RUNNING TITLE: Connectivity-guided dimensions of psychopathology

Linked dimensions of psychopathology and connectivity in functional brain networks

Cedric Huchuan Xia[1], Zongming Ma[2], Rastko Ciric[1], Shi Gu[1,3,4], Richard F. Betzel[3], Antonia N. Kaczkurkin[1], Monica E. Calkins[1], Philip A. Cook[5], Angel Garcia de la Garza[1], Simon Vandekar[6], Tyler M. Moore[1], David R. Roalf[1], Kosha Ruparel[1], Daniel H. Wolf[1], Christos Davatzikos[5], Ruben C. Gur[1,5], Raquel E. Gur[1,5], Russell T. Shinohara[6], Danielle S. Bassett[3,7], & Theodore D. Satterthwaite[1*]

[1]Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA, [2]Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA, 19104, USA, [3]Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA, 19104, USA [4]Department of Computer Science and Engineering, University of Electronic Science and Technology, Chengdu, Sichuan, 611731, China [5]Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA, [6]Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA, [7]Department of Electrical and Systems Engineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA, 19104, USA;

[*]**Correspondence:**

Theodore D. Satterthwaite
Department of Psychiatry
Perelman School of Medicine
University of Pennsylvania
Philadelphia, PA, USA
sattertt@pennmedicine.upenn.edu

ABSTRACT

Neurobiological abnormalities associated with psychiatric disorders do not map well to existing diagnostic categories. High co-morbidity and overlapping symptom domains suggest dimensional circuit-level abnormalities that cut across clinical diagnoses. Here we sought to identify brain-based dimensions of psychopathology using multivariate sparse canonical correlation analysis (sCCA) in a sample of 663 youths imaged as part of the Philadelphia Neurodevelopmental Cohort. This analysis revealed highly correlated patterns of functional connectivity and psychiatric symptoms. We found that four dimensions of psychopathology — mood, psychosis, fear, and externalizing behavior — were highly associated (r=0.68-0.71) with distinct patterns of functional dysconnectivity. Loss of network segregation between the default mode network and executive networks (e.g. fronto-parietal and salience) emerged as a common feature across all dimensions. Connectivity patterns linked to mood and psychosis became more prominent with development, and significant sex differences were present for connectivity patterns related to mood and fear. Critically, findings replicated in an independent dataset (n=336). These results delineate connectivity-guided dimensions of psychopathology that cut across traditional diagnostic categories, which could serve as a foundation for developing network-based biomarkers in psychiatry.

INTRODUCTION

Psychiatry relies on signs and symptoms for clinical decision making, while other branches of medicine are transitioning to the use of biomarkers to aid in diagnosis, prognosis, and treatment selection. The search for biomarkers in psychiatry has intensified,[1] and it is increasingly recognized that existing clinical diagnostic categories could hinder this effort, as they do not pair well with distinct neurobiological abnormalities.[2–4] The high co-morbidity among psychiatric disorders exacerbates this problem.[5] Furthermore, studies have demonstrated common structural, functional, and genetic abnormalities across psychiatric syndromes, potentially explaining such co-morbidity.[6–10] This body of evidence underscores the lack of direct mapping between clinical diagnostic categories and the underlying pathophysiology, potentially leading to dramatic changes to treatment strategies for psychiatric disorders.

This context has motivated the development of the National Institute of Mental Health's Research Domain Criteria, which seek to construct a biologically-grounded framework for neuropsychiatric diseases.[11,12] In such a model, the symptoms of individual patients are conceptualized as the result of mixed dimensional abnormalities of specific brain circuits. While such a model system is theoretically attractive, it has been challenging to implement in practice due to both the multiplicity of clinical symptoms and the many brain systems implicated in psychiatric disorders.[13,14]

Network neuroscience is a powerful approach for examining brain systems implicated in psychopathology.[15–17] One network property commonly evaluated is its community structure, or modular architecture. A network module (also called a sub-network or a community) is a group of densely interconnected nodes, which may form the basis for specialized sub-units of information processing.

Converging results across data sets, methods, and laboratories provide substantial agreement on large-scale functional brain modules such as the somatomotor, visual, default mode, and fronto-parietal control networks.[18–26] Furthermore, multiple studies documented abnormalities within this modular topology in psychiatric disorders.[16,27,28] Specifically, evidence suggests that many psychiatric disorders are associated with abnormalities in network modules subserving higher-order cognitive processes, including the default mode and fronto-parietal control networks.[29]

In addition to such module-specific deficits, studies in mood disorders,[30–32] psychosis,[28,33–35] and other disorders[36,37] have reported abnormal interactions *between* modules that are typically segregated from each other at rest. This is of particular interest as modular segregation of both functional[19,38,39] and structural[40] brain networks is refined during adolescence, a critical period when many neuropsychiatric disorders emerge. Such findings have led many disorders to be considered "neurodevelopmental connectopathies."[41–44] Describing the developmental substrates of neuropsychiatric disorders is a necessary step towards early identification of at-risk youth, and might ultimately allow for interventions that "bend the curve" of maturation to achieve improved functional outcomes.[45]

Despite the increasing interest in describing how abnormalities of brain network development lead to the emergence of neuropsychiatric disorders, existing studies have been limited in several respects. First, most adopted either a categorical case-control approach, or only examined a single dimension of psychopathology. Second, especially in contrast to adult studies, existing work in youth has often used relatively small samples (e.g. dozens of participants). While multivariate techniques could allow examination of both multiple brain systems and clinical dimensions simultaneously, such techniques usually require large samples.

In the current study, we sought to delineate functional network abnormalities associated with a broad array of psychopathology in youth. We capitalized on a large sample of youth from the Philadelphia Neurodevelopmental Cohort (PNC)[46] applying a recently-developed machine learning technique called sparse canonical correlation analysis (sCCA).[47] As a multivariate method, sCCA is capable of discovering complex linear relationships between two high-dimensional datasets.[48,49] Here, we used sCCA to delineate linked dimensions of psychopathology and functional connectivity. As described below, we uncovered dimensions of dysconnectivity that were highly correlated with specific, interpretable dimensions of psychopathology. We found that each psychopathological dimension was associated with a pattern of abnormal connectivity, and that all dimensions were characterized by decreased segregation of default mode and executive networks (fronto-parietal and salience). These network features linked to each dimension of psychopathology showed expected developmental changes and sex differences. Finally, our results were replicated in an independent dataset.

RESULTS

We sought to delineate multivariate relationships between functional connectivity and psychiatric symptoms in a large sample of youth. To do this, we used sCCA, an unsupervised learning technique that seeks to find correlations between two high-dimensional datasets.[47] In total, we studied 999 participants ages 8-22 who completed both functional neuroimaging and a comprehensive evaluation of psychiatric symptoms as part of the PNC.[46,50] We divided this sample into discovery (n=663) and replication datasets (n=336) that were matched on age, sex, race, and overall psychopathology (**Supplementary Fig. 1** and **Supplementary Table 1**). Following pre-processing using a validated pipeline that minimizes the impact of in-scanner motion,[51]

we constructed subject-level functional networks using a 264-node parcellation system[19] that includes an *a priori* assignment of nodes to network communities (**Fig. 1a-c**. e.g. modules or sub-networks; see Online Methods). Prior to analysis with sCCA, we regressed age, sex, race, and motion out of both the connectivity and clinical data to ensure that these potential confounders did not drive results. As features that do not vary across subjects cannot be predictive of individual differences, we limited our analysis of connectivity data to the top 10 percent most variable connections (ranked by median absolute deviation, see Online Methods and **Supplementary Fig. 2)**. The input data thus consisted of 3410 unique functional connections (**Fig. 1b**) and 111 clinical items (**Fig. 1c** and **Supplementary Table 3**). Using elastic net regularization ($L1 + L2$), sCCA was able to obtain a sparse and interpretable model while minimizing over-fitting (**Fig. 1d** and **Supplementary Fig. 3** ; see Online Methods). Ultimately, sCCA identified specific patterns ("canonical variates") of functional connectivity that were linked to distinct combinations of psychiatric symptoms.

Multivariate analysis reveals linked dimensions of psychopathology and connectivity

Based on the scree plot of covariance explained (**Fig. 2a**), we selected the first seven canonical variates for further analysis. Significance of each of these linked dimensions of symptoms and connectivity was assessed using a permutation test (see Online Methods and **Supplementary Fig. 4)**; False Discovery Rate (FDR) was used to control for type I error rate due to multiple testing. Of these seven cannonical variates, three were significant (Pearson correlation $r = 0.71$, $P_{FDR} < 0.001$; $r = 0.70$, $P_{FDR} < 0.001$, $r = 0.68$, $P_{FDR} < 0.01$, respectively) (**Fig. 2b**), with the fourth showing a trend towards significance ($r = 0.68$, $P_{FDR} = 0.07$, $P_{uncorrected} = 0.04$). Notably, these results were robust to many different methodological choices, including the number of features entered into the initial analysis (**Supplementary Fig. 5a**), the parcellation system

(**Supplementary Fig. 5b**), and the use of regularization with elastic net versus data reduction with principal component analysis (**Supplementary Fig. 5c**).

Each canonical variate represented a distinct pattern that relates a weighted set of psychiatric symptoms to a weighted set of functional connections. Inspection of the most heavily weighted clinical symptom for each dimension provided an initial indication regarding their content (**Fig. 2c-f**). For example, "feeling sad" was the most heavily weighted clinical feature in the first dimension, while "auditory perceptions" was the most prominent symptom in the second. Next, we conducted detailed analyses to describe the clinical and connectivity features driving the observed multivariate relationships.

Interpretable, connectivity-guided dimensions of psychopathology cross clinical diagnostic categories

To understand the characteristics of each linked dimension, we used a resampling procedure to identify both clinical and connectivity features that were consistently significant across subsets of the data (Online Methods and see **Supplementary Fig. 6)**. This procedure revealed that 37 out of 111 psychiatric symptoms reliably contributed to at least one of the four dimensions (**Fig. 3)**. Next, we mapped these data-driven items to typical clinical diagnostic categories. This revealed that the features selected by multivariate analyses generally accord with clinical phenomenology. Specifically, despite being selected on the basis of their relationship with functional connectivity, the first three canonical variates delineated dimensions that resemble clinically coherent dimensions of mood, psychosis, and fear (e.g. phobias). The fourth dimension, which was present at an uncorrected threshold, mapped to externalizing behaviors (attention deficit/hyperactivity disorder (ADHD) and oppositional defiant disorder (ODD)).

While each canonical variate mapped onto coherent clinical features, each dimension contained symptoms from several different clinical diagnostic categories. For example, the mood dimension was comprised of symptoms from categorical domains of depression ("feeling sad" received the highest loading), mania ("irritability"), and obsessive-compulsive disorder (OCD; "recurrent thoughts of harming self or others") (**Fig. 3a**). Similarly, while the second dimension mostly consisted of psychosis-spectrum symptoms (such as "auditory verbal hallucinations"), two manic symptoms (i.e. "overly energetic" and "pressured speech") were included as well (**Fig. 3b**). The third dimension was composed of fear symptoms, including both agoraphobia and social phobia (**Fig. 3c**). The fourth dimension was driven primarily by symptoms of both ADHD and ODD, but also included the irritability item from the depression domain (**Fig. 3d**). These data-driven dimensions of psychopathology align with clinical phenomenology, but in a dimensional fashion that does not adhere to discrete categories.

Common and dissociable patterns of dysconnectivity contribute to linked psychopathological dimensions

sCCA identified each dimension of psychopathology through shared associations between clinical data and specific patterns of dysconnectivity. Next, we investigated the loadings of connectivity features that underlie each canonical variate. To aid visualization of the high-dimensional connectivity data, we summarized loading patterns according to network communities established *a priori* by the parcellation system. Specifically, we examined patterns of both *within*-network and *between*-network connectivity (**Supplementary Fig. 7;** Online Methods), as this framework was useful in prior investigations of both brain development[46,52] and psychopathology.[53–55] This procedure revealed that the mood dimension was associated with increased connectivity *within* three networks: default mode, fronto-parietal, and salience

networks (**Fig. 4a,e,i**). However, the most heavily weighted features in the mood dimension reflected abnormalities of connectivity *between* networks. In particular, mood was associated with a lack of segregation between the default mode and both the fronto-parietal and salience networks. The psychosis dimension similarly showed elevated connectivity within the default mode network and its reduced segregation from executive networks (fronto-parietal and salience)(**Fig. 4b,f,j**). The fear dimension also showed elevated connectivity within the fronto-parietal and salience networks, but in contrast showed reduced connectivity within the default mode network itself (**Fig. 4c,g,k**). As was the case for mood and psychosis, the fear dimension exhibited a failure of segregation between the default mode and executive networks (**Fig. 4c,g,k**). Reduced connectivity within the default mode network was also present in the externalizing dimension, as was reduced segregation between default mode and executive networks (**Fig. 4d,h,l**).

The results indicate that while each canonical variate was comprised of unique patterns of dysconnectivity, there were several features that were shared across all dimensions. Such findings agree with accumulating evidence for common circuit-level dysfunction across psychiatric syndromes.[6–10] To quantitatively assess such common features, we compared overlapping results against a null distribution using permutation testing (see Online Methods). This procedure revealed an ensemble of edges that were consistently implicated across all four dimensions. These connections can be mapped to individual nodes, and revealed that the regions most impacted across all dimensions included the frontal pole, superior frontal gyrus, dorsomedial prefrontal cortex, medial temporal gyrus, and amygdala (**Fig. 5a**). Similar analysis at the level of sub-networks (**Fig. 5b**) illustrated that abnormalities of connectivity *within* the default mode and fronto-parietal networks were present in all four psychopathological dimensions (**Fig. 5c**). Furthermore, reduced segregation *between* the default mode and executive networks, such as the fronto-parietal and salience systems, was common to all dimensions. These shared connectivity features complement each

dimension-specific pattern, and offer evidence for both common and dissociable patterns of dysconnectivity associated with psychopathology.

Developmental effects and sex differences are concentrated in specific dimensions

In the above analyses, we examined multivariate associations between connectivity and psychopathology while controlling for participant age. However, given that abnormal neurodevelopment is thought to underlie many psychiatric disorders,[41–44] we next examined whether connectivity patterns significantly associated with psychopathology differ as a function of age or sex in this large developmental cohort. We repeated the analysis conducted above using connectivity and clinical features, but in this case did not regress out age and sex; race and motion were still regressed from both datasets. Notably, the dimensions derived were quite similar, with highly correlated feature weights (**Supplementary Table 2**). As in prior work,[40,56–58] developmental associations were examined using generalized additive models with penalized splines, which allows for statistically rigorous modeling of both linear and non-linear effects while minimizing over-fitting. Using this approach, we found that the brain connectivity patterns associated with both mood and psychosis became significantly more prominent with age (**Fig. 6a,b**, $P_{FDR} < 10^{-13}$, $P_{FDR} < 10^{-6}$, respectively). Additionally, brain connectivity patterns linked to mood and fear were both stronger in female participants than males (**Fig. 6c,d**, $P_{FDR} < 10^{-8}$, $P_{FDR} < 10^{-7}$, respectively). We did not observe age by sex interaction effects in any dimension.

Linked dimensions are replicated in an independent sample

Throughout our analysis of the discovery sample, we used procedures both to guard against over-fitting and to enhance the generalizablity of results (regularization, permutation testing, resampling). As a final step, we tested the replicability of our findings using an independent sample, which was left-out from all analyses described above (n=336, **Supplementary Fig. 1)**. Although this replication sample was half the size of our original discovery sample, sCCA identified four canonical variates that highly resemble the original four linked dimensions of psychopathology (with correlations of loadings between discovery and replication within 0.24 and 0.88; **Fig. 7a,b**). In the replication sample, three out of four dimensions were significant after FDR correction of permutation tests (**Supplementary Fig. 8)**.

DISCUSSION

Leveraging a large neuroimaging data set of youth and recent advances in machine learning, we discovered several multivariate patterns of functional dysconnectivity linked to interpretable dimensions of psychopathology that cross traditional diagnostic categories. These patterns of abnormal connectivity were both highly correlated and replicable in an independent dataset. While each dimension displayed a specific pattern of connectivity abnormalities, loss of network segregation between the default mode and executive networks was common to all dimensions. Furthermore, patterns of dysconnectivity displayed unique developmental effects and sex differences. Together, these results suggest that complex psychiatric symptoms are associated with specific patterns of abnormal connectivity during brain development.

Both the co-morbidity among psychiatric diagnoses and the notable heterogeneity within each diagnostic category suggest that our current symptom-based diagnostic criteria do not "carve nature at its joints".[2–4] Establishing biologically-targeted interventions in psychiatry is predicated upon delineation of the underlying neurobiology. This challenge has motivated the NIMH Research Domain Criteria (RDoC) effort, which seeks to link circuit-level abnormalities in specific brain systems to symptoms that might be present across clinical diagnoses.[11,12] Accordingly, there has been a proliferation of studies that focus on linking specific brain circuit(s) to a specific symptom dimension or behavioral measure across diagnostic categories.[9,10,29,56,59–63] However, by focusing on a single behavioral measure or symptom domain, many studies ignore the co-morbidity among psychiatric symptoms. A common way to attempt to evaluate such co-morbidity is to find latent dimensions of psychopathology using factor analysis or related techniques.[57,64,65] For example, factor analyses of clinical psychopathology have suggested the presence of dimensions including internalizing symptoms, externalizing symptoms, and psychosis symptoms.[50] While such dimensions are reliable, they are drawn entirely from the covariance structure of self-report or interview-based clinical data, and are not informed by neurobiology.

An alternative increasingly pursued is to parse heterogeneity in psychiatric conditions using multivariate analysis of biomarker data such as neuroimaging. For example, researchers have used functional connectivity[59] and gray matter density[66] to study the heterogeneity within major depressive disorder and psychotic disorders, respectively. However, most studies have principally considered only one or two clinical diagnostic categories, and typically the analytic approach yields discrete subtypes (or "biotypes"). By definition, such a design is unable to discover continuous dimensions that span multiple categories. Further, there is tension between the dimensional schema suggested by RDoC and categorical biotypes; as suggested by RDoC, it seems more plausible that psychopathology in an individual results from a mixture

of abnormalities across several brain systems. Finally, unsupervised learning approaches using only imaging data and not considering clinical data may frequently yield solutions that are difficult to interpret, and that do not align with clinical experience.

In contrast, in this study we used a multivariate analysis technique – sCCA – that allowed simultaneous consideration of clinical and functional connectivity data in a large sample with diverse psychopathology. This enabled uncovering linked dimensions of psychopathology and dysconnectivity that cross diagnostic categories yet remain clinically interpretable. In contrast to "one-view" multivariate studies[50,57,64,67] (such as factor analysis of clinical data or clustering of imaging data), the sCCA-derived clinical dimensions were explicitly selected on the basis of co-varying signals that were present as both alterations of connectivity and clinical symptoms. Such a "two-view" approach has been successfully applied in studies of neurodegenerative diseases[48] and normal brain-behavior relationships.[49]

Notably, the brain-driven dimensions described here incorporated symptoms across several diagnostic categories while remaining congruent with prevailing models of psychopathology. For example, the mood dimension was composed of items from five sections of the clinical interview: depression, mania, OCD, suicidality, and psychosis-spectrum. Despite disparate origins, the content of the items forms a clinically coherent picture, including depressed mood, anhedonia, loss of sense of self, recurrent thoughts of self harm, and irritability. Notably, symptoms of irritability were also significantly represented in the externalizing behavior dimension, suggesting that irritability may have heterogeneous, divergent neurobiological antecedents. The fear dimension, on the other hand, represents a more homogeneous picture of various types of phobias (e.g. social phobia and agoraphobia), that had little overlap with other categorical symptoms. Finally, the psychosis dimension (which was only significant in the discovery sample) was

mainly comprised of psychotic symptoms, but also included symptoms of mania. This result accords with studies demonstrating shared inheritance patterns of schizophrenia and bipolar disorder, and findings that specific common genetic variants increase risk of both disorders.[68] Instead of averaging over many clinical features within a diagnostic category, sCCA selected specific items that are most tightly linked to patterns of dysconnectivity. These groups of symptoms remained highly interpretable, and were reproducible in the replication data set.

Each of the clinical dimensions identified was highly correlated with patterns of dysconnectivity. These patterns were summarized according to their location *between* and *within* functional network modules, which has been a useful framework for understanding both brain development and psychopathology.[53–55] While each dimension of psychopathology was associated with a unique pattern of dysconnectivity, one of the most striking findings to emerge was evidence that reduction of functional segregation between the default mode and fronto-parietal networks was a common feature of all dimensions. The exact connections implicated in each dimension might vary, but permutation-based analyses demonstrated that loss of segregation between these two networks was present in all four dimensions. Fox et al.[69] originally demonstrated that the default mode network is anti-correlated with task-positive functional brain systems including the fronto-parietal network. Furthermore, studies of brain maturation have shown that age-related segregation of functional brain modules is a robust and reproducible findings regarding adolescent brain development.[38–40] As part of this process, connections within network modules strengthen and connections between two network modules weaken. This process is apparent using functional connectivity[38,39] as well as structural connectivity.[40] Notably, case-control studies of psychiatric disorders in adults have found abnormalities consistent with a failure of developmental network segregation, in particular between executive networks, such as the fronto-parietal and salience networks, and the default mode network.[27,29]

Using a purely data-driven analysis, our results support the possibility that loss of segregation between the default mode and executive networks may be a common neurobiological mechanism underlying vulnerability to a wide range of psychiatric symptoms.

In addition to such common abnormalities that were present across dimensions, each dimension of psychopathology was associated with a unique, highly correlated pattern of dysconnectivity. For example, connectivity features linked to the mood dimension included hyper-connectivity within the default mode, fronto-parietal and salience networks. These dimensional results from a multivariate analysis are remarkably consistent with prior work, which has provided evidence of default mode hyper-connectivity using conventional case-control designs and univariate analysis.[55,70–73] However, the data-driven approach used here allowed us to discover a combination of novel connectivity features that was more predictive than traditional univariate association analyses. These features included enhanced connectivity between both the dorsal attention and fronto-parietal networks as well as between the ventral attention and salience networks. The fear, externalizing, and psychosis dimensions were defined by a similar mix between novel features and a convergence with prior studies. Specifically, fear was characterized by weakened connectivity within default mode network, enhanced connectivity within fronto-parietal network, and — in contrast to mood — decreased connectivity between ventral attention and salience networks. In contrast to other dimensions, externalizing behavior exhibited increased connectivity in the visual network and decreased connectivity between fronto-parietal and dorsal attention networks.

Importantly, each of these dimensions was initially discovered while controlling for the effects of age and sex. However, given that many psychiatric symptoms during adolescence show a clear evolution with development and marked disparities between males and females,[44,74] we evaluated how the connectivity

features associated with each dimension were correlated with age and sex. We found that the patterns of dysconnectivity that linked to mood and psychosis symptoms strengthened with age during the adolescent period. This finding is consistent with the well-described clinical trajectory of both mood and psychosis disorders, which often emerge in adolescence and escalate in severity during the transition to adulthood.[75,76] In contrast, no age effects were found for externalizing or fear symptoms, which are typically present earlier in childhood and have a more stable time-course.[77,78] Additionally, we observed marked sex differences in the patterns of connectivity that linked to mood and fear symptoms, with these patterns being more prominent in females across the age range studied. This result accords with data from large-scale epidemiological studies, which have documented a far higher risk of mood and anxiety disorders in females.[79,80] Despite marked differences in risk by sex (i.e. double in some samples), the mechanism of such vulnerability has been only sparsely studied in the past.[46,56,62] The present results suggest that sex differences in functional connectivity may in part mediate the risk of mood and fear symptoms.

Although this study benefited from a large sample, advanced multivariate methods, and replication of results in an independent sample, several limitations should be noted. First, although the item-level data used do not explicitly consider clinical diagnostic categories, the items themselves were nonetheless drawn from a standard clinical interview. Incorporating additional data types such as genomics may capture different sources of important biological heterogeneity. Second, while we successfully replicated our findings (except for the psychosis dimension) in an independent sample, the generalizability of the study should be further evaluated in datasets that are acquired in different settings. Third, all data considered in this study were cross-sectional, which has inherent limitations for studies of development. Ongoing follow-up of this cohort will yield informative data that will allow us to evaluate the suitability of these brain-derived dimensions of psychopathology for charting developmental trajectories and prediction of clinical outcome.

In summary, in this study we discovered and replicated multivariate patterns of connectivity that are highly correlated with dimensions of psychopathology in a large sample of youth. These dimensions cross traditional clinical diagnostic categories, yet align with clinical experience. Each dimension was composed of unique features of dysconnectivity, while a lack of functional segregation between the default mode network and executive networks was common to all dimensions. Paralleling the clinical trajectory of each disorder and known disparities in prevalence between males and females, we observed both marked developmental effects and sex differences in these patterns of dysconnectivity. As suggested by the NIMH Research Domain Criteria, our findings demonstrate how specific circuit-level abnormalities in the brain's functional network architecture may give rise to a diverse panoply of psychiatric symptoms. Such an approach has the potential to clarify the high co-morbidity between psychiatric diagnoses and the great heterogeneity within each diagnostic category. Moving forward, the ability of these dimensions to predict disease trajectory and response to treatment should be evaluated, as such a neurobiologically-grounded framework could accelerate the rise of personalized medicine in psychiatry.

End Notes

Acknowledgements Thanks to Chad Jackson for data management and systems support. This study was supported by grants from National Institute of Mental Health: R01MH107703 (TDS), R01MH112847 (RTS TDS), R21MH106799 (DSB TDS), R01MH107235 (RCG), and R01EB022573 (CD). The PNC was supported by MH089983 and MH089924. Additional support was provided by P50MH096891 to REG, R01MH101111 to DHW, K01MH102609 to DRR, K08MH079364 to MEC, R01NS085211 to RTS, and the Dowshen Program for Neuroscience. Data deposition: The data reported in this paper have been deposited in database of Genotypes and Phenotypes (dbGaP), www.ncbi.nlm.nih.gov/gap (accession no. phs000607.v1.p1). All analysis code is available here `https://github.com/cedricx/sCCA/tree/master/sCCA/code/final`.

Author Contributions All authors contributed to this work, approved the final manuscript, and agreed to the copyright transfer policies of the journal.

Financial Disclosure Dr. Shinohara has received legal consulting and advisory board income from Genentech/Roche. All other authors (Mr. Xia, Dr. Ma, Mr. Ciric, Dr. Gu, Dr. Betzel, Dr. Kaczkurkin, Dr. Calkins, Dr. Cook, Mr. Garcia de la Garza, Mr. Vandekar, Dr. Moore, Dr. Roalf, Ms. Ruparel, Dr. Wolf, Dr. Davatzikos, Dr. R.E. Gur, Dr. R.C. Gur, Dr. Bassett, and Dr. Satterthwaite) reported no biomedical financial interests or potential conflicts of interest.

Figure 1 | **Schematic of sparse canonical correlation analysis (sCCA). (a)** Resting-state fMRI data analysis schematic and workflow. After preprocessing, blood-oxygen-level dependent (BOLD) signal time series were extracted from 264 spherical regions of interest distributed across the cortex and subcortical structures. Nodes of the same color belong to the same *a priori* community as defined by Power et al.[19] **(b)** A whole-brain, 264 $\times$ 264 functional connectivity matrix was constructed for each subject in the discovery sample (n =663 subjects). **(c)** Item-level data from a psychiatric screening interview (111 items, based on K-SADS[81]) were entered into sCCA as clinical features. **(d)** sCCA seeks linear combinations of connectivity and clinical symptoms that maximize their correlation. *A priori* community assignment: SMT: somatosensory/motor network; COP: cingulo-opercular network; AUD: auditory network; DMN: default mode network; VIS: visual network; FPT: fronto-parietal network; SAL: salience network; SBC: subcortical network; VAT: ventral attention network; DAT: dorsal attention network; Cerebellar and unsorted nodes not visualized. Psychopathology domains: PSY: psychotic and subthreshold symptoms; DEP: depression; MAN: mania; SUI: suicidality; ADD: attention-deficit hyperactivity disorder; ODD: oppositional defiant disorder; CON: conduct disorder; OCD: obsessive-compulsive disorder; SEP: separation anxiety; GAD: generalized anxiety disorder; PHB: specific phobias; TRT: mental health treatment; PAN: panic disorder; PTSD: post-traumatic stress disorder.

Figure 2 | **sCCA reveals multivariate patterns of linked dimensions of psychopathology and connectivity. (a)** The first seven canonical variates were selected based on covariance explained. Dashed line marks the average covariance explained. **(b)** Three canonical correlations were statistically significant by permutation testing with FDR correction ($q < 0.05$), with the fourth one showing an effect at uncorrected thresholds. Corresponding variates are circled in (a). Error bars denote standard error. Dimensions are ordered by their permutation-based P value. **(c-f)** Scatter plots of brain and clinical scores (linear

combinations of functional connectivity and psychiatric symptoms, respectively) demonstrate the correlated multivariate patterns of connectomic and clinical features. Colored dots in each panel indicate the severity of a representative clinical symptom that contributed the most to this canonical variate. Each insert displays the null distribution of sCCA correlation by permutation testing. Dashed line marks the actual correlation. $^{***}P_{FDR} < 0.001$, $^{**}P_{FDR} < 0.01$, $\dagger P_{uncorrected} = 0.04$.

Figure 3 | **Connectivity-informed dimensions of psychopathology cross clinical diagnostic categories.** **(a)** The mood dimension was composed of a mixture of depressive symptoms, suicidality, irritability, and recurrent thoughts of self harm. **(b)** The psychotic dimension was composed of psychosis-spectrum symptoms, as well as two manic symptoms. **(c)** The fear dimension was comprised of social phobia and agoraphobia symptoms. **(d)** The externalizing behavior dimension showed a mixture of symptoms from attention-deficit and oppositional defiant disorders, as well as irritability from the depression section. The outermost labels are the item-level psychiatric symptoms. The color arcs represent categories from clinical screening interview and the Diagnostic and Statistical Manual of Mental Disorders (DSM). Numbers in the inner rings represent sCCA loadings for each symptom in their respective dimension. Only loadings determined to be statistically significant by a resampling procedure are shown here.

Figure 4 | **Specific patterns of within- and between-network dysconnectivity contribute to linked psychopathological dimensions. (a-d)** Modular (community) level connectivity pattern associated with each psychopathology dimension. Both increased **(e-h)** and diminished **(i-l)** connectivity in specific edges contributed to each dimension of psychopathology. The outer labels represent the anatomical names of nodes. The inner arcs indicate the community membership of nodes. The thickness of the chords represent the loadings of connectivity features.

Figure 5 | **Loss of segregation between default mode and executive networks is shared across all dimensions.** **(a)** By searching for overlap of edges that contributed significantly to each dimension, we found common edges that were implicated across all dimensions of psychopathology. These were then summarized at a nodal level by the sum of their absolute loadings. Nodes that contributed significantly to every dimension included the frontal pole, superior frontal gyrus, dorsomedial prefrontal cortex, medial temporal gyrus, and amygdala. **(b)** Results of a similar analysis conducted at the module level. **(c)** Loss of segregation between the default mode and executive networks were shared across all four dimensions.

Figure 6 | **Developmental effects and sex differences are concentrated in specific dimensions.** Connectivity patterns associated with both the mood **(a)** and psychosis **(b)** dimensions increased significantly with age. Additionally, connectivity patterns associated with both the mood **(c)** and fear **(d)** symptoms were significantly more prominent in females than males. Multiple comparisons were controlled for using the False Discovery Rate ($q < 0.05$).

Figure 7 | **Linked dimensions of psychopathology were replicated in an independent sample.** All procedures were repeated in an independent replication sample of 336 participants. **(a)** The first four canonical variates in the replication sample were selected for further analysis based on covariance explained. Dashed line marks the average covariance explained. **(b)** The mood, fear, and externalizing behavior dimensions were significant by permutation testing. Corresponding variates are circled in (a). Error bars denote standard error. **$P_{FDR} < 0.01$.

Online Methods

Participants Resting-state functional magnetic resonance imaging (rs-fMRI) datasets were acquired as part of the Philadelphia Neurodevelopmental Cohort (PNC), a large community-based study of brain development.[46] In total, 1601 participants completed the cross-sectional neuroimaging protocol (**Supplementary Fig. 1a**). One subject had missing clinical data. To create two independent samples for discovery and replication analyses, we performed random split of the remaining 1600 participants using the CARET package[82] in R. Specifically, using the function CREATEDATAPARTITION, a discovery sample (n=1069) and a replication sample (n=531) were created that were stratified by overall psychopathology. The two samples were confirmed to also have similar distributions in regards to age, sex, and race (**Supplementary Fig. 1b**). The overall psychopathology is the general factor score reported previously from factor analysis of the clinical data alone.[50,57]

Of the discovery sample (n=1069), 111 were excluded due to: gross radiological abnormalities, or a history of medical problems that might affect brain function. Of the remaining 958 participants, 45 were excluded for having low quality T1-weighted images, and 250 were excluded for missing rs-fMRI, incomplete voxelwise coverage, or excessive motion during the functional scan, which is defined as having an average framewise motion more than 0.20mm or more than 20 frames exhibiting over 0.25mm movement. These exclusion criteria produced a final discovery sample consisting of 663 youths (mean age 15.82, SD = 3.32; 293 males and 370 females). Applying the same exclusion criteria to the replication sample produced 336 participants (mean age 15.65, SD = 3.32; 155 males and 181 females). See **Supplementary Table 1** for detailed demographics of each sample.

Psychiatric assessment Psychopathology symptoms were evaluated using a structured screening interview (GOASSESS), which has been described in detail elsewhere.[50] To allow rapid training and standardization across a large number of assessors, GOASSESS was designed to be highly structured, with screen-level symptom and episode information. The instrument is abbreviated and modified from the epidemiologic version of the NIMH Genetic Epidemiology Research Branch Kiddie-SADS.[81] The psychopathology screen in GOASSESS assessed lifetime occurrence of major domains of psychopathology including psychosis spectrum symptoms, mood (major depressive episode, mania), anxiety (agoraphobia, generalized anxiety, panic, specific phobia, social phobia, separation anxiety), behavioral disorders (oppositional defiant, attention deficit/hyperactivity, conduct) disorders, eating disorders (anorexia, bulimia), and suicidal thinking and behavior. The 111 item-level symptoms used in this study were described in prior factor analysis of the clinical data in PNC.[57] For the specific items, see **Supplementary Table 3**.

Image acquisition Structural and functional subject data were acquired on a 3T Siemens Tim Trio scanner with a 32-channel head coil (Erlangen, Germany), as previously described.[46,62] High-resolution structural images were acquired in order to facilitate alignment of individual subject images into a common space. Structural images were acquired using a magnetization-prepared, rapid-acquisition gradient-echo (MPRAGE) T1-weighted sequence ($T_R = 1810$ms; $T_E = 3.51$ms; FoV $= 180 \times 240$mm; resolution $0.9375 \times 0.9375 \times 1$mm). Approximately 6 minutes of task-free functional data were acquired for each subject using a blood oxygen level-dependent (BOLD-weighted) sequence ($T_R = 3000$ms; $T_E = 32$ms; FoV $= 192 \times 192$mm; resolution 3mm isotropic; 124 volumes). Prior to scanning, in order to acclimate subjects to the MRI environment and to help subjects learn to remain still during the actual scanning session, a mock

scanning session was conducted using a decommissioned MRI scanner and head coil. Mock scanning was accompanied by acoustic recordings of the noise produced by gradient coils for each scanning pulse sequence. During these sessions, feedback regarding head movement was provided using the MoTrack motion tracking system (Psychology Software Tools, Inc, Sharpsburg, PA). Motion feedback was only given during the mock scanning session. In order to further minimize motion, prior to data acquisition subjects' heads were stabilized in the head coil using one foam pad over each ear and a third over the top of the head. During the resting-state scan, a fixation cross was displayed as images were acquired. Subjects were instructed to stay awake, keep their eyes open, fixate on the displayed crosshair, and remain still.

Structural Preprocessing A study-specific template was generated from a sample of 120 PNC subjects balanced across sex, race, and age bins using the BUILDTEMPLATEPARALLEL procedure in ANTs.[83] Study-specific tissue priors were created using a multi-atlas segmentation procedure.[84] Subject anatomical images were independently rated by three highly trained image analysts. Any image that did not pass manual inspection was removed from the analysis. Each subject's high-resolution structural image was processed using the ANTs Cortical Thickness Pipeline.[85] Following bias field correction,[86] each structural image was diffeomorphically registered to the study-specific PNC template using the top-performing SYN deformation provided by ANTs.[87] Study-specific tissue priors were used to guide brain extraction and segmentation of the subject's structural image.[88]

Functional Preprocessing Task-free functional images were processed using one of the top-performing pipelines for removal of motion-related artifact.[51] Preprocessing steps included (1) correction for distortions induced by magnetic field inhomogeneities using FSL's FUGUE utility, (2) removal of the 4 initial volumes of

each acquisition, (3) realignment of all volumes to a selected reference volume using MCFLIRT,[89] (4) removal of and interpolation over intensity outliers in each voxel's time series using AFNI's 3DDESPIKE utility, (5) demeaning and removal of any linear or quadratic trends, and (6) co-registration of functional data to the high-resolution structural image using boundary-based registration.[90] The artefactual variance in the data was modelled using a total of 36 parameters, including the 6 framewise estimates of motion, the mean signal extracted from eroded white matter and cerebrospinal fluid compartments, the mean signal extracted from the entire brain, the derivatives of each of these 9 parameters, and quadratic terms of each of the 9 parameters and their derivatives. Both the BOLD-weighted time series and the artefactual model time series were temporally filtered using a first-order Butterworth filter with a passband between 0.01 and 0.08 Hz.[91]

Network construction A functional connectivity network was computed across all parcels of a common parcellation using the residual timeseries following de-noising.[19] The functional connectivity between any pair of brain regions was operationalised as the Pearson correlation coefficient between the mean activation timeseries extracted from those regions.[92] For each parcellation, an $n \times n$ weighted adjacency matrix encoding the connectome was thus obtained, where n represents the total number of nodes (or parcels) in that parcellation. Community boundaries were defined *a priori* for each parcellation scheme. As part of the supplementary analysis to demonstrate the robustness of the results independent of parcellation choices (**Supplementary Fig. 5)**, we also constructed networks based on an alternative system.[24]

To ensure that potential confounders did not drive the canonical correlations, we regressed out relevant covariates out of the input matrices. Specifically, using the `glm` and `residual.glm` functions in `R`, we regressed age, sex, race and in-scanner motion out of the connectivity data, and regressed age, sex and race

out of the clinical data. Importantly, we found that the canonical variates derived from regressed and non-regressed datasets were comparable, with highly correlated feature weights (**Supplementary Table 2**).

Dimensionality reduction Each correlation matrix comprised 34,980 unique connectivity features. We reasoned that since sCCA seeks to capture sources of variation common to both datasets, connectivity features that are most predictive of psychiatric symptoms would be those with high variance across participants. Therefore, to reduce dimensionality of the connectivity matrices, we selected the top edges with the highest median absolute deviation (MAD) (**Supplementary Fig. 2)**. MAD is defined as $median(|\mathbf{X}_i - median(\mathbf{X})|)$, or the median of the absolute deviations from the vector's median. We chose MAD as a measurement for variance estimation, because it is a robust statistic, being more resilient to outliers in a data set than other measures such as the standard deviation. To illustrate which edges were selected based on MAD, we visualized the network adjacency matrix with all edges, at 95th, 90th and 75th percentile (**Supplementary Fig. 2c**).

An alternative approach for dimensionality reduction is performing principal component analysis (PCA), from which we selected the top 111 components (explaining 37% of variance) as connectivity features entered into sCCA. As detailed in **Supplementary Fig. 5,** using PCA yielded similar canonical variates as MAD. We ultimately chose feature selection with MAD because it allowed direct use of individual connectivity strength instead of latent variables (e.g. components from PCA) as the input features to sCCA, thus increasing the interpretability of our results.

sCCA Sparse canonical correlation analysis (sCCA) is a multivariate procedure that seeks maximal correlations between linear combinations of variables in both sets,[93] with regularization to achieve sparsity.[47] In essence, given two matrices, $\mathbf{X}_{n\times p}$ and $\mathbf{Y}_{n\times q}$, where n is the number of observations (e.g.participants), p and q are the number of variables (e.g. clinical and connectivity features, respectively), sCCA involves finding **u** and **v**, which are loading vectors, that maximize cor(**Xu**,**Yv**). Mathematically, this optimization problem can be expressed as

$$\text{maximize}_{\mathbf{u},\mathbf{v}}\mathbf{u}^T\mathbf{X}^T\mathbf{Y}\mathbf{v}, \text{ subject to } \|\mathbf{u}\|_2^2 \leq 1, \|\mathbf{v}\|_2^2 \leq 1, \|\mathbf{u}\|_1 \leq c_1, \|\mathbf{v}\|_1 \leq c_2. \quad (1)$$

Since both L^1 ($\|\mathbf{x}\|_1$) and L^2-norm ($\|\mathbf{x}\|_2$) are used, this is an elastic net regularization that combines the LASSO and ridge penalties. The penalty parameters for the L^2 norm are fixed for both **u** and **v** at 1, but those of L^1 norm, namely c_1 and c_2, are set by the user and need to be tuned (see below).

Grid search for regularization parameters We tuned the L^1 regularization parameters for the connectivity and the clinical features, respectively (see **Supplementary Fig. 3)**. The range of sparsity parameters are constrained to be between 0 and 1 in the PMA package,[47] where 0 indicates the smallest number of features (i.e. highest level of sparsity) and 1 indicates the largest number of features (i.e. lowest level of sparsity). We conducted a grid search in increment of 0.1 to determine the combination of parameters that would yield the highest canonical correlation of the first variate across 10 randomly resampled samples, each consisting of two-thirds of the discovery dataset. Note that the parameters were only tuned on the discovery sample and the same regularization parameters were applied in the replication analysis.

Permutation testing To assess the statistical significance of each canonical variate, we used a permutation testing procedure to create a null distribution of correlations (**Supplementary Fig. 4)**. Essentially, we held the connectivity matrix constant, and then shuffled the rows of the clinical matrix so as to break the linkage of participants' brain features and their symptom features. Then we performed sCCA using the same set of regularization parameters to generate a null distribution of correlations after permuting the input data 1000 times ($\mathbf{B}$). As permutation could induce arbitrary axis rotation, which changes the order of canonical variates, or axis reflection, which causes a sign change for the weights, we matched the canonical variates resulting from permuted data matrices to the ones derived from the original data matrix by comparing the clinical loadings ($\mathbf{v}$).[94] The P_{FDR} value was estimated as the number of null correlations (r_i) that exceeded the average sCCA correlations estimated on the original dataset ($\bar{r}$), with false discovery rate correction (FDR, $q < 0.05$) across the top seven selected canonical variates:

$$P_{\text{permutation}} = \frac{\sum_{i=1}^{B} \begin{cases} 1, & if \quad r_i \geq \bar{r} \\ 0, & if \quad r_i < \bar{r} \end{cases}}{B}. \tag{2}$$

Resampling procedure To further select features that consistently contributed to each canonical variate, we performed a resampling procedure (**Supplementary Fig. 6)**. In each of 1000 samples, we randomly selected two-thirds of the discovery sample and then randomly replaced the remaining one-third from those two-thirds (similar to bootstrapping with replacement). Similar to the permutation procedure, we matched the corresponding canonical variates from resampled matrices to the original one to obtain a set of comparable decompositions.[94] Features whose 95% and 99% confidence intervals (for clinical and connectivity features, respectively) did not cross zero were considered significant, suggesting that they were stable across different sampling cohorts.

Network module analysis To visualize and understand the high dimensional connectivity loading matrix, we summarized it as mean within- and between-module loadings according to the *a priori* community assignment of the Power[19] parcellation (**Supplementary Fig. 7a**). Specifically, *within*-module connectivity loading is defined as

$$\frac{\sum_{i,j \in M} 2W_{ij}}{|M| \times (|M| - 1)} \quad , \tag{3}$$

where $W_{i,j}$ is the sCCA loading of the functional connectivity between nodes i and j, which both belong to the same community m in M. The cardinality of the community assignment vector, $|M|$, represents the number of nodes in each community. *Between*-module connectivity loading is defined as

$$\frac{\sum_{i \in M, j \in N} W_{ij}}{|M| \times |N|} \quad , \tag{4}$$

where $W_{i,j}$ is the sCCA loading of the functional connectivity between nodes i and j, which belong to community m in $|M|$ and community n in $|N|$, respectively.

We used a permutation test based on randomly assigning community memberships to each node while controlling for community size to assess the statistical significance of the mean connectivity loadings (**Supplementary Fig. 7b**). Empirical P-values were calculated similar to Equation 2 and were FDR corrected.

Analysis of common connectivity features across dimensions Each connectivity loading matrix was first binarized based on the presence of a significant edge feature after the resampling procedure in a given

canonical variate. All four binarized matrices were then added and thresholded at 4 (i.e. common to all four dimensions), generating an overlapping edge matrix. Statistical significance was assessed by comparing this concordant feature matrix to a null model. The null model was constructed by computing the overlapping edges, repeated 1000 times, of four randomly generated loading matrices, each preserving the edge density of the original loading matrix. Any edge that appeared at least once in the null model was eliminated from further analysis. With only the statistically significant common edge features, we calculated the mean absolute loading in each edge feature across four dimensions as well as the nodal loading strength using `Brain Connectivity Toolbox`[95] and visualized it with `BrainNet Viewer`[96] both in MATLAB.

Analysis of age effects and sex differences As previously,[40,56–58] generalized additive models (GAMs), using the MGCV package[97,98] in R, were used to characterize age-related effects and sex differences on the specific dysconnectivity pattern associated with each psychopathology dimension. A GAM is similar to a generalized linear model where predictors can be replaced by smooth functions of themselves, offering efficient and flexible estimation of non-linear effects. For each linked dimension i, a GAM was fit:

$$\text{Connectivity Score}_i \sim \text{Sex} + s(\text{Age}). \tag{5}$$

Additionally, we also separately tested whether age by sex interactions were present.

References

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

2. Insel, B. T. R. & Cuthbert, B. N. Brain diorders? Precisely. *Science* **348**, 499–500 (2015).

3. Hyman, S. E. The Diagnosis of Mental Disorders: The Problem of Reification. *Annual Review of Clinical Psychology* **6**, 155–179 (2010).

4. Nestler, E. & Hyman, S. Animal models of neuropsychiatric disorders. *Annual review of clinical psychology* **7**, 39–61 (2011).

5. Jacobi, F. *et al.* Prevalence, co-morbidity and correlates of mental disorders in the general population: results from the German Health Interview and Examination Survey (GHS). *Psychological Medicine* **34**, 597–611 (2004).

6. Goodkind, M. *et al.* Identification of a common neurobiological substrate for mental illness. *JAMA psychiatry* **5797**, 305–315 (2015).

7. McTeague, L. M. *et al.* Identification of Common Neural Circuit Disruptions in Cognitive Control Across Psychiatric Disorders. *American Journal of Psychiatry* **174**, 676–685 (2017).

8. Lee, S. H. *et al.* Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature genetics* **45**, 984–94 (2013).

9. Dichter, G. S., Damiano, C. A. & Allen, J. A. Reward circuitry dysfunction in psychiatric and neurodevelopmental disorders and genetic syndromes: animal models and clinical findings. *Journal of neurodevelopmental disorders* **4**, 19 (2012).

10. Cole, M. W., Repovš, G. & Anticevic, A. The frontoparietal control system: a central role in mental health. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry* **20**, 652–64 (2014).

11. Insel, T. *et al.* Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry* **167**, 748–751 (2010).

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

13. Menon, V. Large-scale brain networks and psychopathology: A unifying triple network model. *Trends in Cognitive Sciences* **15**, 483–506 (2011).

14. Uhlhaas, P. J. & Singer, W. Abnormal neural oscillations and synchrony in schizophrenia. *Nature reviews. Neuroscience* **11**, 100–13 (2010).

15. Bullmore, E. & Sporns, O. Complex brain networks : graph theoretical analysis of structural and functional systems (2009).

16. Bassett, D. S. & Bullmore, E. T. Human brain networks in health and disease. *Current opinion in neurology* **22**, 340–7 (2009).

17. Bassett, D. S. & Sporns, O. Network neuroscience. *Nature Neuroscience* **20**, 353–364 (2017).

18. Vincent, J. L. *et al.* Intrinsic functional architecture in the anaesthetized monkey brain. *Nature* **447**, 83–86 (2007).

19. Power, J. D. *et al.* Functional network organization of the human brain. *Neuron* **72**, 665–78 (2011).

20. Yeo, B. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology* **106**, 1125–1165 (2011).

21. Shen, X., Tokoglu, F., Papademetris, X. & Constable, R. T. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *NeuroImage* **82**, 403–415 (2013).

22. Wig, G. S., Laumann, T. O. & Petersen, S. E. An approach for parcellating human cortical areas using resting-state correlations. *NeuroImage* **93**, 276–291 (2014).

23. Wig, G. S. *et al.* Parcellating an individual subject's cortical and subcortical brain structures using snowball sampling of resting-state correlations. *Cerebral Cortex* **24**, 2036–2054 (2014).

24. Gordon, E. M., Breeden, A. L., Bean, S. E. & Vaidya, C. J. Working memory-related changes in functional connectivity persist beyond task disengagement. *Human brain mapping* **35**, 1004–17 (2014).

25. Wang, D. *et al.* Parcellating cortical functional networks in individuals. *Nature neuroscience* **18**, 1853–60 (2015).

26. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* 1–11 (2016).

27. Woodward, N. D. & Cascio, C. J. Resting-state functional connectivity in psychiatric disorders. *JAMA Psychiatry* **72**, 743–744 (2015).

28. Lynall, M.-E. *et al.* Functional connectivity and brain networks in schizophrenia. *The Journal of neuroscience* **30**, 9477–87 (2010).

29. Satterthwaite, T. D. *et al.* Connectome-wide network analysis of youth with Psychosis-Spectrum symptoms. *Molecular psychiatry* **20**, 1–8 (2015).

30. Ye, M., Qing, P., Zhang, K. & Liu, G. Altered network efficiency in major depressive disorder. *BMC psychiatry* **16**, 450 (2016).

31. Li, W. *et al.* Disrupted small world topology and modular organisation of functional networks in late-life depression with and without amnestic mild cognitive impairment. *Journal of neurology, neurosurgery, and psychiatry* 1–9 (2014).

32. Wang, Y. *et al.* Topologically convergent and divergent functional connectivity patterns in unmedicated unipolar depression and bipolar disorder. *Translational Psychiatry* **7**, e1165 (2017).

33. Alexander-Bloch, A. F. *et al.* Disrupted modularity and local connectivity of brain functional networks in childhood-onset schizophrenia. *Frontiers in systems neuroscience* **4**, 147 (2010).

34. Alexander-Bloch, A. *et al.* The discovery of population differences in network community structure: New methods and applications to brain functional networks in schizophrenia. *NeuroImage* **59**, 3889–3900 (2012).

35. Hadley, J. A. *et al.* Change in brain network topology as a function of treatment response in schizophrenia: a longitudinal resting-state fMRI study using graph theory. *npj Schizophrenia* **2**, 16014 (2016).

36. Rudie, J. D. *et al.* Altered functional and structural brain network organization in autism. *NeuroImage: Clinical* **2**, 79–94 (2013).

37. Fornito, A., Zalesky, A. & Breakspear, M. The connectomics of brain disorders. *Nature Reviews Neuroscience* **16**, 159–172 (2015).

38. Satterthwaite, T. D. *et al.* Heterogeneous impact of motion on fundamental patterns of developmental changes in functional connectivity during youth. *NeuroImage* **83**, 45–57 (2013).

39. Fair, D. a. *et al.* Development of distinct control networks through segregation and integration. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 13507–13512 (2007).

40. Baum, G. L. *et al.* Modular Segregation of Structural Brain Networks Supports the Development of Executive Function in Youth. *Current Biology* **27**, 1561–1572.e8 (2017).

41. Cao, M., Huang, H., Peng, Y., Dong, Q. & He, Y. Toward Developmental Connectomics of the Human Brain. *Frontiers in neuroanatomy* **10**, 25 (2016).

42. Paus, T. Mapping brain maturation and cognitive development during adolescence. *Trends in Cognitive Sciences* **9**, 60–68 (2005).

43. Satterthwaite, T. D. & Baker, J. T. How can studies of resting-state functional connectivity help us understand psychosis as a disorder of brain development? *Current Opinion in Neurobiology* **30**, 85–91 (2015).

44. Casey, B. J., Oliveri, M. E. & Insel, T. A neurodevelopmental perspective on the research domain criteria (RDoC) framework. *Biological Psychiatry* **76**, 350–353 (2014).

45. Insel, T. R. Mental disorders in childhood: shifting the focus from behavioral symptoms to neurodevelopmental trajectories. *JAMA : the journal of the American Medical Association* **311**, 1727–8 (2014).

46. Satterthwaite, T. D. *et al.* Neuroimaging of the Philadelphia Neurodevelopmental Cohort. *NeuroImage* **86**, 544–553 (2014).

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

48. Avants, B. B. *et al.* Sparse canonical correlation analysis relates network-level atrophy to multivariate cognitive measures in a neurodegenerative population. *NeuroImage* **84**, 698–711 (2014).

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

50. Calkins, M. E. *et al.* The Philadelphia Neurodevelopmental Cohort : constructing a deep phenotyping collaborative. *Journal of Child Psychology and Psychiatry* **12**, 1356–1369 (2015).

51. Ciric, R. *et al.* Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *NeuroImage* **154**, 174–187 (2017).

52. Park, H. J. & Friston, K. Structural and functional brain networks: from connections to cognition. *Science* **342**, 579–588 (2013).

53. Sharp, D. J., Scott, G. & Leech, R. Network dysfunction after traumatic brain injury. *Nat Rev Neurol* **10**, 156–166 (2014).

54. Sylvester, C. M. *et al.* Functional network dysfunction in anxiety and anxiety disorders. *Trends in Neurosciences* **35**, 527–535 (2012).

55. Kaiser, R. H., Andrews-Hanna, J. R., Wager, T. D. & Pizzagalli, D. A. Large-Scale Network Dysfunction in Major Depressive Disorder. *JAMA Psychiatry* **72**, 603 (2015).

56. Kaczkurkin, A. N. *et al.* Elevated Amygdala Perfusion Mediates Developmental Sex Differences in Trait Anxiety. *Biological Psychiatry* **80**, 775–785 (2016).

57. Shanmugan, S. *et al.* Common and Dissociable Mechanisms of Executive System Dysfunction Across Psychiatric Disorders in Youth. *The American journal of psychiatry* **173**, 517–526 (2016).

58. Gennatas, E. D. *et al.* Age-related effects and sex differences in gray matter density, volume, mass, and cortical thickness from childhood to young adulthood. *The Journal of Neuroscience* **37**, 3550–16 (2017).

59. Drysdale, A. T. *et al.* Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine* **23**, 28–38 (2016).

60. Adleman, N. E. *et al.* Cross-sectional and longitudinal abnormalities in brain structure in children with severe mood dysregulation or bipolar disorder. *Journal of Child Psychology and Psychiatry and Allied Disciplines* **53**, 1149–1156 (2012).

61. Satterthwaite, T. D. *et al.* Common and Dissociable Dysfunction of the Reward System in Bipolar and Unipolar Depression Reward Dysfunction in Depression. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* **40**, 1–11 (2015).

62. Satterthwaite, T. D. *et al.* Structural Brain Abnormalities in Youth With Psychosis Spectrum Symptoms. *JAMA Psychiatry* **73**, 515–524 (2016).

63. Sharma, A. *et al.* Common Dimensional Reward Deficits Across Mood and Psychotic Disorders: A Connectome-Wide Association Study. *American Journal of Psychiatry* appi.ajp.2016.1 (2017).

64. Pezzoli, P., Antfolk, J. & Santtila, P. Phenotypic factor analysis of psychopathology reveals a new body-related transdiagnostic factor. *PLoS ONE* **12**, 1–18 (2017).

65. Ventura, J., Nuechterlein, K. H., Subotnik, K. L., Gutkind, D. & Gilbert, E. A. Symptom dimensions in recent-onset schizophrenia and mania: A principal components analysis of the 24-item Brief Psychiatric Rating Scale. *Psychiatry Research* **97**, 129–135 (2000).

66. Clementz, B. A. *et al.* Identification of distinct psychosis biotypes using brain-based biomarkers. *American Journal of Psychiatry* **173**, 373–384 (2016).

67. Caspi, A. *et al.* The p Factor. *Clinical Psychological Science* **2**, 119–137 (2014).

68. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460** (2009).

69. Fox, M. D. *et al.* From The Cover: The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences* **102**, 9673–9678 (2005).

70. Sheline, Y. I. *et al.* The default mode network and self-referential processes in depression. *Proceedings of the National Academy of Sciences* **106**, 1942–1947 (2009).

71. Berman, M. G. *et al.* Depression, rumination and the default network. *Social Cognitive and Affective Neuroscience* **6**, 548–555 (2011).

72. Wang, T. *et al.* Disorganized cortical thickness covariance network in major depressive disorder implicated by aberrant hubs in large-scale networks. *Scientific Reports* **6**, 1–12 (2016).

73. Mak, E., Colloby, S. J., Thomas, A. & O'Brien, J. T. The segregated connectome of late-life depression: a combined cortical thickness and structural covariance analysis. *Neurobiology of Aging* **48**, 212–221 (2016).

74. Rapoport, J. L., Giedd, J. N. & Gogtay, N. Neurodevelopmental model of schizophrenia: update 2012. *Molecular Psychiatry* **17**, 1228–1238 (2012).

75. Harrow, M., Carone, B. J. & Westermeyer, J. F. The Course of Psychosis in Eearly Phases of Schizophrenia. *American Journal of Psychiatry* **142**, 702–707 (1985).

76. Richards, D. Prevalence and clinical course of depression: A review. *Clinical Psychology Review* **31**, 1117–1125 (2011).

77. Bongers, I. L., Koot, H. M., Van Der Ende, J. & Verhulst, F. C. Developmental trajectories of externalizing behaviors in childhood and adolescence. *Child Development* **75**, 1523–1537 (2004).

78. Wittchen, H.-U. & Fehm, L. Epidemiology and natural course of social fears and social phobia. *Acta psychiatrica Scandinavica. Supplementum* **108**, 4–18 (2003).

79. Albert, P. R. Why is depression more prevalent in women? *Journal of Psychiatry and Neuroscience* **40**, 219–221 (2015).

80. Kessler, R. C. Epidemiology of women and depression. *Journal of Affective Disorders* **74**, 5–13 (2003).

81. Merikangas, K. R. *et al.* Prevalence and treatment of mental disorders among US children in the 2001–2004 NHANES. *Pediatrics* **125**, 75–81 (2010).

82. Kuhn, M. Building Predictive Models in R Using the caret Package. *Journal Of Statistical Software* **28**, 1–26 (2008).

83. Avants, B. B. *et al.* A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage* **54**, 2033–2044 (2011).

84. Wang, X.-J. & Krystal, J. H. Neuron Perspective Computational Psychiatry. *Neuron* **84**, 638–654 (2014).

85. Tustison, N. J. *et al.* Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *NeuroImage* **99**, 166–179 (2014).

86. Tustison, N. J. *et al.* N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging* **29**, 1310–1320 (2010).

87. Klein, A. *et al.* Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage* **46**, 786–802 (2009).

88. Avants, B. B., Tustison, N. J., Wu, J., Cook, P. A. & Gee, J. C. An open source multivariate framework for N-tissue segmentation with evaluation on public data. *Neuroinformatics* **9**, 381–400 (2011).

89. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage* **17**, 825–841 (2002). `arXiv:1011.1669v3`.

90. Greve, D. N. & Fischl, B. Accurate and robust brain image alignment using boundary-based registration. *NeuroImage* **48**, 63–72 (2009).

91. Hallquist, M. N., Hwang, K. & Luna, B. The nuisance of nuisance regression: Spectral misspecification in a common approach to resting-state fMRI preprocessing reintroduces noise and obscures functional connectivity. *NeuroImage* **82**, 208–225 (2013). NIHMS150003.

92. Biswal, B., Yetkin, F. Z., Haughton, V. M. & Hyde, J. S. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic Resonance in Medicine* **34**, 537–541 (1995).

93. Hotelling, H. Relations between two sets of variates (1936).

94. Mišić, B. *et al.* Network-Level Structure-Function Relationships in Human Neocortex. *Cerebral Cortex* bhw089 (2016).

95. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage* **52**, 1059–1069 (2010).

96. Xia, M., Wang, J. & He, Y. BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics. *PLoS ONE* **8** (2013).

97. Wood, S. N. Fast stable REML and ML estimation of semiparametric GLMs. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **73**, 3–36 (2011).

98. Wood, S. N., Goude, Y. & Shaw, S. Generalized additive models for large data sets. *Journal of the Royal Statistical Society Series C-Applied Statistics* **64**, 139–155 (2015).

99. Gordon, E. M. *et al.* Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cerebral Cortex* **26**, 288–303 (2016).

100. Power, J. D. *et al.* Functional network organization of the human brain. *Neuron* **72**, 665–678 (2011).

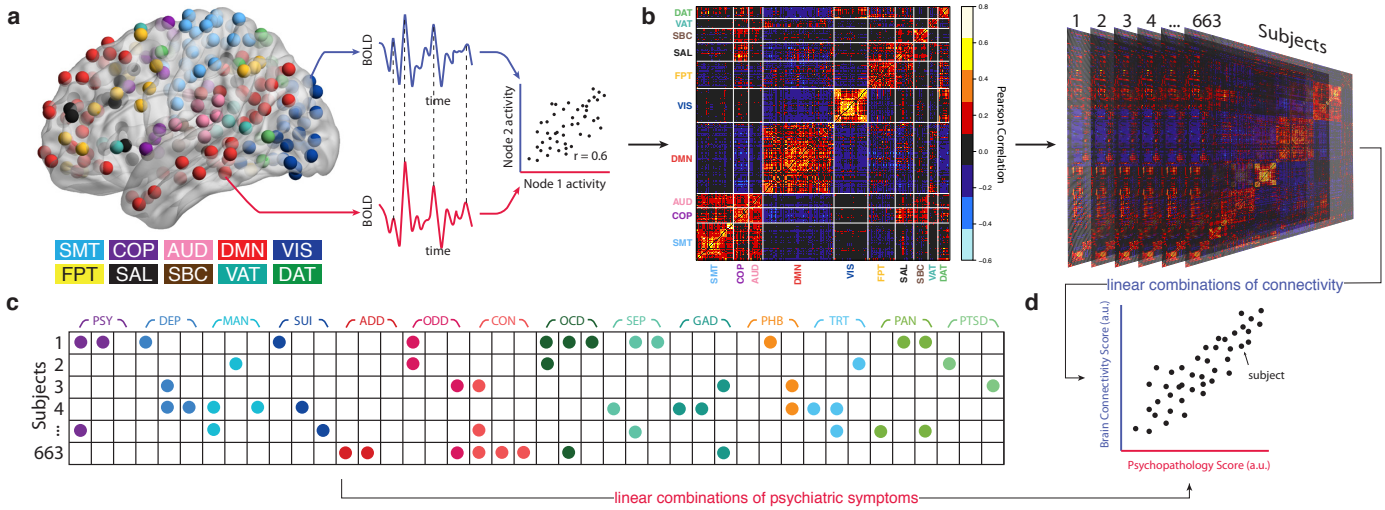


Figure 1 | **Schematic of sparse canonical correlation analysis (sCCA). (a)** Resting-state fMRI data analysis schematic and workflow. After preprocessing, blood-oxygen-level dependent (BOLD) signal time series were extracted from 264 spherical regions of interest distributed across the cortex and subcortical structures. Nodes of the same color belong to the same *a priori* community as defined by Power et al.[19] **(b)** A whole-brain, 264 × 264 functional connectivity matrix was constructed for each subject in the discovery sample (n =663 subjects). **(c)** Item-level data from a psychiatric screening interview (111 items, based on K-SADS[81]) were entered into sCCA as clinical features. **(d)** sCCA seeks linear combinations of connectivity and clinical symptoms that maximize their correlation. *A priori* community assignment: SMT: somatosensory/motor network; COP: cingulo-opercular network; AUD: auditory network; DMN: default mode network; VIS: visual network; FPT: fronto-parietal network; SAL: salience network; SBC: subcortical network; VAT: ventral attention network; DAT: dorsal attention network; Cerebellar and unsorted nodes not visualized. Psychopathology domains: PSY: psychotic and subthreshold symptoms; DEP: depression; MAN: mania; SUI: suicidality; ADD: attention-deficit hyperactivity disorder; ODD: oppositional defiant disorder; CON: conduct disorder; OCD: obsessive-compulsive disorder; SEP: separation anxiety; GAD: generalized anxiety disorder; PHB: specific phobias; TRT: mental health treatment; PAN: panic disorder; PTSD: post-traumatic stress disorder.

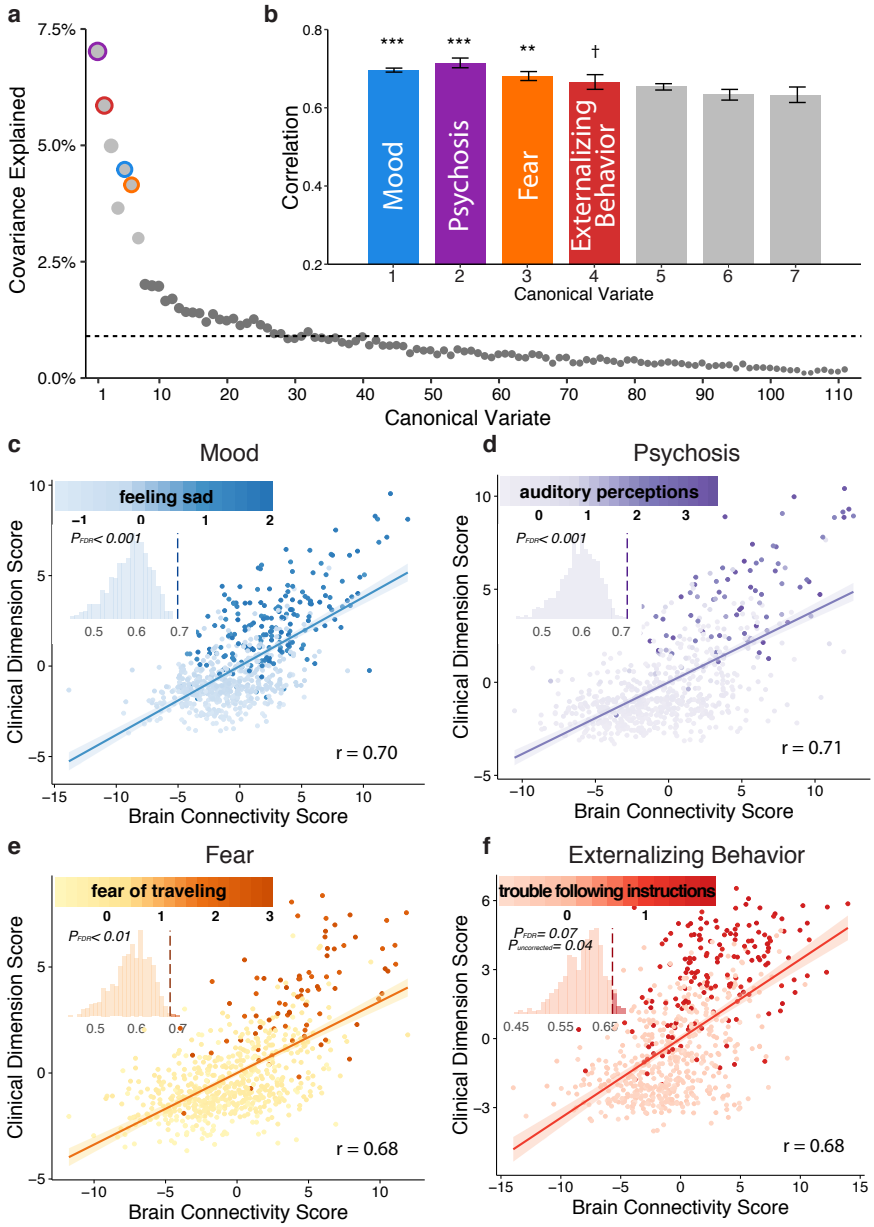


Figure 2 | **sCCA reveals multivariate patterns of linked dimensions of psychopathology and connectivity.** **(a)** The first seven canonical variates were selected based on covariance explained. Dashed line marks the average covariance explained. **(b)** Three canonical correlations were statistically significant by permutation testing with FDR correction ($q < 0.05$), with the fourth one showing an effect at uncorrected thresholds. Corresponding variates are circled in (a). Error bars denote standard error. Dimensions are ordered by their permutation-based P value. **(c-f)** Scatter plots of brain and clinical scores (linear combinations of functional connectivity and psychiatric symptoms, respectively) demonstrate the correlated multivariate patterns of connectomic and clinical features. Colored dots in each panel indicate the severity of a representative clinical symptom that contributed the most to this canonical variate. Each insert displays the null distribution of sCCA correlation by permutation testing. Dashed line marks the actual correlation. $^{***}P_{FDR} < 0.001$, $^{**}P_{FDR} < 0.01$, $\dagger P_{uncorrected} = 0.04$.

Figure 3 | **Connectivity-informed dimensions of psychopathology cross clinical diagnostic categories. (a)** The mood dimension was composed of a mixture of depressive symptoms, suicidality, irritability, and recurrent thoughts of self harm. **(b)** The psychotic dimension was composed of psychosis-spectrum symptoms, as well as two manic symptoms. **(c)** The fear dimension was comprised of social phobia and agoraphobia symptoms. **(d)** The externalizing behavior dimension showed a mixture of symptoms from attention-deficit and oppositional defiant disorders, as well as irritability from the depression section. The outermost labels are the item-level psychiatric symptoms. The color arcs represent categories from clinical screening interview and the Diagnostic and Statistical Manual of Mental Disorders (DSM). Numbers in the inner rings represent sCCA loadings for each symptom in their respective dimension. Only loadings determined to be statistically significant by a resampling procedure are shown here.

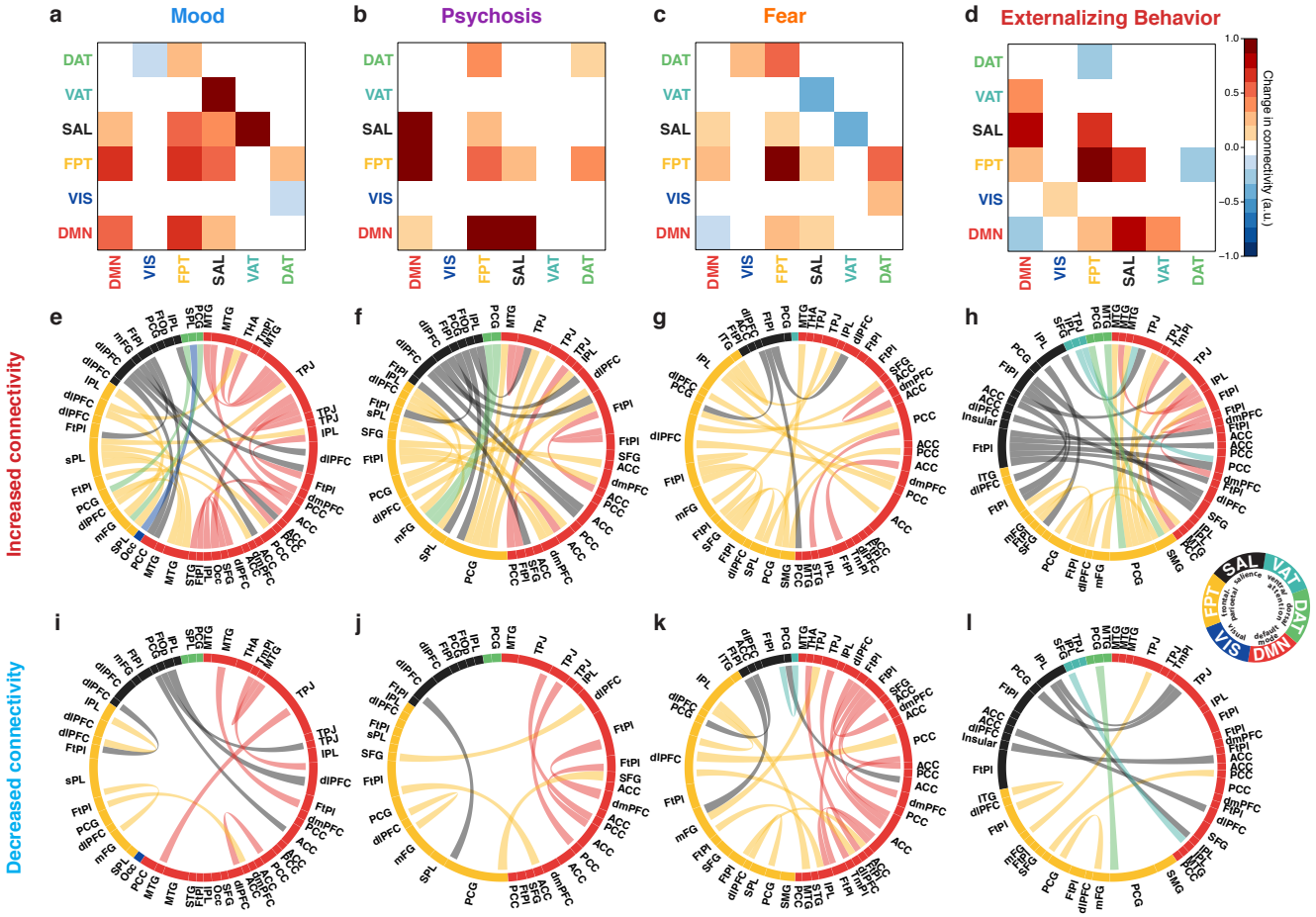


Figure 4 | Specific patterns of within- and between-network dysconnectivity contribute to linked psychopathological dimensions. (a-d) Modular (community) level connectivity pattern associated with each psychopathology dimension. Both increased **(e-h)** and diminished **(i-l)** connectivity in specific edges contributed to each dimension of psychopathology. The outer labels represent the anatomical names of nodes. The inner arcs indicate the community membership of nodes. The thickness of the chords represent the loadings of connectivity features.

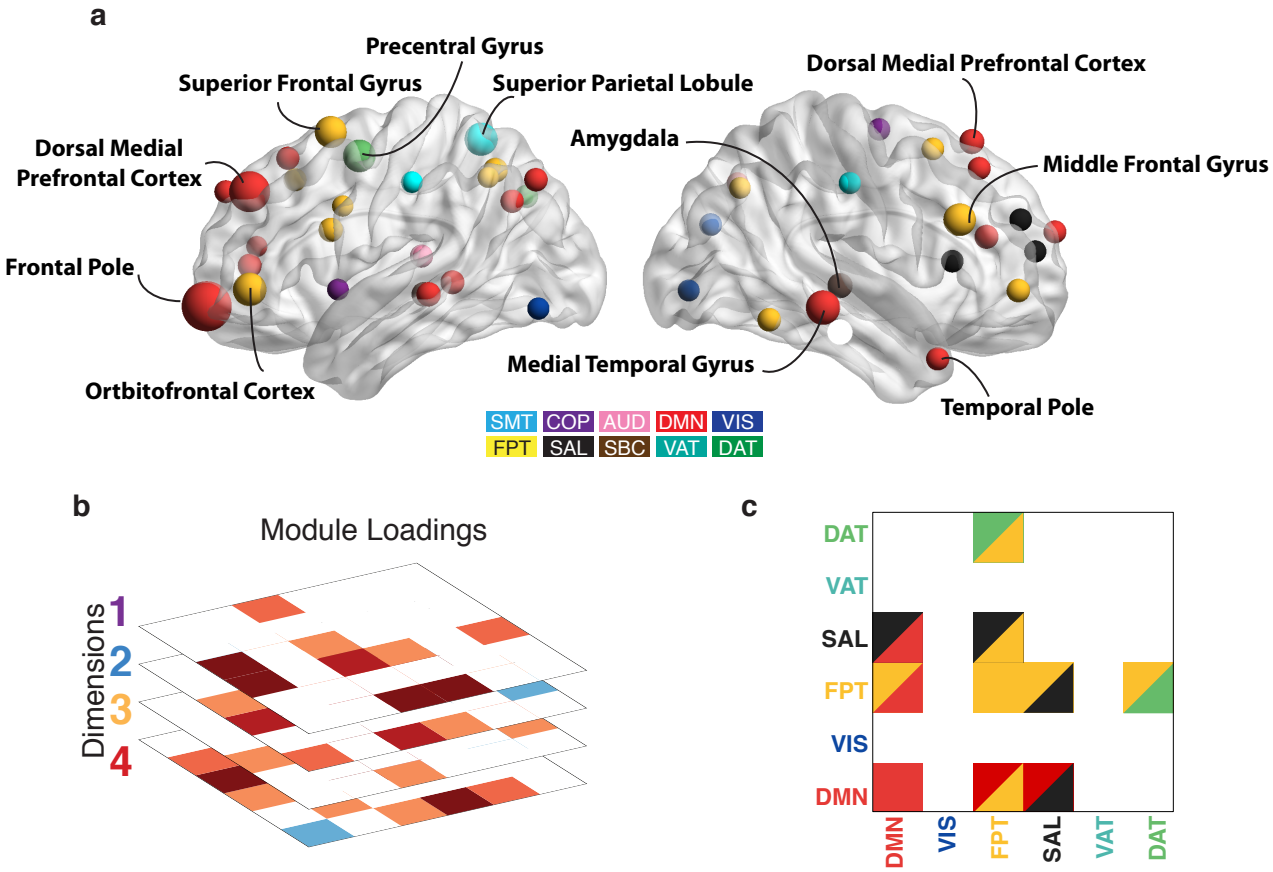


Figure 5 | **Loss of segregation between default mode and executive networks is shared across all dimensions.** **(a)** By searching for overlap of edges that contributed significantly to each dimension, we found common edges that were implicated across all dimensions of psychopathology. These were then summarized at a nodal level by the sum of their absolute loadings. Nodes that contributed significantly to every dimension included the frontal pole, superior frontal gyrus, dorsomedial prefrontal cortex, medial temporal gyrus, and amygdala. **(b)** Results of a similar analysis conducted at the module level. **(c)** Loss of segregation between the default mode and executive networks were shared across all four dimensions.

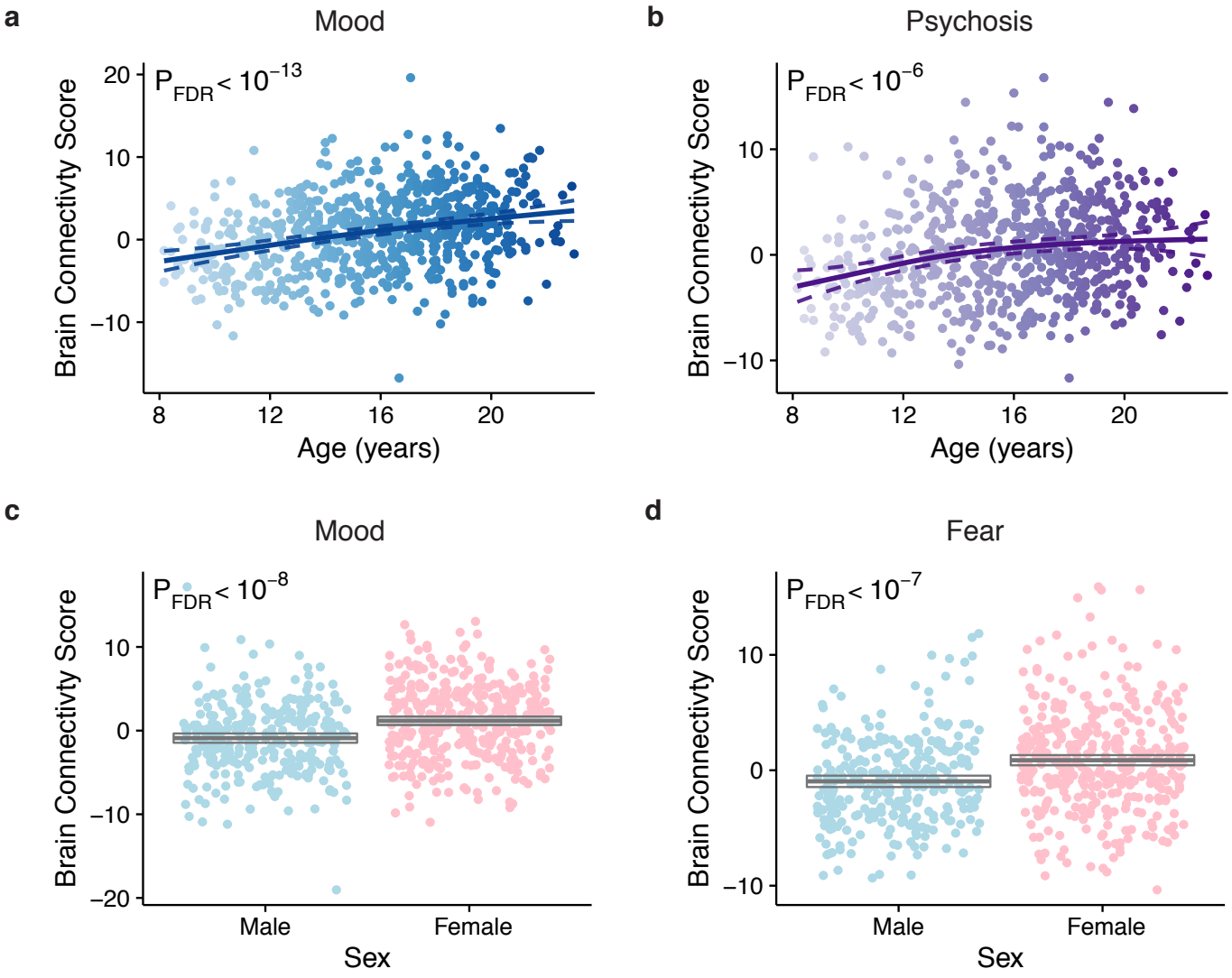


Figure 6 | **Developmental effects and sex differences are concentrated in specific dimensions.** Connectivity patterns associated with both the mood **(a)** and psychosis **(b)** dimensions increased significantly with age. Additionally, connectivity patterns associated with both the mood **(c)** and fear **(d)** symptoms were significantly more prominent in females than males. Multiple comparisons were controlled for using the False Discovery Rate ($q < 0.05$).

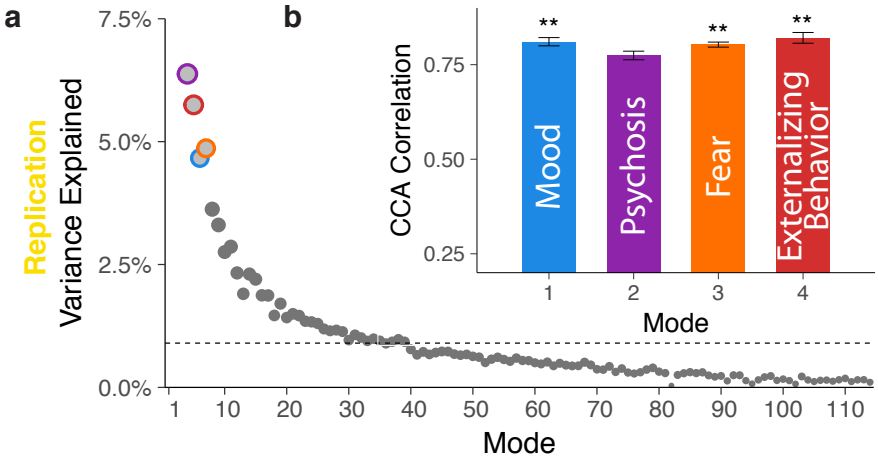


Figure 7 | **Linked dimensions of psychopathology were replicated in an independent sample.** All procedures were repeated in an independent replication sample of 336 participants. **(a)** The first four canonical variates in the replication sample were selected for further analysis based on covariance explained. Dashed line marks the average covariance explained. **(b)** The mood, fear, and externalizing behavior dimensions were significant by permutation testing. Corresponding variates are circled in (a). Error bars denote standard error. $^{**}P_{FDR} < 0.01$.

Supplementary Figure Legends

Supplementary Figure 1 | **Sample Construction. (a)** The cross-sectional sample of the Philadelphia Neurodevelopmental Cohort (PNC) has 1601 participants in total. Excluding the one missing clinical data, 1600 participants were randomly stratified into a discovery (n=1069) and a replication (n=531) sample. Applying health, structural and functional imaging quality exclusion criteria (details in Online Methods), 663 and 336 subjects were included in the final discovery and replication samples, respectively. **(b)** The two samples had similar demographic composition, including distributions of age, race, sex and overall psychopathology.

Supplementary Figure 2 | **Connectivity feature selection using median absolute deviation (MAD).** Since sCCA seeks to capture sources of variation common to both datasets, we selected top 10% or 3410 connectivity features that were variable across the discovery sample. **(a)** The variance was calculated using the median absolute deviation (MAD). It is defined as the median of the difference between each element and the median in a vector. **(b)** MAD of each edge strength in decreasing order. The 95th, 90th, and 75th percentile are labeled, where the 90th corresponds to 3410 edges. **(c)** Average connectivity matrix across all participants of edges with MAD at 100th, 95th, 90th, and 75th percentile levels.

Supplementary Figure 3 | **Grid search for regularization parameters.** We tuned the L^1 regularization parameters for the connectivity and the clinical features in sCCA. The range of sparsity parameters is constrained to be between 0 and 1 in the PMA package,[47] where 0 indicates the smallest number of features (i.e. highest level of sparsity) and 1 indicates the largest number of features (i.e. lowest level of sparsity). We conducted a grid search in increment of 0.1 to determine the combination of parameters that would yield the

highest canonical correlation of the first variate across 10 randomly resampled datasets, each consisting of two-thirds of the discovery dataset.

Supplementary Figure 4 | **Permutation testing to assess significance of linked dimensions. (a)** Schematic of permutation procedure. Connectivity data was held constant, while the rows of the clinical matrix were randomly shuffled, so as to break the linkage of participants' connectivity features and their symptom features. As permutation could induce arbitrary axis rotation, which changes the order of canonical variates, or axis reflection, which causes a sign change for the weights, we matched the canonical variates resulting from permuted data matrices to the ones derived from the original data matrix by comparing the clinical loadings ($\mathbf{v}$).[94] **(b)** Null distributions of correlations generated by the permuted data. Dashed line represents the correlation from the original dataset.

Supplementary Figure 5 | **Patterns of canonical variates were robust to methodological choices.** We found four canonical variates based on covariance explained and correlation across methodological choices, including **(a)** the number of features entered into the analysis (edges with top 5% variance based on MAD), **(b)** an alternative parcellation (Gordon et al.[99]), and **(c)** using alternative techniques of dimensionality reduction (the first 111 principal components). Dashed line marks the average covariance explained. Corresponding variates on the right panels are circled in the left. Error bars denote standard error.

Supplementary Figure 6 | **Resampling procedure to identify stable features contributing to each linked dimension. (a)** Schematic of the resampling procedure. In each sample, two-thirds of the discovery dataset was first randomly selected. The sample size was completed to be the same as the original by replacing with those already selected. **(b)** Resampling distribution for clinical features in each linked dimension. Each bar represents the 95% confidence interval. DSM categories to which each symptom item belongs are shown.

Supplementary Figure 7 | **Network module analysis. (a)** Summarizing loadings on a *between*- and *within*-network basis using *a priori* community assignment from the parcellation of Power et al.[100] **(b)** Schematic for generating null model for modular analysis. Community membership was randomly assigned to each node while controlling for community size. Mean *between*- and *within*-module loadings were then calculated based on these permuted modules, which we used to assess the statistical significance by comparing the orginal values against the null distribution.

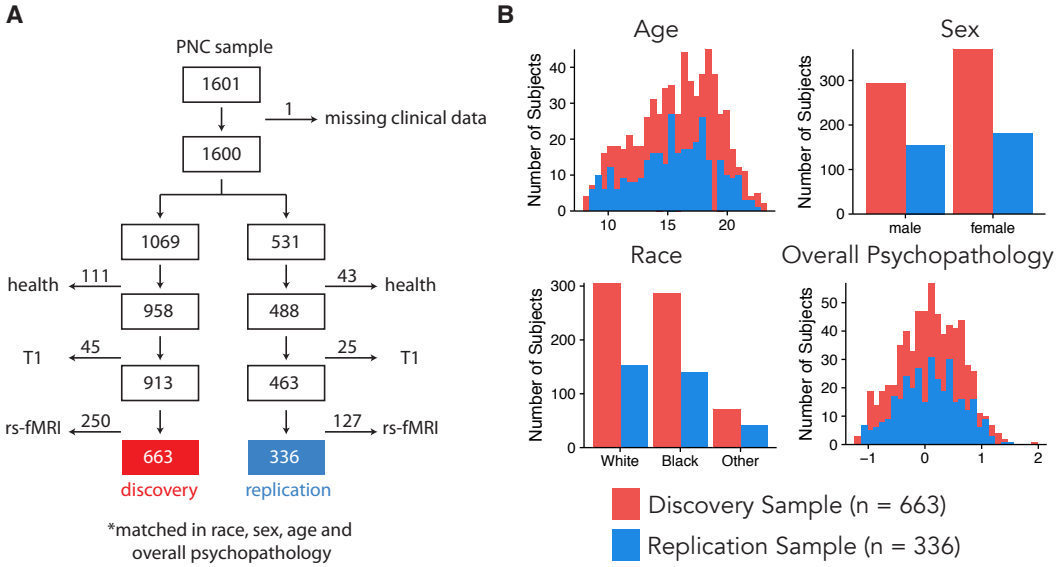
Supplementary Figure 8 | **Canonical variates in the replication sample accord with those found in the discovery sample. (a)** Scatter plots of brain and clinical scores (linear combinations of functional connectivity and psychiatric symptoms, respectively) demonstrate the correlated multivariate patterns of connectomic and clinical features. Colored dots in each panel indicate the severity of a representative clinical symptom that contributed the most to this canonical variate. Each insert displays the null distribution of sCCA correlation by permutation testing. Dashed line marks the actual correlation.

Supplementary Information | Linked dimensions of psychopathology and connectivity in functional brain networks

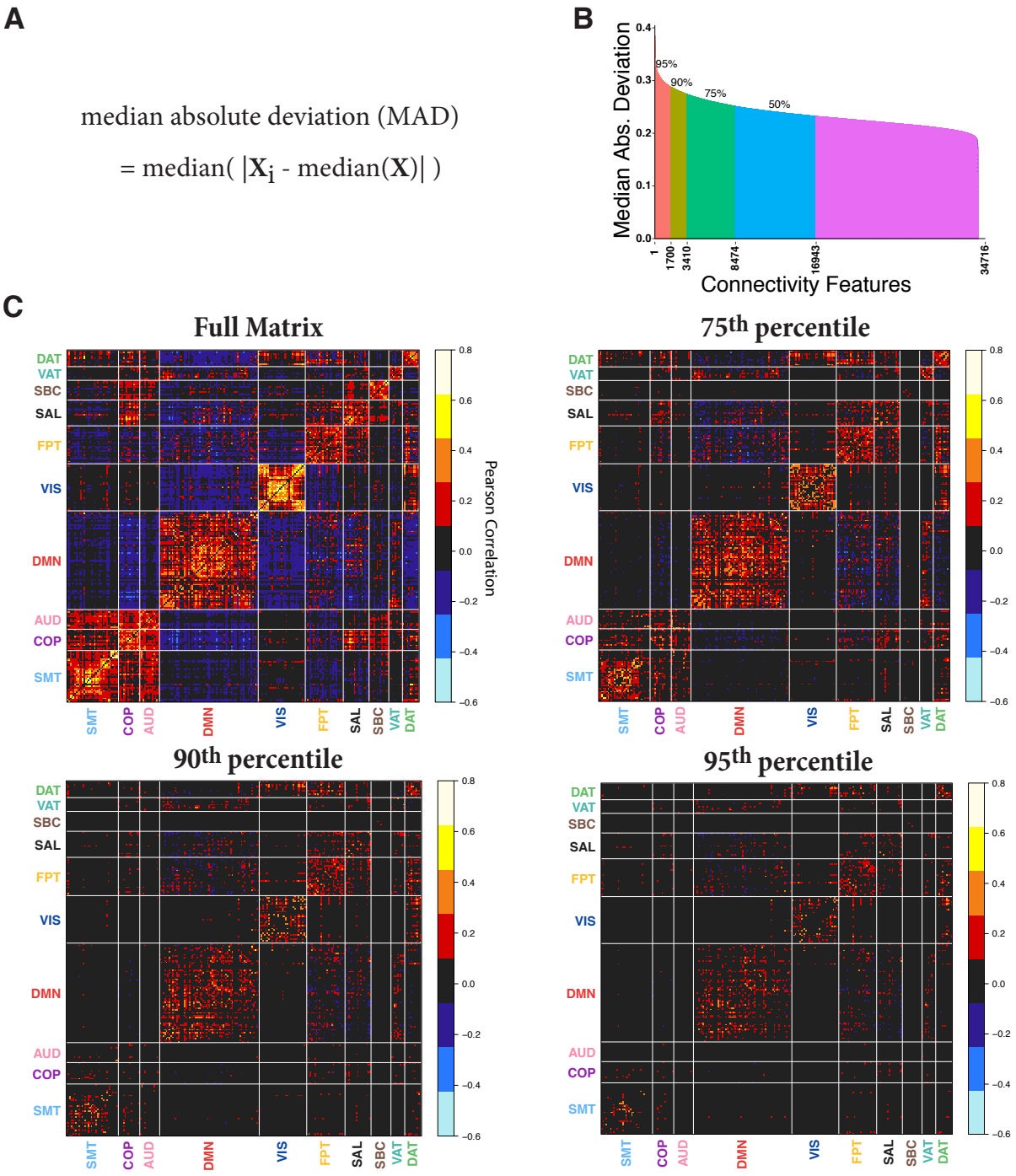
1. **Supplementary Figures and Legends 1–8**;
2. **Supplementary Tables 1-3**;

Supplementary Information

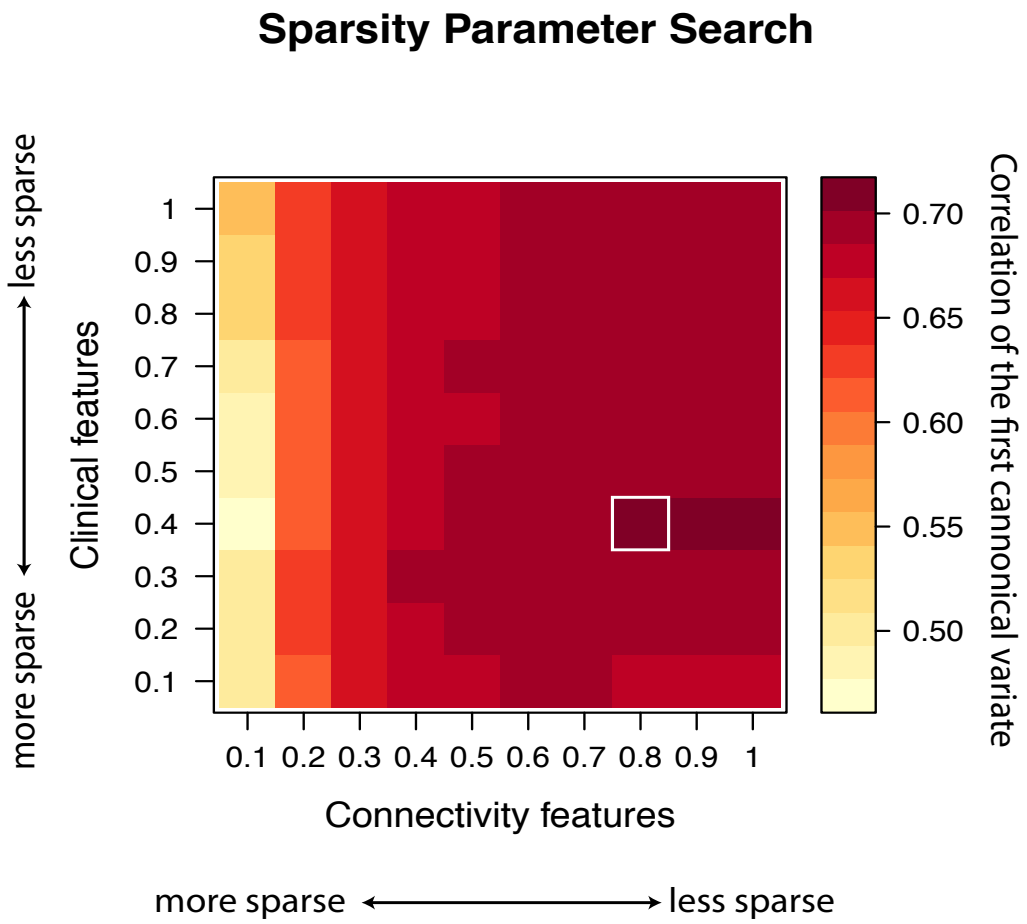
Supplementary Figures



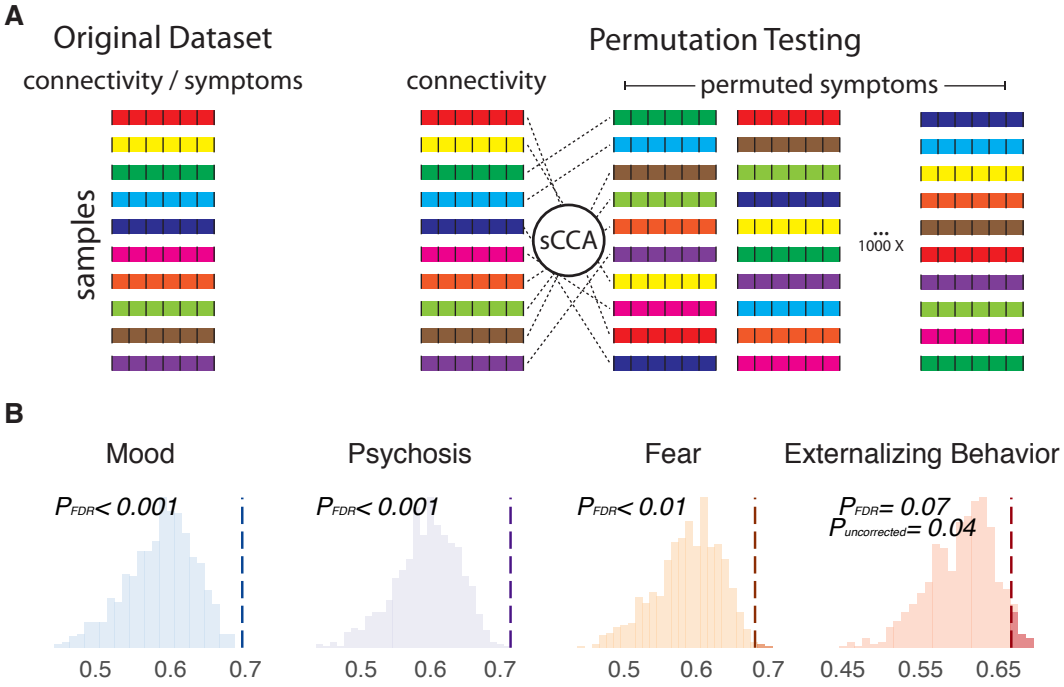
Supplementary Figure 1 | **Sample Construction. (a)** The cross-sectional sample of the Philadelphia Neurodevelopmental Cohort (PNC) has 1601 participants in total. Excluding the one missing clinical data, 1600 participants were randomly stratified into a discovery (n=1069) and a replication (n=531) sample. Applying health, structural and functional imaging quality exclusion criteria (details in Online Methods), 663 and 336 subjects were included in the final discovery and replication samples, respectively. **(b)** The two samples had similar demographic composition, including distributions of age, race, sex and overall psychopathology.



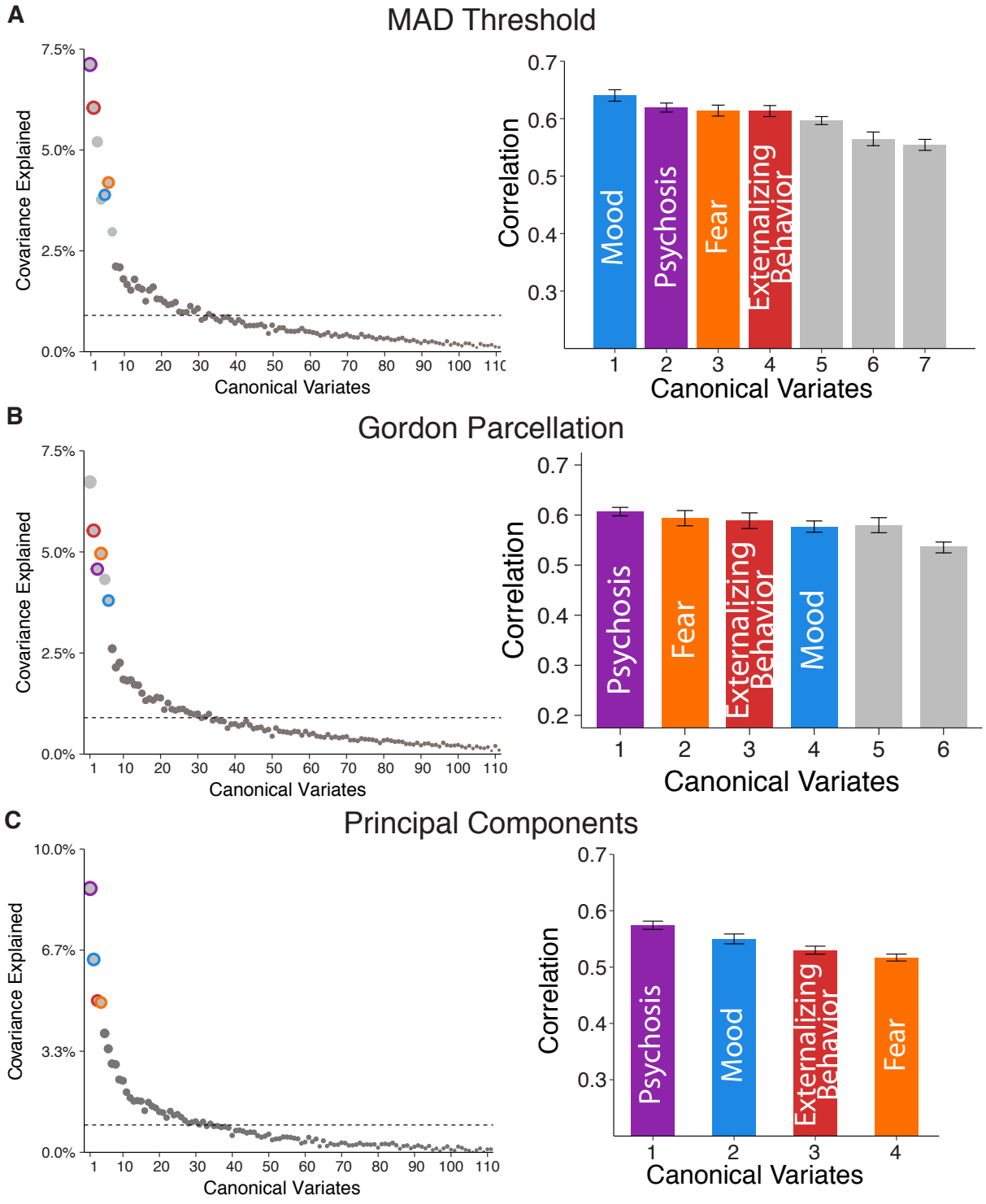
Supplementary Figure 2 | Connectivity feature selection using median absolute deviation (MAD). Since sCCA seeks to capture sources of variation common to both datasets, we selected top 10% or 3410 connectivity features that were variable across the discovery sample. **(a)** The variance was calculated using the median absolute deviation (MAD). It is defined as the median of the difference between each element and the median in a vector. **(b)** MAD of each edge strength in decreasing order. The 95th, 90th, and 75th percentile are labeled, where the 90th corresponds to 3410 edges. **(c)** Average connectivity matrix across all participants of edges with MAD at 100th, 95th, 90th, and 75th percentile levels.



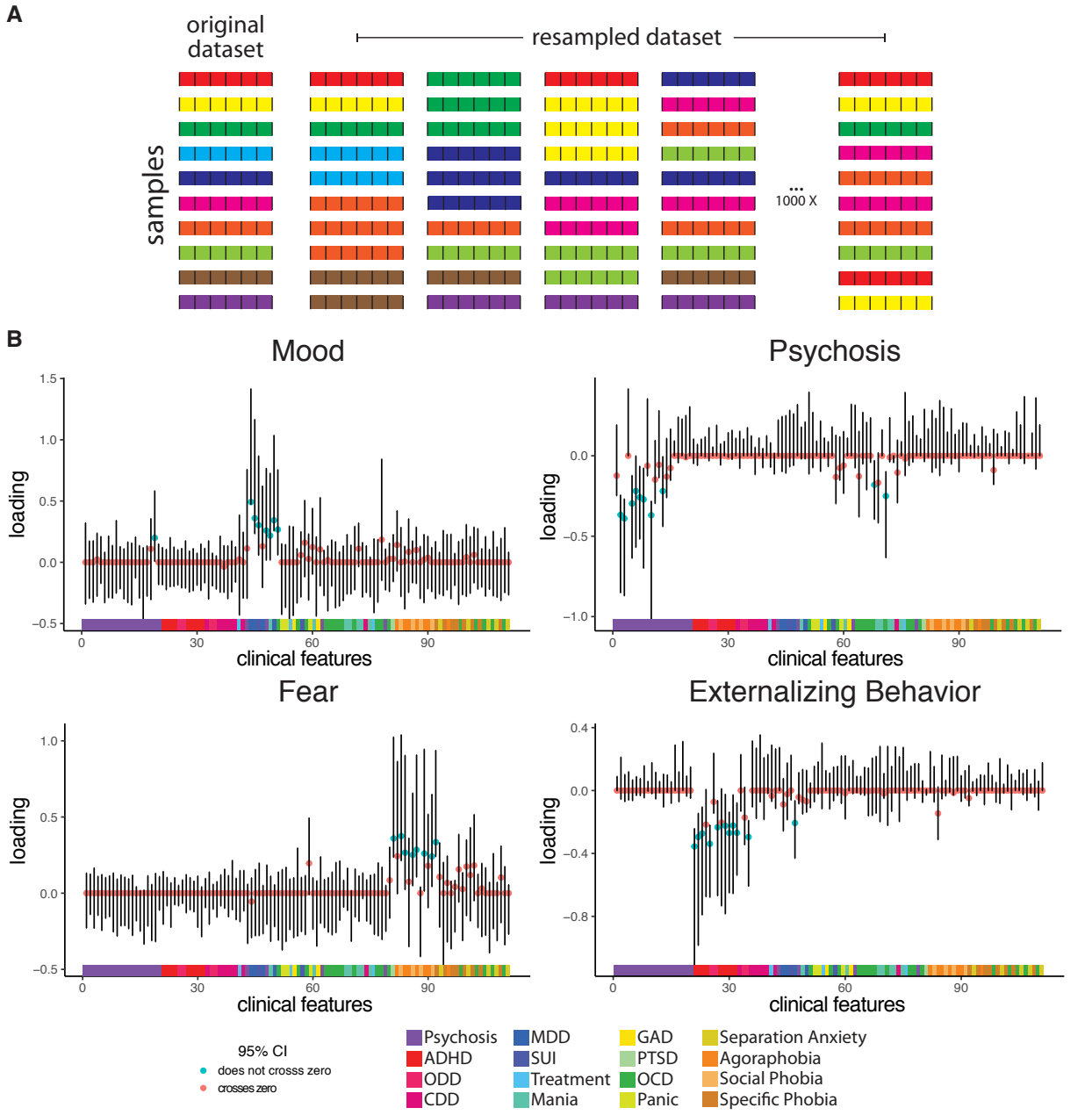
Supplementary Figure 3 | Grid search for regularization parameters. We tuned the L^1 regularization parameters for the connectivity and the clinical features in sCCA. The range of sparsity parameters is constrained to be between 0 and 1 in the PMA package,[47] where 0 indicates the smallest number of features (i.e. highest level of sparsity) and 1 indicates the largest number of features (i.e. lowest level of sparsity). We conducted a grid search in increment of 0.1 to determine the combination of parameters that would yield the highest canonical correlation of the first variate across 10 randomly resampled datasets, each consisting of two-thirds of the discovery dataset.



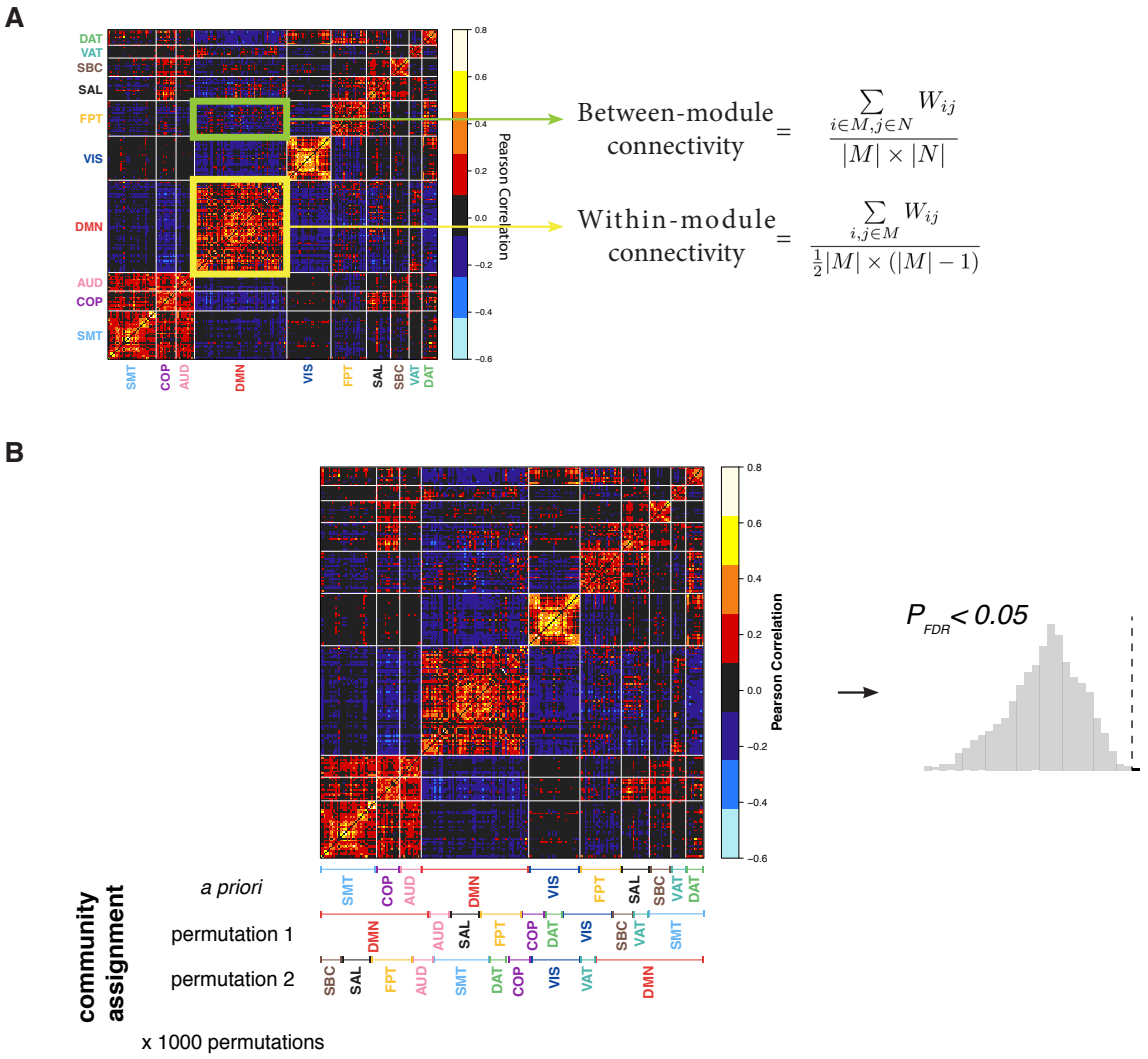
Supplementary Figure 4 | **Permutation testing to assess significance of linked dimensions. (a)** Schematic of permutation procedure. Connectivity data was held constant, while the rows of the clinical matrix were randomly shuffled, so as to break the linkage of participants' connectivity features and their symptom features. As permutation could induce arbitrary axis rotation, which changes the order of canonical variates, or axis reflection, which causes a sign change for the weights, we matched the canonical variates resulting from permuted data matrices to the ones derived from the original data matrix by comparing the clinical loadings ($\mathbf{v}$).[94] **(b)** Null distributions of correlations generated by the permuted data. Dashed line represents the correlation from the original dataset.



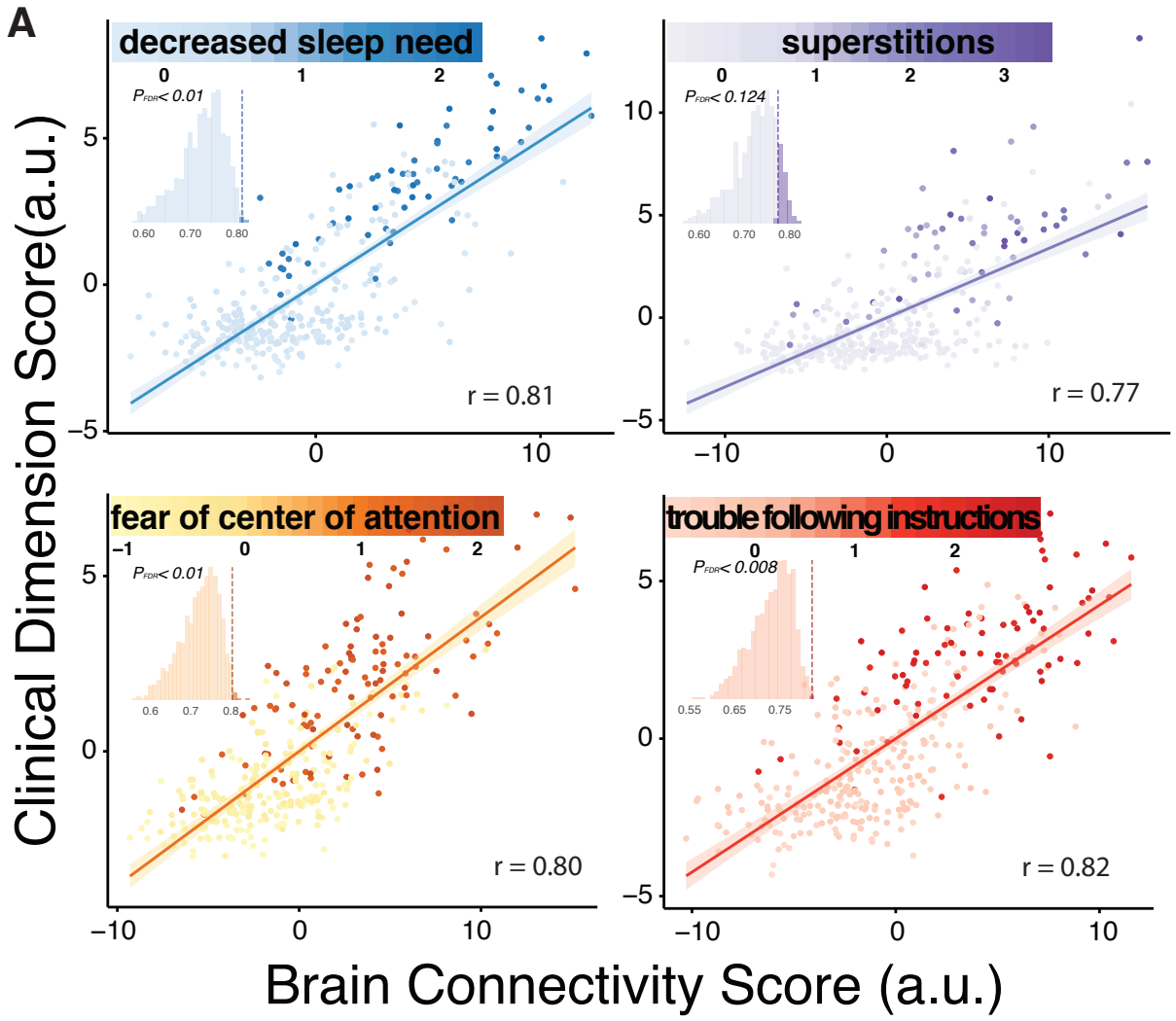
Supplementary Figure 5 | Patterns of canonical variates were robust to methodological choices. We found four canonical variates based on covariance explained and correlation across methodological choices, including **(a)** the number of features entered into the analysis (edges with top 5% variance based on MAD), **(b)** an alternative parcellation (Gordon et al.[99]), and **(c)** using alternative techniques of dimensionality reduction (the first 111 principal components). Dashed line marks the average covariance explained. Corresponding variates on the right panels are circled in the left. Error bars denote standard error.



Supplementary Figure 6 | **Resampling procedure to identify stable features contributing to each linked dimension.** **(a)** Schematic of the resampling procedure. In each sample, two-thirds of the discovery dataset was first randomly selected. The sample size was completed to be the same as the original by replacing with those already selected. **(b)** Resampling distribution for clinical features in each linked dimension. Each bar represents the 95% confidence interval. DSM categories to which each symptom item belongs are shown.



Supplementary Figure 7 | **Network module analysis.** **(a)** Summarizing loadings on a *between-* and *within*-network basis using *a priori* community assignment from the parcellation of Power et al.[100] **(b)** Schematic for generating null model for modular analysis. Community membership was randomly assigned to each node while controlling for community size. Mean *between-* and *within*-module loadings were then calculated based on these permuted modules, which we used to assess the statistical significance by comparing the orginal values against the null distribution.



Supplementary Figure 8 | Canonical variates in the replication sample accord with those found in the discovery sample. (a) Scatter plots of brain and clinical scores (linear combinations of functional connectivity and psychiatric symptoms, respectively) demonstrate the correlated multivariate patterns of connectomic and clinical features. Colored dots in each panel indicate the severity of a representative clinical symptom that contributed the most to this canonical variate. Each insert displays the null distribution of sCCA correlation by permutation testing. Dashed line marks the actual correlation.

Supplementary Tables

Supplementary Table 1 | Demographics in each sample

		Discovery	Replication	Total
n		663	336	999
Sex	Male	293	155	448
	Female	370	181	551
Race	White	306	153	459
	Black	286	141	427
	Other	71	42	113
Age	8-10	70	40	110
	11-13	125	63	188
	14-16	195	102	297
	17-19	206	100	306
	20-22	58	30	88
	>22	9	1	10
	mean	15.82 ± 3.32	15.65 ± 3.32	15.76 ± 3.32

Supplementary Table 2 | Correlations of loadings between covariate-regressed and non-regressed features

	Connectivity	Symptoms
Mood	0.73	0.54
Psychosis	0.95	0.88
Fear	0.70	0.35
Externalizing behavior	0.98	0.97

Loadings of both connectivity and clinical features across dimensions were highly correlated between input data that had age and sex regressed out of and those that had not. All correlations were statistically significant ($P_{FDR} < 0.001$).

Supplementary Table 3 | Clinical Assessment

Questions from the GOASSESS Semi-Structured Interview

DSM	Label	Question
Attention Deficit Disorder	ADD011	Did you often have trouble paying attention or keeping your mind on your school, work, chores, or other activities that you were doing? **(trouble paying attention)**
	ADD012	Did you often have problems following instructions and often fail to finish school, work, or other things you meant to get done?
	ADD013	Did you often dislike, avoid, or put off school or homework (or any other activity requiring concentration) **(problems following instructions)**
	ADD014	Did you often lose things you needed for school or projects at home (assignments or books) or make careless mistakes in school work or other activities? **(making careless mistakes)**
	ADD015	Did you often have trouble making plans, doing things that had to be done in a certain kind of order, or that had a lot of different steps? **(trouble making plans)**
	ADD016	Did you often have people tell you that you did not seem to be listening when they spoke to you or that you were daydreaming? **(trouble listening)**
	ADD020	Did you often have difficulty sitting still for more than a few minutes at a time, even after being asked to stay seated, or did you often fidget with your hands or feet or wiggle in your seat or were you "always on the go"? **(difficulty sitting still)**
	ADD021	Did you often blurt out answers to other people's questions before they finished speaking or interrupt people abruptly?
	ADD022	Did you often join other people's conversations or have trouble waiting your turn (e.g., waiting in line, waiting for a teacher to call on you in class)? **(difficulty waiting turns)**
Agoraphobia	AGR001	Looking at this card, have you ever been very nervous or afraid of being in crowds (for example, a classroom, cafeteria, restaurant, or movie theater)?
	AGR002	Looking at this card, have you ever been very nervous or afraid of going to public places (such as a store or shopping mall)?
	AGR003	Looking at this card, have you ever been very nervous or afraid of being in an open field?
	AGR004	Looking at this card, have you ever been very nervous or afraid of going over bridges or through tunnels? **(bridges/tunnels)**
	AGR005	Looking at this card, have you ever been very nervous or afraid of traveling by yourself? **(solo travel)**
	AGR006	Looking at this card, have you ever been very nervous or afraid of traveling away from home? **(leaving home)**
	AGR007	Looking at this card, have you ever been very nervous or afraid of traveling in a car?
	AGR008	Looking at this card, have you ever been very nervous or afraid of using public transportation like a bus or SEPTA? **(public transit)**

Continued on next page

Supplementary Table 3 – *Continued from previous page*

DSM	Label	Question
Conduct Disorder	CDD001	Was there ever a time when you often did things that got you into trouble with adults like lying or stealing (something worth more than $5), from family, others, or stores?
	CDD002	Did you ever skip school, stay out at night later than you were supposed to (more than 2 hours), or run away from home overnight?
	CDD003	Did you ever set fires, break into cars, or destroy someone else's property on purpose?
	CDD004	Do you have a probation officer or have you ever been on probation?
	CDD005	Did you often bully others (hitting, threatening or scaring someone who was younger or smaller), threaten or frighten someone on purpose, or often start physical fights with others?
	CDD006	Have you ever been physically cruel to an animal or person (on purpose)?
	CDD007	Did you ever try to hurt someone with a weapon (a bat, brick, broken bottle, knife, or gun)?
	CDD008	Did you ever threaten someone?
Depression	DEP001	Has there ever been a time when you felt sad or depressed most of the time? **(feeling sad)**
	DEP002	Has there ever been a time when you cried a lot, or felt like crying? **(crying)**
	DEP004	Has there ever been a time when you felt grouchy, irritable or in a bad mood most of the time; even little things would make you mad? **(irritability)**
	DEP006	Has there ever been a time when nothing was fun for you and you just weren't interested in anything? **(anhedonia)**
Generalized Anxiety	GAD001	Have you ever been a worrier?
	GAD002	Did you worry a lot more than most children/people your age?
Manic Disorder	MAN001	Have there been times when you were much more active, excited or energetic than usual, had problems sitting still, or needed to move around a lot? **(overly energetic)**
	MAN002	Has there ever been a time when you felt so full of energy that you couldn't stop doing things and didn't get tired?
	MAN003	Has there ever been a time when you felt like you hardly needed sleep?
	MAN004	Have there been times when you kept talking a lot, couldn't stop talking, talked faster than usual, had thoughts faster than usual, or had so many ideas in your head that you could hardly keep track of them? **(pressured speech)**
	MAN005	Have you ever had a time when you felt much more happy or excited than you usually do when there was nothing special going on?
	MAN006	Have you ever had a time when you felt like you could do almost anything?
	MAN007	Has there ever been a time when you felt unusually grouchy, cranky, or irritable; when the smallest things would make you really mad? **(irritability)**

Continued on next page

Supplementary Table 3 – *Continued from previous page*

DSM	Label	Question
Obsessive Compulsive Disorder	OCD001	Have you ever been bothered by thoughts that don't make sense to you, that come over and over again and won't go away, such as concern with harming others/self? **(thoughts of harming)**
	OCD002	Have you ever been bothered by thoughts that don't make sense to you, that come over and over again and won't go away, such as pictures of violent things?
	OCD003	Have you ever been bothered by thoughts that don't make sense to you, that come over and over again and won't go away, such as thoughts about contamination/germs/illness?
	OCD004	Have you ever been bothered by thoughts that don't make sense to you, that come over and over again and won't go away, such as fear that you would do something/say something bad without intending to?
	OCD005	Have you ever been bothered by thoughts that don't make sense to you, that come over and over again and won't go away, such as feelings that bad things that happened were your fault?
	OCD006	Have you ever been bothered by thoughts that don't make sense to you, that come over and over again and won't go away, such as forbidden/bad thoughts?
	OCD007	Have you ever been bothered by thoughts that don't make sense to you, that come over and over again and won't go away, such as need for symmetry/exactness?
	OCD008	Have you ever been bothered by thoughts that don't make sense to you, that come over and over again and won't go away, such as religious thoughts?
	OCD011	Have you ever had to do something over and over again - that would have made you feel really nervous if you couldn't do it, like cleaning or washing (for example, your hands, house)?
	OCD012	Have you ever had to do something over and over again - that would have made you feel really nervous if you couldn't do it, like counting?
	OCD013	Have you ever had to do something over and over again - that would have made you feel really nervous if you couldn't do it, like checking (for example, doors, locks, ovens)?
	OCD014	Have you ever had to do something over and over again - that would have made you feel really nervous if you couldn't do it, like getting dressed over and over again?
	OCD015	Have you ever had to do something over and over again - that would have made you feel really nervous if you couldn't do it, like going in and out a door over and over again?
	OCD016	Have you ever had to do something over and over again - that would have made you feel really nervous if you couldn't do it, like ordering or arranging things?
	OCD017	Have you ever had to do something over and over again - that would have made you feel really nervous if you couldn't do it, like doing things over and over again at bedtime, like arranging the pillows, sheets, or other things?

Continued on next page

Supplementary Table 3 – *Continued from previous page*

DSM	Label	Question
	OCD018	Have you ever saved up so many things that people complained or they got in the way?
	OCD019	Do you feel the need to do things just right (like they have to be perfect)?
Oppositional Defiant Disorder	ODD001	Was there a time when you often did things that got you into trouble with adults such as losing your temper, arguing with or talking back to adults, or being grouchy or irritable with them? **(losing temper)**
	ODD002	Was there a time when you often got into trouble with adults for refusing to do what they told you to do or for breaking rules at home/school? **(breaking rules)**
	ODD003	Did you often annoy other people on purpose or blame other people for your mistakes (excluding siblings)?
	ODD005	Did you ever get into trouble for getting even with other people by doing things to hurt them, telling lies about them, or messing up their things?
	ODD006	Were you often irritable or grouchy, or did you often get angry because you thought that things were unfair? **(irritability due to unfairness)**
Panic Disorder	PAN001	Have you ever had an attack like this?
	PAN003	Has there ever been a time when all of a sudden you felt very, very scared or uncomfortable - and your chest hurt, you couldn't catch your breath, your heart beat very fast, you felt very shaky, and sweaty/tingly/numb in your hands or feet?
	PAN004	Has there ever been a time when all of a sudden, you felt that you were losing control, something terrible was going to happen, that you were going crazy, or going to die?
Specific Phobia	PHB001	Looking at this card, have you ever been very nervous or afraid of animals or bugs, like dogs, snakes, or spiders?
	PHB002	Looking at this card, have you ever been very nervous or afraid of being in really high places, like a roof or tall building?
	PHB003	Looking at this card, have you ever been very nervous or afraid of water or situations involving water, such as a swimming pool, lake, or ocean?
	PHB004	Looking at this card, have you ever been very nervous or afraid of storms, thunder, or lightning?
	PHB005	Looking at this card, have you ever been very nervous or afraid of doctors, needles, or blood?
	PHB006	Looking at this card, have you ever been very nervous or afraid of closed spaces, like elevators or closets?
	PHB007	Looking at this card, have you ever been very nervous or afraid of flying or airplanes?
	PHB008	Looking at this card, have you ever been very nervous or afraid of any other things or situations?
Psychosis	PSY001	Have you ever heard voices when no one was there? **(auditory verbal hallucination)**
	PSY029	Have you ever seen visions or seen things which other people could not see?
	PSY050	Have you ever smelled strange odors other people could not smell?

Continued on next page

Supplementary Table 3 – *Continued from previous page*

DSM	Label	Question
	PSY060	Have you ever had strange feelings in your body like things were crawling on you or someone touching you and nothing or no one was there?
	PSY070	Have you ever believed in things that most other people or your parents don't believe in?
	PSY071	Have you ever believed in things and later found out they weren't true, like people being out to get you, or talking about you behind your back, or controlling what you do or think? **(persecutory/suspicious)**
PTSD	PTD001	Have you ever been very upset by seeing a dead body or by seeing pictures of the dead body of somebody you knew well?
Treatment Seeking	SCR001	Have you ever talked to a counselor, psychologist, social worker, psychiatrist or some other professional about your feelings or problems with your mood or behaviors?
	SCR006	Are you currently taking medication because of your emotions and/or behaviors?
	SCR007	Have you ever had to go to a hospital and stay overnight because of problems with your mood, feelings, or how you were acting?
	SCR008	Have you or anyone else (like your friends, parents, or teachers) ever thought you needed help because of problems with your mood, feelings, or how you were acting?
Separation Anxiety	SEP500	Since you were 5 years old, has there ever been a time when you had a lot of worries about your (attachment figures) and were very upset or got sick (for example, felt sick to your stomach, headaches, thrown-up) when you were away from him/her?
	SEP508	Has there ever been a time when you wanted to stay home from school or not go to other places (for example, sleep-overs) without your (attachment figures)?
	SEP509	When you knew that you were going to be away from home or (attachment figure(s)), did you get very upset and worry (e.g., when you learned (attachment figure(s)) were going on an upcoming trip or night out)?
	SEP510	Did you ever worry/have bad dreams about something terrible happening to you or your (attachment figures) so that you would not see them again?
	SEP511	Were you scared to be alone in your room (or any place in your house) or did you need your (attachment figure(s)) to stay with you while you fell asleep?
Subthreshold Psychosis	SIP003	I think that I have felt that there are odd or unusual things going on that I can't explain. **(odd/unusual thoughts)**
	SIP004	I think that I might be able to predict the future.
	SIP005	I may have felt that there could possibly be something interrupting or controlling my thoughts, feelings, or actions. **(thought control)**
	SIP006	I have had the experience of doing something differently because of my superstitions. **(superstitions)**

Continued on next page

Supplementary Table 3 – *Continued from previous page*

DSM	Label	Question
	SIP007	I think I may get confused at times whether something I experience or perceive may be real or may be just part of my imagination or dreams. **(reality confusion)**
	SIP008	I have thought that it might be possible that other people can read my mind, or that I can read others' minds
	SIP009	I wonder if people may be planning to hurt me or even may be about to hurt me.
	SIP010	I believe that I have special natural or supernatural gifts beyond my talents and natural strengths.
	SIP011	I think I might feel like my mind is "playing tricks" on me. **(mind tricks)**
	SIP012	I have had the experience of hearing faint or clear sounds of people or a person mumbling or talking when there is no one near me. **(auditory perception)**
	SIP013	I think that I may hear my own thoughts being said out loud. **(audible thoughts)**
	SIP014	I have been concerned that I might be "going crazy."
	SIP027	Do people ever tell you that they can't understand you?
	SIP028	Do people ever seem to have difficulty understanding you?
	SIP032	Do you ever feel a loss of sense of self or feel disconnected from yourself or your life? **(loss sense of self)**
	SIP033	Has anyone pointed out to you that you are less emotional or connected to people than you used to be?
	SIP038	Within the past 6 months, are you having a harder time getting your work or schoolwork done?
	SIP039	Within the past 6 months, are you having a harder time getting normal activities done?
Social Phobia	SOC001	Looking at this card, was there ever a time in your life when you felt afraid or uncomfortable or really, really shy with people, like meeting new people, going to parties, or eating or drinking, writing or doing homework in front of others? **(focus of social situation)**
	SOC002	Looking at this card, was there ever a time in your life when you felt afraid or uncomfortable talking on the telephone or with people your own age who you don't know very well? **(novel social situations)**
	SOC003	Looking at this card, was there ever a time in your life when you felt afraid or uncomfortable when you had to do something in front of a group of people, like speaking in class?
	SOC004	Looking at this card, was there ever a time in your life when you felt afraid or uncomfortable acting, performing, giving a talk/speech, playing a sport or doing a musical performance, or taking an important test or exam (even though you studied enough)? **(public performance)**
	SOC005	Looking at this card, was there ever a time in your life when you felt afraid or uncomfortable because you were the center of attention and were concerned something embarrassing might happen and you felt very afraid or felt uncomfortable? **(center of attention)**

Continued on next page

Supplementary Table 3 – *Continued from previous page*

DSM	Label	Question
Suicidality	SUI001	Have you ever thought a lot about death or dying?
	SUI002	Have you ever thought about killing yourself? **(suicidality)**