

## ***Connectomic Alterations Linked to Transdiagnostic Risk for Psychopathology at the Transition to Adolescence***

*Chandra Sripada MD, PhD<sup>1</sup>\*, Mike Angstadt MS<sup>1</sup>, Saige Rutherford BS<sup>1</sup>, Aman Taxali MS<sup>1</sup>, Tristan Greathouse BS<sup>1</sup>, D. Angus Clark PhD<sup>1</sup>, Luke Hyde PhD<sup>2</sup>, Alex Weigard PhD<sup>1</sup>, Sarah Brislin PhD<sup>1</sup>, Brian Hicks PhD<sup>1</sup>, Mary Heitzeg PhD<sup>1</sup>*

<sup>1</sup>*Department of Psychiatry, University of Michigan, Ann Arbor, <sup>2</sup>Department of Psychology, University of Michigan, Ann Arbor, MI*

*\*Correspondence:*

*Chandra Sripada  
4250 Plymouth Road  
Ann Arbor MI 48109  
[sripada@umich.edu](mailto:sripada@umich.edu)*

### ***Word Count***

*Abstract: 350*

*Manuscript Text: 3000*

***Keywords:*** *P factor, general factor of psychopathology, resting state fMRI, connectomics, Adolescent Brain Cognitive Development (ABCD) Study, predictive modeling, intrinsic connectivity networks, default network, fronto-parietal network, control networks*

## Abstract

**BACKGROUND** Convergent research identifies a general factor (“P factor”) that confers transdiagnostic risk for psychopathology. However, brain functional connectivity patterns that underpin the P factor remain poorly understood, especially at the transition to adolescence when many serious mental disorders have their onset.

**OBJECTIVE:** Identify a distributed connectome-wide neurosignature of the P factor and assess the generalizability of this neurosignature in held out samples.

**DESIGN, SETTING, AND PARTICIPANTS** This study used data from the full baseline wave of the Adolescent Brain and Cognitive Development (ABCD) national consortium study, a prospective, population-based study of 11,875 9- and 10-year olds. Data for this study were collected from September 1, 2016 to November 15, 2018 at 21 research sites across the United States.

**MAIN OUTCOMES AND MEASURES** We produced whole brain functional connectomes for 5,880 youth with high quality resting state scans. We then constructed a low rank basis set of 250 components that captures interindividual connectomic differences. Multi-level regression modeling was used to link these components to the P factor, and leave-one-site-out cross-validation was used to assess generalizability of P factor neurosignatures to held out subjects across 19 ABCD sites.

**RESULTS** The set of 250 connectomic components was highly statistically significantly related to the P factor, over and above nuisance covariates alone (ANOVA nested model comparison, incremental R-squared 6.05%,  $\chi^2(250) = 412.1$ ,  $p < 4.6 \times 10^{-10}$ ). In addition, two individual connectomic components were statistically significantly related to the P factor after Bonferroni correction for multiple comparisons ( $t(5511) = 4.8$ ,  $p < 1.4 \times 10^{-6}$ ;  $t(5121) = 3.9$ ,  $p < 9.7 \times 10^{-5}$ ). Functional connections linking control networks and default mode network were prominent in the P factor neurosignature. In leave-one-site-out cross-validation, the P factor neurosignature generalized to held out subjects (average correlation between actual and predicted P factor scores across 19 held out sites = 0.13;  $p_{\text{PERMUTATION}} < 0.0001$ ). Additionally, results remained significant after a number of robustness checks.

**CONCLUSIONS AND RELEVANCE:** The general factor of psychopathology is associated with connectomic alterations involving control networks and default mode network. Brain imaging combined with network neuroscience can identify distributed and generalizable signatures of transdiagnostic risk for psychopathology during emerging adolescence.

## 1. Introduction

Recent investigations into patterns of covariance across psychiatric symptoms consistently find a general factor of psychopathology, termed the “P factor”, that is associated with most prevalent psychiatric symptoms<sup>1–4</sup>. Concurrently, categorical diagnostic approaches that currently predominate encounter serious issues. Tellingly, one persistent problem has been excess overlap across disorders in symptoms<sup>5,6</sup>, neural mechanisms<sup>7,8</sup>, and genetic risk factors<sup>9,10</sup>—a problem that could be readily explained if a domain general P factor drives co-occurrence of symptoms irrespective of diagnostic boundaries. Despite these compelling features of the P factor model, key gaps in knowledge remain, especially regarding the neural mechanisms that produce broad liabilities to diverse psychopathologies and the developmental pathways through which these broad liabilities operate.

Network neuroscience<sup>11–14</sup> is well positioned to help fill in this gap in knowledge. The human brain is organized into a number of large-scale connectivity networks<sup>15,16</sup>, and there is growing understanding of distinct information processing functions implemented by these networks and by interacting network ensembles. Recent psychological models of the P factor emphasize executive dysfunction and impulsivity<sup>3,17,18</sup>, suggesting the possibility that deficits in control networks (e.g., fronto-parietal network, dorsal attention network)<sup>19,20</sup>, involved in attention and cognitive control, are implicated in producing the P factor—a possibility that awaits detailed investigation.

Network neuroscience can also illuminate the developmental pathways that lead to psychopathology<sup>21–23</sup>. Brain networks undergo massive maturation during adolescence<sup>24–26</sup>. Importantly, this is also the time that many serious mental disorders first emerge<sup>5,21,27</sup>. A critical goal for psychiatric neuroscience is to develop brain-based objective markers of risk that can identify vulnerable youth who can most benefit from targeted interventions<sup>21</sup>. Because of the breadth of the P factor’s posited effects on diverse forms of psychiatric symptomatology, identification of network alterations during youth that are associated with the P factor is particularly pressing.

The current study examines connectome-wide functional connectivity patterns associated with the P factor in a sample of 11,875 9 and 10-year olds in the Adolescent Brain and Cognitive Development (ABCD) national consortium study, Release 2.1<sup>28,29</sup>. Recently, our group constructed and validated a P factor model<sup>30</sup> in ABCD from the Child Behavior Checklist<sup>31</sup> (CBCL)-parent report using bifactor modeling. For the present study, we produced resting state connectomes for 5,880 youth who met stringent neuroimaging quality control standards. We next applied a multivariate approach, brain basis set (BBS)<sup>32–34</sup>, to link whole-brain functional connectivity patterns to the P factor. BBS takes advantage of the fact that, although functional connectomes are large and complex, encompassing tens of thousands of connections, there is massive redundancy in the set of connections that differ across people. This allows a small set of components—we used 250 in the present study—to capture most meaningful inter-individual variation<sup>32,33</sup>. To establish multivariate relationships between this basis set of connectomic components and the P factor, we adopted a two-pronged analysis approach that

leverages standard parametric statistical methods (well suited for complex multilevel designs with covariates and sample weighting) and newer cross-validation approaches (well suited for assessing generalizability to unseen data). Both analyses convergently demonstrate that the P factor is associated with brain-wide altered connectivity patterns prominently implicating connections within and between control networks and processing networks.

## 2. Methods

### 2.1. Sample and Data

The ABCD study is a multisite longitudinal study with 11,875 children between 9-10 years of age from 21 sites across the United States. The study conforms to the rules and procedures of each site's Institutional Review Board, and all participants provide informed consent (parents) or assent (children). Detailed description of recruitment procedures<sup>35</sup>, assessments<sup>36</sup>, and imaging protocols<sup>37</sup> are available elsewhere. The ABCD data used in this report came from NDA Study 721, 10.15154/1504041, which can be found at <https://nda.nih.gov/study.html?id=721>.

### 2.2. Data Acquisition, fMRI Preprocessing, and Connectome Generation

Imaging protocols were harmonized across sites and scanners. High spatial (2.4 mm isotropic) and temporal resolution (TR=800 ms) resting state fMRI was acquired in four separate runs (5min per run, 20 minutes total, full details are described in <sup>38</sup>). The entire data pipeline described below was run through automated scripts on the University of Michigan's high-performance cluster, and is described below, with additional detailed methods automatically generated by fMRIprep software provided in the Supplement. Code for running the analyses can be found at [https://github.com/SripadaLab/ABCD\\_Resting\\_Psychopathology](https://github.com/SripadaLab/ABCD_Resting_Psychopathology).

Preprocessing was performed using fMRIprep version 1.5.0<sup>39</sup>, a Nipype<sup>40</sup> based tool. Full details of the fMRIprep analysis can be found in supplemental materials. Briefly, T1-weighted (T1w) and T2-weighted images were run through recon-all using FreeSurfer v6.0.1. T1w images were also spatially normalized nonlinearly to MNI152NLin6Asym space using ANTs 2.2.0. Each functional run was corrected for fieldmap distortions, rigidly coregistered to the T1, motion corrected, and normalized to standard space. ICA-AROMA was run to generate aggressive noise regressors. Anatomical CompCor was run and the top 5 principal components of both CSF and white matter were retained. Functional data were transformed to CIFTI space using the method provided on the HCP Wiki (<https://wiki.humanconnectome.org/display/PublicData/HCP+Users+FAQ#HCPUsersFAQ-9.HowdoImapdatabetweenFreeSurferandHCP>). All preprocessed data were visually inspected at two separate stages to ensure only high-quality data was included: After co-registration of the functional data to the structural data and after registration of the functional data to MNI template space.

Connectomes were generated for each functional run using the Gordon 333 parcel atlas<sup>41</sup>, augmented with parcels from high-resolution subcortical<sup>42</sup> and cerebellar<sup>43</sup> atlases. Volumes exceeding a framewise displacement threshold of 0.5mm were marked to be censored. Covariates were regressed out of the time series in a single step<sup>44</sup>, including: linear trend, 24

motion parameters (original translations/rotations + derivatives + quadratics, aCompCorr 5 CSF and 5 WM components and ICA-AROMA aggressive components, high pass filtering at 0.008Hz, and censored volumes. Next, correlation matrices were calculated for each run. Each matrix was then Fisher r-to-z transformed, and then averaged across runs for each subject yielding their final connectome.

### **2.3. Constructing Structural Model of Psychopathology**

The general psychopathology factor (P-factor) used here is based on the parent-rated Child Behavior Checklist (CBCL; age 6 to 18 form<sup>31</sup>). A bifactor model was fit to eight CBCL scales, with a general P factor that all scales loaded onto (average scale loading = .69) and two specific factors. This model is described in detail in the Supplement as well our previous studies, where we compared this model to 14 alternatives<sup>30</sup> (finding consistent P factors across diverse model specifications,  $rs > 0.90$ ) and we delineated nomological networks of the P factor across diverse ABCD variables<sup>45</sup>.

### **2.4. Inclusion/Exclusion**

There are 11,875 subjects in the ABCD Release 2.0.1 dataset. Screening was initially done using ABCD raw QC to limit to subjects with 2 or more good runs of resting data as well as a good T1 and T2 image (QC score, protocol compliance score, and complete all =1). This resulted in 9580 subjects with 2 or more runs that entered preprocessing. Each run was subsequently visually inspected for registration and warping quality, and only those subjects who still had 2 or more good runs were retained (N=8858). After connectome generation, runs were excluded if they had less than 4 minutes of uncensored data, and next subjects were retained only if they had to 2 or more good runs (N=6568). Finally, subjects who were missing data required for regression modeling (P factor scores or any of the nuisance covariates) were dropped, and then sites with fewer than 75 subjects were dropped. This left us with N=5880 subjects across 19 sites for the whole sample multiple regression analysis and leave-one-site-out cross-validation.

### **2.5. Constructing a Brain Basis Set (BBS)**

BBS is a validated multivariate method that uses principal components dimensionality reduction to produce a basis set of components that are then associated with phenotypes<sup>46,32</sup>. We select the top 250 components for our basis set based on previous work showing that 50-100 components per 1000 subjects captures most meaningful variance without overfitting<sup>32</sup>.

### **2.6. Whole Sample Multiple Regression Analysis**

To assess the multivariate relationship between the 250-component brain basis set and the P factor, we used multi-level multiple regression modeling, with the P factor as outcome variable and expression scores for the 250 components as predictors. Gender, race, age, mean FD, and mean FD squared were entered as fixed effect covariates, with family id and ABCD site entered as random effects (family nested within site).

To help convey overall patterns across the entire multiple regression model, we constructed a “consensus” component map. We multiplied each connectomic component with its associated

beta from the preceding multiple regression model. Next, we summed across all 250 components yielding a single map, and thresholded the entries at  $z=2$ .

### 2.7. Leave-One-Site-Out Cross Validation

To assess generalizability of BBS-based regression models, we used leave-one-site-out cross-validation. In each fold of the cross-validation, data from one of the 19 sites served as the held-out test dataset and data from the other 18 sites served as the training dataset. Additionally, to ensure separation of train and test datasets, at each fold of the cross-validation, a new PCA was performed on the training dataset yielding a 250-component basis set.

### 2.8. Accounting for Covariates in Cross-Validation Framework

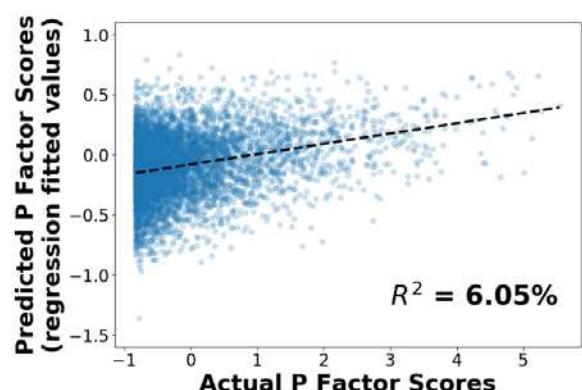
In each fold of cross-validation, BBS models were trained in the train partition with the following covariates: gender, race, age, age squared, mean FD and mean FD squared. To maintain strict separation between training and test datasets, regression coefficients for the covariates learned from the training sample are applied to the test sample, and the variance they explain is subtracted away. This procedure, described in detail in our previous publication<sup>34</sup>, yields an estimate of the contribution of brain components alone in predicting test subject P factor scores, excluding the contribution of the nuisance covariates. Note that by employing leave-one-site-out, members of twinships and sibships are never present in both training and test samples.

### 2.9. Permutation Testing

We assessed the significance of all cross-validation-based correlations with non-parametric permutation tests in which we randomly permuted the 5,880 subjects' P factor scores 10,000 times, as described in detail in the Supplement.

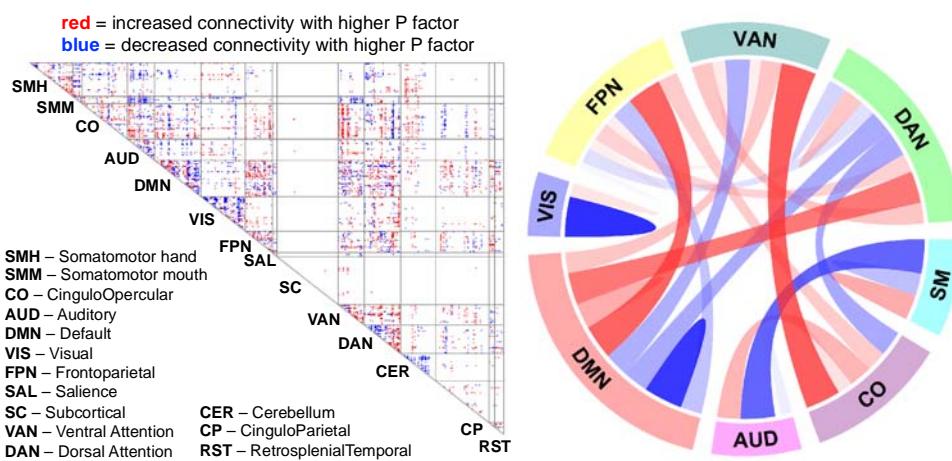
## 3. Results

### 3.1. A Distributed Pattern of Network Alterations Is Highly Statistically Significantly Related to the P Factor



**Figure 1. Fitted Relationship Between 250 Connectivity Components and the P Factor in 5,880 Youths.** We conducted a multi-level regression with 250 connectivity components as predictors and the P Factor as outcome variable, controlling for demographic covariates and clustering due

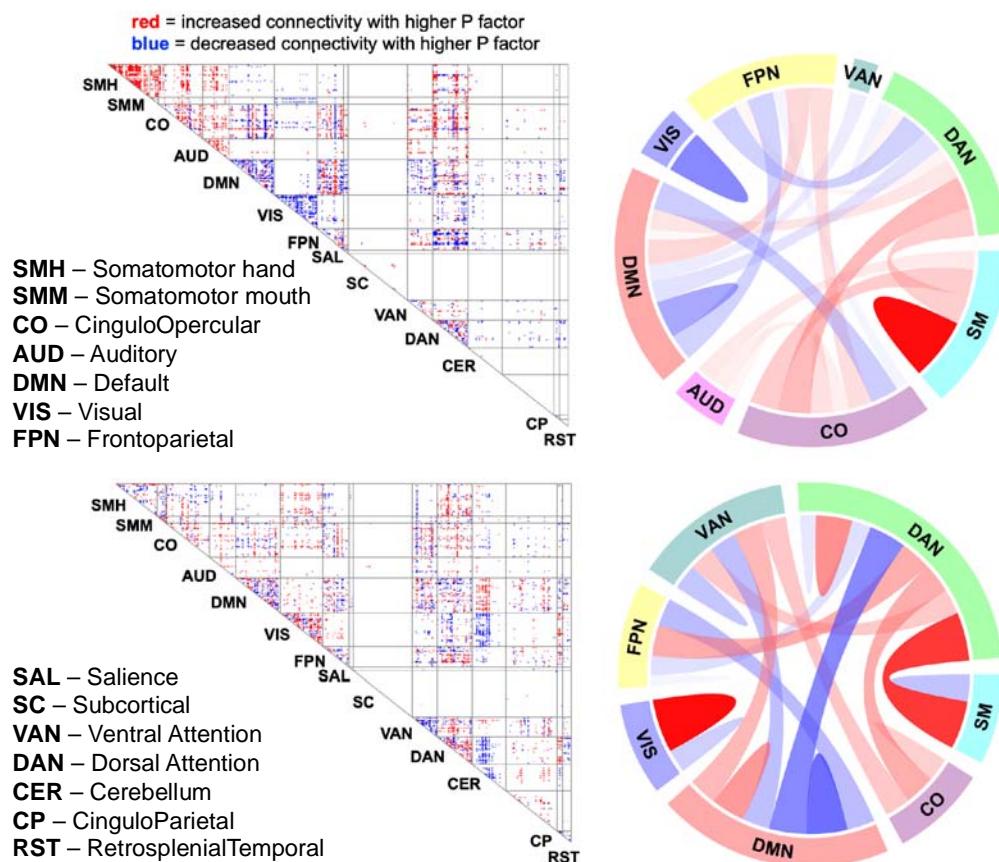
to family and site. The scatter plot reflects the fitted fixed effect relationship between predictors and the P factor (incremental  $R^2$  of 250 components over and above a covariates-only model is 6.05%).



**Figure 2: Multivariate Connectomic Neurosignature of the P Factor.** Using multi-level regression modeling, we observed a highly statistically significant relationship between an individual differences basis set of 250 connectivity components and the P factor. Based on the fitted model coefficients, we created a consensus connectome, an importance-weighted composite of components, that represents a multivariate neurosignature of the P factor. (Left Panel) Connection-level map of the consensus connectome. This map shows altered within-network connectivity (triangles along the diagonal of the left panel) at several networks including DMN. It also shows altered connectivity involving control networks (FPN, CO, VAN, DAN) and DMN. (Right Panel) Network-level map of the consensus connectome, in which width of arcs reflects number of suprathreshold connections. Note, to enhance interpretability, some networks are not shown (see Supplement).

We used multiple regression analysis to quantify the relationship between brain connectivity patterns and the P factor. In a multi-level regression model, we entered 250 brain components as predictors and the P factor as outcome variable, with race, gender, head motion, ABCD site, and family id as nuisance covariates. We found that compared to a covariates-only model, the 250 brain components yielded an 6.05% increase in variance explained, a highly statistically significant increase ( $\chi^2(250) = 412.1, p < 4.6 \times 10^{-10}$ ; Figure 1). We in addition created a consensus connectome, an importance-weighted composite of components, that represents a multivariate neurosignature of the P factor (Figure 2). Two motifs are particularly prominent in the neurosignature: 1) altered connectivity within the 14 large-scale networks, especially in DMN and sensory networks; and 2) altered connectivity within control networks and between control networks and DMN. In particular, connections within large-scale networks (along the diagonal of the figure) represent 8.48% of the connectome, but 23.65% of the suprathreshold connections in the neurosignature. Additionally, cross-network connections involving control networks (FPN, CO, DAN, VAN) and DMN represent 11.58% of the connectome, but 31.83% of the suprathreshold connections in the neurosignature.

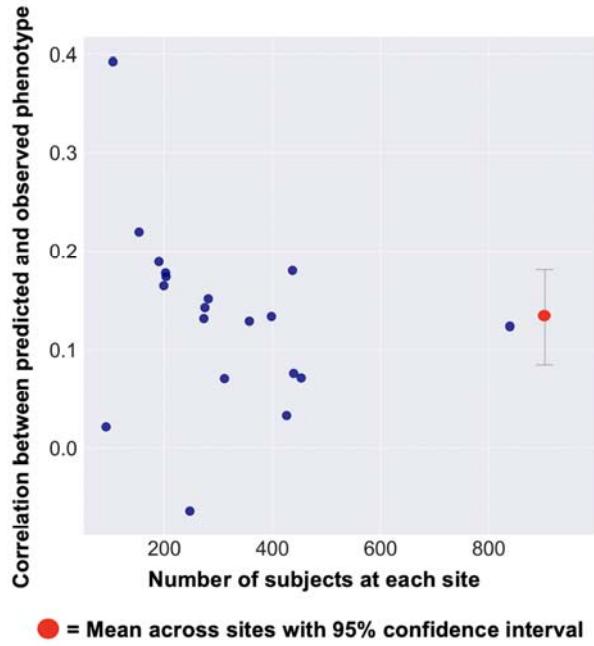
### 3.2. Two Connectivity Components are Statistically Significantly Related to the P factor



**Figure 3: Connectivity Components Associated with the P Factor.** Two individual connectivity components were statistically significantly related to the P factor (after Bonferroni correction for 250 comparisons). Both components involve alterations in within-network connectivity, especially in SMH, DMN, VN, and DAN, as well as altered connectivity involving control networks and DMN. (Left Side) Connection-level map showing suprathreshold connections. (Right Side) Network-level map, in which width of arcs reflects number of suprathreshold connections. Note, to enhance interpretability, some networks are not shown (see Supplement).

We next examined statistical significance of individual connectivity components within the multiple regression model, using Bonferroni correction to control family-wise error while making 250 statistical comparisons. After this correction, we found that two components remained statically significantly associated with the P factor ( $t(5511) = 4.8$ ,  $p < 1.4 \times 10^{-6}$ ;  $t(5121) = 3.9$ ,  $p < 9.7 \times 10^{-5}$ ). Connectivity patterns from these components are shown in Figure 3 (see its associated figure caption for descriptions of these findings). Importantly, there were 27 components that were significantly associated the P factor at  $p < 0.05$ , uncorrected (where only 12.5 are expected by chance), indicating that the P factor is associated with numerous relatively subtle connectome-wide alterations.

### 3.3. Multivariate Neurosignatures of the P factor Generalize to Held Out Samples



**Figure 4. Site-Specific Correlations Between Predicted and Actual P Factor Scores in Leave-One-Site-Out Cross-Validation Analysis.**

We next examined the relationship between these 250 components and the P factor using a predictive modeling framework with leave-one-site-out cross-validation. A regression model was trained for connectome-based prediction of the P factor at all sites except one, this model was used to derive predictions of the P factor at the held out site, and this process was repeated with each site held out. To maintain separation between training and testing samples, we learned new connectivity components and covariate regression coefficients in each training sample and applied them to the testing sample. The correlation between actual and predicted P factor scores was 0.13, a value higher than all 10,000 values in a permutation-based null distribution ( $p_{\text{PERM}} < 0.0001$ ; Figure 4).

### 3.4. Multivariate Neurosignatures Remain Significantly Related to the P factor Across a Number of Tests of Robustness

We next assessed the robustness of our multiple regression analysis and leave-one-site-out cross-validation analysis by changing key elements of the analysis streams. We tested models that utilized: 1) Log-transformed P factor scores to improve their distributional properties; 2) Additional ABCD demographic covariates (household income, highest parental education, household marital status); 3) An ultra-low head motion sample (mean FD<0.2;  $N = 2,721$ ); 4) P factor models learned in the training sample and applied to the held out sample (to create total separation of train and test partitions); 5) Sampling weights, based on American Community Survey proportions<sup>47</sup>, which make findings more generalizable by approximating a

representative population. As shown in Table 1, all models remained highly statistically significant across all these analyses, confirming the robustness of our analysis.

Analysis	Whole Sample Regression	Leave-One-Site-Out Cross-Validation
Main Analysis	$R^2=6.05$ ; $p<4.6 \times 10^{-10}$	$r=0.13$ ; $p_{PERM}<0.0001^*$
1. Log Transform P Factor	$R^2=6.20$ ; $p<1.3 \times 10^{-11}$	$r=0.13$ ; $p_{PERM}<0.0001^*$
2. Additional Covariates	$R^2=5.53$ ; $p<4.2 \times 10^{-8}$	$r=0.11$ ; $p_{PERM}<0.0001^*$
3. Low Motion Subsample	$R^2=5.31$ ; $p<4.3 \times 10^{-6}$	$r=0.10$ ; $p_{PERM}=0.0003$
4. Leave-One-Site Out P Factor	--	$r=0.12$ ; $p_{PERM}<0.0001^*$
5. With Sample Weights	$R^2=8.14$ ; $p<1.9 \times 10^{-12}$	--

**Table 1: Summary of Additional Analyses to Assess Robustness.** We assessed the sensitivity of our main analyses (top row) to modeling choices with a number of robustness checks (rows labeled 1-5). We found results remained statistically significant across all these analyses. \* = observed correlation was higher than all 10,000 correlations in the permutation distribution.

#### 4. Discussion

This study investigated brain-wide connectomic alterations associated with the general factor of psychopathology (“P factor”), a broad liability for diverse forms of psychopathology, in 5,880 9- and 10-year olds in the ABCD multi-site sample. By combining large sample size, reliable multivariate methods, cross-validation to assess generalizability, sampling weights to approximate representativeness, and a number of additional tests of model robustness, we provide the strongest evidence yet that the P factor is associated with distributed changes in connectivity patterns of the brain’s intrinsic functional architecture. We found especially prominent alterations in connections within large-scale networks and connections linking control networks and DMN. Overall, our results highlight the potential for neuroimaging to detect and delineate network abnormalities associated with broad liabilities for psychopathology during the transition to adolescence, a key developmental window in which many serious psychopathologies first emerge.

Previous studies tended to use case control designs to characterize brain network abnormalities associated with individual disorders<sup>48</sup>, where these abnormalities were assumed to reflect disorder-specific pathophysiology. Accumulated results, however, suggest observed network alterations often lack specificity. For example, two prominent network motifs we observed in the present study, reduced connectivity within DMN and altered control networks/DMN connectivity, have elsewhere been demonstrated in a number of individual psychiatric

disorders, including hypoconnectivity of DMN in autism<sup>49</sup>, schizophrenia<sup>50</sup>, and ADHD<sup>46,51,52</sup>, and reduced DMN/TPN anti-correlation in schizophrenia<sup>53-55</sup>, bipolar disorder<sup>56</sup>, and ADHD<sup>46,52</sup>. The present study potentially explains this lack of specificity by linking motifs such as these instead to the P factor. The P factor represents a broad liability to diverse forms of psychopathology, and thus P-linked connectomic motifs would be expected to show up non-specifically across diverse case-control disorder comparisons.

It is noteworthy that these same two connectivity motifs, i.e., connectivity within DMN and control networks/DMN connectivity, are known to undergo intense maturation during childhood and adolescence<sup>26</sup>. In particular, during this developmental period, links within DMN exhibit massive increased connectivity<sup>57,46</sup> and links involving control networks and DMN exhibit substantial reconfiguration<sup>57-59,46</sup>, with a predominant pattern of segregation (increasing anti-correlation)<sup>58,46</sup>. These observations are suggestive of graphical colocation of P-effects and maturation-effects in the connectome, with P associated with a more immature connectivity pattern. Future studies should quantitatively examine this hypothesis, by using formal tests for graphical colocation in cross-sectional data<sup>59</sup> or by quantifying P-associated maturation patterns in future waves of ABCD longitudinal data. Studies such as these could provide direct evidence that the P factor reflects neurodevelopmental dysmaturation, inviting further research into how to mitigate “miswired” connectivity patterns<sup>60</sup> or prospectively block their emergence.

It is interesting to interpret our results in light of recent proposals for the psychological basis of the P factor. We recently examined<sup>45</sup> the nomological networks of the P factor in the ABCD sample ( $n=11,875$ ), finding that P is associated with three psychological variables: higher levels of fear/distress emotions, higher impulsivity, and reduced neurocognition. This aligns with proposals from other researchers<sup>3,17,18</sup> that P represents heightened generation of impulses (both negative fear/distress emotions as well as reward-seeking impulses) and reduced executive functioning. Consistent with this idea, our connectomic results highlight altered connectivity involving control networks (FPN, CO, DAN, VAN) and processing networks (DMN, visual network, auditory network). Notably other recent functional connectivity studies of the P factor also found alterations in processing networks, including visual network<sup>61</sup> and somatomotor network<sup>62</sup>. Control networks are involved in working memory and attention and are sources of cognitive control signals<sup>19,20</sup>, while processing networks are involved in stimulus processing and spontaneous thought<sup>63</sup> and are targets of cognitive control<sup>64,65</sup>. Thus, known functional profiles of networks found here to be associated with the P factor align well with current models of the psychological factors that contribute to higher levels of P.

While the brain/behavior associations identified in this study are statistically reliable and generalize to held out samples, they are nonetheless modest in size, and several points are relevant. First, we systematically assessed the ability of the multivariate neurosignature of the P factor to generalize to brain imaging data from 19 held out sites, each employing their own fMRI scanners. Few studies of psychopathology seek to systematically assess cross-site generalizability in this way (cf. <sup>33</sup>). Direct comparison of our multivariate cross-validation effect size with within-sample effect sizes from other studies would thus be inappropriate. Such

comparison would be especially inapt if the comparison study used more standard mass univariate analysis methods (rather than multivariate methods), because conducting thousands of statistical tests produces inflated p-values (which is widely known) as well as inflated effect sizes<sup>66</sup> (which is less well appreciated). Second, this study was conducted in 9- and 10-year old youth, many of whom had relatively low levels of psychiatric symptomatology. It is expected that subjects' psychopathology load will rise during the course of adolescence<sup>27</sup>, and it is possible that brain/behavior relationships will correspondingly increase, a hypothesis that can be directly tested in future waves of longitudinal ABCD data. Third, this study exclusively examined brain/behavior relationships with resting state functional connectomes. Other lines of work link the P factor to structural alterations<sup>67</sup> (gray matter thinning) and white matter tract changes<sup>68,69</sup>. Thus, it is possible that multi-modal fusion methods that aggregate information across modalities will further boost the size of brain/behavior relationships.

In sum, using a rigorous multivariate approach combined with cross-validation in held out samples, we established that during emerging adolescence, the P factor is associated with distributed alterations in control networks and DMN, critical elements of the brain's intrinsic functional architecture. These findings set the stage for future studies in the ABCD sample, leveraging longitudinal waves of data to trace the psychological and neural progression of the P factor during a critical window of vulnerability to mental illness that spans adolescence to young adulthood.

## Competing Interests

The authors declare no conflicts of interest.

## Acknowledgements

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 U01DA041022, U01DA041028, U01DA041048, U01DA041089, U01DA041106, U01DA041117, U01DA041120, U01DA041134, U01DA041148, U01DA041156, U01DA041174, U24DA041123, and U24DA041147. A full list of supporters is available at <https://abcdstudy.org/nih-collaborators>. A listing of participating sites and a complete listing of the study investigators can be found at <https://abcdstudy.org/principal-investigators.html>. 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 721, 10.15154/1504041, which can be found at <https://nda.nih.gov/study.html?id=721>.

This work was supported by the following grants from the United States National Institutes of Health, the National Institute on Drug Abuse, and the National Institute on Alcohol Abuse and

Alcoholism: R01MH107741 (CS), U01DA041106 (CS, LH, MH, BH), T32 AA007477 (DC, AW). In addition, CS was supported by a grant from the Dana Foundation David Mahoney Neuroimaging Program. This research was supported in part through computational resources and services provided by Advanced Research Computing at the University of Michigan, Ann Arbor.

## References

1. Lahey BB, Applegate B, Hakes JK, Zald DH, Hariri AR, Rathouz PJ. Is there a general factor of prevalent psychopathology during adulthood? *Journal of abnormal psychology*. 2012;121(4):971.
2. Caspi A, Houts RM, Belsky DW, et al. The p factor: one general psychopathology factor in the structure of psychiatric disorders? *Clinical Psychological Science*. 2014;2(2):119-137.
3. Caspi A, Moffitt TE. All for one and one for all: Mental disorders in one dimension. *American Journal of Psychiatry*. 2018;175(9):831-844.
4. Smith GT, Atkinson EA, Davis HA, Riley EN, Oltmanns JR. The general factor of psychopathology. *Annual Review of Clinical Psychology*. 2020;16:75-98.
5. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of general psychiatry*. 2005;62(6):593-602.
6. Krueger RF, Markon KE. Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annu Rev Clin Psychol*. 2006;2:111-133.
7. McTeague LM, Goodkind MS, Etkin A. Transdiagnostic impairment of cognitive control in mental illness. *Journal of psychiatric research*. 2016;83:37-46.
8. McTeague LM, Huemer J, Carreon DM, Jiang Y, Eickhoff SB, Etkin A. Identification of common neural circuit disruptions in cognitive control across psychiatric disorders. *American Journal of Psychiatry*. 2017;174(7):676-685.
9. Gandal MJ, Haney JR, Parikshak NN, et al. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science*. 2018;359(6376):693-697.
10. Serretti A, Fabbri C. Shared genetics among major psychiatric disorders. *The Lancet*. 2013;381(9875):1339-1341.
11. Bassett DS, Sporns O. Network neuroscience. *Nature neuroscience*. 2017;20(3):353.
12. Sporns O. The human connectome: a complex network. *Annals of the New York Academy of Sciences*. 2011;1224(1):109-125. doi:10.1111/j.1749-6632.2010.05888.x
13. Sporns O. From simple graphs to the connectome: Networks in neuroimaging. *Neuroimage*. Published online September 10, 2011. doi:10.1016/j.neuroimage.2011.08.085
14. Sporns O. Contributions and challenges for network models in cognitive neuroscience. *Nat Neurosci*. 2014;17(5):652-660. doi:10.1038/nn.3690

15. Bressler SL, Menon V. Large-scale brain networks in cognition: emerging methods and principles. *Trends in Cognitive Sciences*. 2010;14(6):277-290. doi:10.1016/j.tics.2010.04.004
16. Menon V. Large-scale brain networks and psychopathology: a unifying triple network model. *Trends Cogn Sci (Regul Ed)*. 2011;15(10):483-506. doi:10.1016/j.tics.2011.08.003
17. Carver CS, Johnson SL, Timpano KR. Toward a functional view of the p factor in psychopathology. *Clinical Psychological Science*. 2017;5(5):880-889.
18. Deutz MH, Geeraerts SB, Belsky J, et al. General psychopathology and dysregulation profile in a longitudinal community sample: Stability, antecedents and outcomes. *Child Psychiatry & Human Development*. 2020;51(1):114-126.
19. Cole MW, Schneider W. The cognitive control network: Integrated cortical regions with dissociable functions. *NeuroImage*. 2007;37(1):343-360. doi:10.1016/j.neuroimage.2007.03.071
20. Gratton C, Sun H, Petersen SE. Control networks and hubs. *Psychophysiology*. 2018;55(3):e13032.
21. Casey BJ, Oliveri ME, Insel T. A Neurodevelopmental Perspective on the Research Domain Criteria (RDoC) Framework. *Biological Psychiatry*. 2014;76(5):350-353. doi:10.1016/j.biopsych.2014.01.006
22. Menon V. Developmental pathways to functional brain networks: emerging principles. *Trends Cogn Sci (Regul Ed)*. 2013;17(12):627-640. doi:10.1016/j.tics.2013.09.015
23. Paus T. Mapping brain maturation and cognitive development during adolescence. *Trends in Cognitive Sciences*. 2005;9(2):60-68. doi:10.1016/j.tics.2004.12.008
24. Casey BJ, Tottenham N, Liston C, Durston S. Imaging the developing brain: what have we learned about cognitive development? *Trends in Cognitive Sciences*. 2005;9(3):104-110. doi:10.1016/j.tics.2005.01.011
25. Durston S, Casey BJ. What have we learned about cognitive development from neuroimaging? *Neuropsychologia*. 2006;44(11):2149-2157.
26. Grayson DS, Fair DA. Development of large-scale functional networks from birth to adulthood: A guide to the neuroimaging literature. *NeuroImage*. 2017;160:15-31. doi:10.1016/j.neuroimage.2017.01.079
27. Paus T, Keshavan M, Giedd JN. Why do many psychiatric disorders emerge during adolescence? *Nat Rev Neurosci*. 2008;9(12):947-957. doi:10.1038/nrn2513

28. Volkow ND, Koob GF, Croyle RT, et al. The conception of the ABCD study: From substance use to a broad NIH collaboration. *Developmental cognitive neuroscience*. 2018;32:4-7.
29. Karcher NR, Barch DM. The ABCD study: understanding the development of risk for mental and physical health outcomes. *Neuropsychopharmacology*. Published online 2020:1-13.
30. Clark DA, Hicks BM, Angstadt M, et al. The General Factor of Psychopathology in the Adolescent Brain Cognitive Development (ABCD) Study: A Comparison of Alternative Modeling Approaches. *Clinical Psychological Science*. Published online forthcoming.
31. Achenbach TM, Ruffle TM. The Child Behavior Checklist and related forms for assessing behavioral/emotional problems and competencies. *Pediatrics in review*. 2000;21(8):265-271.
32. Sripada C, Angstadt M, Rutherford S, et al. Basic Units of Inter-Individual Variation in Resting State Connectomes. *Scientific Reports*. 2019;9(1):1900. doi:10.1038/s41598-018-38406-5
33. Sripada C, Rutherford S, Angstadt M, et al. Prediction of Neurocognition in Youth From Resting State fMRI. *Molecular Psychiatry*. Published online 2019:1-19. doi:10.1101/495267
34. Sripada C, Angstadt M, Rutherford S, Taxali A, Shedd K. Toward a “treadmill test” for cognition: Improved prediction of general cognitive ability from the task activated brain. *Human Brain Mapping*. Published online 2020.
35. Garavan H, Bartsch H, Conway K, et al. Recruiting the ABCD sample: design considerations and procedures. *Developmental cognitive neuroscience*. 2018;32:16-22.
36. Barch DM, Albaugh MD, Avenevoli S, et al. Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: Rationale and description. *Developmental cognitive neuroscience*. Published online 2017.
37. Casey BJ, Cannonier T, Conley MI, et al. The adolescent brain cognitive development (ABCD) study: imaging acquisition across 21 sites. *Developmental cognitive neuroscience*. Published online 2018.
38. Hagler DJ, Hatton SN, Makowski C, et al. Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *bioRxiv*. Published online November 4, 2018:457739. doi:10.1101/457739
39. Esteban O, Markiewicz CJ, Blair RW, et al. fMRIprep: a robust preprocessing pipeline for functional MRI. *Nature methods*. 2019;16(1):111.

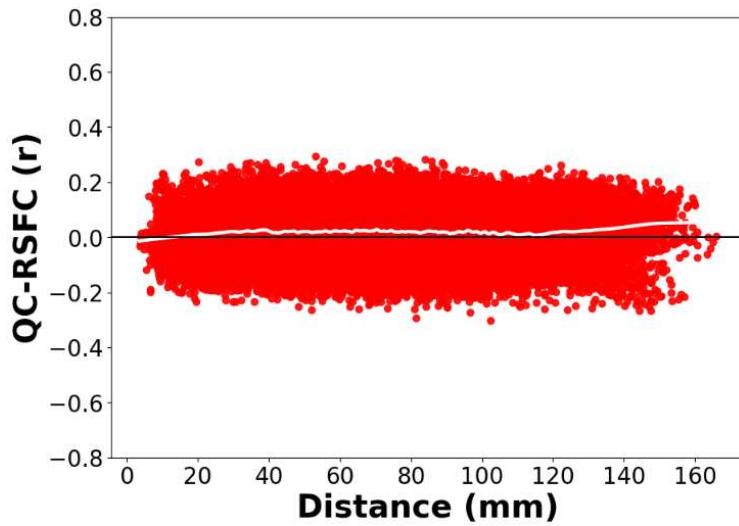
40. Gorgolewski K, Burns CD, Madison C, et al. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in neuroinformatics*. 2011;5:13.
41. Gordon EM, Laumann TO, Adeyemo B, Huckins JF, Kelley WM, Petersen SE. Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cereb Cortex*. 2016;26(1):288-303. doi:10.1093/cercor/bhu239
42. Tian Y, Margulies DS, Breakspear M, Zalesky A. Hierarchical organization of the human subcortex unveiled with functional connectivity gradients. *bioRxiv*. Published online 2020.
43. Diedrichsen J, Maderwald S, Küper M, et al. Imaging the deep cerebellar nuclei: a probabilistic atlas and normalization procedure. *Neuroimage*. 2011;54(3):1786-1794.
44. Lindquist MA, Geuter S, Wager TD, Caffo BS. Modular preprocessing pipelines can reintroduce artifacts into fMRI data. *Human brain mapping*. 2019;40(8):2358-2376.
45. Brislin S, Martz ME, Joshi S, et al. Differentiated Nomological Networks of Internalizing, Externalizing, and the General Factor of Psychopathology ("P factor") in Emerging Adolescence in the ABCD study. Published online 2020.
46. Kessler D, Angstadt M, Sripada C. Brain Network Growth Charting and the Identification of Attention Impairment in Youth. *JAMA Psychiatry*. 2016;73(5):481-489.
47. Heeringa SG, Berglund PA. A guide for population-based analysis of the Adolescent Brain Cognitive Development (ABCD) Study baseline data. *BioRxiv*. Published online 2020.
48. Castellanos FX, Di Martino A, Craddock RC, Mehta AD, Milham MP. Clinical applications of the functional connectome. *NeuroImage*. 2013;80:527-540. doi:10.1016/j.neuroimage.2013.04.083
49. Assaf M, Jagannathan K, Calhoun VD, et al. Abnormal functional connectivity of default mode sub-networks in autism spectrum disorder patients. *Neuroimage*. 2010;53(1):247-256.
50. Pankow A, Deserno L, Walter M, et al. Reduced default mode network connectivity in schizophrenia patients. *Schizophrenia research*. 2015;165(1):90-93.
51. Fair DA, Posner J, Nagel BJ, et al. Atypical Default Network Connectivity in Youth with Attention-Deficit/Hyperactivity Disorder. *Biological Psychiatry*. 2010;68(12):1084-1091. doi:10.1016/j.biopsych.2010.07.003
52. Sripada C, Kessler D, Fang Y, Welsh R, Kumar KP, Angstadt M. Disrupted network architecture of the resting brain in attention-deficit/hyperactivity disorder. *Human Brain Mapping*. 2014;35(9):4693-4705.

53. Zhou Y, Liang M, Tian L, et al. Functional disintegration in paranoid schizophrenia using resting-state fMRI. *Schizophrenia research*. 2007;97(1-3):194-205.
54. Whitfield-Gabrieli S, Thermenos HW, Milanovic S, et al. Hyperactivity and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia. *Proceedings of the National Academy of Sciences*. 2009;106(4):1279-1284.
55. Whitfield-Gabrieli S, Ford JM. Default mode network activity and connectivity in psychopathology. *Annual review of clinical psychology*. 2012;8:49-76.
56. Chai XJ, Whitfield-Gabrieli S, Shinn AK, et al. Abnormal medial prefrontal cortex resting-state connectivity in bipolar disorder and schizophrenia. *Neuropsychopharmacology*. 2011;36(10):2009-2017.
57. Fair DA, Cohen AL, Dosenbach NUF, et al. The maturing architecture of the brain's default network. *PNAS*. 2008;105(10):4028-4032. doi:10.1073/pnas.0800376105
58. Fair DA, Dosenbach NUF, Church JA, et al. Development of distinct control networks through segregation and integration. *Proc Natl Acad Sci USA*. 2007;104(33):13507-13512. doi:10.1073/pnas.0705843104
59. Sripada C, Kessler D, Angstadt M. Lag in maturation of the brain's intrinsic functional architecture in attention-deficit/hyperactivity disorder. *PNAS*. 2014;111(39):14259-14264. doi:10.1073/pnas.1407787111
60. Di Martino A, Fair DA, Kelly C, et al. Unraveling the Miswired Connectome: A Developmental Perspective. *Neuron*. 2014;83(6):1335-1353. doi:10.1016/j.neuron.2014.08.050
61. Elliott ML, Romer A, Knodt AR, Hariri AR. A connectome-wide functional signature of transdiagnostic risk for mental illness. *Biological psychiatry*. 2018;84(6):452-459.
62. Kebets V, Holmes AJ, Orban C, et al. Somatosensory-Motor Dysconnectivity Spans Multiple Transdiagnostic Dimensions of Psychopathology. *Biological Psychiatry*. Published online 2019.
63. Andrews-Hanna JR, Smallwood J, Spreng RN. The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Ann NY Acad Sci*. 2014;1316(1):29-52. doi:10.1111/nyas.12360
64. Cole MW, Repovš G, Anticevic A. The frontoparietal control system: a central role in mental health. *The Neuroscientist*. 2014;20(6):652-664.

65. Cocchi L, Zalesky A, Fornito A, Mattingley JB. Dynamic cooperation and competition between brain systems during cognitive control. *Trends in cognitive sciences*. 2013;17(10):493-501.
66. Lindquist MA, Gelman A. Correlations and multiple comparisons in functional imaging: a statistical perspective (Commentary on Vul et al., 2009). *Perspectives on Psychological Science*. 2009;4(3):310-313.
67. Romer AL, Elliott ML, Knott AR, et al. Pervasively thinner neocortex as a transdiagnostic feature of general psychopathology. *American Journal of Psychiatry*. Published online 2020:appi. *ajp*. 2020.19090934.
68. Kaufmann T, Alnæs D, Doan NT, Brandt CL, Andreassen OA, Westlye LT. Delayed stabilization and individualization in connectome development are related to psychiatric disorders. *Nature neuroscience*. 2017;20(4):513-515.
69. Alnæs D, Kaufmann T, Doan NT, et al. Association of heritable cognitive ability and psychopathology with white matter properties in children and adolescents. *JAMA psychiatry*. 2018;75(3):287-295.

# Supplement

## Quality Control-Resting State Functional Connectivity Plot



We used multiple procedures to control the effect of head motion on brain-behavior relationships, which are listed in §2.2 in the Main Manuscript. To assess the effectiveness of these procedures, we produced a quality control resting state functional connectivity (QC-RSFC) plot<sup>1,2</sup>. This plot shows the relationship between mean framewise displacement and connectivity for edges binned by distance. Motion effects produce a sloped line (distance-dependent artifact), while a flat line is indicative of minimal motion-related effects. The RSFC-QC plot for our ABCD resting state data showed a flat line, providing additional evidence that our stringent motion correction strategies were effective.

### P Factor Modeling

A general P-factor with two orthogonal specific factors were modeled using the parent-rated Child Behavior Checklist (CBCL; age 6 to 18 form<sup>3</sup>). We fit a bifactor model to the eight CBCL scales (Withdrawn, Somatic Complaints, Anxious/Depressed, Social Problems, Thought Problems, Attention Problems, Delinquent Behavior, and Aggressive Behavior). In this model there was a general P factor that all scales loaded onto (average scale loading on P = .69), and internalizing and externalizing specific factors (average scale loading = .43). This model fit well based on conventional fit thresholds ( $\chi^2 = 747.73$ ,  $df = 16$ ,  $p < .001$ ; RMSEA = .062; CFI = .985; TLI = .974; SRMR = .015) and was chosen for its good model fit and theoretical interpretability. Importantly, we elsewhere demonstrate that across a variety of alternative specifications of hierarchical models of psychopathology in ABCD, the resulting P factors that emerge yield highly similar rank ordering of subjects ( $rs > 0.90$ )<sup>4</sup>, rendering decisions about which specific modeling strategy to adopt less consequential.

## Network Visualizations

We created circular network visualizations of whole brain connectomes (see Figures 2 and 3 in the Main Manuscript) using the circlize software library in R. To increase interpretability, SMH and SMM were combined into a single somatomotor network (SM) and five networks that were not extensively implicated were omitted (SAL, SC, CER, CP, RST). In addition, cells in which suprathreshold connections were less than 7.5% of the total possible connections were omitted.

## Permutation Framework

The distribution under chance of correlations between BBS-based predictions of neurocognitive scores and observed neurocognitive scores was generated by randomly permuting the 5,880 subjects' P factor scores 10,000 times. At each iteration, we performed the leave-one-site out cross validation procedure described in the main manuscript, including refitting BBS models at each fold of the cross-validation. We then recalculated the average correlation across folds between predicted versus actual P factor scores. The average correlation across folds that was actually observed was located in this null distribution in terms of rank, and statistical significance was set as this rank value divided by 10,000.

Since the BBS models fit at each iteration of the permutation test included covariates, the procedure of Freedman and Lane<sup>5</sup> was followed. In brief, a BBS model was first estimated with nuisance covariates alone, residuals were formed and were permuted. The covariate effect of interest was then included in the subsequent model, creating an approximate realization of data under the null hypothesis, and the statistical test of interest was calculated on this data (see FSL Randomise <http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Randomise/Theory> for a neuroimaging implementation).

## Robustness Checks

We performed a number of additional analyses to assess the robustness of our results under alternative model specifications. One alternative approach derived P factor scores with strict separation between training and testing partitions. In particular, for the train partition subjects, we fit a bifactor model to CBCL subscales as described in the Main Manuscript (§2.3) and derived P factor scores. We then applied this same fitted model to the held-out test subjects to obtain their P factor scores. This process was repeated for all folds of the leave-one-site-out cross-validation.

	Included	Excluded
N	5880	5995
Age (mean (s.d.))	9.97 (0.62)	9.88 (0.62)
Female (%)	2973 (50.6)	2708 (45.2)
Race Ethnicity (%)		
White	3543 (60.2)	2581 (43.1)
Black	663 (11.3)	1095 (18.3)
Hispanic	1036 (17.6)	1343 (22.4)
Asian	86 (1.5)	177 (3.0)
Other	552 (9.4)	642 (10.7)
No answer	--	157 (2.6)
Highest Parental Education (%)		
< HS Diploma	173 (2.9)	410 (6.8)
Bachelor	1649 (28.0)	1339 (22.3)
HS Diploma/GED	386 (6.6)	734 (12.2)
Post Graduate Degree	2223 (37.8)	1782 (29.7)
Some College	1449 (24.6)	1590 (26.5)
No answer	--	140 (2.3)
Household Marital Status – Married (%)	4305 (73.2)	3604(60.1)
Household Income (%)		
<50K	1486 (25.3)	1692 (28.2)
>=100k	2617 (44.5)	1941 (31.9)
>=50k & <100K	1777 (30.2)	1261 (21.0)
No answer	--	1128 (18.8)

**Supplemental Table 1: Demographic Characteristics of Included Versus Excluded Subjects**

## Supplemental Methods

### fMRIprep Methods

The following was generated automatically by fMRIprep software and is copied here unchanged:

Results included in this manuscript come from preprocessing performed using *fMRIprep* 1.5.0 (Esteban, Markiewicz, et al. (2018); Esteban, Blair, et al. (2018); RRID:SCR\_016216), which is based on *Nipype* 1.2.2 (Gorgolewski et al. (2011); Gorgolewski et al. (2018); RRID:SCR\_002502).

#### Anatomical data preprocessing

The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al. 2010), distributed with ANTs 2.2.0 (Avants et al. 2008, RRID:SCR\_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a *Nipype* implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTS as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL 5.0.9, RRID:SCR\_002823, Zhang, Brady, and Smith 2001). Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1, RRID:SCR\_001847, Dale, Fischl, and Sereno 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR\_002438, Klein et al. 2017). Volume-based spatial normalization to one standard space (MNI152NLin6Asym) was performed through nonlinear registration with antsRegistration (ANTs 2.2.0), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: *FSL's MNI ICBM 152 non-linear 6th Generation Asymmetric Average Brain Stereotaxic Registration Model* [Evans et al. (2012), RRID:SCR\_002823; TemplateFlow ID: MNI152NLin6Asym].

#### Functional data preprocessing

For each of the 10 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIprep*. A deformation field to correct for susceptibility distortions was estimated based on two echo-planar imaging (EPI) references with opposing phase-encoding directions, using 3dQwarp Cox and Hyde (1997) (AFNI 20160207). Based on the estimated susceptibility distortion, an unwarped BOLD reference was calculated for a more accurate co-registration with the anatomical reference. The BOLD reference was then co-registered to the T1w reference using

bbregister (FreeSurfer) which implements boundary-based registration (Greve and Fischl 2009). Co-registration was configured with six degrees of freedom. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL 5.0.9, Jenkinson et al. 2002). The BOLD time-series, were resampled to surfaces on the following spaces: *fsaverage5*. The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time-series will be referred to as *preprocessed BOLD in original space*, or just *preprocessed BOLD*. The BOLD time-series were resampled into standard space, generating a *preprocessed BOLD run in ['MNI152NLin6Asym'] space*. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIprep*. Automatic removal of motion artifacts using independent component analysis (ICA-AROMA, Pruim et al. 2015) was performed on the *preprocessed BOLD on MNI space* time-series after removal of non-steady state volumes and spatial smoothing with an isotropic, Gaussian kernel of 6mm FWHM (full-width half-maximum). Corresponding “non-aggressively” denoised runs were produced after such smoothing. Additionally, the “aggressive” noise-regressors were collected and placed in the corresponding confounds file. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* (following the definitions by Power et al. 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (*CompCor*, Behzadi et al. 2007). Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 5% variable voxels within a mask covering the subcortical regions. This subcortical mask is obtained by heavily eroding the brain mask, which ensures it does not include cortical GM regions. For aCompCor, components are calculated within the intersection of the aforementioned mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the  $k$  components with the largest singular values are retained, such that the retained components’ time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al. 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. All resamplings can be performed with *a single*

*interpolation step* by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using `antsApplyTransforms` (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos 1964). Non-gridded (surface) resamplings were performed using `mri_vol2surf` (FreeSurfer).

Many internal operations of `fMRIPrep` use *Nilearn* 0.5.2 (Abraham et al. 2014, RRID:SCR\_001362), mostly within the functional processing workflow. For more details of the pipeline, see [the section corresponding to workflows in `fMRIPrep`'s documentation](#).

## Copyright Waiver

The above boilerplate text was automatically generated by `fMRIPrep` with the express intention that users should copy and paste this text into their manuscripts *unchanged*. It is released under the [CC0](#) license.

## References

Abraham, Alexandre, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gael Varoquaux. 2014. “Machine Learning for Neuroimaging with Scikit-Learn.” *Frontiers in Neuroinformatics* 8. <https://doi.org/10.3389/fninf.2014.00014>.

Avants, B.B., C.L. Epstein, M. Grossman, and J.C. Gee. 2008. “Symmetric Diffeomorphic Image Registration with Cross-Correlation: Evaluating Automated Labeling of Elderly and Neurodegenerative Brain.” *Medical Image Analysis* 12 (1): 26–41. <https://doi.org/10.1016/j.media.2007.06.004>.

Behzadi, Yashar, Khaled Restom, Joy Liau, and Thomas T. Liu. 2007. “A Component Based Noise Correction Method (CompCor) for BOLD and Perfusion Based fMRI.” *NeuroImage* 37 (1): 90–101. <https://doi.org/10.1016/j.neuroimage.2007.04.042>.

Cox, Robert W., and James S. Hyde. 1997. “Software Tools for Analysis and Visualization of fMRI Data.” *NMR in Biomedicine* 10 (4-5): 171–78. [https://doi.org/10.1002/\(SICI\)1099-1492\(199706/08\)10:4/5<171::AID-NBM453>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1099-1492(199706/08)10:4/5<171::AID-NBM453>3.0.CO;2-L).

Dale, Anders M., Bruce Fischl, and Martin I. Sereno. 1999. “Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction.” *NeuroImage* 9 (2): 179–94. <https://doi.org/10.1006/nimg.1998.0395>.

Esteban, Oscar, Ross Blair, Christopher J. Markiewicz, Shoshana L. Berleant, Craig Moodie, Feilong Ma, Ayse Ilkay Isik, et al. 2018. “`fMRIPrep`.” *Software*. Zenodo. <https://doi.org/10.5281/zenodo.852659>.

Esteban, Oscar, Christopher Markiewicz, Ross W Blair, Craig Moodie, Ayse Ilkay Isik, Asier Erramuzpe Aliaga, James Kent, et al. 2018. “fMRIprep: A Robust Preprocessing Pipeline for Functional MRI.” *Nature Methods*. <https://doi.org/10.1038/s41592-018-0235-4>.

Evans, AC, AL Janke, DL Collins, and S Baillet. 2012. “Brain Templates and Atlases.” *NeuroImage* 62 (2): 911–22. <https://doi.org/10.1016/j.neuroimage.2012.01.024>.

Gorgolewski, K., C. D. Burns, C. Madison, D. Clark, Y. O. Halchenko, M. L. Waskom, and S. Ghosh. 2011. “Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python.” *Frontiers in Neuroinformatics* 5: 13. <https://doi.org/10.3389/fninf.2011.00013>.

Gorgolewski, Krzysztof J., Oscar Esteban, Christopher J. Markiewicz, Erik Ziegler, David Gage Ellis, Michael Philipp Notter, Dorota Jarecka, et al. 2018. “Nipype.” *Software*. Zenodo. <https://doi.org/10.5281/zenodo.596855>.

Greve, Douglas N, and Bruce Fischl. 2009. “Accurate and Robust Brain Image Alignment Using Boundary-Based Registration.” *NeuroImage* 48 (1): 63–72. <https://doi.org/10.1016/j.neuroimage.2009.06.060>.

Jenkinson, Mark, Peter Bannister, Michael Brady, and Stephen Smith. 2002. “Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images.” *NeuroImage* 17 (2): 825–41. <https://doi.org/10.1006/nimg.2002.1132>.

Klein, Arno, Satrajit S. Ghosh, Forrest S. Bao, Joachim Giard, Yrjö Häme, Eliezer Stavsky, Noah Lee, et al. 2017. “Mindboggling Morphometry of Human Brains.” *PLOS Computational Biology* 13 (2): e1005350. <https://doi.org/10.1371/journal.pcbi.1005350>.

Lanczos, C. 1964. “Evaluation of Noisy Data.” *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis* 1 (1): 76–85. <https://doi.org/10.1137/0701007>.

Power, Jonathan D., Anish Mitra, Timothy O. Laumann, Abraham Z. Snyder, Bradley L. Schlaggar, and Steven E. Petersen. 2014. “Methods to Detect, Characterize, and Remove Motion Artifact in Resting State fMRI.” *NeuroImage* 84 (Supplement C): 320–41. <https://doi.org/10.1016/j.neuroimage.2013.08.048>.

Pruim, Raimon H. R., Maarten Mennes, Daan van Rooij, Alberto Llera, Jan K. Buitelaar, and Christian F. Beckmann. 2015. “ICA-AROMA: A Robust ICA-Based Strategy for Removing Motion Artifacts from fMRI Data.” *NeuroImage* 112 (Supplement C): 267–77. <https://doi.org/10.1016/j.neuroimage.2015.02.064>.

Satterthwaite, Theodore D., Mark A. Elliott, Raphael T. Gerraty, Kosha Ruparel, James Loughead, Monica E. Calkins, Simon B. Eickhoff, et al. 2013. “An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-

state functional connectivity data.” *NeuroImage* 64 (1): 240–56.  
<https://doi.org/10.1016/j.neuroimage.2012.08.052>.

Tustison, N. J., B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee. 2010. “N4ITK: Improved N3 Bias Correction.” *IEEE Transactions on Medical Imaging* 29 (6): 1310–20.  
<https://doi.org/10.1109/TMI.2010.2046908>.

Zhang, Y., M. Brady, and S. Smith. 2001. “Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm.” *IEEE Transactions on Medical Imaging* 20 (1): 45–57. <https://doi.org/10.1109/42.906424>.