Brain Functional Connectome Defines a Transdiagnostic Dimension Shared by Cognitive Dysfunction and Psychopathology in Preadolescents

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

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

[2]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, Baltimore, Maryland, USA.

* To whom correspondence may be addressed:

Yihong Yang, Ph.D.

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

Email: yihongyang@intra.nida.nih.gov

Keywords: Adolescence; Brain Connectome; Cognitive Function; Psychopathology; Transdiagnostic Factor

Abstract
The search for biomarkers of psychiatric disorders has remained elusive, in part, due to high comorbidity, low specificity, and poor concordance between neurobiological abnormalities and existing diagnostic categories. Developmental factors that have impacts on symptom expression and brain function further complicate biomarker identification. Growing evidence suggests that incorporating cognition into studies of psychopathology may be a path forward, as cognitive dysfunction is a common feature across psychiatric disorders. Recent neuroimaging advances have allowed for characterization of functional connectomes, the collective set of functional connectivity across the whole-brain, using resting-state fMRI. Functional connectomes may be useful psychiatric biomarkers as they have been shown to underlie individual differences in cognition and explain variance in psychopathology across individuals. In the present study, we sought to identify brain-based dimensions that are associated with general cognitive capacity and psychopathology using canonical correlation analysis in a sample of 7,383 preadolescents from the Adolescent Brain Cognitive Development study. Our analysis revealed patterns of functional connectivity correlated with cognitive control capacity and psychopathology. In particular, we identified a principal connectome-based latent brain variate that was positively correlated with cognitive measures across domains and negatively correlated with parent-reported psychopathology across diagnoses and domains. Functional connectivity loadings of the brain variate were across distributed cortical and subcortical brain networks and showed a dose-dependent relationship with the cumulative number of current psychiatric disorders in present. These findings provide preliminary evidence for a connectome-based biomarker that underlies individual differences in cognitive function and predicts transdiagnostic psychopathology in a dose-dependent fashion.

Significance Statement

Adolescence is a critical developmental window when most psychiatric disorders emerge, meanwhile comorbidity among disorders is prevalent. Using functional MRI and behavioral assessments from a large community-based sample of preadolescents aged 9-10, we identified a specific pattern of functional connectome (the regional activity synchronization across brain at rest) that showed directionally opposite associations with cognitive capacity and mental health problems. Specifically, the connectome pattern predicted individual differences in the performance on a range of cognitive tasks and the severity of parent-reported psychopathological problems. Notably, it predicted the comorbidity of current psychiatric disorders, pointing to a dose-response relationship. Our findings provide preliminary support that this connectome-based brain measure could represent a resiliency biomarker for protection against transdiagnostic psychopathology during preadolescence.

Main Text

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 for psychiatric conditions has limited progress in the field (1). Further complicating psychiatric nosology are issues of comorbidity and specificity (2). There is a high degree of comorbidity across psychiatric disorders and many symptoms are present across diagnostic categories (3). 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 (4–7). 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 (8). 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 (9, 10). These findings highlight the importance of taking age- and developmental-stage into account in biomarker identification. Additionally, they suggest that there is value in 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 (8).

Given the problems identifying biomarkers of specific psychiatric disorders, the field is shifting towards transdiagnostic and dimensional investigations, spearheaded by the NIMH's Research Domain Criteria (RDoC) (11). 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 (12). Recent neuroimaging techniques, such as resting-state fMRI, and resting-state functional connectivity (rsFC) analysis, enable noninvasive investigation of the system-level organization of brain circuits via the temporal synchrony between brain regions (13). The functional connectome, the collective set of functional connectivity in the brain, can reliably discriminate one brain from another like a fingerprint(14). and is thought to underlie individual differences in cognitive and affective functions (14, 15) or in expression or regulation of the psychopathological symptoms (16). Altered functional connectomes have been suggested to associate with a number of different psychiatric disorders (17).

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 incorporating cognition into studies of psychopathology may be a path forward. Cognitive dysfunction is one common feature across psychiatric disorders (18, 19). Studies of cognition suggest a hierarchical framework with both higher order processes (e.g., executive function and planning) and lower order processes (e.g., motor skills, perception, and memory), and an underlying, largely heritable, latent factor reflecting general cognition (g-factor) (20). Measures of general cognition and higher order cognitive processes have been shown to predict socio-occupational stability, academic success, and quality of life (20, 21). Additionally, recent research has identified genetic, epigenetic, developmental and environmental factors that collectively affect the configuration of brain networks and their efficiency in modulating cognition (22–24). Individual differences in cognition

may reflect variability in the connectivity of underlying neurocognitive brain networks, and predict psychopathology (25, 26). Relationships between cognition and psychopathology are complex and may be bidirectional (27). Problematically, to date, much of the research in this area has focused on cognitive functioning within specific psychiatric disorder categories (e.g., schizophrenia) (28) without examining relative differences in cognition across comorbid psychiatric disorders.

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 and conduct disorder) and internalizing (e.g., depression and anxiety) domains (29). 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 (30). 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 (30, 31). In fact, general cognition and psychopathology scores are commonly anticorrelated (27, 30). 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 (30).

In the present study, we seek to identify latent brain-behavior associations between the functional connectome and a broad set of behavioral assessments spanning the cognition and psychopathology domains in preadolescent youth. We analyzed data from the Adolescent Brian Cognitive Development (ABCD) study (32, 33), which includes brain imaging and comprehensive behavioral assessments from a large community-based sample of 9-10-year-old children in U.S., using canonical correlation analysis (CCA), a multi-view machine learning approach (34, 35). The CCA identifies latent components from two high-dimensional data sets that show maximized cross-set correlations (34, 35). As a novel tool for brain-behavior association analyses, CCA delineates whole-brain connectivity patterns associated with a set of behavior assessments without *a prior* assumptions (36–38). In this study, a single functional connectivity pattern was identified that showed a significant and generalizable association with the behavioral assessments, positively correlating with cognitive functions while negatively correlating with psychopathological measures in a transdiagnostic manner. Furthermore, the rsFC pattern showed a dose-dependent relationship to the cumulative number of psychiatric diagnoses present.

Results

The overarching goal of the study was to characterize latent sources of brain-behavior associations linking individual differences in functional brain connectome and multiple behavioral assessments spanning cognitive functions and psychopathology related constructs in the preadolescent population.

Brain and behavioral data used in the current analysis were from the baseline assessment and first fMRI scan visit of the ABCD study collected in a nationally-representative community sample of preadolescent participants (32, 39). From a total of 11,875 participants, 7,382 (3,714 females, aged 9.95 ± 0.62 y/o) were included in the current analysis (see Figure S1 for exclusion criteria, and Table S1 for demographic information of the included participants). Individual functional connectomes were constructed from 20-min resting-state fMRI data using resting-state functional connectivity (rsFC) between 352 regions-of-interest(40) (ROIs) across the brain including 333 cortical areas and 19 subcortical areas. The behavioral data included 20 assessments of cognitive function (Table S2) and 31 dimensional assessments of psychopathology-related constructs (Table S3).

The multivariate statistical method of CCA (41) was used to identify latent sources of the brain-behavior associations (Figure 1). Before submitted to CCA, covariates related to the data acquisition (including scanner type, head motion, and number of frames after motion censoring) were regressed from both the rsFC and the behavioral datasets. Dimension of the two data sets were reduced using principal component analysis (PCA). For the behavioral set, 49 dimensions explaining 100% of the variance were kept, and for the rsFC set, the number of reduced dimensions were tuned as a hyperparameter (See method for details).

The whole CCA procedure described above was performed in a 5-fold cross-validation framework (Figure S2, Figure S3). In each fold, the CCA model was generated from a training set (5906 participants, 80% of the total samples used in the analysis) with a nested 5-fold cross-validation for tuning the PCA-based dimension reduction, and subsequently tested for its generalizability on the test set (1476 participants, 20% of the total samples used in the analysis). The optimal number of PC varied from 650 to 1050 among the 5 outer folds as listed in Table S4.

Determined by the rank of the behavioral set, 49 modes of associated canonical variates (CVs) were identified between the functional connectome and the multi-dimensional behavioral assessments. Among them, 5 modes of connectome-behavior associations showed both significant canonical correlation and out-of-sample generalizability. The first CCA mode further showed its unique features compared to the other 4 modes in that it extracted a substantially large proportion of the population variance (characterized by normalized redundant variance(42, 43) and was the only mode that exceeded the null confidence interval of 0.01 based on a permutation test. As shown in Table S4, the first mode accounted for the dominant proportion of the population variance in the connectome set (26.32%~28.34% variance across folds for mode 1 vs. 4.57%~6.78% variance across folds, for the largest of any other mode) and in the behavioral set (23.84%~29.22% variance for mode 1 vs. 5.50%~8.05% variance, for any other modes). A substantial proportion of the extracted population variance in a CCA mode would facilitate its cross-set prediction (42, 43). In the following, we present results of the first (or principal) mode in the main text and those of other 4 modes (modes 2-5) are detailed in Table S4, Figure S4 and S5.

The Principal Mode of Connectome-Behavior Association

As shown in a typical fold of the cross validation (Figure 2), the canonical correlation of the principal mode exhibits a large effect size ($\rho = 0.72$, family-wise error rate [fwer] adjusted $p_{fwer} < 10^{-3}$) in the training set and can be generalized to the hold-out test set ($\rho = 0.56$, $p_{fwer} < 10^{-3}$). See Table S4 for the result of other folds.

To understand the behavioral relevance of the principal CCA mode, we examined its behavioral loading. Behavioral loading, defined as the correlation between the behavioral CV to each behavioral assessment, reflects the variance of each behavioral assessment that is extracted by the latent behavioral CV of the CCA (Figure 3A). For the principal mode, nearly all assessments related to cognitive ability show positive loadings on the latent behavioral CV, whereas most of the psychopathology-related constructs show negative loadings. Furthermore, differences were observed across externalizing and internalizing domains of psychopathology. Measures indexing externalizing 'spectrum' problems generally exhibited (e.g., $r = -0.27$ for CBCL Rule Breaking and $r = -0.25$ for CBCL Conduct Problem), compared to measures indexing internalizing 'spectrum' problems (e.g., $r = -0.008$ for CBCL Anxious-Depression and $r = -0.06$ for CBCL Withdrawn-Depression) exhibits lower loadings .

Figure 3B shows the FC loading, defined as correlation between each FC and the connectome CV of the principal mode. A positive loading indicates that the rsFC of the ROI pair contributed positively to the association in which higher brain connectome variate was correlated with better performance in cognitive tasks and lower severity of parent-reported psychopathological symptoms. A measurement of ROI loading was defined by summing the rsFC loadings of that ROI (i.e., each row in the loading matrix in Figure 3B), indicating the overall contribution of the

rsFC associated with the given ROI to the identified brain connectome CV. Inequality of ROI loadings was tested for the positive and negative loading matrices, separately, and both negative and positive loadings distributed unequally across ROIs (see Figure S6). Accordingly, ROIs across the brain formed two clusters, each being dominated with significant positive or negative rsFC loadings (Figure 3C), compared to the null distribution of ROI loading generated by permutating the loading matrix of Figure 3B.

ROIs with significant positive loadings include cortical areas of the anterior insula/frontal operculum (aI/fO), dorsal anterior cingulate cortex (dACC), ventral lateral prefrontal cortex (vlPFC), dorsal medial prefrontal cortex (dmPFC), superior temporal gyrus (STG), as well as subcortical regions of the caudate, putamen, accumbens, thalamus and brain stem. ROIs with significant negative loadings encompass cortical regions of the intra-parietal sulcus and superior parietal lobule (IPS/SPL), frontal eye field (FEF), as well as the pre-motor area (PMA), and associative visual areas in the occipital lobe (Figure 3C).

Furthermore, the loading of the ROIs significantly differed by their affiliation to large-scale functional networks (Figure 3D; one-way ANOVA, $F(7,344) = 42.58$, $p < 0.001$). Specifically, negative loadings appeared mainly in the visual (VIS) and the dorsal attention network (DAN), while the largest positive loadings were mainly presented in the salience/ventral attention network (SAL/VAN), default mode network (DMN), and subcortical network (SBC).

In summary, the principal CCA mode revealed a specific pattern of rsFC, that is associated with behaviors spanning across cognition and psychopathology related constructs. FC loadings revealed that a cortico-subcortical system, mainly encompassing frontal and parietal cortex and subcortical regions contributed the most to the connectome CV. Associations between the rsFC component and the multi-dimensional behavior set showed an interesting pattern, with the corresponding behavioral CV positively correlating with cognitive measures and negatively correlating with psychopathological measures.

Cross-domain Consistency of the Principal Connectome-Behavior Association Mode

As revealed by the principal CCA mode, the single rsFC pattern covaried with behavioral CV that correlated positively with cognitive functions and negatively with psychopathology related constructs. Such a transdiagnostic association implies that deficit of cognitive functions and severity of psychopathological problems may proxy the functioning of a common neural process. Or alternatively, such an association may be driven by only one domain of the behaviors. For example, the principal CCA mode may simply reflect the association between cognitive functions and rsFC. And in such a case, psychopathological measures may still show negative loadings due to the negative correlation between p- and g- factors.

To exclude the alternative hypothesis, we tested the cross-domain consistency of the CCA modes. Specifically, the same CCA procedure described above was conducted, but with behavioral assessments from only a single domain of cognition or psychopathology at a time. In both cases, the principal mode survived from both tests of within-sample significance and out-of-sample generalizability. Though the values of canonical correlation differ between the CCA models estimated from cognition, psychopathology (Figure S7) or their combination (Figure 2), the connectome CV showed very high consistency among the three cases (combined vs. cognition only, $r = 0.99$; combined vs. psychopathology only, $r = 0.95$; Figure 4). Note that cross-domain consistency with perfect correlation ($|r| > 0.9$) was found to be specific to the principal mode (see Table S5 for other significant modes).

These results confirm the non-trivial transdiagnostic nature of the connectome variate of the principal mode and suggest that it may act as the core neural substrate shared by the cognitive impairment and psychopathology in youth population.

Connectome Variate Predictive of Multi-domain Behavioral Assessments
Our CCA identified a robust connectome CV associated with a latent mixture of multi-domain behavioral assessments. However, interpreting the clinical relevance of brain-behavior associations in the latent space is difficult. We further assessed the utility of the identified connectome CV for predicting individual behavioral assessments in their original feature space (See Method for details).

As shown in Figure 5, by projecting onto the principal connectome-behavior association mode, rsFC alone is capable of predicting a wide range of univariate behavior assessments across the domains of cognition and psychopathology. Specifically, all assessments in the cognitive domain, except for the cash choice task and stop-signal reaction time, are significantly predicted (FDR-adjusted $p < 0.05$).

The significantly predicted psychopathology related constructs (FDR-adjusted $p < 0.05$) crossed the externalizing problems, e.g., CBCL scales of Rule-breaking Behavior and Conduct Problems; other CBCL syndrome of Social Problem and Oppositional Defiant Problem, ADHD; internalizing problems, e.g. Stress, Anxiety and Somatic Problems; impulsivity related scales of Positive Urgency, Negative Urgency, Reward Responses, and Fun Seeking; the scale of Mania, and the scale of Psychosis Severity. While the analysis showed that the connectome CV predicted psychopathology across domains, an externalizing 'spectrum' predominance was observed compared to the internalizing 'spectrum'.

In line with the cross-domain consistency of the principal connectome CV, this prediction of behaviors holds for the connectome CV identified with CCA applied on single behavioral domain of cognition or psychopathology-related construct (Figures S7- S9).

Connectome Variate Predictive of Clinical Diagnoses
To validate the connectome variate of the principal mode as a transdiagnostic factor, we assessed the association between the predicted connectome variate score with clinical diagnoses.

When all the 7382 participants were stratified into clinical diagnostic groups, a common pattern was identified whereby preadolescents with current psychiatric disorders had lower connectome CV scores compared to those without any current diagnosis (Figure 6A).). In examining diagnosis-specific associations, our results showed that groups with Conduct Disorder (childhood and adolescent onset), Separation Anxiety, Unspecified Bipolar and Related Disorder, ADHD, Social Anxiety Disorder, Obsessive Compulsive Disorder, Oppositional Defiant Disorder reached statistical significance (FDR-adjustment $p < 0.05$).

In analyses examining the impact of severity of psychopathology, a main effect of number of current psychiatric disorders on the connectome variate score was found (One-way ANOVA, $F(3,7378) = 4.62$, $p < 0.01$; Figure. 6B). Post-hoc tests revealed that preadolescents with 2 and $\geqslant$3 psychiatric disorders showed significantly lower connectome variate scores compared to the children with no diagnoses (2 vs. 0: $t(378) = 2.68$, FDR-adjusted $p = 0.04$; $\geqslant$3 vs. 0 $t(259) = 5.21$, FDR-adjusted $p < 0.001$), and children with $\geqslant$3 diagnoses showed a significant lower connectome variate score compared to the children with 1 or 2 diagnoses ($\geqslant$3 vs. 2: $t(474) = 2.65$, FDR-adjusted $p < 0.05$; $\geqslant$3 vs. 1: $t(474) = 3.83$, FDR-adjusted $p < 0.001$).

Moreover, participants with low connectome CV scores (≤1 SD below MEAN) had 1.32 times the risk of having one or more current psychiatric disorders compared to those with high connectome variate scores ($\geqslant$1 SD above MEAN) (Risk Ratio = 1.32, CI:1.18-1.48, $p < 0.001$). The contingency table is shown in Table S6.

Again, above associations between clinical diagnoses and the connectome CV score holds, when CCA was performed associating the connectome set with behaviors within single domain of cognition assessments (Figure S10).

Heritability of the Connectome Canonical Variate

It has been found children's cognitive ability and psychopathology are both heritable (44–46). And the anticorrelation between the two domains of behavior are genetically rooted (47). A possibility is that these two domains of behaviors are driven by a shared brain system which is heritable per se. We hypothesize this can be measured with the principal connectome CV.

We tested the heritability of the principal mode connectome CV using the twin data from the ABCD study. Heritability of the connectome CV was estimated using classic twin design and structural equation modeling (66). From the 7,832 participants, 38 pairs of monozygotic (MZ) twins and 62 pairs of dizygotic (DZ) twins were included in the analysis.

Intrapair correlation for the connectome variate score were $r_{MZ} = 0.71$ for MZ twins and $r_{DZ} = 0.39$ for DZ twins. Mean value and variances of rsFC variate score did not differ significantly between MZ and DZ twins. Because $0.5 \times r_{MZ} < r_{DZ}$, a full Cholesky ACE (**A**dditive genetic, **C**ommon environmental, and Unique **E**nvironmental) model and its sub-models were tested, which attributes the phenotypical variance of the connectome variate into three sources, additive genetic, common environmental, and non-shared environmental. The ACE model can be reduced to an AE mode ($\Delta\chi^2 = 0.02$, df = 1, $p = 0.99$, ACE vs. AE), indicating a non-significant influence of the common environmental factor. However, the non-genetic model E model was rejected ($\Delta\chi^2 = 36.93$, df = 1, $p < 10^{-4}$, AE vs. E), indicating the significant influence of additive genetic factor. Finally, the result indicated a heritability of the rsFC variate: $h^2 = 74.42\%$ (95% CI: 56.76%-85.42%).

Impact of Sociodemographic Factors

Finally, we conducted sensitivity tests to evaluate how the principal connectome-behavior association was affected by sociodemographic factors. We focused on the key sociodemographic variables(48), including sex, race/ethnicity, household income, and parent education. We regressed out these sociodemographic variables as covariates one at a time, in addition to the Tier-1 nuisances, i.e. scanner type, mean frame displacement and remaining number of frames after censoring. As shown in Figure S11, the principal connectome-behavior association was reproduced with similar connectome variate loading (r = 0.86~0.99), behavioral loading (r = 0.98~0.99) and generalized canonical correlation (ρ = 0.48~0.58).

Discussion

Applying CCA to a multidimensional dataset that included comprehensive measurements of rsFC, cognition, and psychopathology from a large community-based sample of U.S. preadolescents, we identified a single connectome-based latent brain variate that covaried with a broad number of cognitive and psychopathological measurements. The connectome CV had rsFC loadings distributed in attention and cognitive control networks and was positively correlated with cognitive performance across domains and negatively correlated with parent-reported psychopathology measures across dimensions and diagnoses. In a held-out sample of unseen participants, the brain variate predicted a range of behavioral measures including cumulative number of current psychiatric disorders, which showed a dose-response relationship. Together, these findings provide preliminary evidence for a connectome-based biomarker that indexes individual differences in cognitive performance and transdiagnostic vulnerability to psychopathology across multiple psychiatric disorders.

Identification of psychiatric biomarkers is essential for both understanding the ontology of psychiatric disorders and for developing precise treatments for them. Traditional efforts to identify psychiatric biomarkers have relied predominantly on case-control designs and focused on neural

alterations within a single disease. However, few clinically useful disorder-specific brain-behavior associations have been discovered using these approaches. This has led to a shift in approach towards identifying core brain-behavior associations that cross traditional diagnostic boundaries as promoted by the RDoC initiative (12, 49). Neuroimaging studies using the transdiagnostic framework have demonstrated more consistent findings showing overlapping neural alteration deficits, including lower gray matter volume (6) and impaired activation during cognitive tasks (50), across diagnostic categories. Consistent with these findings, results of our current study identified the altered FC across large-scale brain networks, suggestive of maladaptive connectomes, that are associated with a broad spectrum of psychopathology and cognitive processes, and may represent a transdiagnostic neural marker for psychiatric disorders. However, while the RDoC approach seeks associations with specific cognitive and affective functions across diagnoses, we found a single CV that related to most aspects of both cognition and psychopathology, irrespective of diagnosis. In addition, given the ABCD sample, our findings extended the developmental window for observing the transdiagnostic neural deficits from adulthood down to middle childhood, a critical period for the onset of mental disorders. As such, our connectome-based variate may provide unique predictive or prognostic value related to the emergence and progression of psychiatric disorders from preadolescence to young adulthood, providing a roadmap for developing brain-based treatment targets.

Our findings parallel prior epidemiologic research showing that general cognition and psychopathology factors are anticorrelated (31, 51). Furthermore, our results extend this work by showing that measures of cognition and psychopathology that have traditionally been viewed as separate constructs, load on the positive and negative ends of the same dimensional factor (rather than orthogonal ones) and covary with the same brain variate. It is particularly notable that we derived essentially the same brain variate when using the cognitive or behavioral features in isolation as when using them in combination, suggesting the brain variate captures the essential shared neural process underlying the two domains of behavior. The general/non-specific association between our connectome-based brain variate and wide array of psychopathology measures and psychiatric diagnoses, suggests that this brain signal indexes transdiagnostic rather than disorder-specific vulnerability in preadolescents. That the brain variate exhibits a dose-response relationship with cumulative number of current psychiatric disorders suggests potential clinical applicability as a diagnostic-severity staging biomarker, and warrants further exploration. While our findings require cautious interpretation until out of sample validation and prospective testing is performed, they provide early support for the latent connectome-based brain variate being a candidate biomarker indexing individual differences in risk for transdiagnostic psychopathology and cognition in preadolescents.

In addition to its efficiency in explaining individual variance shared by the multiple behaviors, our identified brain-based variate mirrors the general behavior-derived factors like g- (44, 47, 52) and p- (45, 46, 53) in their heritability. Note that the general behavioral factors are genetically associated (47). In the current twin data, the connectome CV exhibited a high heritability, which is significant and with a comparable extent to the general cognitive function and externalizing problems. This again confirms that the identified connectome CV may reflect the same genetically rooted individual difference that has been commonly observed with p- and g-. It suggests that the specific pattern of rsFC may participant in the active gene-environment correlation (rGE) process as a mediator (54), by facilitating youth's capability to select, modify and create their own experiences , especially in interaction with negative environment during the development.

Our rsFC loadings identified diverse cortical and subcortical brain regions including the aIns/fO, dlPFC, vmPFC, STG, IPS/SPL, FEF, PMA, striatum, thalamus and brain stem that contributed significant variance to the connectome-based brain variate. Notably, many of these cortical regions map onto a frontoparietal brain system which includes the VAN/SAL and DAN, which are implicated in attention-based cognitive control.(55–59) As such, proper functioning of this network may be required to support healthy cognitive performance and protect against the development of

psychiatric disorders,(16) and alterations of rsFC in these ROIs may result in both cognitive dysfunction and the emergence of psychopathology. Interestingly, we found that ROIs of significant loadings form two clusters according to the direction of their loadings. Specifically, negative loadings were mainly located in the DAN while positive ones in the VAN/SAL, which may be aligned with the separated functions of these two subsystems in supporting the attention-based cognitive control.(56, 60) Subcortical regions including the striatum, thalamus, and brain stem also had significant positive loadings. This may indicate the critical role of these regions in supporting cognition by relaying, shifting, and sustaining functional information from their cortical counterparts.(61)

This study has some noteworthy limitations. As the study is a secondary analysis of ABCD data, we were reliant on the assessments, data collection procedures, imaging parameters, and acquisition and harmonization choices made by the primary study team. While broad psychopathology indices were collected, they were restricted to parent-report measures at this timepoint which is limiting and should be cross-validated for accuracy with multi-informant assessments from subsequent data waves. The ABCD data used in the present study are cross-sectional and when combined with our correlational analytic approach precludes causal interpretation of our findings. However, given the longitudinal design of the ABCD study, we will be able to test the ability of the candidate biomarker identified in this study to prospectively predict future psychopathology by incorporating later waves of ABCD data in subsequent analyses. It is also important to point out that our data-driven analytic approach using CCA was designed to detect linear associations between the two multi-dimensional datasets of functional connectome and behavioral assessments. It was unexpected that one mode accounted for the majority of variance across both cognitive and psychopathological measures in the study. The four other modes, while accounting for much less variance than our primary mode, did show some distinct brain-behavior associations with specific cognitive and psychopathological measures that warrant further investigation. Although psychopathological scales showed generally negative correlations with the connectome variate, scales within the internalizing spectrum had lower extent than those in the externalizing spectrum. This suggests that association of psychopathology with functional connectome might not be fully captured by the current linear model of CCA, especially for those in the internalizing spectrum. Non-linear models such as kernel CCA(34) might be valuable for identifying other linkages between functional connectome and the multi-domain behaviors, which would be important for finding biotypes within specific clinical categories or cognitive domains.

Materials and Methods

Participants

Neuroimaging data and behavioral assessments of 11,875 children aged 9- to 10-years were obtained from the ABCD study (32). The large and long-term ongoing project aimed to characterize psychological and neurobiological development from pre-adolescence to young adulthood. Participants and their families were recruited through school and community settings in 21 centers across the US, following locally and centrally approved Institutional Review Board procedures as detailed in Garavan et al., 2018 (62). Participants were excluded due to missing data, having a neurological condition or poor data quality. Exclusion criteria are detailed in Figure S1.

Assessments of Multi-domain Behavior

The behavioral dataset comprised assessments spanning two domains including cognitive functions and psychopathology-related constructs. The cognitive ability of the participants was quantified with 20 scores derived from their performance on the ABCD 15-test neuro-cognitive battery (63) and behavior from the 3 neuroimaging tasks(39). Cognitive tests and measures are detailed in supplementary materials. Psychopathology-related constructs were measured with 31 dimensional scores including the parent-reported Child Behavioral Checklist (CBCL), scale of

mania, youth reported scales of impulsivity, and psychosis risk. Measurements of the psychopathology-related constructs are detailed in Supplementary Materials.

Clinical Diagnoses of Psychiatric Disorders
The clinical diagnostic assessment was measured with the computerized Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS) for DSM-5 (KSADS-5) (64, 65).

Resting-State Functional Connectome
In the ABCD Study, four 5-min resting-state functional image series and a T1-weighted structural image were collected for each participant with 3T scanners. Scanning parameters slightly differed across scanning sites and are detailed in Supplementary Materials.

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 pipeline included Human Connectome Project (HCP)'s minimal preprocessing pipeline(66) and the DCAN BOLD Processing (DBP) software (67). Processed data were obtained from collection 3165 provided by DCAN Labs. Descriptions of the pre-processing and access of data are detailed in Supplementary Materials.

The functional connectomes of individual brains were constructed from resting-state functional connectivity between 352 regions-of-interest (ROIs) across the brain. The ROIs were defined by the parcellation scheme from Gordon et al 2016 (40), which included 333 cortical areas and 19 subcortical areas. The ROIs were originally assigned into 12 functional communities (40) and further assembled into 7 large-scale networks according to the guideline by Uddin et al. (68). Functional connectivity was indexed with z-transformed Pearson's correlation of pre-processed BOLD time series between ROIs. From the 352-by-352 matrix, the 61,776 upper-triangular values were used to describe the individual's functional connectome.

Covariates
In the current study, we considered two tiers of covariates that may affect the CCA analysis (48). The Tier-1 covariates included the scanner type, head motion measured with mean framewise displacement (mean FD), and the number of retained frames after motion censoring. These Tier-1 covariates could induce artificial rsFC values, and the head motion during scanning could be associated with behavioral measures. Therefore, they might bias the CCA results with a spurious brain-behavior association. To eliminate such potential effect, we regressed out Tier-1 covariates from both brain and behavioral data in all our CCA analyses.

The Tier-2 covariates included 5 major sociodemographic variables, i.e., sex, age, race/ethnicity, household income, and parental education. These variables may have effects on the brain-behavior association, however a causal mechanism is unclear. These variables might also proxy for both confounding factors and mediators or colliders simultaneously (48). Therefore, we conducted our original analyses without considering these variables and then conducted subsequent CCA analyses regressing out the Tier-2 covariates one at a time in addition to Tier-1 covariates, in order to investigate the impact of these Tier-2 covariates on the identified brain-behavior association.

Discovery of Brain-Behavior Associated Dimensions with CCA
To identify associations between the functional connectome of the youth brain and multi-domain behavioral assessments, we conducted a CCA (41) on the two datasets. To ensure the generalizability of the multi-variable result, the CCA was conducted in a hold-out framework(69). The 7,382 participants were randomly split into a training set of 5906 subjects (80% of the total samples) for discovery and a test set of 1476 subjects (20% of the total samples) for validation. Within the training set, a nested 5-fold hold out validation (80% vs. 20% for each fold) was performed to tune the hyperparameter, the number of principle components retained after the dimension reduction procedure (see more details below).

Before submitting the brain features into the CCA, two preprocessing steps were performed at the group level. First, Tier-1 covariates were regressed out from the connectome and behavior dataset separately, and the residuals were transformed into a lower dimensional basis where data are exchangeable (70, 71). Second, a procedure of Principal Component Analysis (PCA) was applied on both datasets of functional connectome and the multi-domain behaviors, respectively. Note that CCA is invariant to linear transformations (70). Therefore, PCA does not change the result of CCA applied on original feature space. However, removing the low-ranked PCs prevents small perturbations in the original data from causing instability in the CCA solutions and therefore leading to overfitting. Though the PCA-based dimension reduction is a common practice used for CCA, the number of PCs to retain has not been fully addressed. Therefore, we tuned this hyperparameter for a highest generalizability of the CCA results through a nested 5-fold cross-validation procedure within the training set. Note that, in the cross-validation, PCA was only performed on the training dataset, and then the resultant coefficients were applied to the test dataset.

CCA simultaneously identifies orthogonal latent variates from the brain and behavior datasets, while ensuring a maximized correlation between the two variates paired by their orders (modes). To identify meaningful brain-behavior associations, we tested these modes for their statistical significance (70, 71), and generalizability (69), and redundancy index (36, 42) via permutation tests (see supplementary materials for more details).

Association Between Connectome Variate and Behavioral Assessments

To assess the association between the connectome variate and individual differences in cognitive performance and psychopathological measures, we conducted a connectome-based prediction. The prediction was performed in a hold-out manner. Coefficients of the first CCA mode estimated from the training set, produced by multiplying the weighting matrices given by PCA and CCA, were applied to the connectome data in the test set. Correlations between the predicted connectome variate score to each behavioral assessment were calculated and tested with Pearson's correlation. P values were corrected for multiple comparisons using the Benjamini-Hochberg false discovery rate (FDR) method (72).

Association Between Connectome Variate and Clinical Diagnoses

To validate the connectome variate as a transdiagnostic factor, the association between the connectome variate and clinical diagnoses was assessed. First, all the 7382 participants with psychiatric disorders were grouped by their K-DSADS diagnoses, and the connectome variate score of each diseased group was compared with the group with no current psychiatric diagnosis. Second, participants with psychiatric disorders were grouped by 0, 1, 2, or ≥3 comorbid diagnoses. In these two analyses, groups with insufficient size (<20) were excluded. Welch two-sample t-tests(73) were used to compare differences between groups. Finally, the risk of at least one diagnosis was compared between participants with rsFC score above MEAN+1SD vs. under MEAN-1SD in the whole cohort.

Heritability of Connectome Variate

To estimate the heritability of the connectome variate of the principal mode, we fitted linear structural equation model with a classic twin study design (CTS) (74) using the R openMx package (75). In CTS, the information of MZ and DZ twin pairs were used to disentangle the influence of genetic and environmental factors on phenotypical trait, i.e., the connectome variate score in the current case. The total phenotypical variance, P, was attributed into a sum of genetic and environmental factors (P = A + C + D + E): additive genetic influences (A), non-additive genetic influences (dominance, D), environmental influence shared by family members (common environmental variance, C) and unique environmental influence (E). The CTS assumed that (1) MZ and DZ twin pairs shared their environments to the same extent; (2) Random mating and (3) Minimal gene-by-environmental interaction. Therefore, MZ and DZ twins had the same degree of C and E, but varied degree of A and D, and such varied structures allow the different variance components to be estimated via a path analysis (74). Though C and D were included in the CTS,

they cannot be estimated simultaneously. In the current case, ACE was selected over ADE because $r_{DZ} > 0.5^* r_{MZ}$.

Goodness of fit (GOF) was tested for the full ACE model and its sub-models (models restricting one or more path to be zero). Significance of each path was assessed by comparing the fit of the restricted model to the full model. Chi-square test was used for statistical assessment. Relatively, Akaike's information criterion (AIC), AIC = χ^2 -2df was used to determine the best fit model considering both the GOF and the parsimony of the model.

Finally, heritability (h^2) was estimated as the ratio of the genetics related variance to the total phenotypical variance. In the current case of AE model, $h^2 = (a^2)/(a^2+e^2)$.

Acknowledgments

This research was supported by the Intramural Research Program of the National Institute on Drug Abuse, National Institutes of Health. Grant number: 1-ZIADA000469.

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 aged 9-10 years 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).

References

1. I. Singh, N. Rose, Biomarkers in psychiatry. *Nature* **460**, 202–207 (2009).

2. H. A. Pincus, J. D. Tew, M. B. First, Psychiatric comorbidity: is more less? *World Psychiatry* **3**, 18–23 (2004).

3. H. M. van Loo, J.-W. Romeijn, Psychiatric comorbidity: fact or artifact? *Theor. Med. Bioeth.* **36**, 41–60 (2015).

4. S. H. Lee, *et al.*, Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).

5. L. M. McTeague, M. S. Goodkind, A. Etkin, Transdiagnostic impairment of cognitive control in mental illness. *J. Psychiatr. Res.* **83**, 37–46 (2016).

6. M. Goodkind, *et al.*, Identification of a common neurobiological substrate for mental Illness. *JAMA Psychiatry* **72**, 305–315 (2015).

7. A. Sharma, *et al.*, Common Dimensional Reward Deficits Across Mood and Psychotic Disorders: A Connectome-Wide Association Study. *Am. J. Psychiatry* **174**, 657–666 (2017).

8. M. Solmi, *et al.*, Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. *Mol. Psychiatry* (2021) https:/doi.org/10.1038/s41380-021-01161-7.

9. T. Paus, M. Keshavan, J. N. Giedd, Why do many psychiatric disorders emerge during adolescence? *Nat. Rev. Neurosci.* **9**, 947–957 (2008).

10. N. Gogtay, *et al.*, Dynamic mapping of human cortical development during childhood through early adulthood. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8174–8179 (2004).

11. L. M. Mcteague, Reconciling RDoC and DSM approaches in clinical psychophysiology and neuroscience. *Psychophysiology* **53**, 323–327 (2016).

12. B. N. Cuthbert, T. R. Insel, Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Med.* **11**, 126 (2013).

13. B. Biswal, Y. FZ, H. VM, H. JS, - Functional connectivity in the motor cortex of resting human brain using. *Magn Reson Med* **34**, 537–541 (1995).

14. E. S. Finn, *et al.*, Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* **18**, 1664–1671 (2015).

15. M. P. Van Den Heuvel, C. J. Stam, R. S. Kahn, H. E. Hulshoff Pol, Efficiency of functional brain networks and intellectual performance. *J. Neurosci.* **29**, 7619–7624 (2009).

16. M. W. Cole, G. Repovš, A. Anticevic, The frontoparietal control system: A central role in mental health. *Neuroscientist* **20**, 652–664 (2014).

17. J. W. Buckholtz, A. Meyer-Lindenberg, Psychopathology and the Human Connectome: Toward a Transdiagnostic Model of Risk For Mental Illness. *Neuron* **74**, 990–1004 (2012).

18. M. J. Millan, *et al.*, Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. *Nat. Rev. Drug Discov.* **11**, 141–168 (2012).

19. C. T. Gualtieri, D. W. Morgan, The frequency of cognitive impairment in patients with anxiety, depression, and bipolar disorder: an unaccounted source of variance in clinical trials. *J. Clin. Psychiatry* **69**, 1122–1130 (2008).

20. A. K. Barbey, Network Neuroscience Theory of Human Intelligence. *Trends Cogn. Sci.* **22**, 8–20 (2018).

21. H. R. Snyder, A. Miyake, B. L. Hankin, Advancing understanding of executive function impairments and psychopathology: Bridging the gap between clinical and cognitive approaches. *Front. Psychol.* **6**, 328 (2015).

22. N. P. Friedman, A. Miyake, Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex.* **86**, 186–204 (2017).

23. N. P. Friedman, *et al.*, Stability and change in executive function abilities from late adolescence to early adulthood: A longitudinal twin study. *Dev. Psychol.* **52**, 326–340 (2016).

24. Z. Sha, T. D. Wager, A. Mechelli, Y. He, Common Dysfunction of Large-Scale Neurocognitive Networks Across Psychiatric Disorders. *Biol. Psychiatry* **85**, 379–388 (2019).

25. A. E. Reineberg, J. R. Andrews-Hanna, B. E. Depue, N. P. Friedman, M. T. Banich, Resting-state networks predict individual differences in common and specific aspects of executive function. *Neuroimage* **104**, 69–78 (2015).

26. S. Shanmugan, *et al.*, Common and dissociable mechanisms of executive system dysfunction across psychiatricdisorders in youth. *Am. J. Psychiatry* **173**, 517–526 (2016).

27. A. L. Romer, D. A. Pizzagalli, 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®. *Dev. Cogn. Neurosci.* **51**, 100994 (2021).

28. T. A. Lesh, T. A. Niendam, M. J. Minzenberg, C. S. Carter, Cognitive Control Deficits in Schizophrenia: Mechanisms and Meaning. *Neuropsychopharmacology* **36**, 316–338 (2011).

29. H. Kim, N. R. Eaton, The hierarchical structure of common mental disorders: Connecting multiple levels of comorbidity, bifactor models, and predictive validity. *J. Abnorm. Psychol.* **124**, 1064–1078 (2015).

30. A. Caspi, *et al.*, The p factor: One general psychopathology factor in the structure of psychiatric disorders? *Clin. Psychol. Sci.* **2**, 119–137 (2014).

31. A. Caspi, T. E. Moffitt, All for one and one for all: Mental disorders in one dimension. *Am. J. Psychiatry* **175**, 831–844 (2018).

32. N. D. Volkow, *et al.*, The conception of the ABCD study: From substance use to a broad NIH collaboration. *Dev. Cogn. Neurosci.* **32**, 4–7 (2018).

33. N. R. Karcher, D. M. Barch, The ABCD study: understanding the development of risk for mental and physical health outcomes. *Neuropsychopharmacology* **46**, 1–12 (2021).

34. H. T. Wang, *et al.*, Finding the needle in a high-dimensional haystack: Canonical correlation analysis for neuroscientists. *Neuroimage* **216**, 116745 (2020).

35. D. M. Witten, R. Tibshirani, T. Hastie, A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* **10**, 515–534 (2009).

36. S. M. Smith, *et al.*, A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nat. Neurosci.* **18**, 1565–1567 (2015).

37. V. Kebets, *et al.*, Somatosensory-Motor Dysconnectivity Spans Multiple Transdiagnostic Dimensions of Psychopathology. *Biol. Psychiatry* **86**, 779–791 (2019).

38. C. H. Xia, *et al.*, Linked dimensions of psychopathology and connectivity in functional brain networks. *Nat. Commun.* **9**, 1–14 (2018).

39. B. J. Casey, *et al.*, The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Dev. Cogn. Neurosci.* **32**, 43–54 (2018).

40. E. M. Gordon, *et al.*, Generation and Evaluation of a Cortical Area Parcellation from

Resting-State Correlations. *Cereb. Cortex* **26**, 288–303 (2016).

41. H. Hotelling, Relations Between Two Sets of Variates. *Biometrika* **28**, 321 (1936).

42. D. Stewart, W. Love, A general canonical correlation index. *Psychol. Bull.* **70**, 160–163 (1968).

43. K. E. Muller, Relationships between redundancy analysis, canonical correlation, and multivariate regression. *Psychometrika* **46**, 139–142 (1981).

44. R. Plomin, Genetics and general cognitive ability. *Nature* **402**, C25–C29 (1999).

45. I. Pappa, *et al.*, Single Nucleotide Polymorphism Heritability of Behavior Problems in Childhood: Genome-Wide Complex Trait Analysis. *J. Am. Acad. Child Adolesc. Psychiatry* **54**, 737–744 (2015).

46. S. H. Lee, *et al.*, Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).

47. K. P. Harden, *et al.*, Genetic Associations Between Executive Functions and a General Factor of Psychopathology. *J. Am. Acad. Child Adolesc. Psychiatry* **59**, 749–758 (2020).

48. A. S. Dick, *et al.*, Meaningful associations in the adolescent brain cognitive development study. *Neuroimage* **239**, 118262 (2021).

49. T. Insel, *et al.*, Research Domain Criteria (RDoC): Toward a. *Am. J. Psychiatry Online*, 748–751 (2010).

50. L. M. McTeague, *et al.*, Identification of common neural circuit disruptions in cognitive control across psychiatric disorders. *Am. J. Psychiatry* **174**, 676–685 (2017).

51. K. M. Keyes, J. Platt, A. S. Kaufman, K. A. McLaughlin, Association of fluid intelligence and psychiatric disorders in a population-representative sample of US adolescents. *JAMA Psychiatry* **74**, 179–188 (2017).

52. R. Plomin, S. Von Stumm, The new genetics of intelligence. *Nat. Rev. Genet.* **19**, 148–159 (2018).

53. A. Neumann, *et al.*, Single Nucleotide Polymorphism Heritability of a General Psychopathology Factor in Children. *J. Am. Acad. Child Adolesc. Psychiatry* **55**, 1038–1045.e4 (2016).

54. E. Sprooten, B. Franke, C. U. Greven, The P-factor and its genomic and neural equivalents: an integrated perspective. *Mol. Psychiatry* **27**, 38–48 (2022).

55. V. Menon, M. D'Esposito, The role of PFC networks in cognitive control and executive function. *Neuropsychopharmacology*, 1–14 (2021).

56. S. E. Petersen, M. I. Posner, The attention system of the human brain: 20 years after. *Annu. Rev. Neurosci.* **35**, 73–89 (2012).

57. N. U. F. Dosenbach, *et al.*, A Core System for the Implementation of Task Sets. *Neuron* **50**, 799–812 (2006).

58. J. D. Power, S. E. Petersen, Control-related systems in the human brain. *Curr. Opin. Neurobiol.* **23**, 223–228 (2013).

59. J. Duncan, The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn. Sci.* **14**, 172–179 (2010).

60. M. I. Posner, S. E. Petersen, The attention system of the human brain. *Annu. Rev. Neurosci.* **13**, 25–42 (1990).

61. M. M. Halassa, S. Kastner, Thalamic functions in distributed cognitive control. *Nat. Neurosci.* **20**, 1669–1679 (2017).

62. H. Garavan, *et al.*, Recruiting the ABCD sample: Design considerations and procedures. *Dev. Cogn. Neurosci.* **32**, 16–22 (2018).

63. M. Luciana, *et al.*, Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery. *Dev. Cogn. Neurosci.* **32**, 67–79 (2018).

64. K. A. Kobak, C. J. Kratochvil, C. Stanger, J. Kaufman, Computerized screening of comorbidity in adolescents with substance or psychiatric disorders. *Anxiety Disord. Depress. Jolaa, CA)* (2013).

65. D. M. Barch, *et al.*, Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: Rationale and description. *Dev. Cogn. Neurosci.* **32**, 55–66 (2018).

66. M. F. Glasser, *et al.*, The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, 105–124 (2013).

67. D. A. Fair, *et al.*, Correction of respiratory artifacts in MRI head motion estimates. *Neuroimage* **208**, 116400 (2020).

68. L. Q. Uddin, B. T. T. Yeo, R. N. Spreng, Towards a Universal Taxonomy of Macro-scale Functional Human Brain Networks. *Brain Topogr.* **32**, 926–942 (2019).

69. A. Mihalik, *et al.*, Multiple Holdouts With Stability: Improving the Generalizability of Machine Learning Analyses of Brain–Behavior Relationships. *Biol. Psychiatry* **87**, 368–376 (2020).

70. A. M. Winkler, O. Renaud, S. M. Smith, T. E. Nichols, Permutation inference for canonical correlation analysis. *Neuroimage* **220**, 117065 (2020).

71. J. O. Linke, *et al.*, Shared and Anxiety-Specific Pediatric Psychopathology Dimensions Manifest Distributed Neural Correlates. *Biol. Psychiatry* **89**, 579–587 (2021).

72. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).

73. M. Delacre, D. Lakens, C. Leys, Why psychologists should by default use welch's t-Test instead of student's t-Test. *Int. Rev. Soc. Psychol.* **30**, 92–101 (2017).

74. F. V. Rijsdijk, P. C. Sham, Analytic approaches to twin data using structural equation models. *Brief. Bioinform.* **3**, 119–133 (2002).

75. M. C. Neale, *et al.*, OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika* **81**, 535–549 (2016).

76. B. T. Thomas Yeo, *et al.*, The organization of the human cerebral cortex estimated by

intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).

Figures and Tables

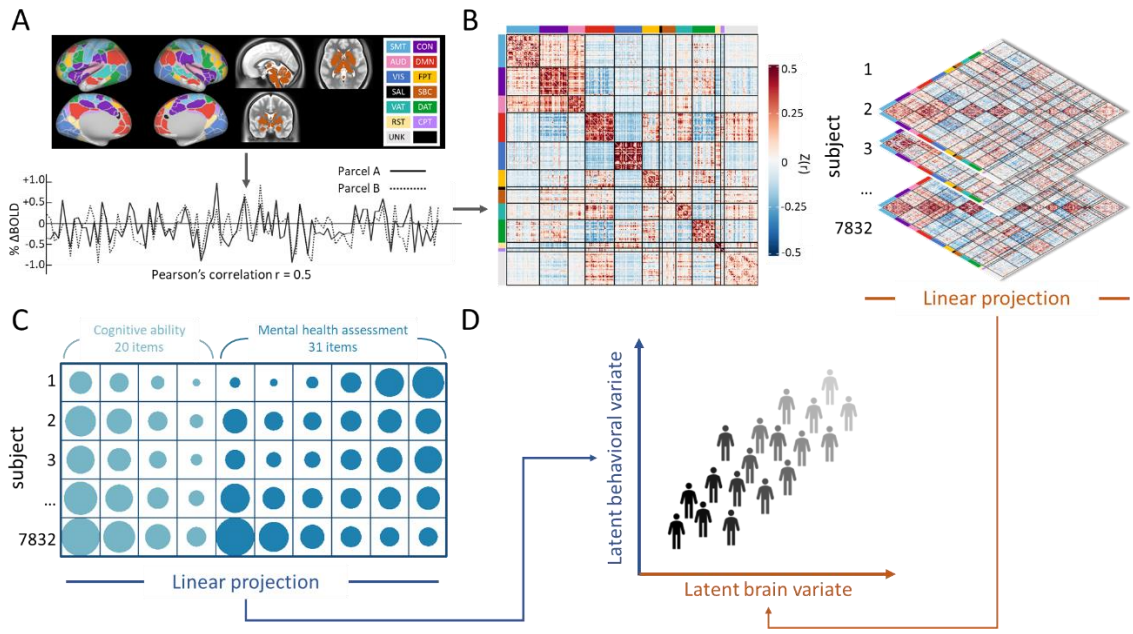

Figure 1. Scheme of canonical correlation analysis (CCA). **(A) Construction of functional connectome.** Preprocessed time-series were extracted from 352 parcels defined by the Gordon atlas(40). Then z-transformed Pearson's correlation of the time series were calculated for each pair of the brain parcels. Network abbreviations modules are listed in the SI. **(B) Connectome dataset.** Functional connectome data of 7382 young participants (including 5906 as the training set and 1476 as the test set) were included as the brain feature set**. (C) Behavioral-assessment dataset.** The behavioral assessments consisted of 20 scales for cognitive ability and 31 scales for mental-health related constructs. **(D) Youth population in the aligned latent space identified by CCA.** CCA identifies linear subspaces for the two datasets. The two datasets are ensured to be maximally correlated when projected to the latent spaces, and thus reveal latent brain-behavior associations in the youth population.

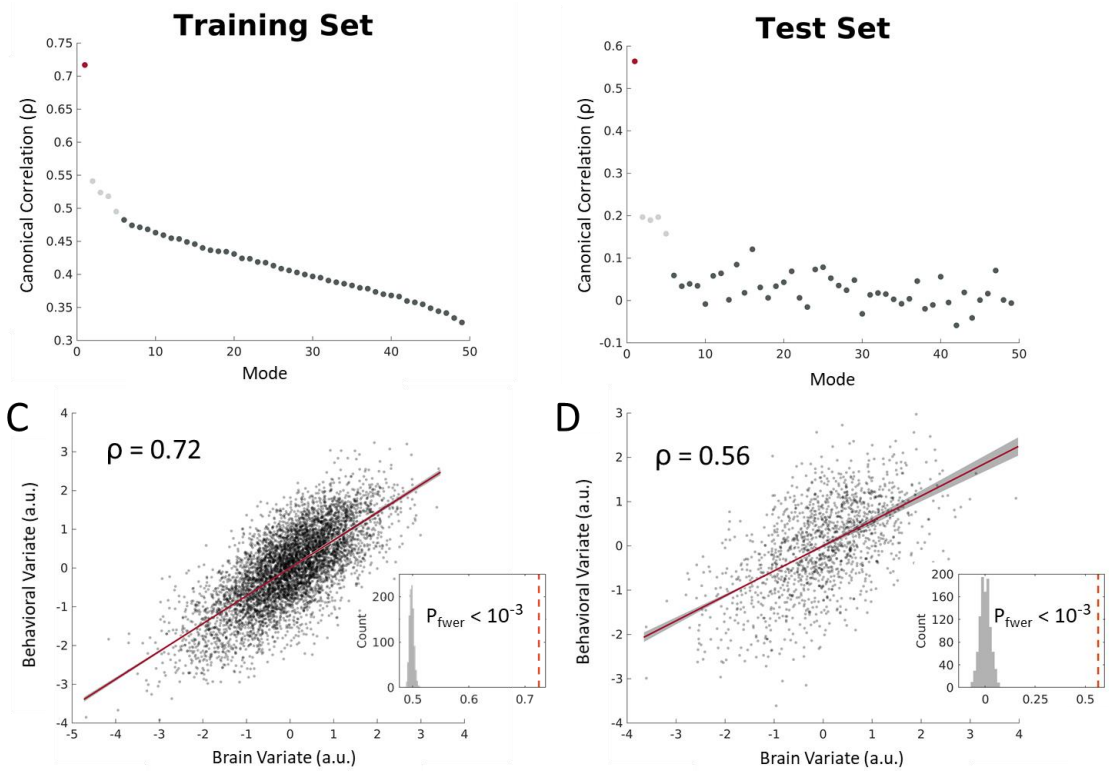

Figure 2. Canonical correlations (CCs) between the brain and behavioral datasets. (A) CCs in the training set. Dots in the plot indicate CCs of the 49 CCA modes from the training set. In the training set, top five (red and light gray points) modes show significant correlations in comparison to the null distribution. **(B) CCs in the test set.** CCA modes in the test set were obtained from the hold-out validation, i.e., by applying the training-set CCA coefficients on the data in the test set. Dots in the plot indicate CCs of the 49 modes from the training set. When generalizing to the test set, the top five CCA mode shows significant CCs in comparison to the null distributions five (red and light gray points). **(C) and (D) Scatter plots of the associated brain and behavior variate scores in mode 1, for training and test sets 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 ($\rho = 0.72$, $p < 10^{-3}$, N = 5906) and the test set ($\rho = 0.56$, $p < 10^{-3}$, N = 1476). Inset plots represent permutation test results conducted for CC in the training and test sets

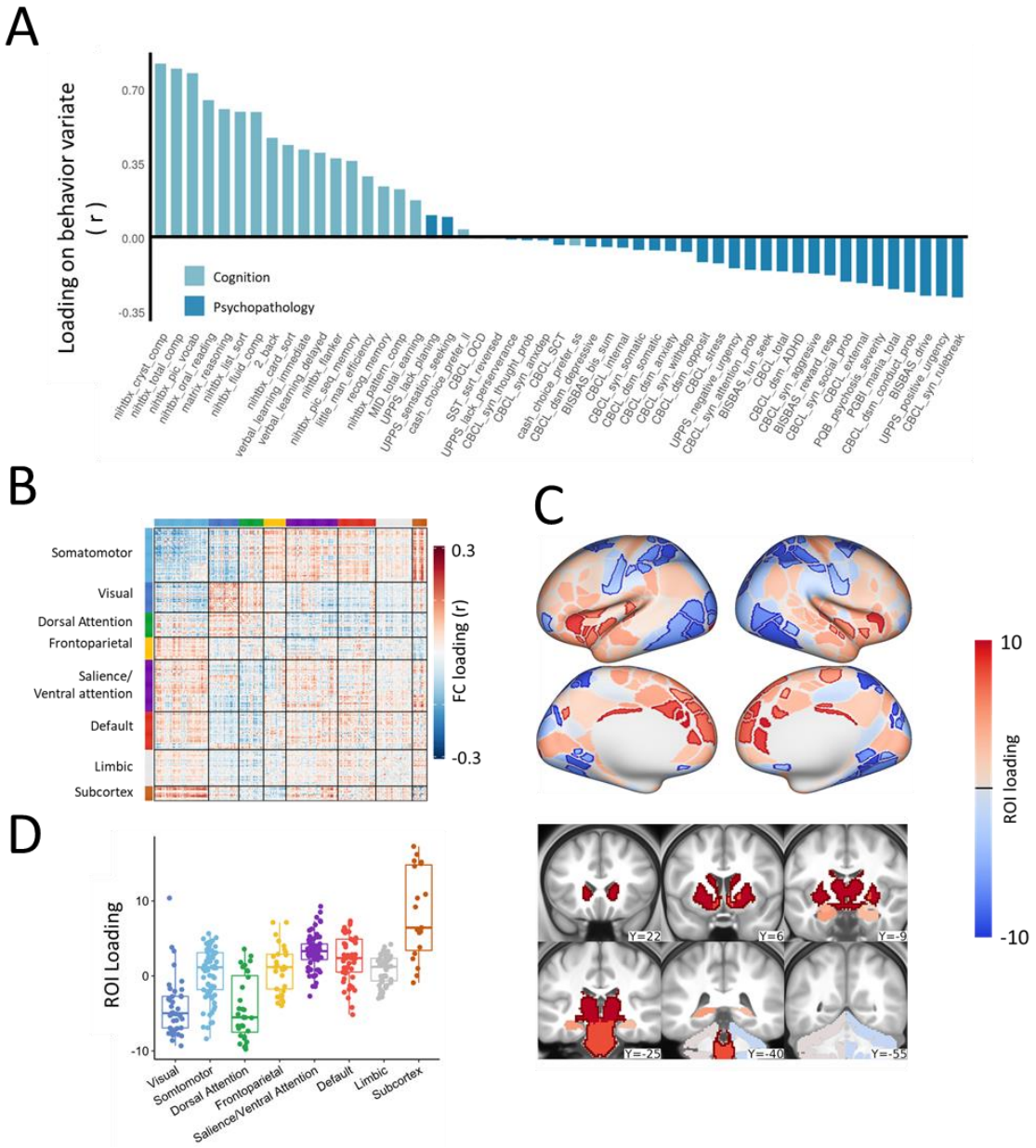

Figure 3 Brain and behavioral loadings of the principal CCA mode. (A) Loading of each behavioral assessment on the behavioral variate. Abbreviates are listed in Table S2 and S3 **(B) Loading of rsFC on the brain connectome variate. (C) Map of ROI loading**. Highlighted borders indicate ROIs with significant positive (dark red) or significant negative (dark blue) loadings compared with a permutation-generated null distribution . The significance level was set at $p < 0.01$ and was FDR-corrected for multiple comparison. **(D) Distribution of ROI loadings across large-scale functional brain networks.** The 352 parcels were grouped into 7 cortical networks and the subcortex (76).

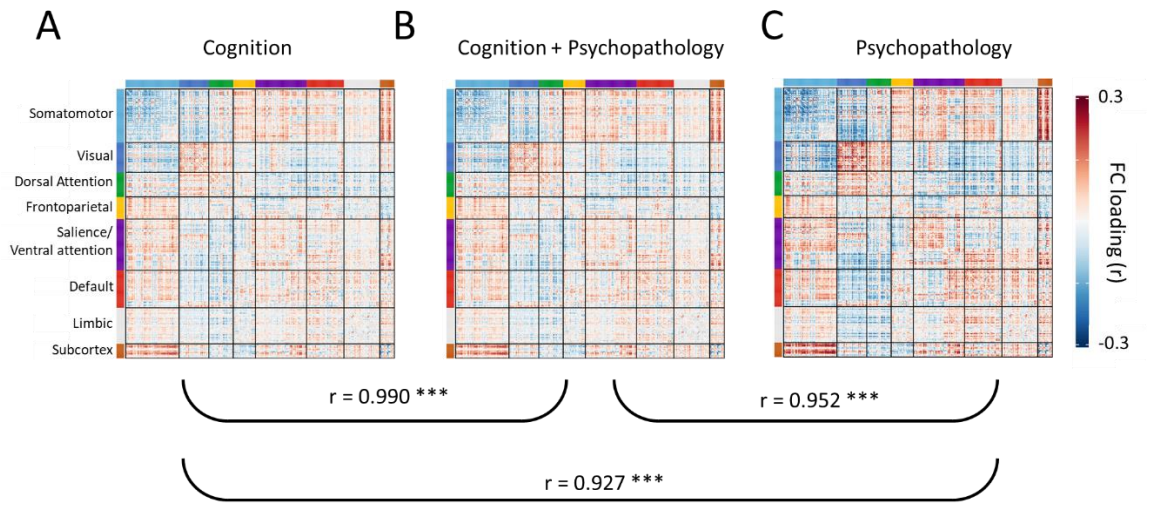

Figure 4 Cross-domain consistent connectome variate of the principal mode. Loading map of the principal CCA mode obtained by associating the connectome data set to the **(A)** Cognition only **(B)** Combined cognition and psychopathology and **(C)** Psychopathology only behavioral sets show high consistency (spatial Pearson's correlation) between them. Loading maps are averaged by the cross-validation folds. The resultant p values are corrected for multiple comparison using the FDR procedure, *** FDR-adjusted $p < 0.001$.

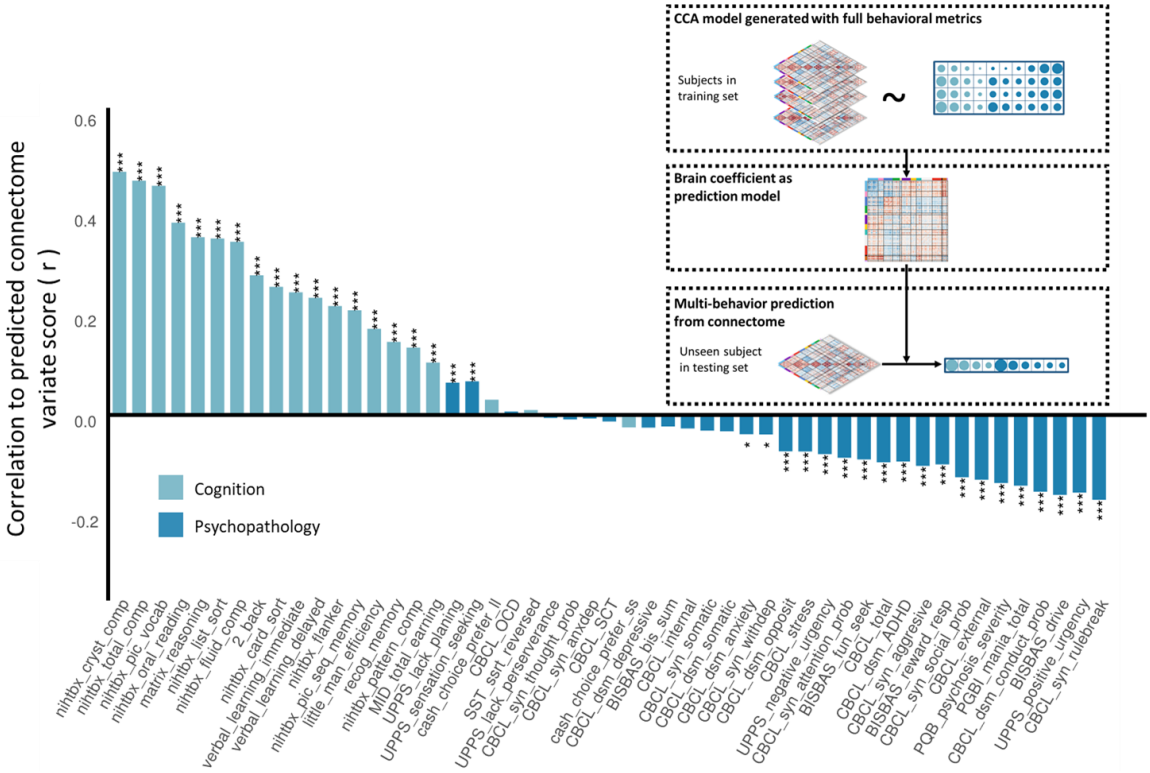

Figure 5 Behavioral assessment prediction from the connectome variate score. The bar plot shows the performance of using the CCA estimated connectome variate to predict behavior assessment scores in an unseen population. Behavioral assessments are ordered by the rank of their behavioral loadings. Inset plot shows the cross-population prediction procedure. The predictive ability of behavioral assessments was assessed with Pearson's correlation, with p-values FDR corrected. * FDR-adjusted $p < 0.05$, ** FDR-adjusted $p < 0.01$, *** FDR-adjusted $p < 0.001$

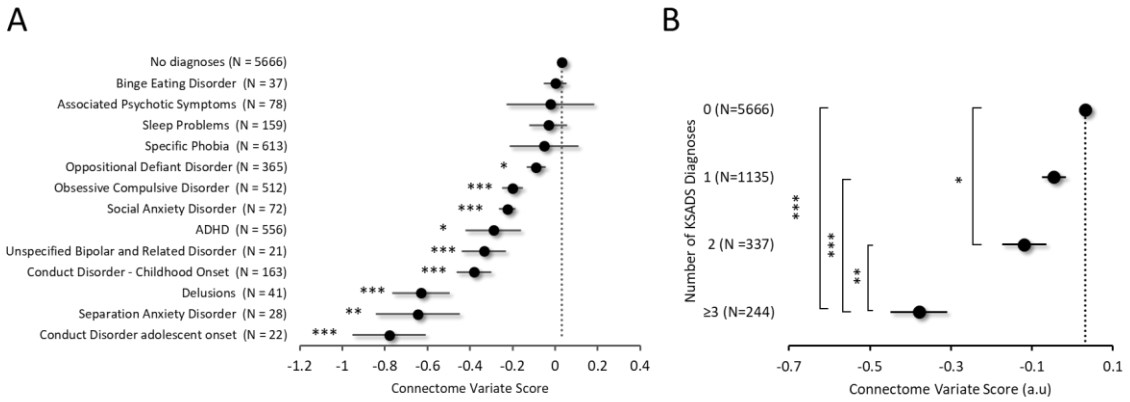

Figure 6 Relationships between connectome variate scores and current psychiatric disorders in the entire sample. (A) Group-wise comparison of mean connectome variate scores in preadolescents with different psychiatric diagnoses. Dots and lines in the plot indicate mean and standard error of the connectome variate score in each group of KSADS diagnosis. The participant group with no current psychiatric diagnoses was used as the baseline. Welch's *t*-tests were used to compare differences in mean connectome variate scores in each diagnostic subgroup (e.g., participants with conduct disorder) compared to the baseline group. *P*-values were FDR-adjusted for multiple comparisons. **(B) Differences in mean connectome variate scores in preadolescent participants as a function of cumulative number of current psychiatric diagnoses identified via the KSADS.** Pair-wise group differences were post-hoc compared using the Games-Howell procedure. * Adjusted $p < 0.05$, ** Adjusted $p < 0.01$, *** Adjusted $p < 0.001$

Table 1 Intrapair twin correlations, influence of genetic/environmental factors, heritability of connectome variate

Twin Correlations		Variance		Heritability
r_{MZ} (N)	r_{DZ} (N)	a^2 (95% CI)	e^2 (95% CI)	h^2 (95% CI)
0.71 (76)	0.39 (126)	0.60 (0.42 -0.82)	0.20 (0.13 - 0.32)	0.74 (0.56-0.86)