiang, R., Bustillo, J., & Calhoun, V. (2020). Neuroimaging-based Individualized Prediction of Cognition and Behavior for Mental Disorders and Health: Methods and Promises. *Biological Psychiatry*, *88*(11), 818–828. https://doi.org/10.1016/j.biopsych.2020.02.016

Thapar, A., & Riglin, L. (2020). The importance of a developmental perspective in Psychiatry: what do recent genetic-epidemiological findings show? *Molecular Psychiatry*, *25*(8), 1631–1639. https://doi.org/10.1038/s41380-020-0648-1

Uddin, L. Q. (2015). Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience*, *16*(1), 55–61. https://doi.org/10.1038/nrn3857

van Loo, H. M., & Romeijn, J.-W. (2015). Psychiatric comorbidity: fact or artifact? *Theoretical Medicine and Bioethics*, *36*(1), 41–60. https://doi.org/10.1007/s11017-015-9321-0

Volkow, N. D., Koob, G. F., Croyle, R. T., Bianchi, D. W., Gordon, J. A., Koroshetz, W. J., … Weiss, S. R. B. (2018). The conception of the ABCD study: From substance use to a broad NIH collaboration. *Developmental Cognitive Neuroscience*, *32*(October 2017), 4–7. https://doi.org/10.1016/j.dcn.2017.10.002

Wang, H.-T., Smallwood, J., Mourao-Miranda, J., Xia, C. H., Satterthwaite, T. D., Bassett, D. S., & Bzdok, D. (2018). Finding the needle in high-dimensional haystack: A tutorial on canonical correlation analysis. Retrieved from http://arxiv.org/abs/1812.02598

Wang, H. T., Smallwood, J., Mourao-Miranda, J., Xia, C. H., Satterthwaite, T. D., Bassett, D. S., & Bzdok, D. (2020). Finding the needle in a high-dimensional haystack: Canonical correlation analysis for neuroscientists. *NeuroImage*, *216*(March), 116745. https://doi.org/10.1016/j.neuroimage.2020.116745

Watts, A. L., Lane, S. P., Bonifay, W., Steinley, D., & Meyer, F. A. C. (2020). Building Theories on Top of, and Not Independent of, Statistical Models: The Case of the p-factor. *Psychological Inquiry*,

31(4), 310–320. https://doi.org/10.1080/1047840X.2020.1853476

Witten, D. M., Tibshirani, R., & Hastie, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, *10*(3), 515–534. https://doi.org/10.1093/biostatistics/kxp008

Wulfert, E., Block, J. A., Santa Ana, E., Rodriguez, M. L., & Colsman, M. (2002). Delay of gratification: Impulsive choices and problem behaviors in early and late adolescence. *Journal of Personality*, *70*(4), 533–552. https://doi.org/10.1111/1467-6494.05013

Xia, C. H., Ma, Z., Ciric, R., Gu, S., Betzel, R. F., Kaczkurkin, A. N., … Satterthwaite, T. D. (2018). Linked dimensions of psychopathology and connectivity in functional brain networks. *Nature Communications*, *9*(1), 1–14. https://doi.org/10.1038/s41467-018-05317-y

Yan, C. G., Cheung, B., Kelly, C., Colcombe, S., Craddock, R. C., Di Martino, A., … Milham, M. P. (2013). A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. *NeuroImage*, *76*, 183–201. https://doi.org/10.1016/j.neuroimage.2013.03.004

Youngstrom, E. A., Findling, R. L., Danielson, C. K., & Calabrese, J. R. (2001). Discriminative validity of parent report of hypomanic and depressive symptoms on the general behavior inventory. *Psychological Assessment*, *13*(2), 267–276. https://doi.org/10.1037/1040-3590.13.2.267

Youngstrom, E. A., Frazier, T. W., Demeter, C., Calabrese, J. R., & Findling, R. L. (2008). Developing a 10-Item Mania Scale from the Parent General Behavior Inventory for Children and Adolescents. *Journal of Clinical Psychiatry*, *69*(5), 831–839. https://doi.org/10.4088/JCP.v69n0517

Zapolski, T. C. B., Stairs, A. M., Settles, R. F., Combs, J. L., & Smith, G. T. (2010). The measurement of dispositions to rash action in children. *Assessment*, *17*(1), 116–125. https://doi.org/10.1177/1073191109351372

Acknowledgement

This research was supported by the Intramural Research Program of the National Institute on Drug Abuse, National Institutes of Health.

Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9-10 and follow them over 10 years into early adulthood. The ABCD Study is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators.

The ABCD data repository grows and changes over time. The ABCD data used in this report came from NDA Study 901. DOIs can be found at DOI 10.15154/1519007.

This work utilized the computational resources of the NIH HPC Biowulf cluster. (http://hpc.nih.gov)