University of Pennsylvania
ScholarlyCommons

Publicly Accessible Penn Dissertations

2019

Linking Functional Brain Networks To Psychopathology And Beyond

Huchuan Xia
University of Pennsylvania, cedrichxia@gmail.com

Follow this and additional works at: https://repository.upenn.edu/edissertations

 Part of the Behavioral Neurobiology Commons, and the Biostatistics Commons

Recommended Citation

Xia, Huchuan, "Linking Functional Brain Networks To Psychopathology And Beyond" (2019). *Publicly Accessible Penn Dissertations*. 3540.
https://repository.upenn.edu/edissertations/3540

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/edissertations/3540
For more information, please contact repository@pobox.upenn.edu.

Linking Functional Brain Networks To Psychopathology And Beyond

Abstract

Neurobiological abnormalities associated with neuropsychiatric disorders do not map well to existing diagnostic categories. High co-morbidity suggests dimensional circuit-level abnormalities that cross diagnoses. As neuropsychiatric disorders are increasingly reconceptualized as disorders of brain development, deviations from normative brain network reconfiguration during development are hypothesized to underlie many illness that arise in young adulthood. In this dissertation, we first applied recent advances in machine learning to a large imaging dataset of youth (n=999) to delineate brain-guided dimensions of psychopathology across clinical diagnostic boundaries. Specifically, using sparse Canonical Correlation Analysis, an unsupervised learning method that seeks to capture sources of variation common to two high-dimensional datasets, we discovered four linked dimensions of psychopathology and connectivity in functional brain networks, namely, mood, psychosis, fear, and externalizing behavior. While each dimension exhibited an unique pattern of functional brain connectivity, loss of network segregation between the default mode and executive networks emerged as a shared connectopathy common across four dimensions of psychopathology.

Building upon this work, in the second part of the dissertation, we designed, implemented, and deployed a new penalized statistical learning approach, Multi-Scale Network Regression (MSNR), to study brain network connectivity and a wide variety of phenotypes, beyond psychopathology. MSNR explicitly respects both edge- and community-level information by assuming a low rank and sparse structure, both encouraging less complex and more interpretably modeling. Capitalizing on a large neuroimaging cohort (n=1,051), we demonstrated that MSNR recapitulated interpretably and statistically significant associations between functional connectivity patterns with brain development, sex differences, and motion-related artifacts. Compared to common single-scale approaches, MSNR achieved a balance between prediction performance and model complexity, with improved interpretability.

Together, integrating recent advances in multiple disciplines across machine learning, network science, developmental neuroscience, and psychiatry, this body of work fits into the broader context of computational psychiatry, where there is intense interest in the quest of delineating brain network patterns associated with psychopathology, among a diverse range of phenotypes.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Neuroscience

First Advisor

Theodore D. Satterthwaite

Second Advisor

Danielle S. Bassett

Keywords

Biostatistics, Functional Connectivity, Machine Learning, Network Neuroscience, Psychopathology, RDoC

Subject Categories
Behavioral Neurobiology | Biostatistics | Neuroscience and Neurobiology

This dissertation is available at ScholarlyCommons: https://repository.upenn.edu/edissertations/3540

LINKING FUNCTIONAL BRAIN NETWORKS TO PSYCHOPATHOLOGY AND BEYOND

Huchuan Xia

A DISSERTATION

in

Neuroscience

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

Supervisor of Dissertation

_____

Theodore D. Satterthwaite, M.D., M.A.

Assistant Professor of Psychiatry

Co-Supervisor of Dissertation

_____

Danielle S. Bassett, Ph.D.

Associate Professor of Bioengineering, Electrical & Systems Engineering, Psychiatry, Physics and Astronomy, and Neurology

Graduate Group Chairperson

_____

Joshua I. Gold, Ph.D.

Professor of Neuroscience

Dissertation Committee

Geoffrey K. Aguirre, M.D., Ph.D., Professor of Neurology

Frances E. Jensen, M.D., Professor of Neurology

Lyle H. Ungar, Ph.D., Professor of Computer and Information Science

Russell T. Shinohara, Ph.D., Associate Professor of Biostatistics, Epidemiology, and Informatics

LINKING FUNCTIONAL BRAIN NETWORKS TO PSYCHOPATHOLOGY AND BEYOND

COPYRIGHT

2019

HUCHUAN XIA

This work is licensed under the
Creative Commons Attribution-
NonCommercial-ShareAlike 3.0
License

To view a copy of this license, visit

https://creativecommons.org/licenses/by-nc-sa/3.0/us/

This dissertation is dedicated to

my parents, who courageously green-lighted and selflessly supported my idea of coming to America for higher education;

my three living grandparents, who always silently cheered me on, even when I didn't call them as much as I should have;

and the loving memory of my grandmother and my aunt, who passed away during my graduate studies. Both teachers at an elementary school in rural China, they taught me some of the first words and numbers I know.

谨以本博士毕业论文，献给

我的父母， 他们勇敢地同意了并无私地支持了我来美国留学的想法；

我的奶奶、爷爷、外公，尽管我没有常常给他们通电话，他们仍然悄悄地给我加油；

在我读博期间去世的外婆和大姨，作为中国乡村的小学教师，她们是我的启蒙老师。

ACKNOWLEDGMENT

I am incredibly grateful to my adviser, Ted Satterthwaite, who is among the most outstanding scientists, effective teachers, and compassionate mentors I have ever worked with. As an aspiring physician-scientist who is working towards a career in translational psychiatry, I see Ted as my most influential role model. I owe much debt for the progress in my research and achievement to Ted's selflessness in giving his time and advice, and allowing plenty of room for flexibility and individual development. I also have enormous gratitude for my co-adviser, Dani Bassett, whose wisdom, insights, and genuine warmth have benefited me so much throughout my thesis work. Furthermore, I would like to express my sincere gratitude to Zongming Ma, whose crucial collaboration made much of this dissertation possible. I am very much thankful to the members of my thesis committee, including Geoffrey Aguirre, Francis Jensen, Lyle Ungar, and Taki Shinohara, for always making the time to offer me their invaluable feedback and constructive input. Finally, I would like to thank everyone at the Brain Behavior Laboratory for providing me enrichment in and outside of the lab, the Neuroscience Graduate Group, particularly Josh Gold, Yale Cohen, and Christine Clay, for helping me navigate administrative challenges, the Medical Scientist Training Program Office for always keeping me on track, Deutsche Forschungsgemeinschaft for making my dream of studying abroad possible, and the Blavatnik Family Foundation for financially supporting my last year of dissertation.

ABSTRACT

LINKING FUNCTIONAL BRAIN NETWORKS TO PSYCHOPATHOLOGY AND BEYOND

Huchuan Xia

Theodore D. Satterthwaite, M.D.,M.A., Danielle S. Bassett, Ph.D.

Neurobiological abnormalities associated with neuropsychiatric disorders do not map well to existing diagnostic categories. High co-morbidity suggests dimensional circuit-level abnormalities that cross diagnoses. As neuropsychiatric disorders are increasingly reconceptualized as disorders of brain development, deviations from normative brain network reconfiguration during development are hypothesized to underlie many illness that arise in young adulthood. In this dissertation, we first applied recent advances in machine learning to a large imaging dataset of youth (n=999) to delineate brain-guided dimensions of psychopathology across clinical diagnostic boundaries. Specifically, using sparse Canonical Correlation Analysis, an unsupervised learning method that seeks to capture sources of variation common to two high-dimensional datasets, we discovered four linked dimensions of psychopathology and connectivity in functional brain networks, namely, mood, psychosis, fear, and externalizing behavior. While each dimension exhibited an unique pattern of functional brain connectivity, loss of network segregation between the default mode and executive networks emerged as a shared connectopathy common across four dimensions of psychopathology.

Building upon this work, in the second part of the dissertation, we designed, implemented, and deployed a new penalized statistical learning approach, Multi-Scale Network Regression (MSNR), to study brain network connectivity and a wide variety of phenotypes, beyond psychopathology. MSNR explicitly respects both edge- and community-level information by assuming a low rank and sparse structure, both encouraging less complex and more interpretably modeling. Capitalizing on a large neuroimaging cohort (n=1,051), we demonstrated that MSNR recapitulated interpretably and statistically significant associations between functional connectivity patterns with brain development, sex differences, and motion-related artifacts. Compared to common single-scale approaches, MSNR achieved a balance between prediction performance and model complexity, with improved interpretability.

Together, integrating recent advances in multiple disciplines across machine learning, network science, developmental neuroscience, and psychiatry, this body of work fits into the broader context of computational psychiatry, where there is intense interest in the quest of delineating brain network patterns associated with psychopathology, among a diverse range of phenotypes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

General Introduction

Heterogeneity and Comorbidity in Neuropsychiatric Illness

It is increasingly clear that psychiatric diagnostic labels do not "carve nature at its joint." (Singh & Rose, 2009) In other words, clinical boundaries do not map cleanly onto the underlying neurobiology of mental disorders (B. T. R. Insel & Cuthbert, 2015). Two phenomena highlight such mismatch between existing diagnostic categories and distinct neurobiological abnormalities: 1) the marked levels of heterogeneity within an individual diagnosis (Hodgkinson et al., 1987), and 2) co-morbidity across diagnoses (Jacobi et al., 2004). Accordingly, studies have demonstrated different "subtypes" within discrete psychiatric disorders, potentially explaining such heterogeneity (Clementz et al., 2016; Drysdale et al., 2016). Similarly, research has also reported common structural, functional, and genetic abnormalities across psychiatric syndromes, potentially explaining such co-morbidity (Goodkind et al., 2015; Lee et al., 2013; McTeague et al., 2017). This large body of literature gives prominence to the lack of direct correspondence between clinical diagnostic categories and the underlying pathophysiology.

Neurodevelopmental Model of Psychopathology

Another important observation regarding psychopathology is the fact that many major neuropsychiatric disorders first begin in adolescence, with as much as 75% of symptom onset occur before the age of 25 (Tomás Paus, Keshavan, &

Giedd, 2008). This early age of onset, together with insufficient therapeutic interventions, contributes to the tremendous lifetime burden of psychiatric illness, which routinely ranks as having the greatest impact on quality of life worldwide (Whiteford et al., 2013). Not coincidentally, throughout adolescence and early adulthood, the brain undergoes dramatic and complex changes (Cao, Huang, Peng, Dong, & He, 2016; Giedd & Rapoport, 2010; Tomáš Paus, 2005). These evidence indicates that abnormal brain maturation during critical phases of development may be associated with psychopathology (Bassett, Xia, & Satterthwaite, 2018; Rapoport, Giedd, & Gogtay, 2012). Despite the growing appreciation that abnormal neurodevelopment is involved in many psychiatric disorders, much is still unknown about how specific abnormalities of brain development are associated with psychopathology.

These contexts have strongly motivated the goal to identify common circuit-level abnormalities, especially in the developing brain, that may give rise to the heterogeneous psychiatric symptoms across clinical diagnostic categories (Cuthbert & Insel, 2013). Broadly, this is supported by an initiative championed by the Research Domain Criteria (RDoC) of the National Institute of Mental Health (T. Insel et al., 2010). RDoC seeks to construct a biologically-grounded research framework for investigating psychiatric diseases. Critically, RDoC aims to “explore dimensions of functioning that span the full range of human behavior from normal to abnormal”, by integrating multimodal data, including genetic, imaging, and behavior (Casey, Oliveri, & Insel, 2014).

Network Neuroscience of Neurodevelopment and Disease

Network neuroscience is a powerful approach to study the myriad brain systems implicated in psychopathology (Bassett & Sporns, 2017; Bassett et al., 2018). Research in the past two decades in this emerging field has found converging patterns of normal neurodevelopment, using both functional connectivity (e.g. temporal correlation of blood-oxygen-level-dependent, or BOLD, signals) (Gu et al., 2015; Power, Fair, Schlaggar, & Petersen, 2010; Satterthwaite et al., 2013), and structural connectivity (e.g. estimation of white matter tract based fractional anisotropy) networks (Baum et al., 2016). A commonly studied network feature is the connectivity *within*- and *between*-community of the network, also called network modules (Sporns & Betzel, 2016). A network community is a collection of brain regions that are highly connected to each other, but form sparse connections with regions outside of the community. In other words, network community is internally dense, and externally sparse. During normative development, within-community connectivity tend to strengthen with age; whereas between-community connectivity tend to weaken with age (Baum et al., 2016; Power et al., 2010; Satterthwaite et al., 2013). This pattern of network reconfiguration suggests that the developing brain becomes more segregated and specialized during this critical period of plasticity. Given the neurodevelopmental model of psychopathology, this widely replicated network findings during development suggests that deviations from this normative

network could underlie much vulnerability to psychopathology (Bassett et al., 2018; Casey et al., 2014).

Indeed, prior studies using human imaging data and animal models have found brain network patterns do not neatly respect the clinical categories defined in the Diagnostic and Statistical Manual. For example, abnormalities of within- and between-community connectivity of the default mode network and executive networks have been implicated in a diverse range of psychopathology, including both internalizing symptoms (e.g., mood and anxiety) (Berman et al., 2011; Greicius, Supekar, Menon, & Dougherty, 2009; Skudlarski et al., 2010; Whitfield-Gabrieli et al., 2009) and externalizing symptoms (e.g., attention deficit and misconduct behaviors) (Castellanos et al., 2008; Skudlarski et al., 2010; Uddin et al., 2010; von Rhein et al., 2016). In animal studies, local and long-range synchrony of neuronal activity, such as local field potential activity in the γ-band, has been shown to exhibit common abnormal patterns in animal models of a wide range of neuropsychiatric disorders (Adhikari, Topiwala, & Gordon, 2010; Grayson et al., 2016; Hultman et al., 2016; Sigurdsson, Stark, Karayiorgou, Gogos, & Gordon, 2010; Uhlhaas & Singer, 2010).

Despite the increasing recognition that brain network abnormalities do not map cleanly to current clinical categories, existing studies taking a trans-diagnostic approach have been limited in several respects. First, most were restricted to one single dimension of psychopathology, missing the opportunity to

parse heterogeneity across the multiplicity of diagnoses (Satterthwaite et al., 2015). Second, dimensions of psychopathology derived from traditional factor analyses only examined the clinical symptomatology. While such approach, including our prior work (Calkins et al., 2015; A N Kaczkurkin et al., 2017; Antonia N. Kaczkurkin et al., 2016; Shanmugan et al., 2016), exploited a diverse range of psychiatric symptoms, the lack of guidance by brain features limited its impact to delineate the underlying neurobiology. Third, the vast majority of past research efforts have focused on adults, unable to answer the prevailing hypothesis of psychopathology as disorders of brain development (T. R. Insel, 2014). Finally, existing work that were able to study the critical window of brain development unfortunately suffered from small sample size, with dozens of participants. Modern multivariate analysis often requires much larger sample sizes to have the power to link high-dimensional brain patterns to complex behavioral and clinical measures (Bzdok & Yeo, 2017).

Multi-Scale Brain networks

Without a doubt, investigating how complex brain connectivity patterns are associated with neuropsychiatric illness has been an active area of research in the neuroscience community (Bassett & Sporns, 2017; Bzdok et al., 2016; Rubinov & Sporns, 2009). More broadly, the availability of large, open neuroimaging datasets as well as modern analytical tools and computational power have empowered scientists to uncover brain-phenotype relationships

across many domains, including development and aging, cognition, and neuropsychiatric illness (Biswal et al., 2010; Jernigan et al., 2016; Schumann et al., 2010; Van Essen et al., 2012).

However, most of these studies used general purpose statistical tools, without explicitly taking advantage of or taking into account of features that are unique to brain connectivity networks. This gap between the abundance of brain network data and shortage of appropriate analytical tools remains largely unfilled today. The ongoing quest to extract meaningful brain-phenotype relationships using connectomic data demands a network-specific approach (Craddock, Tungaraza, & Milham, 2015; Varoquaux & Craddock, 2013).

In modern network neuroscience, brain networks are represented by nodes, which denote the anatomical brain regions, and edges, which represent the connections between any pair of nodes (Rubinov & Sporns, 2009). As a stereotypical network can be made up of hundreds of nodes, and in turn, tens of thousands edges, one can investigate the properties of the network at different scales. At the *micro*-scale, one can investigate the individual edges (Craddock et al., 2015). At the *meso*-scale, assemble of edges form communities or modules, which are internal sparse and external dense structures that are thought to form the basis for specialized sub-units of information processing (Betzel, Medaglia, & Bassett, 2018). Finally, at the *macro*-level, networks can be studied using global summary statistics from classical graph theory measurement, including global

efficiency, characteristic path length, and clustering coefficient (Rubinov & Sporns, 2009).

Histological tracing and brain imaging studies have extensively documented these scales of network architecture in the nervous systems of humans and other species. This large body of work includes *C. elegans* (Sohn, Choi, Ahn, Lee, & Jeong, 2011)*, Drosophila* (Shih et al., 2015)*,* mouse (Wang, Sporns, & Burkhalter, 2012), rat (Bota, Sporns, & Swanson, 2015), cat (de Reus & van den Heuvel, 2013), and macaque (Modha & Singh, 2010). Additionally, prior work has also demonstrated that brain network architecture present on these different scales are associated with development, aging, learning, memory, cognition, neurological, and psychiatric illness (Bassett et al., 2018; Betzel et al., 2014; Braun et al., 2016; Bressler & Menon, 2010; Crossley et al., 2013; Fornito, Zalesky, & Breakspear, 2015; Grillon et al., 2013; Gu et al., 2015; Kernbach et al., 2018; Park & Friston, 2013; Power et al., 2010; Xia et al., 2018; Yu et al., 2019).

Single-Scale Approaches to Study Brain-Phenotype Relationships

Common strategies for studying brain connectivity and phenotype relationship consider brain network features one individual scale at a time, either with edge, community, or global statistics alone. For example, researchers have found that patients with schizophrenia had elevated global efficiency, a macro-

scale measure, in their functional brain networks compared to healthy controls (Lynall et al., 2010). While this approach has shown to be powerful in a great number of studies at demonstrating network abnormalities in neurological and psychiatric disorders, global network measures at the macro-scale inevitably fail to capture a large amount of information about complex brain systems at smaller scales.

Alternatively, on the micro-scale, there exist strategies that focus on group-level comparisons of individual edges. It takes in the form of mass univariate analysis, where a statistical test, such as a linear model, is applied to every edge (Craddock et al., 2015; Varoquaux & Craddock, 2013). While this procedure is methodologically easy to implement, a few drawbacks make it less practical. Chief among these caveats is the need to correct for a larger number of multiple comparisons, which ultimately dampens power for discovering potentially weak relationships between individual edge and phenotypes (Storey, 2002). In the process of reducing type I error, this approach can be very conservative and result in high type II error rates. To achieve balance between detection power and false discovery, alternative edge-based methods have been developed, such as the network based statistic (Zalesky, Fornito, & Bullmore, 2010) and multivariate distance matrix regression (Zapala & Schork, 2012). While these methods largely address the need for accounting for multiple comparison testing on each edge through family wise error rate correction procedures similar to those employed by conventional fMRI studies, they nonetheless focus

exclusively on the microscale of edges while ignoring innate data structures in the brain network, producing edge-level results that are difficult to interpret. With a select attention on each element of the adjacency matrix without appreciation of information present at a larger scale, edge-only approaches cannot see the forest for the trees.

Another equally problematic caveat of the edge-based approach is that it requires first vectorizing the connectivity matrix. This manipulation of the data structure transforms the original symmetric adjacency matrices into a wide feature table, where each column represents the edge strength across subjects. This unavoidably disrupts structures in the data, most notably block structures that represent meso-scale network features. To explicitly respect this community-level network information, one could calculate the *within-* and *between-*community connectivity as dependent variables in the linear models similar to the mass univariate analysis using edges (Betzel et al., 2014; Crossley et al., 2013; Gu et al., 2015). However, analogous to a high-order smoothing operation, extracting the mean connectivity of community pairs mixes disparate signals and also misses microscale information. While optimized for interpretability and low dimensionality in an attempt to improve signal to noise ratio, the community-based approach could be throwing the baby (signal) out with the bathwater (noise).

All told, single-scale approaches to study connectome-phenotype relationship, whether on a microscale (edge) or mesoscale (community), present with their own unique set of challenges of statistical power and interpretability. Thus, a regression method that integrates information across multiple scales with proper constraints could potentially achieve the best from both worlds. Indeed, recent theoretical and experimental studies have described many complex systems, including the financial system (Fenn et al., 2011), protein structure (Delmotte, Tate, Yaliraki, & Barahona, 2011), physical particles (Bassett, Owens, Porter, Manning, & Daniels, 2015), and the brain (Bassett & Siebenhühner, 2013; Betzel & Bassett, 2017) from a multi-scale perspective. However, this body of literature mostly concerns itself with network characterization and multi-scale community detection, rather than how to extract relationship between brain network and phenotypes in a multi-scale manner.

In this dissertation, the overall arching goal is to study complex connectivity patterns in functional brain networks that are linked to a diverse range of measurement, in particular, psychopathology. In both studies that follow, we applied advanced machine learning techniques to delineate multivariate patterns of functional connectivity.

In Chapter 2, we set out to map out linked dimensions of psychopathology that are highly associated with patterns of functional connectivity. Specifically, to delineate brain-guided dimensions that cut across existing diagnostic categories,

we applied sparse canonical correlational analysis (sCCA) (Witten, Tibshirani, & Hastie, 2009) to the Philadelphia Neurodevelopmental Cohort (PNC) (Satterthwaite et al., 2014), a large cohort of youth with multimodal imaging and item-wise psychiatric symptoms. We discovered four linked dimensions of psychopathology and brain connectivity patterns, namely mood, psychosis, fear, and externalizing behavior. These brain-guided psychopathological dimensions cross traditional categorical boundaries while concurring with clinical experience. Each dimension exhibited unique brain connectivity patterns; however, across all psychopathology, loss of normative segregation between the default mode and executive networks emerged as a common feature of connectivity dysfunction.

In Chapter 3, we introduce a new regression method specifically designed to analyze the associations between high-dimensional connectomic data and phenotypes of interest, which we refer to as Multi-Scale Network Regression (MSNR). Specifically, we designed a penalized multivariate approach to explicitly exploit both edge and community level information to extract brain-phenotype relationships. By constraining a low rank and sparse structure on edges and community level information, respectively, MSNR encourages less complex and more interpretable modeling while achieves higher out-of-sample prediction performance and statistical significance via permutation tests. We applied MSNR to PNC and found that MSNR recapitulated known functional brain connectivity patterns in association with age, sex, and in-scanner motion. In a head-to-head comparison with common single-scale approaches that consider either edges or

community connectivity alone, MSNR achieved a balance between out-of-sample prediction and model complexity, with improved interpretability.

References

Adhikari, A., Topiwala, M. A., & Gordon, J. A. (2010). Synchronized Activity between the Ventral Hippocampus and the Medial Prefrontal Cortex during Anxiety. *Neuron*, *65*(2), 257–269. https://doi.org/10.1016/j.neuron.2009.12.002

Bassett, D. S., Owens, E. T., Porter, M. A., Manning, M. L., & Daniels, K. E. (2015). Extraction of force-chain network architecture in granular materials using community detection. *Soft Matter*, *11*(14), 2731–2744. https://doi.org/10.1039/C4SM01821D

Bassett, D. S., & Siebenhühner, F. (2013). Multiscale Network Organization in the Human Brain. In *Multiscale Analysis and Nonlinear Dynamics* (pp. 179–204). Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. https://doi.org/10.1002/9783527671632.ch07

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, *20*(3), 353–364. https://doi.org/10.1038/nn.4502.Network

Bassett, D. S., Xia, C. H., & Satterthwaite, T. D. (2018). Understanding the Emergence of Neuropsychiatric Disorders With Network Neuroscience. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(9), 742–753. https://doi.org/10.1016/J.BPSC.2018.03.015

Baum, G. L., Ciric, R., Roalf, D. R., Betzel, R. F., Moore, T. M., Shinohara, R. T., … Satterthwaite, T. D. (2016). Modular Segregation of Structural Brain Networks Supports the Development of Executive Function in Youth.

Berman, M. G., Peltier, S., Nee, D. E., Kross, E., Deldin, P. J., & Jonides, J. (2011). Depression, rumination and the default network. *Social Cognitive and Affective Neuroscience*, *6*(5), 548–555. https://doi.org/10.1093/scan/nsq080

Betzel, R. F., & Bassett, D. S. (2017). Multi-scale brain networks. *NeuroImage*, *160*, 73–83. https://doi.org/10.1016/J.NEUROIMAGE.2016.11.006

Betzel, R. F., Byrge, L., He, Y., Goñi, J., Zuo, X.-N., & Sporns, O. (2014).

Changes in structural and functional connectivity among resting-state networks across the human lifespan. *NeuroImage*, *102*, 345–357. https://doi.org/10.1016/J.NEUROIMAGE.2014.07.067

Betzel, R. F., Medaglia, J. D., & Bassett, D. S. (2018). Diversity of meso-scale architecture in human and non-human connectomes. *Nature Communications*, *9*(1), 346. https://doi.org/10.1038/s41467-017-02681-z

Biswal, B. B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C., Smith, S. M., … Milham, M. P. (2010). Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(10), 4734–4739. https://doi.org/10.1073/pnas.0911855107

Bota, M., Sporns, O., & Swanson, L. W. (2015). Architecture of the cerebral cortical association connectome underlying cognition. *Proceedings of the National Academy of Sciences*, *112*(16), E2093–E2101. https://doi.org/10.1073/pnas.1504394112

Braun, U., Schäfer, A., Bassett, D. S., Rausch, F., Schweiger, J. I., Bilek, E., … Tost, H. (2016). Dynamic brain network reconfiguration as a potential schizophrenia genetic risk mechanism modulated by NMDA receptor function. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(44), 12568–12573. https://doi.org/10.1073/pnas.1608819113

Bressler, S. L., & Menon, V. (2010). Large-scale brain networks in cognition: emerging methods and principles. *Trends in Cognitive Sciences*, *14*(6), 277–290. https://doi.org/10.1016/J.TICS.2010.04.004

Bzdok, D., Varoquaux, G., Grisel, O., Eickenberg, M., Poupon, C., & Thirion, B. (2016). Formal Models of the Network Co-occurrence Underlying Mental Operations. *PLOS Computational Biology*, *12*(6), e1004994. https://doi.org/10.1371/journal.pcbi.1004994

Bzdok, D., & Yeo, B. T. T. (2017). Inference in the age of big data: Future perspectives on neuroscience. *NeuroImage*, *155*, 549–564. https://doi.org/10.1016/j.neuroimage.2017.04.061

Calkins, M. E., Merikangas, K. R., Moore, T. M., Burstein, M., Behr, M. A., Satterthwaite, T. D., … Gur, R. E. (2015). The Philadelphia Neurodevelopmental

Cohort : constructing a deep phenotyping collaborative. *Journal of Child Psychology and Psychiatry*, *12*, 1356–1369. https://doi.org/10.1111/jcpp.12416

Cao, M., Huang, H., Peng, Y., Dong, Q., & He, Y. (2016). Toward Developmental Connectomics of the Human Brain. *Frontiers in Neuroanatomy*, *10*(March), 25. https://doi.org/10.3389/fnana.2016.00025

Casey, B. J., Oliveri, M. E., & Insel, T. (2014). A neurodevelopmental perspective on the research domain criteria (RDoC) framework. *Biological Psychiatry*, *76*(5), 350–353. https://doi.org/10.1016/j.biopsych.2014.01.006

Castellanos, F. X., Margulies, D. S., Kelly, C., Uddin, L. Q., Ghaffari, M., Kirsch, A., … Milham, M. P. (2008). Cingulate-Precuneus Interactions: A New Locus of Dysfunction in Adult Attention-Deficit/Hyperactivity Disorder. *Biological Psychiatry*, *63*(3), 332–337. https://doi.org/10.1016/j.biopsych.2007.06.025

Clementz, B. A., Sweeney, J. A., Hamm, J. P., Ivleva, E. I., Ethridge, L. E., Pearlson, G. D., … Tamminga, C. A. (2016). Identification of distinct psychosis biotypes using brain-based biomarkers. *American Journal of Psychiatry*, *173*(4), 373–384. https://doi.org/10.1176/appi.ajp.2015.14091200

Craddock, R. C., Tungaraza, R. L., & Milham, M. P. (2015). Connectomics and new approaches for analyzing human brain functional connectivity. *GigaScience*, *4*(1), 13. https://doi.org/10.1186/s13742-015-0045-x

Crossley, N. A., Mechelli, A., Vertes, P. E., Winton-Brown, T. T., Patel, A. X., Ginestet, C. E., … Bullmore, E. T. (2013). Cognitive relevance of the community structure of the human brain functional coactivation network. *Proceedings of the National Academy of Sciences*, *110*(28), 11583–11588. https://doi.org/10.1073/pnas.1220826110

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

de Reus, M. A., & van den Heuvel, M. P. (2013). Rich Club Organization and Intermodule Communication in the Cat Connectome. *Journal of Neuroscience*, *33*(32), 12929–12939. https://doi.org/10.1523/JNEUROSCI.1448-13.2013

Delmotte, A., Tate, E. W., Yaliraki, S. N., & Barahona, M. (2011). Protein multi-scale organization through graph partitioning and robustness analysis: application to the myosin–myosin light chain interaction. *Physical Biology*, *8*(5), 055010. https://doi.org/10.1088/1478-3975/8/5/055010

Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., … Liston, C. (2016). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine*, *23*(1), 28–38. https://doi.org/10.1038/nm.4246

Fenn, D. J., Porter, M. A., Williams, S., McDonald, M., Johnson, N. F., & Jones, N. S. (2011). Temporal evolution of financial-market correlations. *Physical Review E*, *84*(2), 026109. https://doi.org/10.1103/PhysRevE.84.026109

Fornito, A., Zalesky, A., & Breakspear, M. (2015). The connectomics of brain disorders. *Nature Reviews Neuroscience*, *16*(3), 159–172. https://doi.org/10.1038/nrn3901

Giedd, J. N., & Rapoport, J. L. (2010). Structural MRI of Pediatric Brain Development: What Have We Learned and Where Are We Going? *Neuron*, *67*(5), 728–734. https://doi.org/10.1016/j.neuron.2010.08.040

Goodkind, M., Eickhoff, S. B., Oathes, D. J., Jiang, Y., Chang, A., Jones-hagata, L. B., … Etkin, A. (2015). Identification of a common neurobiological substrate for mental illness. *JAMA Psychiatry*, *5797*(4), 305–315. https://doi.org/10.1001/jamapsychiatry.2014.2206

Grayson, D. S., Bliss-Moreau, E., Machado, C. J., Bennett, J., Shen, K., Grant, K. A., … Amaral, D. G. (2016). The Rhesus Monkey Connectome Predicts Disrupted Functional Networks Resulting from Pharmacogenetic Inactivation of the Amygdala. *Neuron*, *91*(2), 453–466. https://doi.org/10.1016/j.neuron.2016.06.005

Greicius, M. D., Supekar, K., Menon, V., & Dougherty, R. F. (2009). Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cerebral Cortex*, *19*(1), 72–78. https://doi.org/10.1093/cercor/bhn059

Grillon, M.-L., Oppenheim, C., Varoquaux, G., Charbonneau, F.,

Devauchelle, A.-D., Krebs, M.-O., … Huron, C. (2013). Hyperfrontality and hypoconnectivity during refreshing in schizophrenia. *Psychiatry Research: Neuroimaging*, *211*(3), 226–233. https://doi.org/10.1016/J.PSCYCHRESNS.2012.09.001

Gu, S., Satterthwaite, T. D., Medaglia, J. D., Yang, M., Gur, R. E. R. C. R. E., Gur, R. E. R. C. R. E., & Bassett, D. S. (2015). Emergence of system roles in normative neurodevelopment. *Proceedings of the National Academy of Sciences*, *112*(44), 201502829. https://doi.org/10.1073/pnas.1502829112

Hodgkinson, S., Sherrington, R., Gurling, H., Marchbanks, R., Reeders, S., Mallet, J., … Brynjolfsson, J. (1987). Molecular genetic evidence for heterogeneity in manic depression. *Nature*, *325*(6107), 805–806. https://doi.org/10.1038/325805a0

Hultman, R., Mague, S. D., Li, Q., Katz, B. M., Michel, N., Lin, L., … Hastie, T. (2016). Dysregulation of Prefrontal Cortex-Mediated Slow-Evolving Limbic Dynamics Drives Stress-Induced Emotional Pathology. *Neuron*, *0*(0), 257–269. https://doi.org/10.1016/j.neuron.2016.05.038

Insel, B. T. R., & Cuthbert, B. N. (2015). Brain diorders? Precisely. *Science*, *348*(6234), 499–500.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., … Wang, P. (2010). Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, *167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Insel, T. R. (2014). Mental disorders in childhood: shifting the focus from behavioral symptoms to neurodevelopmental trajectories. *JAMA : The Journal of the American Medical Association*, *311*(17), 1727–1728. https://doi.org/10.1001/jama.2014.1193

Jacobi, F., Wittchen, H.-U., Hölting, C., Höfler, M., Pfister, H., Müller, N., & Lieb, R. (2004). 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*(4), 597–611. https://doi.org/10.1017/S0033291703001399

Jernigan, T. L., Brown, T. T., Hagler, D. J., Akshoomoff, N., Bartsch, H., Newman, E., … Pediatric Imaging, Neurocognition and Genetics Study. (2016). The Pediatric Imaging, Neurocognition, and Genetics (PING) Data Repository. *NeuroImage*, *124*(Pt B), 1149–1154. https://doi.org/10.1016/j.neuroimage.2015.04.057

Kaczkurkin, A N, Moore, T. M., Calkins, M. E., Ciric, R., Detre, J. A., Elliott, M. A., … Satterthwaite, T. D. (2017). Common and dissociable regional cerebral blood flow differences associate with dimensions of psychopathology across categorical diagnoses. *Molecular Psychiatry*. https://doi.org/10.1038/mp.2017.174

Kaczkurkin, Antonia N., Moore, T. M., Ruparel, K., Ciric, R., Calkins, M. E., Shinohara, R. T., … Satterthwaite, T. D. (2016). Elevated Amygdala Perfusion Mediates Developmental Sex Differences in Trait Anxiety. *Biological Psychiatry*, *80*(10), 775–785. https://doi.org/10.1016/j.biopsych.2016.04.021

Kernbach, J. M., Satterthwaite, T. D., Bassett, D. S., Smallwood, J., Margulies, D., Krall, S., … Bzdok, D. (2018). Shared endo-phenotypes of default mode dysfunction in attention deficit/hyperactivity disorder and autism spectrum disorder. *Translational Psychiatry*, *8*(1), 133. https://doi.org/10.1038/s41398-018-0179-6

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

Lynall, M.-E., Bassett, D. S., Kerwin, R., McKenna, P. J., Kitzbichler, M., Muller, U., & Bullmore, E. (2010). Functional connectivity and brain networks in schizophrenia. *The Journal of Neuroscience*, *30*(28), 9477–9487. https://doi.org/10.1523/jneurosci.0333-10.2010

McTeague, L. M., Huemer, J., Carreon, D. M., Jiang, Y., Eickhoff, S. B., & Etkin, A. (2017). Identification of Common Neural Circuit Disruptions in Cognitive Control Across Psychiatric Disorders. *American Journal of Psychiatry*, *174*(7), 676–685. https://doi.org/10.1176/appi.ajp.2017.16040400

Modha, D. S., & Singh, R. (2010). Network architecture of the long-

distance pathways in the macaque brain. *Proceedings of the National Academy of Sciences*, *107*(30), 13485–13490. https://doi.org/10.1073/pnas.1008054107

Park, H. J., & Friston, K. (2013). Structural and functional brain networks: from connections to cognition. *Science*, *342*(November), 579–588. https://doi.org/10.1126/science.1238411

Paus, Tomáš. (2005). Mapping brain maturation and cognitive development during adolescence. *Trends in Cognitive Sciences*, *9*(2), 60–68. https://doi.org/10.1016/j.tics.2004.12.008

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

Power, J. D., Fair, D. A., Schlaggar, B. L., & Petersen, S. E. (2010). The Development of Human Functional Brain Networks. *Neuron*, *67*(5), 735–748. https://doi.org/10.1016/J.NEURON.2010.08.017

Rapoport, J. L., Giedd, J. N., & Gogtay, N. (2012). Neurodevelopmental model of schizophrenia: Update 2012. *Molecular Psychiatry*, *17*(12), 1228–1238. https://doi.org/10.1038/mp.2012.23

Rubinov, M., & Sporns, O. (2009). Complex network measures of brain connectivity : Uses and interpretations. *NeuroImage*, *52*(3), 1059–1069. https://doi.org/10.1016/j.neuroimage.2009.10.003

Satterthwaite, T. D., Elliott, M. A., Ruparel, K., Loughead, J., Prabhakaran, K., Calkins, M. E., … Gur, R. E. (2014). Neuroimaging of the Philadelphia Neurodevelopmental Cohort. *NeuroImage*, *86*(2014), 544–553. https://doi.org/10.1016/j.neuroimage.2013.07.064

Satterthwaite, T. D., Vandekar, S. N., Wolf, D. H., Bassett, D. S., Ruparel, K., Shehzad, Z., … Gur, R. E. (2015). Connectome-wide network analysis of youth with Psychosis-Spectrum symptoms. *Molecular Psychiatry*, *20*(February), 1–8. https://doi.org/10.1038/mp.2015.66

Satterthwaite, T. D., Wolf, D. H., Ruparel, K., Erus, G., Elliott, M. A.,

Eickhoff, S. B., … Gur, R. C. (2013). Heterogeneous impact of motion on fundamental patterns of developmental changes in functional connectivity during youth. *NeuroImage*, *83*(2013), 45–57. https://doi.org/10.1016/j.neuroimage.2013.06.045

Schumann, G., Loth, E., Banaschewski, T., Barbot, A., Barker, G., Büchel, C., … IMAGEN consortium. (2010). The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Molecular Psychiatry*, *15*(12), 1128–1139. https://doi.org/10.1038/mp.2010.4

Shanmugan, S., Wolf, D. H., Calkins, M. E., Moore, T. M., Ruparel, K., Hopson, R. D., … Satterthwaite, T. D. (2016). Common and Dissociable Mechanisms of Executive System Dysfunction Across Psychiatric Disorders in Youth. *The American Journal of Psychiatry*, *173*(5), 517–526. https://doi.org/10.1176/appi.ajp.2015.15060725

Shih, C.-T., Sporns, O., Yuan, S.-L., Su, T.-S., Lin, Y.-J., Chuang, C.-C., … Chiang, A.-S. (2015). Connectomics-Based Analysis of Information Flow in the Drosophila Brain. *Current Biology*, *25*(10), 1249–1258. https://doi.org/10.1016/j.cub.2015.03.021

Sigurdsson, T., Stark, K. L., Karayiorgou, M., Gogos, J. A., & Gordon, J. A. (2010). Impaired hippocampal–prefrontal synchrony in a genetic mouse model of schizophrenia. *Nature*, *464*(7289), 763–767. https://doi.org/10.1038/nature08855

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

Skudlarski, P., Jagannathan, K., Anderson, K., Stevens, M. C., Calhoun, V. D., Skudlarska, B. A., & Pearlson, G. (2010). Brain Connectivity Is Not Only Lower but Different in Schizophrenia: A Combined Anatomical and Functional Approach. *Biological Psychiatry*, *68*(1), 61–69. https://doi.org/10.1016/j.biopsych.2010.03.035

Sohn, Y., Choi, M.-K., Ahn, Y.-Y., Lee, J., & Jeong, J. (2011). Topological Cluster Analysis Reveals the Systemic Organization of the Caenorhabditis elegans Connectome. *PLoS Computational Biology*, *7*(5), e1001139. https://doi.org/10.1371/journal.pcbi.1001139

Sporns, O., & Betzel, R. F. (2016). Modular Brain Networks. *Annual Review of Psychology*, *67*(1), 613–640. https://doi.org/10.1146/annurev-psych-122414-033634

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(3), 479–498. https://doi.org/10.1111/1467-9868.00346

Uddin, L. Q., Supekar, K., Amin, H., Rykhlevskaia, E., Nguyen, D. A., Greicius, M. D., & Menon, V. (2010). Dissociable connectivity within human angular gyrus and intraparietal sulcus: Evidence from functional and structural connectivity. *Cerebral Cortex*, *20*(11), 2636–2646. https://doi.org/10.1093/cercor/bhq011

Uhlhaas, P. J., & Singer, W. (2010). Abnormal neural oscillations and synchrony in schizophrenia. *Nature Reviews. Neuroscience*, *11*(2), 100–113. https://doi.org/10.1038/nrn2774

Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E. J., Bucholz, R., … Yacoub, E. (2012). The Human Connectome Project: A data acquisition perspective. *NeuroImage*, *62*(4), 2222–2231. https://doi.org/10.1016/J.NEUROIMAGE.2012.02.018

Varoquaux, G., & Craddock, R. C. (2013). Learning and comparing functional connectomes across subjects. *NeuroImage*, *80*, 405–415. https://doi.org/10.1016/J.NEUROIMAGE.2013.04.007

von Rhein, D., Oldehinkel, M., Beckmann, C. F., Oosterlaan, J., Heslenfeld, D., Hartman, C. A., … Mennes, M. (2016). Aberrant local striatal functional connectivity in attention-deficit/hyperactivity disorder. *Journal of Child Psychology and Psychiatry*, *57*(6), 697–705. https://doi.org/10.1111/jcpp.12529

Wang, Q., Sporns, O., & Burkhalter, A. (2012). Network Analysis of Corticocortical Connections Reveals Ventral and Dorsal Processing Streams in Mouse Visual Cortex. *Journal of Neuroscience*, *32*(13), 4386–4399. https://doi.org/10.1523/JNEUROSCI.6063-11.2012

Whiteford, H. A., Degenhardt, L., Rehm, J., Baxter, A. J., Ferrari, A. J., Erskine, H. E., … Vos, T. (2013). Global burden of disease attributable to mental

and substance use disorders: Findings from the Global Burden of Disease Study 2010. *The Lancet*, *382*(9904), 1575–1586. https://doi.org/10.1016/S0140-6736(13)61611-6

Whitfield-Gabrieli, S., Thermenos, H. W., Milanovic, S., Tsuang, M. T., Faraone, S. V, McCarley, R. W., … Seidman, L. J. (2009). Hyperactivity and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(4), 1279–1284. https://doi.org/10.1073/pnas.0809141106

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

Xia, C. H., Ma, Z., Ciric, R., Gu, S., Betzel, R. F., Kaczkurkin, A. N., … Satterthwaite, T. D. (2018). Linked dimensions of psychopathology and connectivity in functional brain networks. *Nature Communications*, *9*(1), 3003. https://doi.org/10.1038/s41467-018-05317-y

Yu, M., Linn, K. A., Shinohara, R. T., Oathes, D. J., Cook, P. A., Duprat, R., … Sheline, Y. I. (2019). Childhood trauma history is linked to abnormal brain connectivity in major depression. https://doi.org/10.1073/pnas.1900801116

Zalesky, A., Fornito, A., & Bullmore, E. T. (2010). Network-based statistic: Identifying differences in brain networks. *NeuroImage*, *53*(4), 1197–1207. https://doi.org/10.1016/J.NEUROIMAGE.2010.06.041

Zapala, M. A., & Schork, N. J. (2012). Statistical Properties of Multivariate Distance Matrix Regression for High-Dimensional Data Analysis. *Frontiers in Genetics*, *3*, 190. https://doi.org/10.3389/fgene.2012.00190

CHAPTER 2

Linked Dimensions Of Psychopathology And Connectivity In Functional Brain Networks

This chapter has been published:

Xia C.H., Ma Z., Ciric R., Gu S., Betzel R., Kaczkurkin A.N., Calkins M.E., Cook P.A., García de la Garza A., Vandekar S.N., Cui Z., Moore T.M., Roalf D.R., Ruparel K., Wolf D.H., Davatzikos C., Gur R.C., Gur R.E., Shinohara R.T., Bassett D.S., Satterthwaite T.D.. 2018 “Linked Dimensions Of Psychopathology And Connectivity In Functional Brain Networks”. *Nature Communications*, *9: 3003*

Abstract

Neurobiological abnormalities associated with psychiatric disorders do not map well to existing diagnostic categories. High co-morbidity suggests dimensional circuit-level abnormalities that cross diagnoses. Here we seek to identify brain-based dimensions of psychopathology using sparse canonical correlation analysis in a sample of 663 youths. This analysis reveals correlated patterns of functional connectivity and psychiatric symptoms. We find that four dimensions of psychopathology – mood, psychosis, fear, and externalizing behavior – are associated ($r = 0.68–0.71$) with distinct patterns of connectivity. Loss of network segregation between the default mode network and executive networks emerges as a common feature across all dimensions. Connectivity linked to mood and psychosis becomes more prominent with development, and sex differences are present for connectivity related to mood and fear. Critically, findings largely replicate in an independent dataset ($n = 336$). These results delineate connectivity-guided dimensions of psychopathology that cross clinical 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, without the use of established biomarkers to aid in diagnosis, prognosis, and treatment selection. It is increasingly recognized that existing clinical diagnostic categories could hinder the search for biomarkers in psychiatry (Singh & Rose, 2009), as they are not clearly associated with distinct neurobiological abnormalities (B. T. R. Insel & Cuthbert, 2015). The high co-morbidity among psychiatric disorders exacerbates this problem (Jacobi et al., 2004). Furthermore, studies have demonstrated common structural, functional, and genetic abnormalities across psychiatric syndromes, potentially explaining such co-morbidity (Goodkind et al., 2015; Lee, Ripke, Neale, Faraone, Purcell, Perlis, Mowry, Wray, et al., 2013; McTeague et al., 2017). This body of evidence underscores the lack of direct mapping between clinical diagnostic categories and the underlying pathophysiology.

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 psychiatric diseases (Cuthbert & Insel, 2013). 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.

Network neuroscience is a powerful approach for examining brain systems implicated in psychopathology (Bassett & Sporns, 2017; Bullmore & Sporns, 2009). 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 (Gordon et al., 2016; Power et al., 2011; Yeo et al., 2011). Furthermore, multiple studies documented abnormalities within this modular topology in psychiatric disorders (Bassett & Bullmore, 2009; Lynall et al., 2010). 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 (Bassett, Xia, & Satterthwaite, 2018; Satterthwaite, Vandekar, et al., 2015).

In addition to such module-specific deficits, studies in mood disorders (Kaiser, Andrews-Hanna, Wager, & Pizzagalli, 2015; Li et al., 2014), psychosis (Alexander-Bloch et al., 2012; Lynall et al., 2010), and other disorders (Fornito, Zalesky, & Breakspear, 2015; Rudie et al., 2013) 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 (Power, Fair, Schlaggar, & Petersen, 2010) and structural (Baum et al., 2017) brain networks is refined during adolescence, a critical period when many psychiatric disorders emerge. Such findings have led many disorders to be considered “neurodevelopmental connectopathies.” (Paus, 2005) Describing the developmental substrates of psychiatric 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 (T. R. Insel, 2014).

Despite the increasing interest in describing how abnormalities of brain network development lead to the emergence of psychiatric disorders, existing studies have been limited in several respects. First, most have adopted a categorical case-control approach, or only examined a single dimension of psychopathology (Satterthwaite, Vandekar, et al., 2015), and are therefore unable to capture heterogeneity across diagnoses. Second, dimensional psychopathology derived from factor analyses, including our prior work (Calkins et al., 2015; A N Kaczkurkin et al., 2017; Antonia N. Kaczkurkin et al., 2016; Shanmugan et al., 2016), were solely driven by covariance in the clinical symptomatology, rather than being guided by both brain and behavior features. Third, especially in contrast to adult studies, existing work in youth has often used relatively small samples (e.g., dozens of participants). While multivariate

techniques allow the examination of both multiple brain systems and clinical dimensions simultaneously, such techniques usually require large samples (Bzdok & Yeo, 2017).

In the current study, we seek to delineate functional network abnormalities associated with a broad array of psychopathology in youth. We have capitalized on a large sample of youth from the Philadelphia Neurodevelopmental Cohort (PNC) (Satterthwaite et al., 2014) by applying a recently-developed machine learning technique called sparse canonical correlation analysis (sCCA) (Witten, Tibshirani, & Hastie, 2009). As a multivariate method, sCCA is capable of discovering complex linear relationships between two high-dimensional datasets (Avants et al., 2014; Smith et al., 2015). It should be noted that the approach of the current study is distinct from prior work discovering biotypes within categories of psychopathology, based purely on imaging features themselves (e.g., functional connectivity (Drysdale et al., 2016) and gray matter density (Clementz et al., 2016)). In contrast, we seek to link a broad range of symptoms that are present across categories to individual differences in functional brain networks. Such an approach has been successfully applied in prior work on neurodegenerative diseases (Avants et al., 2014) as well as normal brain-behavior relationships (Smith et al., 2015).

Here, we use sCCA to delineate linked dimensions of psychopathology and functional connectivity. As described below, we uncover dimensions of connectivity that are highly correlated with specific, interpretable dimensions of psychopathology. We find that each psychopathological dimension is associated with a distinct pattern of abnormal connectivity, and that all dimensions are characterized by decreased segregation of default mode and executive networks (fronto-parietal and salience). These network features linked to each dimension of psychopathology show expected developmental changes and sex differences. Finally, our results are largely replicated in an independent dataset.

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 (Satterthwaite et al., 2014). In total, 1601 participants completed the cross-sectional neuroimaging protocol (**Table 2-1, Figure 2-1**). One subject had missing clinical data. To create two independent samples for discovery and replication analyses, we performed a random split of the remaining 1600 participants using the CARET package 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 (**Supplementary Figure 2-1**). The two samples were confirmed to also have similar distributions in regards to age, sex, and race (**Figure 2-1**), as well as motion (**Supplementary Figure 2-2**). Overall psychopathology is the general factor score reported previously from factor analysis of the clinical data alone (Calkins et al., 2015; Shanmugan et al., 2016).

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 image coverage, or excessive motion during the functional scan, which is defined as having an average framewise motion more than 0.20 mm or more than 20 frames exhibiting over 0.25 mm movement (using the Jenkinson calculation (Jenkinson, Bannister, Brady, & Smith, 2002)). 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 **Table 2-1** and **Figure 2-1** 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 (Calkins et al., 2015). 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 (Merikangas et al., 2010). 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), eating disorders (anorexia, bulimia), and suicidal thinking and behavior (**Supplementary Table 2-1**). The 111 item-level symptoms used in this study were described in prior factor analysis of the clinical data in PNC (Shanmugan et al., 2016). For the specific items, see **Supplementary Data 2-1**.

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 (Satterthwaite et al., 2014). 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 = 1810ms; T_E = 3.51 ms; FoV = 180 × 240 mm; resolution 0.9375 × 0.9375 × 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 × 192 mm; resolution 3 mm 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 (Avants, Tustison, Song, et al., 2011). Study-specific tissue priors were created using a multi-atlas segmentation procedure (Wang et al., 2014). 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 (Tustison et al., 2014). Following bias field correction (Tustison et al.,

2010), each structural image was diffeomorphically registered to the study-specific PNC template using the top-performing SYN deformation provided by ANTS (Klein et al., 2009). Study-specific tissue priors were used to guide brain extraction and segmentation of the subject's structural image (Avants, Tustison, Wu, Cook, & Gee, 2011).

FUNCTIONAL PREPROCESSING

Task-free functional images were processed using one of the top-performing pipelines for removal of motion-related artifact (Ciric et al., 2017). 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` (Jenkinson et al., 2002), (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 (Greve & Fischl, 2009). The artefactual variance in the data was modelled using a total of 36 parameters, including the six 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 nine parameters, and quadratic terms of each of the nine parameters and their

derivatives. Importantly, our findings are robust to the methodological choice of regressing out global signal (**Supplementary Figure 2-3** and **Supplementary Figure 2-4**). 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 (Hallquist, Hwang, & Luna, 2013).

NETWORK CONSTRUCTION

We built a functional connectivity network using the residual timeseries (following de-noising) of all parcels of a common parcellation (Power et al., 2011). The parcellation used in the main analysis consists of 264 spherical nodes of 20 mm diameter distributed across the brain (Power et al., 2011). The a priori communities for this set of nodes were originally delineated using the Infomap algorithm (Rosvall & Bergstrom, 2008) and were replicated in an independent sample. This parcellation was particularly suitable for our analysis as it has been previously used for studying developmental changes in connectivity and network modularity (Power et al., 2010) and has been used as part of several studies in this dataset in the past (Chai et al., 2017; Ciric et al., 2017; Gu et al., 2015). As part of the supplementary analysis to demonstrate the robustness of the results independent of parcellation choices (**Supplementary Figure 2-8**), we also constructed networks based on an alternative parcellation developed by Gordon et al. (2016). This set of nodes was derived using edge detection and boundary mapping to define areal

parcels. The functional connectivity between any pair of brain regions was operationalized as the Pearson correlation coefficient between the mean activation timeseries extracted from those regions. 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.

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 2-2**).

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 Figure 2-5**). MAD is defined as $median(|\boldsymbol{X}_i - median(\boldsymbol{X})|)$, 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 Figure 2-5c**).

An alternative approach for dimensionality reduction is 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 Figure 2-8**, 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.

SPARSE CANONICAL CORRELATION ANALYSIS

sCCA is a multivariate procedure that seeks maximal correlations between linear combinations of variables in both sets, with regularization to achieve sparsity (Witten et al., 2009). In essence, given two matrices, $\boldsymbol{X}_{n\times p}$ and $\boldsymbol{Y}_{n\times q}$, 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 $\boldsymbol{u}$ and $\boldsymbol{v}$, which are loading vectors, that maximize $cor(\boldsymbol{Xu}, \boldsymbol{Yv})$. Mathematically, this optimization problem can be expressed as

$$\begin{aligned} &\text{maximize } \boldsymbol{u}^T \boldsymbol{X}^T \boldsymbol{Y} \boldsymbol{v}, \\ &\text{subject to } \|\boldsymbol{u}\|_2^2 \leq 1, \|\boldsymbol{v}\|_2^2 \leq 1, \|\boldsymbol{u}\|_1 \leq c_1, \|\boldsymbol{v}\|_1 \leq c_2. \end{aligned} \tag{1}$$

Since both $l_1(\|\cdot\|_1)$ and $l_2(\|\cdot\|_2)$-norm 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 $\boldsymbol{u}$ and $\boldsymbol{v}$ 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).

We chose a linear kernel over non-linear implementations of sCCA for two reasons. First, while a more complex model may potentially better fit the data, increased model complexity often results in reduced interpretability. Secondly, a non-linear model may require a larger sample size to accurately estimate the increased number of parameters.

GRID SEARCH FOR REGULARIZATION PARAMETERS

We tuned the l_1 regularization parameters for the connectivity and the clinical features, respectively (see **Supplementary Figure 2-6**). The range of sparsity parameters are constrained to be between 0 and 1 in the PMA

package (Witten et al., 2009), 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 increments 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 Figure 2-7**). 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 (*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 ($\boldsymbol{v}$) (Mišić et al., 2016). 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:

$$p_{permutation} = \frac{\sum_{i=1}^{B} \begin{cases} 1, if\ r_i \geq \bar{r} \\ 0, if\ r_i < \bar{r} \end{cases}}{B}. \quad (2)$$

In other words, we randomly assigned subjects' clinical features to other subjects' connectivity features, therefore breaking up the internal co-varying structures of the original dataset. The canonical variates resulting from these re-aligned datasets with preserved data distribution will then serve as the null distribution against which the real correlations are compared. The logic is that any significant co-varying relationships will have to be greater than the signals in a permuted data structure.

RESAMPLING PROCEDURE

To further select features that consistently contributed to each canonical variate, we performed a resampling procedure (**Supplementary Figure 2-9**). 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 (Mišić et al., 2016). 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 parcellation (**Supplementary Figure 2-10a**) (Power et al., 2011). Specifically, within-module connectivity loading is defined as

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

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

$$\frac{\sum_{i\in m, j\in n} W_{ij}}{|\boldsymbol{M}| \times (|\boldsymbol{N}|)} \quad , \qquad (4)$$

where W_{ij} is the sCCA loading of the functional connectivity between nodes i and j, which belong to community m in $\boldsymbol{M}$ and community n in $\boldsymbol{N}$, 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 Figure 2-10b**). Empirical *p*-values were calculated similar to **Eq. (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 (Rubinov & Sporns, 2009) and visualized it with BrainNet Viewer (Xia, Wang, & He, 2013) both in MATLAB.

ANALYSIS OF AGE EFFECTS AND SEX DIFFERENCES

As previously (Baum et al., 2017; Nassar et al., 2018; Shanmugan et al., 2016), generalized additive models (GAMs), using the MGCV package 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:

$$Connectivity\ Score_i \sim sex + s(age). \quad (5)$$

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

DATA AVAILABILITY

The data reported in this paper have been deposited in database of Genotypes and Phenotypes (dbGaP): accession no. [phs000607.v3.p2]. (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000607.v3.p2)

CODE AVAILABILITY

All analysis code is available here: https://github.com/cedricx/sCCA/tree/master/sCCA/code/final.

Results

LINKED DIMENSIONS OF PSYCHOPATHOLOGY AND CONNECTIVITY

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 (Witten et al., 2009). In total, we studied 999 participants of ages 8–22 who completed both functional neuroimaging and a comprehensive evaluation of psychiatric symptoms as part of the PNC (Calkins et al., 2015; Satterthwaite et al., 2014) (**Table 2-1 and Figure 2-1**). Participants in the PNC were recruited from Children's Hospital of Philadelphia pediatric network in the greater Philadelphia area. In this community-based study, participants were not recruited from psychiatric services. As such, the prevalence of screening into specific psychopathology categories generally aligned with epidemiologically ascertained samples, as previously described (Calkins et al., 2015) (see **Supplementary Table 2-1**). We divided this sample into discovery ($n = 663$) and replication datasets ($n = 336$) that were matched on age, sex, race, and overall psychopathology (**Figure 2-1 and Supplementary Figure 2-1**). Following pre-processing using a validated pipeline that minimizes the impact of in-scanner motion (Ciric et al., 2017) (see **Supplementary Figure 2-2, Supplementary Figure 2-3,** and

Supplementary Figure 2-4), we constructed subject-level functional networks using a 264-node parcellation system that includes an a priori assignment of nodes to network communities (Power et al., 2011) (**Figure 2-2a–c**, e.g., modules or sub-networks; see Methods section). 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 ten percent most variable connections, as measured by median absolute deviation, which is more robust against outliers than standard deviation (**Supplementary Figure 2-5**). The input data thus consisted of 3410 unique functional connections (**Figure 2-2b**) and 111 clinical items (**Figure 2-2c**). The clinical items were drawn from the structured GOASSESS interview (Calkins et al., 2015), and covers a diverse range of psychopathological domains, including mood and anxiety disorders, psychosis-spectrum symptoms, attention-deficit/hyperactivity disorder (ADHD), and other disorders (see details in **Supplementary Data 2-1**). Using elastic net regularization (l_1 + l_2) and parameter tuning over both the clinical and connectivity features, sCCA was able to obtain a sparse and interpretable model while minimizing over-fitting (**Figure 2-2d** and **Supplementary Figure 2-6**). Ultimately, sCCA identified specific patterns ("canonical variates") of functional connectivity that were linked to distinct combinations of psychiatric symptoms.

Based on the scree plot of covariance explained (**Figure 2-3a**), 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, which compares the canonical correlate of each variate to a null distribution built by randomly re-assigning subjects' brain and clinical features (see Methods section and **Supplementary Figure 2-7**); False Discovery Rate (FDR) was used to control for type I error rate due to multiple testing. Of these seven canonical 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) (**Figure 2-3b**), with the fourth showing a trend toward 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 Figure 2-8a**), the parcellation system (**Supplementary Figure 2-8b**), and the use of regularization with elastic net versus data reduction with principal component analysis (**Supplementary Figure 2-8c**).

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 (**Figure 2-3c–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.

BRAIN-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 (see Methods section and **Supplementary Figure 2-9**). This procedure revealed that 37 out of 111 psychiatric symptoms reliably contributed to at least one of the four dimensions (**Figure 2-4**). 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 (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”) (**Figure 2-4a**). 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 (**Figure 2-4b**). The third dimension was composed of fear symptoms, including both agoraphobia and social phobia (**Figure 2-4c**). The fourth dimension was driven primarily by symptoms of both ADHD and ODD, but also included the irritability item from the depression domain (**Figure 2-4d**). The connectivity-guided clinical dimensions were significantly correlated with, but not identical to, previous factor models such as the bifactor models (Shanmugan et al., 2016) (see **Supplementary Figure 2-12**). 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 CONNECTIVITY

sCCA identified each dimension of psychopathology through shared associations between clinical data and specific patterns of connectivity. 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 Figure 2-10**; see Methods section), as this framework has been useful in prior investigations of both brain development (Power et al., 2010; Satterthwaite et al., 2013) and psychopathology (Alexander-Bloch et al., 2012; Kaiser et al., 2015; Sharp, Scott, & Leech, 2014; Sylvester et al., 2012). This procedure revealed specific patterns of network-level connectivity that were related to the four dimensions of psychopathology (**Figure 2-5**). For example, the mood dimension was characterized by a marked increase in connectivity between the ventral attention and salience networks (**Figure 2-5a, e, i**), while the psychosis dimension received the highest loadings in connectivity between the default mode and executive systems (salience and fronto-parietal networks (**Figure 2-5b, f, j**). In contrast, increased within-network connectivity of the fronto-parietal network was most evident in the fear dimension (**Figure 2-5c, g, k**). Alterations of the salience system were particularly prominent for the externalizing behavior dimension, including lower connectivity with the default mode network and greater connectivity with the fronto-parietal control network (**Figure 2-5d, h, l**). Quantitatively, the specific loadings of within- and between-network connectivity in each dimension did not significantly correlate with each other (all $p > 0.05$), demonstrating that each dimension of psychopathology was characterized by a unique pattern of network connectivity.

The results indicate that while each canonical variate was comprised of unique patterns of connectivity, there were several features that were shared across all dimensions. Such findings agree with accumulating evidence for common circuit-level dysfunction across psychiatric syndromes (Goodkind et al., 2015; Lee, Ripke, Neale, Faraone, Purcell, Perlis, Mowry, International Inflammatory Bowel Disease Genetics Consortium (IIBDGC), et al., 2013). To quantitatively assess such common features, we compared overlapping results against a null distribution using permutation testing (see Methods section). 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 (**Figure 2-6a**). Similar analysis at the level of sub-networks (**Figure 2-6b**) illustrated that abnormalities of connectivity within the default mode and fronto-parietal networks were present in all four psychopathological dimensions (**Figure 2-6c**). 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 connectivity associated with psychopathology.

DEVELOPMENTAL EFFECTS AND SEX DIFFERENCES

In the above analyses, we examined multivariate associations between connectivity and psychopathology while controlling for participant age. However, given that abnormal brain development is thought to underlie many psychiatric disorders (T. R. Insel, 2014; Paus, 2005), 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 regressed as prior. Notably, the dimensions derived were quite similar, with highly correlated feature weights (**Supplementary Table 2-2**). As in prior work (Baum et al., 2017; Gennatas et al., 2017), 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 (**Figure 2-7a, 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 (**Figure 2-7c, 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 generalizability 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 ($n = 336$, **Table 2-1**, **Figure 2-1**, and **Supplementary Figure 2-1**). 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. Specifically, the correlations between the clinical loadings in the discovery sample and those in the replication sample were $r = 0.85$ for psychosis ($P_{FDR} < 4.4 \times 10^{-16}$), $r = 0.73$ for externalizing ($P_{FDR} < 4.4 \times 10^{-16}$), $r = 0.59$ for fear ($P_{FDR} = 8.43 \times 10^{-12}$), and $r = 0.23$ for mood ($P_{FDR} = 0.01$). In the replication sample, three out of four dimensions were significant after FDR correction of permutation tests (**Figure 2-8** and **Supplementary Figure 2-11**). While the bootstrap analysis identified 37 out of 111 symptoms in the discovery sample to consistently contribute to the four linked-dimensions (**Figure 2-4**), the same analysis in the replication sample yielded similar sets of symptoms (80%, 64%, 63%, and 50% overlapping for psychosis, externalizing behavior, fear, and mood, respectively). Additionally, connectivity patterns associated with mood symptoms increased significantly with age ($P_{FDR} = 0.0082$), while connectivity patterns associated with psychosis symptoms showed a trend towards increasing with age ($P_{uncorrected} = 0.027$, $P_{FDR} = 0.053$). As in the discovery

sample, connectivity patterns associated with fear ($P_{FDR} = 0.039$) and mood ($P_{FDR} = 0.0083$) were both elevated in females in the replication sample.

Discussion

Leveraging a large neuroimaging data set of youth and recent advances in machine learning, we discovered several multivariate patterns of functional connectivity linked to interpretable dimensions of psychopathology that cross traditional diagnostic categories. These patterns of abnormal connectivity were largely 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 connectivity 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” (B. T. R. Insel & Cuthbert, 2015). 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 (Cuthbert & Insel, 2010). 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 (Satterthwaite, Kable, et al., 2015; Sharma et al., 2017). 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. For example, factor analyses of clinical psychopathology have suggested the presence of dimensions including internalizing symptoms, externalizing symptoms, and psychosis symptoms (Calkins et al., 2015; Shanmugan et al., 2016). 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 and increasingly pursued approach is to parse heterogeneity in psychiatric conditions using multivariate analysis of biomarker data such as neuroimaging. For example, researchers have used functional connectivity (Drysdale et al., 2016) and gray matter density (Clementz et al., 2016) 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 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 method allowed us to uncover linked dimensions of psychopathology and connectivity that cross diagnostic categories yet remain clinically interpretable. Compared to supervised classification methods (e.g., case-control, or multi-class), where each subject is categorized into one discrete class, unsupervised sCCA overcomes the inherent limitation of using discrete diagnostic categories (such as those provided by the Diagnostic and Statistical Manual of Mental Disorders) and allows continuous dimensions of psychopathology to be present in an individual to a varying degree. In addition, in contrast to “one-view” multivariate studies (such as factor analysis of clinical data or clustering of imaging data) (Calkins et al., 2015; Shanmugan et al., 2016), the sCCA-derived clinical dimensions were explicitly selected on the basis of co-varying signals that were present as both individual differences of connectivity and clinical symptoms. Such an unsupervised “two-view” approach

has been successfully applied in studies of neurodegenerative diseases (Avants et al., 2014) and normal brain-behavior relationships (Smith et al., 2015). In this dimensional, trans-diagnostic approach, the psychopathology of an individual is represented as a mixture of dimensional brain circuit abnormalities, which together produce a specific combination of psychiatric symptoms.

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 (Purcell et al., 2009). Instead of averaging over many clinical features within a diagnostic category, sCCA selected specific items that were most tightly linked to patterns of connectivity. These groups of symptoms remained highly interpretable, and were largely 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 (Alexander-Bloch et al., 2012; Satterthwaite et al., 2013). 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. (2005) 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 finding

regarding adolescent brain development (Baum et al., 2017; Satterthwaite et al., 2013). As part of this process, connections within network modules strengthen and connections between two network modules weaken. This process is apparent using functional connectivity (Power et al., 2010; Satterthwaite et al., 2013) as well as structural connectivity (Baum et al., 2017). 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 (Woodward & Cascio, 2015). 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, lending new evidence for the triple-network model of psychiatric disorders (Lefebvre et al., 2016; Menon, 2011).

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 (Berman et al., 2011; Sheline et al., 2009). 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. Finally, the psychosis dimension exhibited stronger connectivity in default mode network and reduced segregation from executive networks (fronto-parietal and salience). Notably, while prior studies have focused on the central role of default mode dysconnectivity in schizophrenia (Whitfield-Gabrieli & Ford, 2012) with mixed evidence for hyper-connectivity (Zhou et al., 2007) and hypo-connectivity (Pankow et al., 2015), in the present data the effect within default mode network itself was not nearly as strong as its reduced segregation from the executive networks. Indeed, this finding is consistent with recent data that in psychosis the disruption of segregation between the default mode and task

positive networks is a more consistent feature than dysconnectivity within the default mode itself (Lefort-Besnard et al., 2018).

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 (Casey, Oliveri, & Insel, 2014) and marked disparities between males and females (Rapoport, Giedd, & Gogtay, 2012), 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 (Harrow, Carone, & Westermeyer, 1985). 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 (Bongers, Koot, Van Der Ende, & Verhulst, 2004). 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 (Albert, 2015; Kessler, 2003). 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 (Satterthwaite, Wolf, et al., 2015). 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, it should be emphasized that our approach did not seek to define biotypes within clinical diagnostic categories in a fully data-driven manner, as in influential prior work (Clementz et al., 2016; Drysdale et al., 2016). Rather, here we sought to provide complementary understanding of heterogeneity by linking symptoms that are present across clinical diagnostic categories to alterations of functional connectivity, uncovering dimensions of psychopathology that are guided by and linked to underlying network abnormalities. However, this approach necessarily is limited by the clinical data being used, in this case item-level data from a structured clinical interview. 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. Fourth, our replication sample was constructed from the PNC data. Using an independently acquired dataset to validate our findings would provide evidence of greater generalizability than splitting the original data into two samples. However, this approach was dictated by the lack of correspondence with clinical instruments used in other large-scale developmental imaging studies. This limitation underscores the need for harmonization of not just imaging data but also clinical measures across studies moving forward. Finally, our current analysis only considered functional connectivity and clinical psychopathology. Future research could incorporate rich multi-modal imaging data, cognitive measures, and genomics.

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 connectivity, 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 connectivity. 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.

Tables

Table 2-1: Philadelphia neurodevelopmental cohort (PNC)

		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

Table 2-1 The cross-sectional sample of the PNC has 1601 participants in total. After applying health, structural, and functional imaging quality exclusion criteria (details in Online Methods section), 663 and 336 subjects were included in the final discovery and replication samples, respectively.

Figures

Figure 2-1:

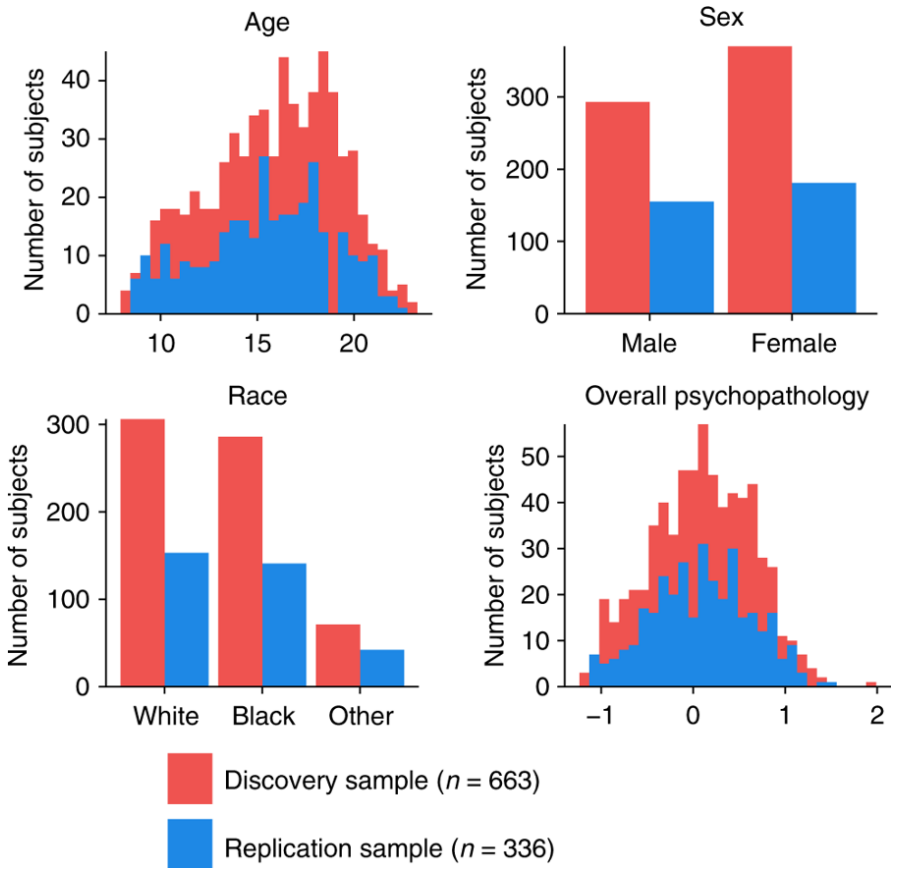


Figure 2-1 Participants demographics. The discovery and replication samples had similar demographic composition, including similar distributions of age, race, sex, and overall psychopathology.

Figure 2-2

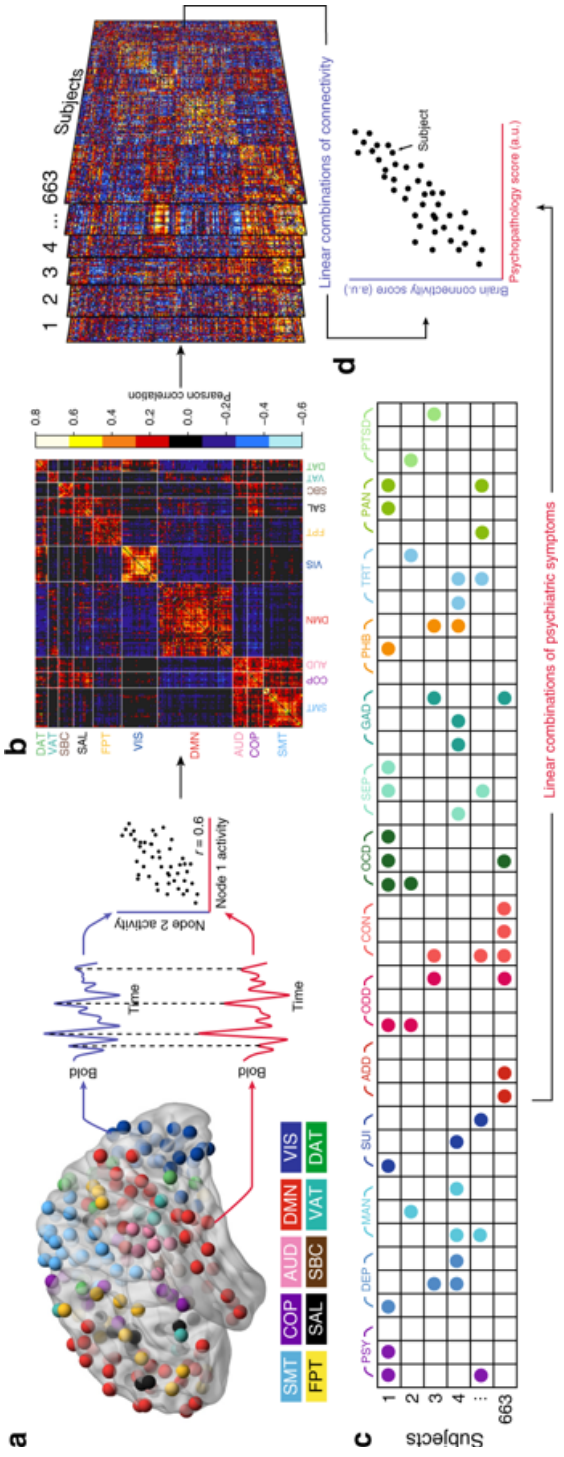

a
b
c
d
Bold
Time
Node 1 activity
Node 2 activity
r = 0.6
Pearson correlation
Subjects
Linear combinations of connectivity
Brain connectivity score (a.u.)
Psychopathology score (a.u.)
Subject
Linear combinations of psychiatric symptoms
DAT
VAT
SBC
SAL
FPT
VIS
DMN
AUD
COP
SMT
PSY
DEP
MAN
SUI
ADD
OBD
CON
OCD
SEP
GAD
PHB
TRT
PAN
PTSD

Figure 2-2: 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. (2011) (**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 (Merikangas et al., 2010) were entered into sCCA as clinical features (see details in **Supplementary Data 2-1**). (**d**) sCCA seeks linear combinations of connectivity and clinical symptoms that maximize their correlation. A priori community assignment: somatosensory/motor network (SMT), cingulo-opercular network (COP), auditory network (AUD), default mode network (DMN), visual network (VIS), fronto-parietal network (FPT), salience network (SAL), subcortical network (SBC), ventral attention network (VAT), dorsal attention network (DAT), Cerebellar and unsorted nodes not visualized. Psychopathology domains: psychotic and subthreshold symptoms (PSY), depression (DEP), mania (MAN), suicidality (SUI), attention-deficit hyperactivity disorder (ADD), oppositional defiant disorder (ODD), conduct disorder (CON), obsessive-compulsive disorder (OCD), separation anxiety (SEP), generalized anxiety disorder (GAD), specific phobias (PHB),

mental health treatment (TRT), panic disorder (PAN), post-traumatic stress disorder (PTSD).

Figure 2-3

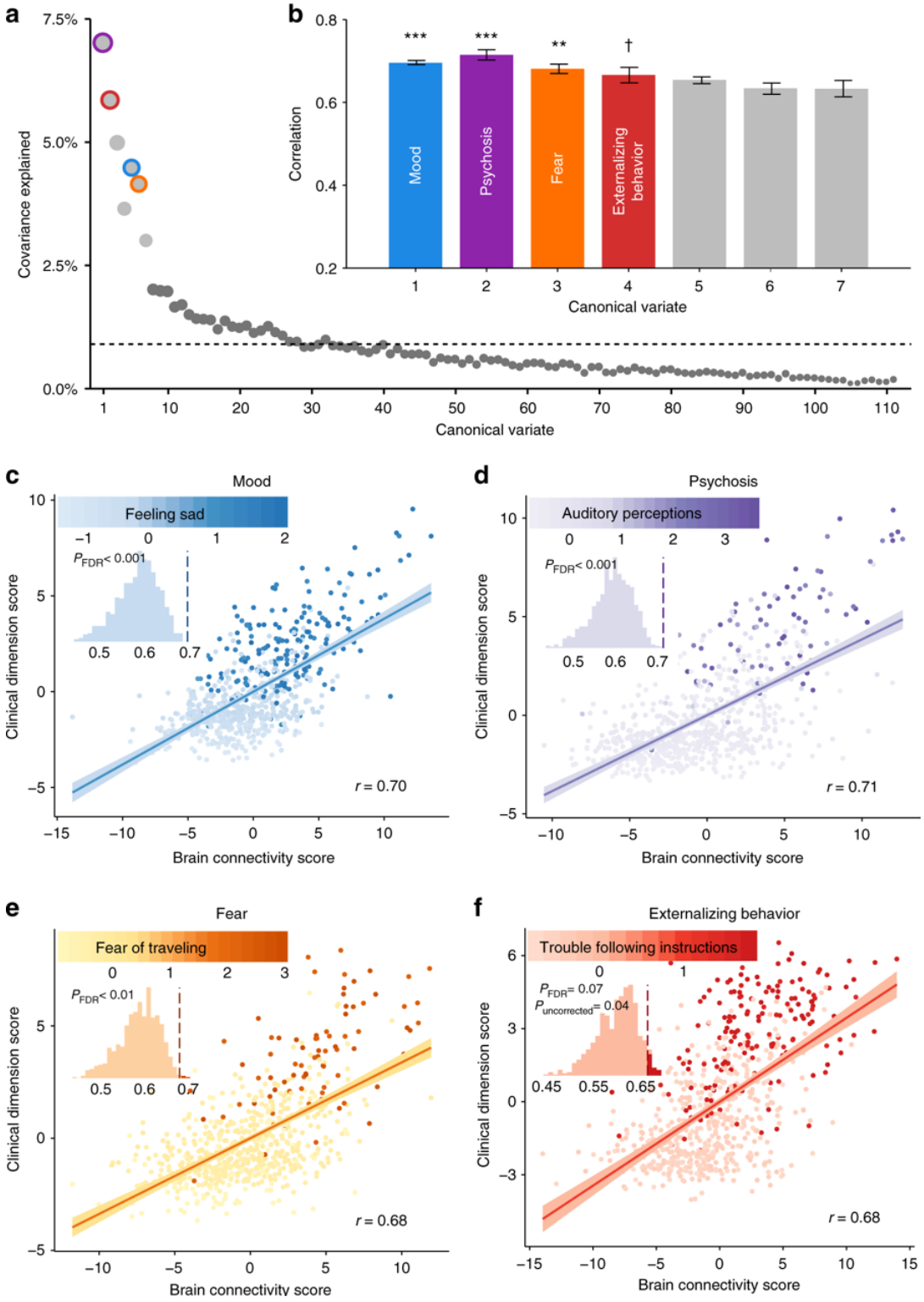

a
b
Covariance explained
Canonical variate
Correlation
Mood
Psychosis
Fear
Externalizing behavior
c
Mood
Feeling sad
Clinical dimension score
Brain connectivity score
r = 0.70
d
Psychosis
Auditory perceptions
r = 0.71
e
Fear
Fear of traveling
r = 0.68
f
Externalizing behavior
Trouble following instructions
r = 0.68

Figure 2-3 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. ***$P_{FDR} < 0.001$, **$P_{FDR} < 0.01$, †$P_{uncorrected} = 0.04$

Figure 2-4


Dimension
A
B
C
D
Oppo.Defiant
Attention Deficit
Social phobia
Agoraphobia
OCD
SUI
Depression
Mania
Psychosis
Losing temper
Irritability due to unfairness
Breaking rules
Trouble making plans
Making careless mistakes
Difficulty waiting turn
Difficulty sitting still
Trouble listening
Trouble paying attention
Problems following instructions
Public performance
Center of attention
Focus of social situations
Novel social situations
Bridges/tunnels
Public transit
Leaving home
Solo travel
Thoughts of harming
Suicidality
Irritability
Anhedonia
Crying
Feeling sad
Irritability
Pressured speech
Overly energetic
Persecutory/suspicious
Auditory verbal hallucinations
Loss sense of self
Mind tricks
Thought control
Superstitions
Audible thoughts
Reality confusion
Odd/unusal thoughts
Auditory perceptions
0.4 0.38 0.36 0.3 0.27 0.26 0.22 0.27 0.24 0.2 0.25 0.18 0.21 0.49 0.38 0.35 0.31 0.28 0.21 0.27 0.27 0.28 0.34 0.24 0.26 0.36 0.37 0.36 0.29 0.28 0.24 0.23 0.22 0.35 0.3 0.27

Figure 2-4 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 (see details in **Supplementary Data 2-1**). 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 2-5

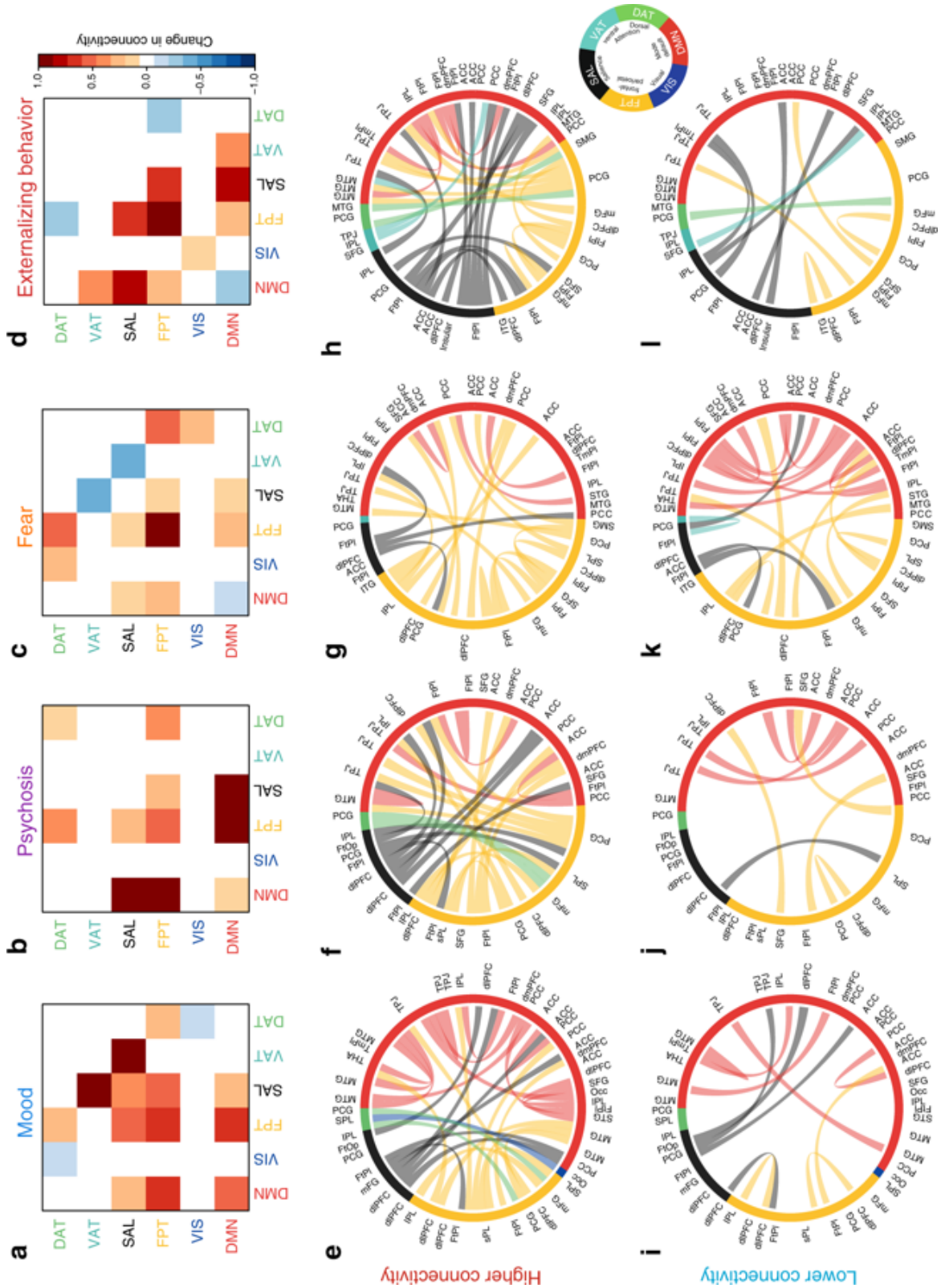

Change in connectivity
1.0
0.5
0.0
-0.5
-1.0
a
Mood
b
Psychosis
c
Fear
d
Externalizing behavior
DAT
VAT
SAL
FPT
VIS
DMN
Higher connectivity
Lower connectivity

Figure 2-5 Patterns of within- and between-network connectivity 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 represents the loadings of connectivity features.

Figure 2-6

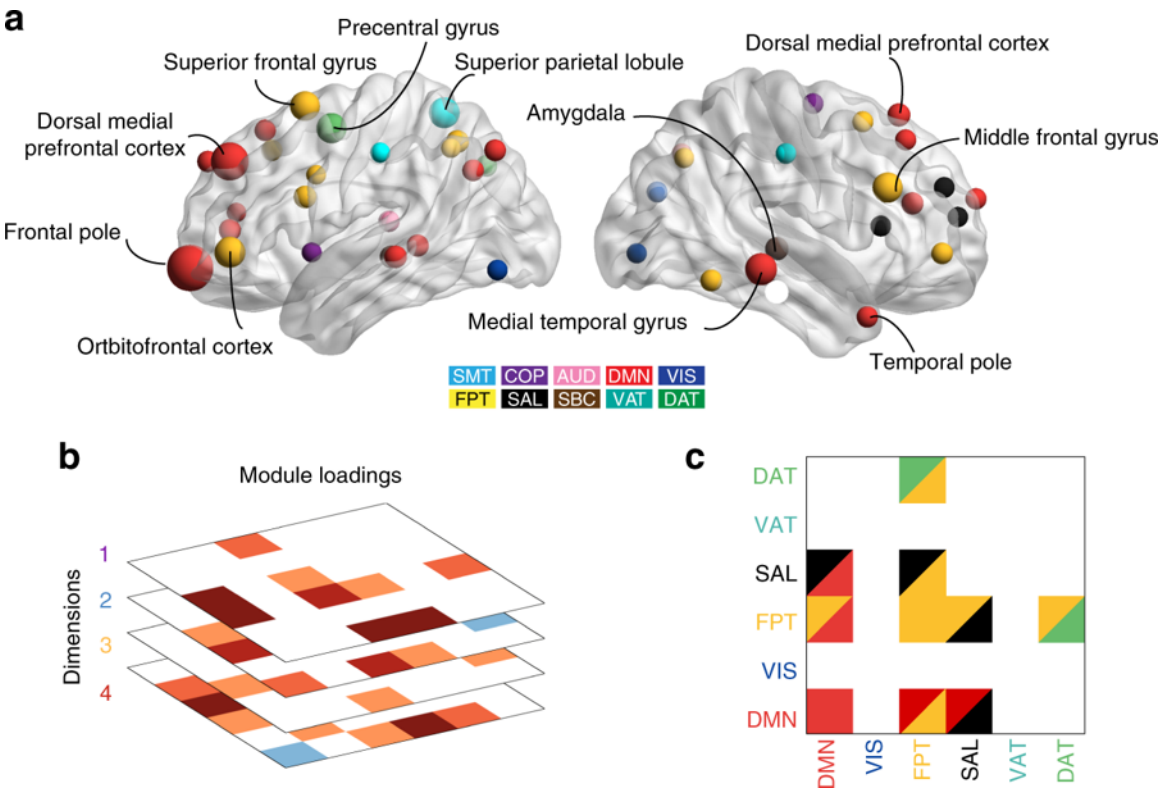


Figure 2-6 Loss of segregation between default mode and executive networks is shared across dimensions. (**a**) By searching for overlap of edges that contributed significantly to each dimension, we found common edges that were implicated across all dimensions of psychopathology. These were then summarized 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 was shared across all four dimensions.

Figure 2-7

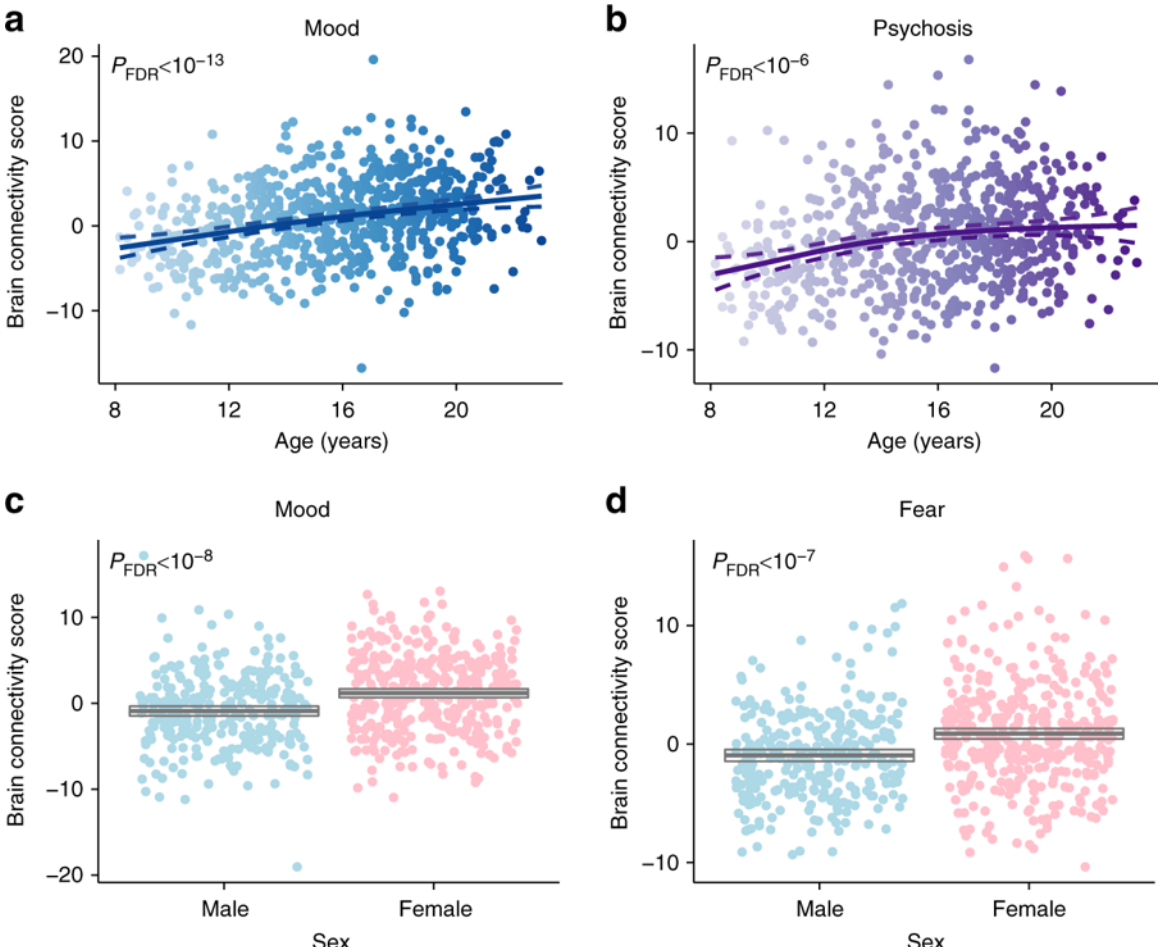


Figure 2-7 Developmental effects and sex differences are concentrated in specific dimensions. Connectivity patterns associated with both the mood (**a**) and psychosis (**b**) dimensions increased significantly with age. Additionally, connectivity patterns associated with both the mood (**c**) and fear (**d**) dimensions were significantly more prominent in females than males. Multiple comparisons were controlled for using the False Discovery Rate ($q < 0.05$). Dashed lines and boxes indicate the 95% confidence interval.

Figure 2-8

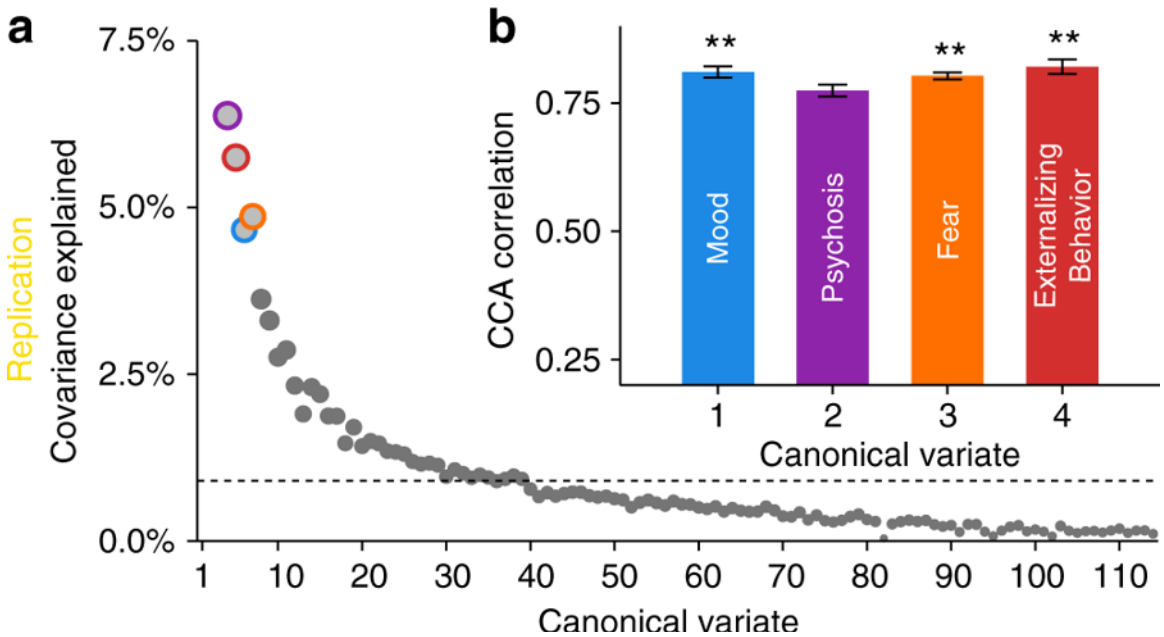
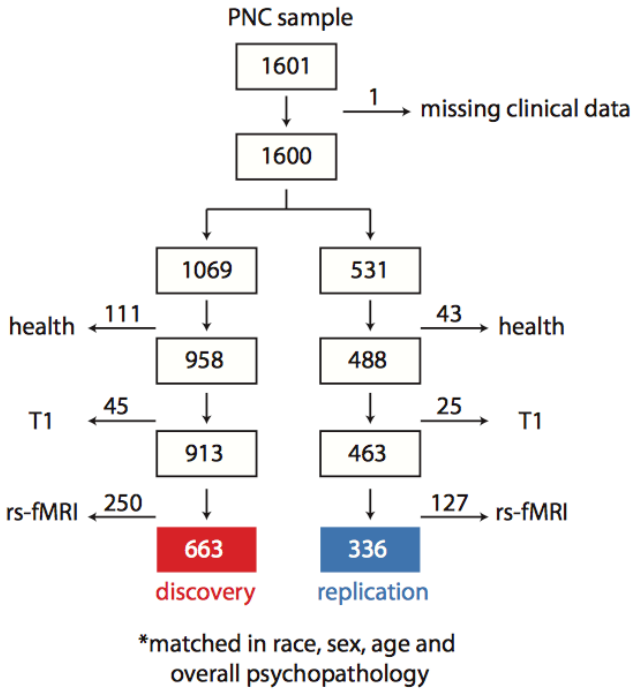


Figure 2-8 Linked dimensions of psychopathology were replicated in an independent sample. All procedures were repeated in an independent replication sample of 336 participants. (**a**) The first four canonical variates in the replication sample were selected for further analysis based on covariance explained. 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. $^{**}P_{FDR} < 0.01$

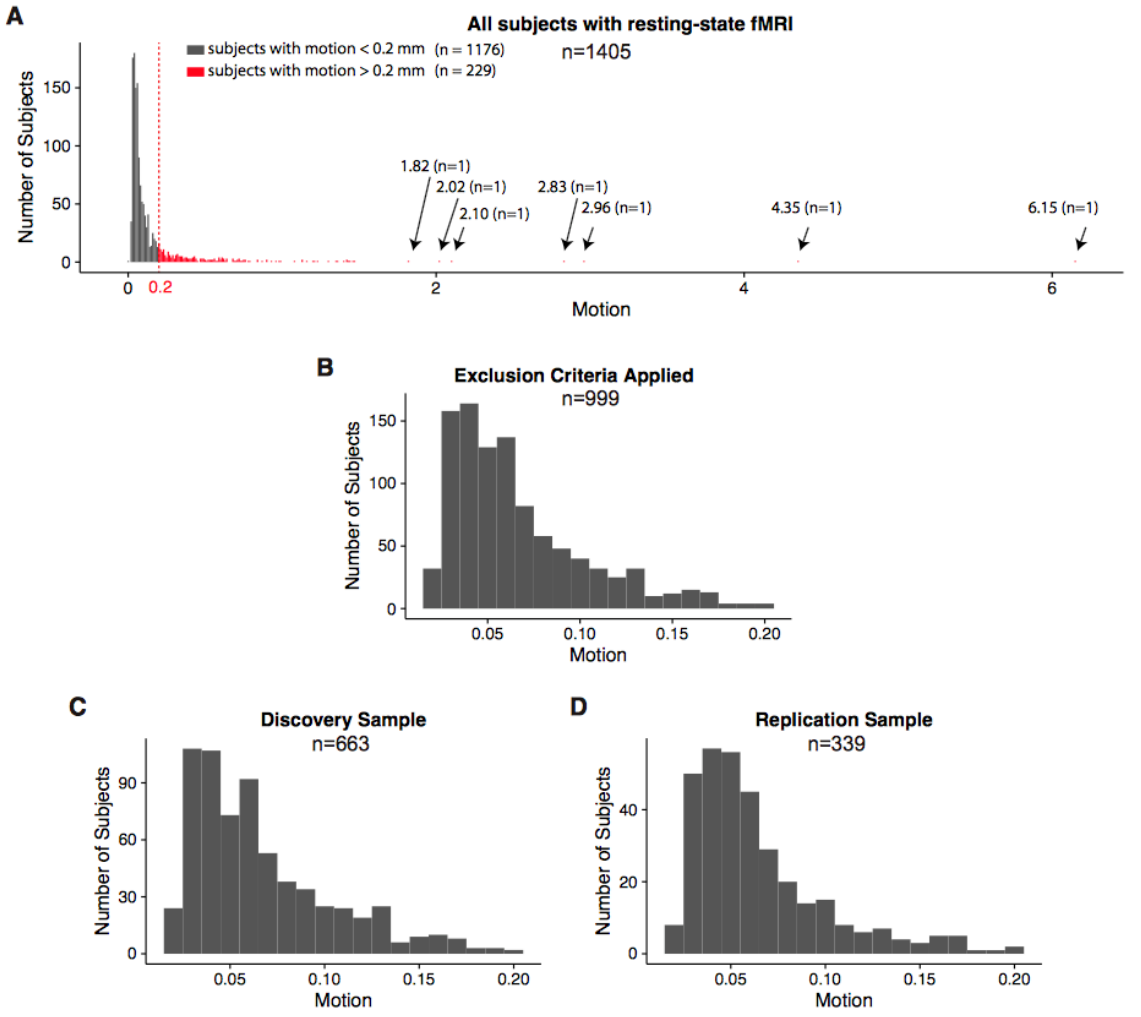
Supplementary Information

Supplementary Figure 2-1



Supplementary Figure 2-1 Sample Construction. 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 sample (n=531). Applying quality exclusion criteria for health, structural imaging, and functional imaging (details in Methods), 663 and 336 subjects were included in the final discovery and replication samples, respectively.

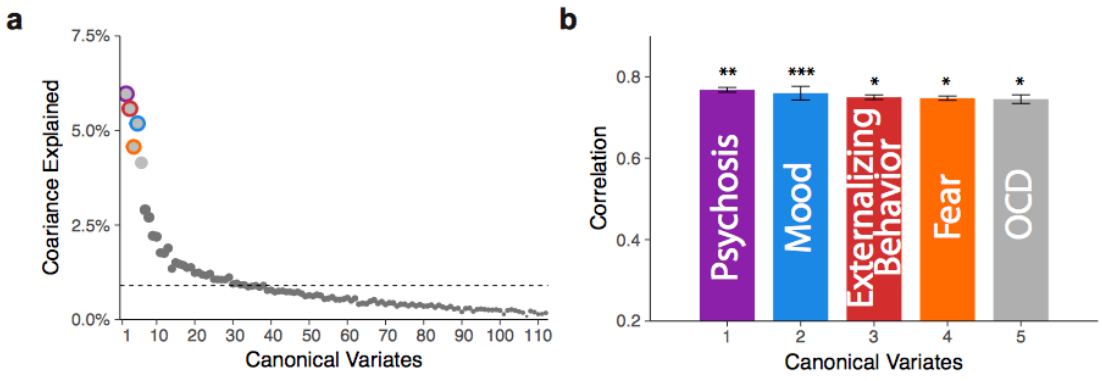
Supplementary Figure 2-2



Supplementary Figure 2-2 In-scanner motion of subjects. (**a**) 1405 out of 1601 participants of PNC had acquired resting- state fMRI. The histogram shows the distribution of mean framewise displacement using the Jenkinson calculation. The exclusion criteria of motion for the final sample is 0.2mm or greater, which is colored in red (n=229). (**b**) After applying all exclusion criteria, including health, structural and functional imaging quality exclusion criteria, 999 subjects were

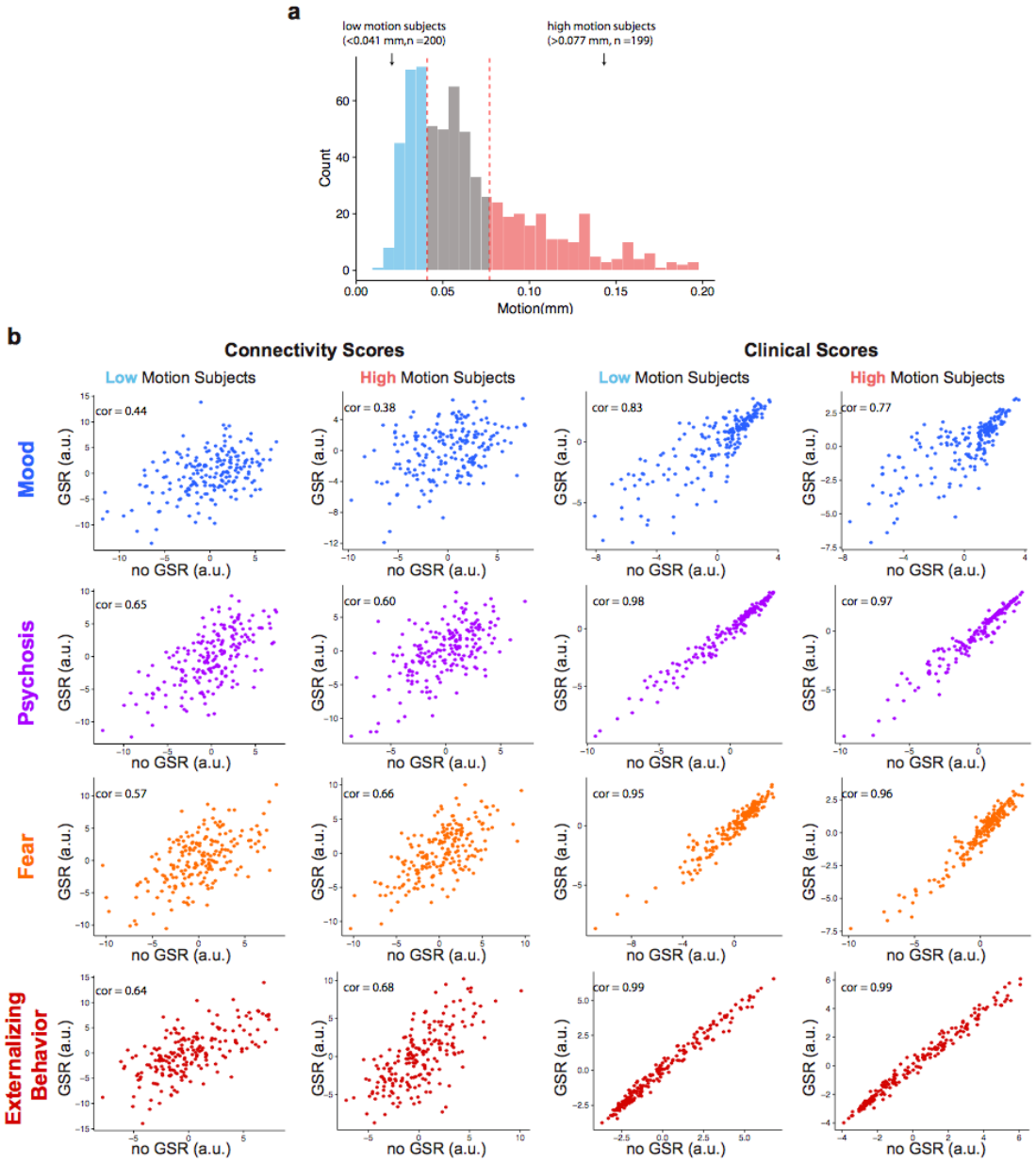
included in the final sample. The histogram shows the head motion distribution of the final sample, which consists of a discovery sample (**c**), and a replication sample (**d**).

Supplementary Figure 2-3



Supplementary Figure 2-3 Pre-processed data without global signal regression (GSR). (**a**) We preprocessed the functional data with 12 parameter + `aCompCor`, which is the one of the best performing preprocessing procedures to correct for motion without GSR (Behzadi, Restom, Liau, & Liu, 2007; Ciric et al., 2017; Muschelli et al., 2014; Parkes, Fulcher, Yücel, & Fornito, 2018). Subsequently, we followed the same procedures as in the main analysis. The first five canonical variates were selected for further analysis based on covariance explained. Dashed line marks the average covariance explained. (**b**) The original four dimensions -— psychosis, mood, fear, and externalizing behavior — and a fifth dimension (corresponding to OCD-spectrum symptoms) were significant by permutation testing. Corresponding variates are circled in panel (**a**). Error bars denote standard error. *** $P_{FDR} < 0.001$, ** $P_{FDR} < 0.01$, * $P_{FDR} < 0.05$.

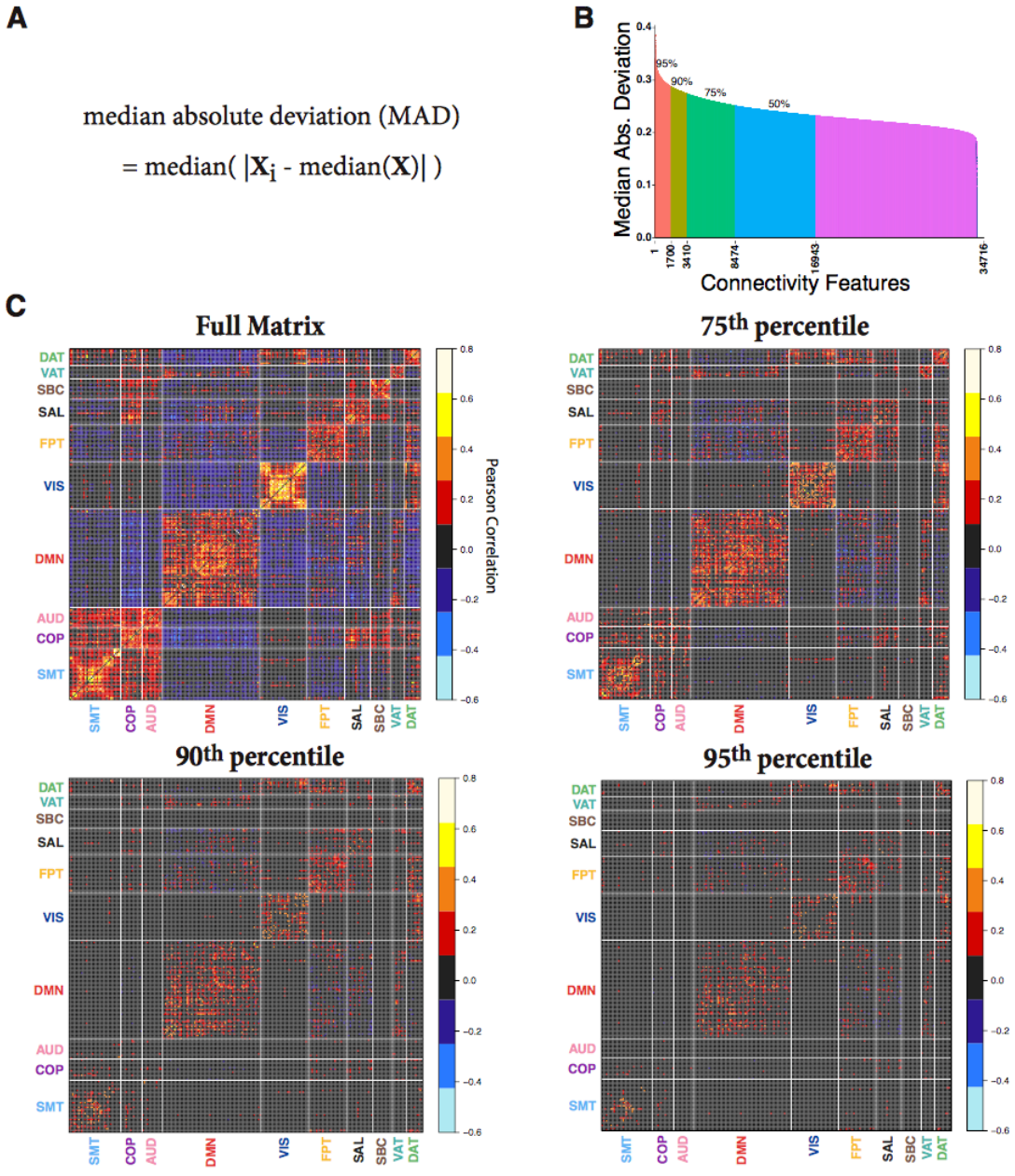
Supplementary Figure 2-4



Supplementary Figure 2-4 Comparison of GSR effects in low and high motion subjects. (**a**) Histogram of subject in-scanner motion in the discovery cohort (n=663), of which those with the lowest motion (< 0.041 mm, n = 200) and

those with the highest motion (> 0.077 mm, n =199) were selected for the comparison of their CCA dimensional scores processed with and without GSR. (**b**) We calculated the correlation coefficient between the CCA dimensional scores (i.e. connectivity and clinical scores) processed without GSR (x axis) and those processed with GSR (y axis) in each motion group for each of the four canonical dimensions. All correlation coefficients were highly significant ($P < 2.2 \times 10-16$).

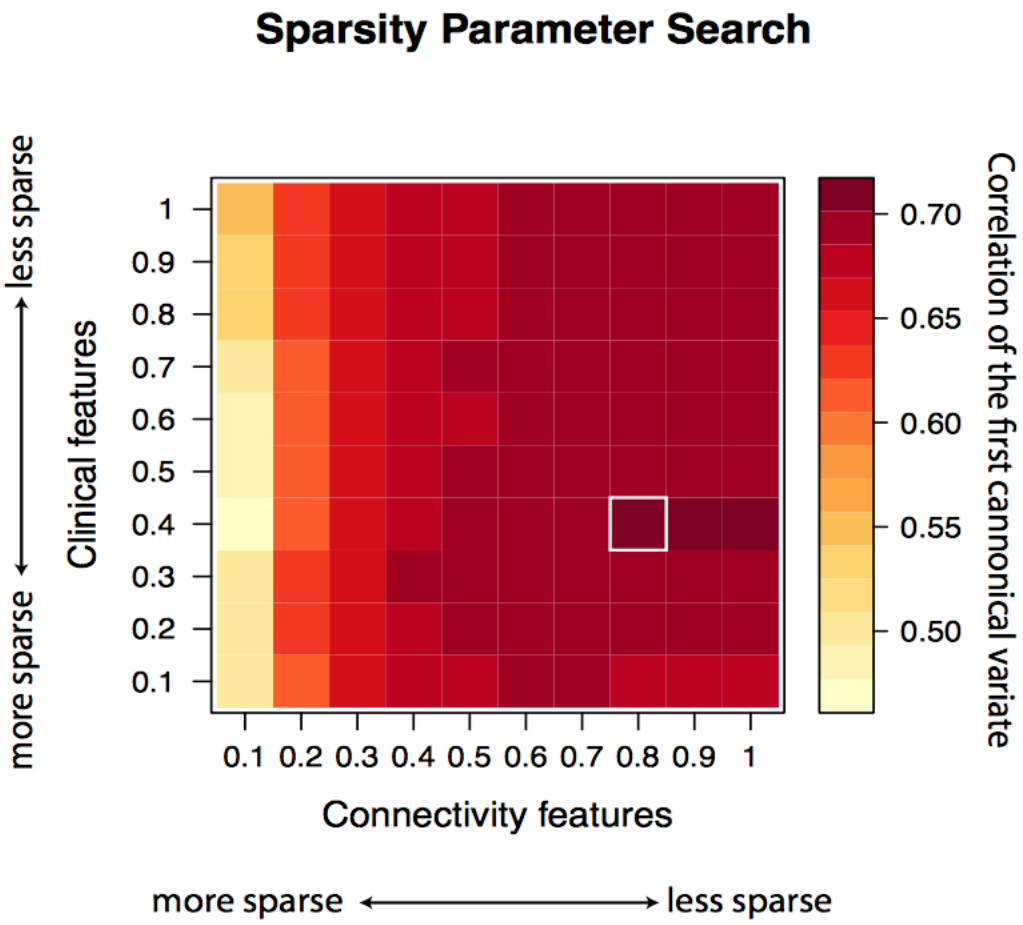
Supplementary Figure 2-5



Supplementary Figure 2-5 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.

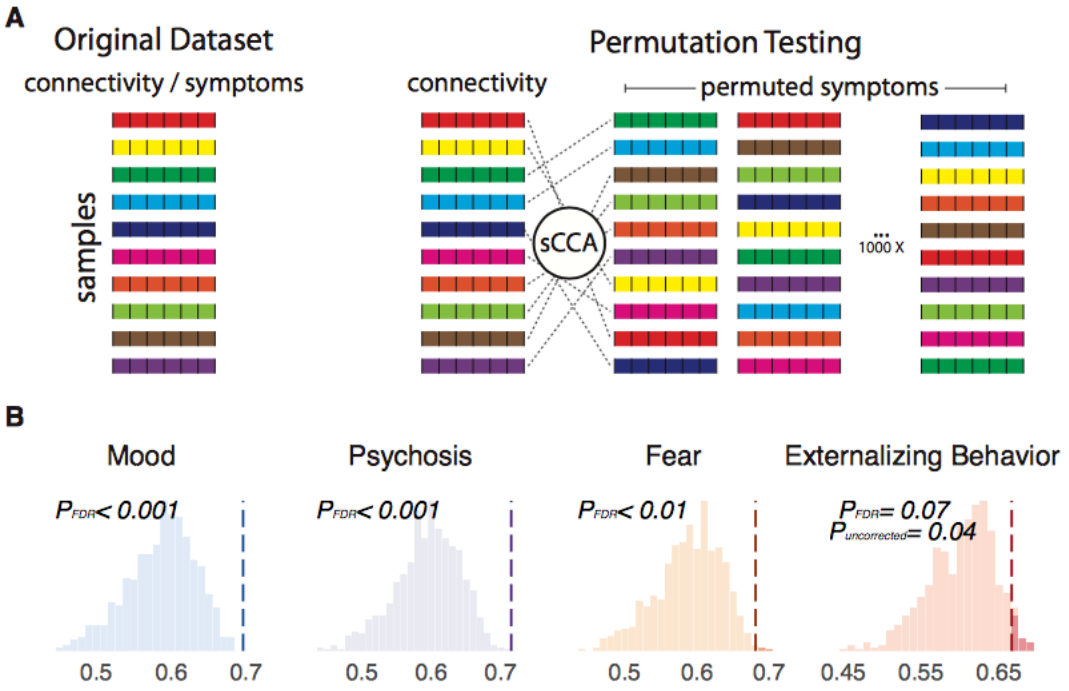
Supplementary Figure 2-6



Supplementary Figure 2-6 Grid search for regularization parameters. We tuned the L1 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 (Witten et al., 2009), 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.

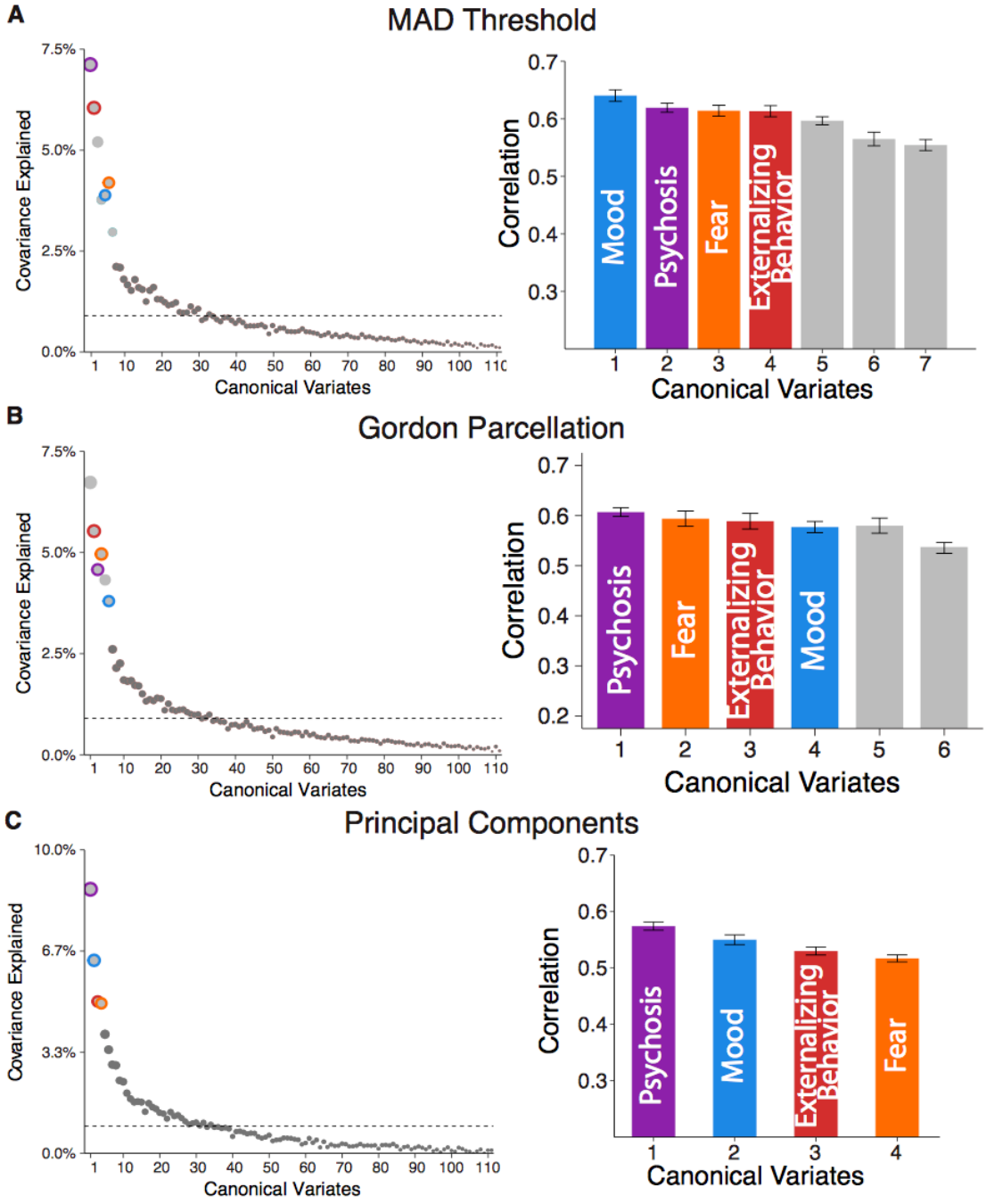
Supplementary Figure 2-7



Supplementary Figure 2-7 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 ($\boldsymbol{v}$) (Mišić et al., 2016). (**b**) Null distributions of

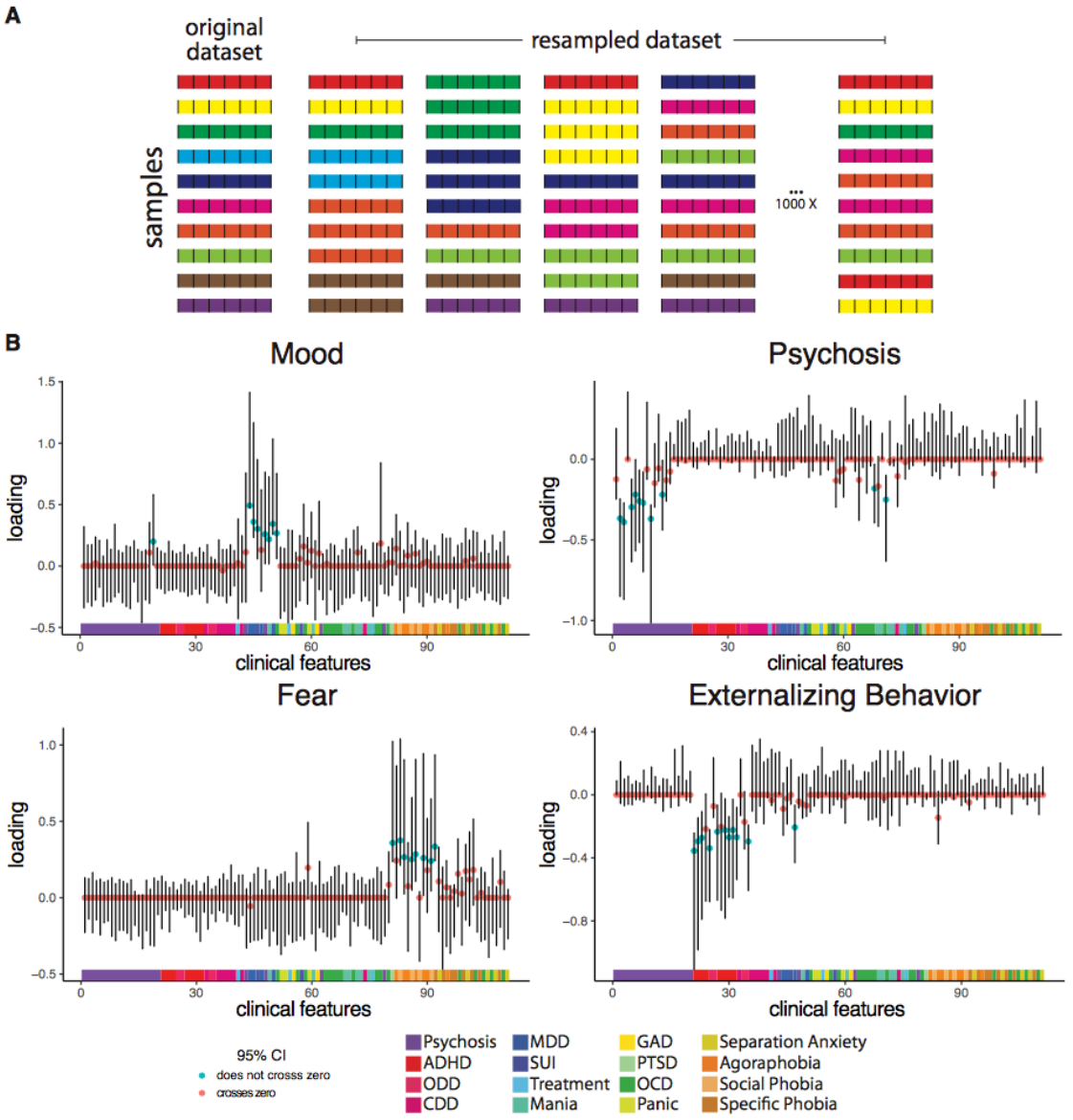
correlations generated by the permuted data. Dashed line represents the correlation from the original dataset.

Supplementary Figure 2-8



Supplementary Figure 2-8 Patterns of canonical variates were robust to methodological choices. We found four canon- ical 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., 2016), 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.

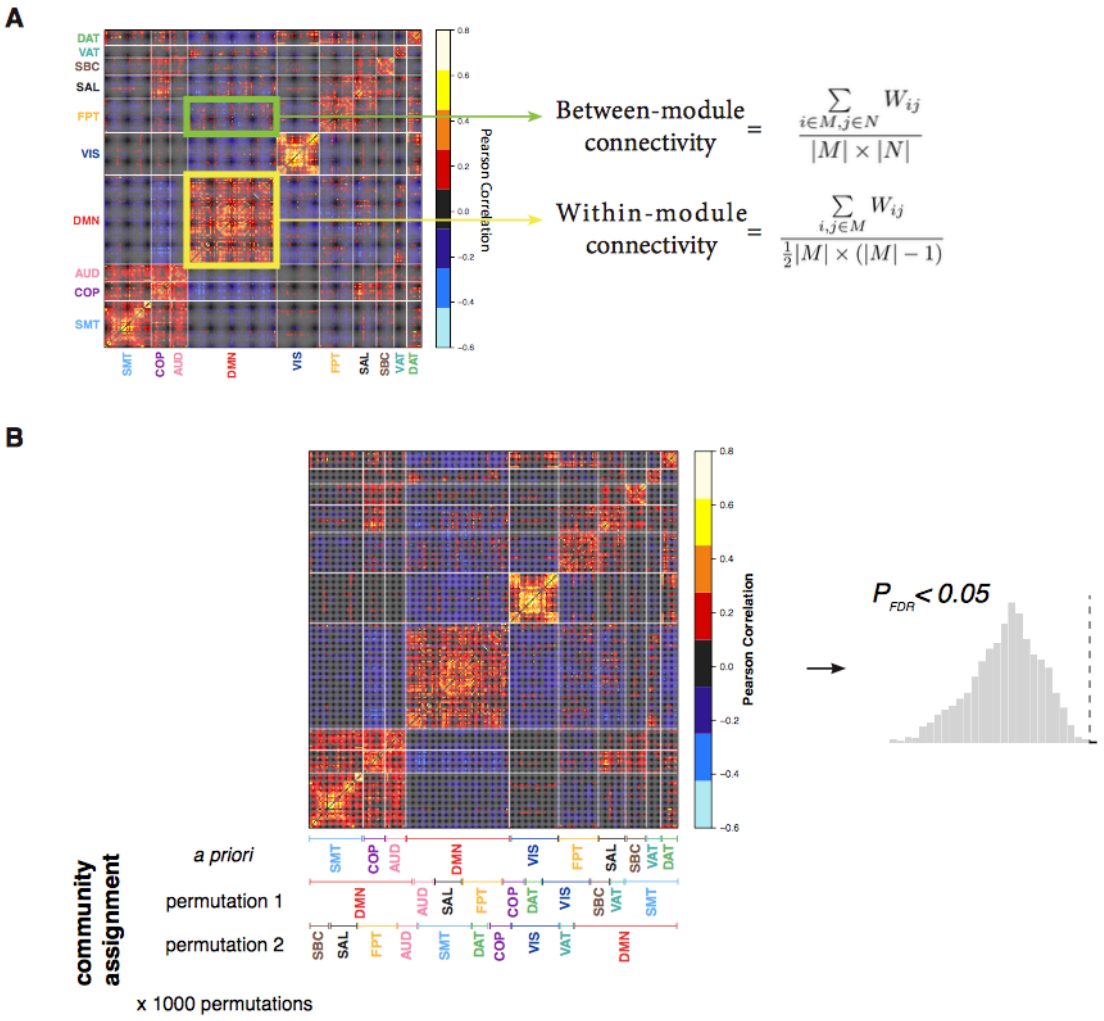
Supplementary Figure 2-9



Supplementary Figure 2-9 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.

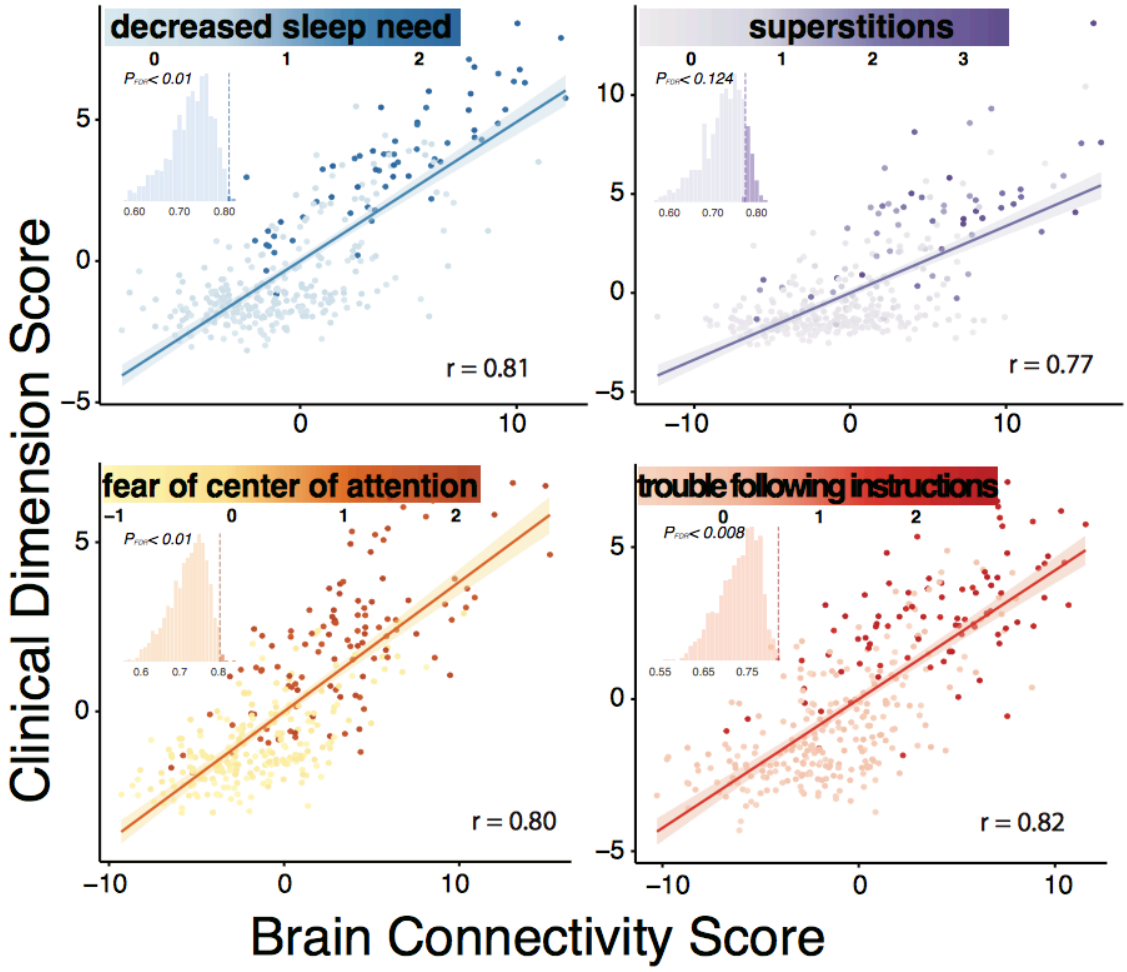
Supplementary Figure 2-10



Supplementary Figure 2-10 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. (2011) (**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 original values against the null distribution.

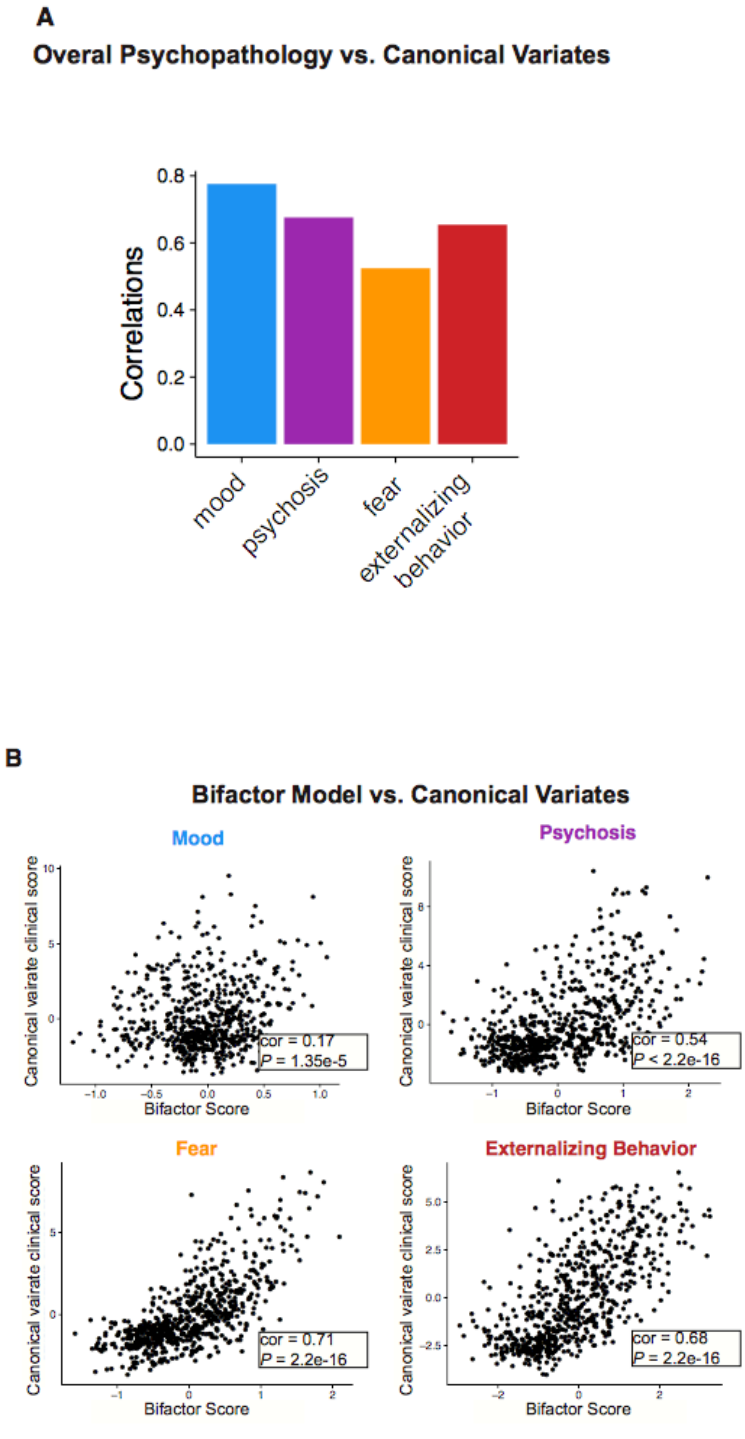
Supplementary Figure 2-11



Supplementary Figure 2-11 Canonical variates in the replication sample accord with findings in the discovery sample. 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 Figure 2-12



Supplementary Figure 2-12 Correlations between canonical variates and previous factor analysis model. To under- stand how similar connectivity-guided dimensions of psychopathology are to those derived from pure clinical items reported before, we examined the correlation between the canonical variate clinical scores and (**a**) overall psychopathol- ogy score, and (**b**) dimensional bifactor models scores, both initially reported in Shanmugan et al., (2016).

Supplementary Table 2-1

Psychopathology Categories	All sample (n=1601)	Total analyzed sample (n=999)	Discovery (n=663)	Replication (n=336)
Mania	1.2%	1.3%	1.1%	1.8%
Depression	13.2%	15.3%	16.3%	13.4%
Bulimia	0.4%	0.3%	0.5%	0.0%
Anorexia	1.0%	1.4%	1.8%	1.2%
Generalized Anxiety Disorder	1.7%	1.6%	1.8%	1.2%
Separation Anxiety Disorder	4.7%	3.9%	4.1%	3.6%
Social Phobia	23.1%	24.8%	25.2%	24.1%
Panic Disorder	1.0%	0.8%	1.1%	0.3%
Agoraphobia	5.7%	5.8%	6.5%	4.5%
Obsessive-Compulsive Disorder	2.8%	2.7%	2.6%	3.0%
Post-Traumatic Stress Disorder	11.7%	12.4%	12.4%	12.5%
Psychosis	6.7%	6.6%	7.5%	4.8%
Attention-Deficit Disorder	17.3%	15.3%	15.2%	15.5%
Conduct Disorder	8.9%	8.6%	7.8%	10.1%

Supplementary Table 2-1 Clinical Psychopathology Levels in the PNC.

Supplementary Table 2-2

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

Supplementary Table 2-2 Correlations of loadings between covariate-regressed and non-regressed features. 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 2-3

Questions from the GOASESS 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 listen- ing 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)**
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)**
Gen erali zed	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 en- ergetic 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)**
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 forbid- den/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?
	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 often annoy other people on purpose or blame other people for your mistakes (excluding siblings)?
	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?
Pho bia	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?
	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)**
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)**
	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?
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 (at- tachment 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 happen- ing 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?
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, play- ing 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)**
Suicidality	SUI001	Have you ever thought a lot about death or dying?
	SUI002	Have you ever thought about killing yourself? **(suicidality)**

Supplementary Table 2-3 Item-wise psychiatric symptoms included as part of the data analysis. We included 111 psychiatric symptoms as input data to the sCCA. Item with bolded abbreviation were the symptoms highlighted in **Figure 2-4**.

References

Albert, P. R. (2015). Why is depression more prevalent in women? *Journal of Psychiatry and Neuroscience*, *40*(4), 219–221. https://doi.org/10.1503/jpn.150205

Alexander-Bloch, A. F., Lambiotte, R., Roberts, B., Giedd, J., Gogtay, N., & Bullmore, E. (2012). The discovery of population differences in network community structure: New methods and applications to brain functional networks in schizophrenia. *NeuroImage*, *59*(4), 3889–3900. https://doi.org/10.1016/j.neuroimage.2011.11.035

Avants, B. B., Libon, D. J., Rascovsky, K., Boller, A., Mcmillan, C. T., Massimo, L., … Grossman, M. (2014). Sparse canonical correlation analysis relates network-level atrophy to multivariate cognitive measures in a neurodegenerative population. *NeuroImage*, *84*, 698–711. https://doi.org/10.1016/j.neuroimage.2013.09.048

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*, *54*(3), 2033–2044. https://doi.org/10.1016/J.NEUROIMAGE.2010.09.025

Avants, B. B., Tustison, N. J., Wu, J., Cook, P. A., & Gee, J. C. (2011). An Open Source Multivariate Framework for n-Tissue Segmentation with Evaluation on Public Data. *Neuroinformatics*, *9*(4), 381–400. https://doi.org/10.1007/s12021-011-9109-y

Bassett, D. S., & Bullmore, E. T. (2009). Human brain networks in health and disease. *Current Opinion in Neurology*, *22*(4), 340–7. https://doi.org/10.1097/WCO.0b013e32832d93dd

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, *20*(3), 353–364. https://doi.org/10.1038/nn.4502

Bassett, D. S., Xia, C. H., & Satterthwaite, T. D. (2018). Understanding the Emergence of Neuropsychiatric Disorders With Network Neuroscience. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. https://doi.org/10.1016/J.BPSC.2018.03.015

Baum, G. L., Ciric, R., Roalf, D. R., Betzel, R. F., Moore, T. M., Shinohara, R. T., … Satterthwaite, T. D. (2017). Modular Segregation of Structural Brain Networks Supports the Development of Executive Function in Youth. *Current Biology*, *27*(11), 1561–1572.e8. https://doi.org/10.1016/j.cub.2017.04.051

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Berman, M. G., Peltier, S., Nee, D. E., Kross, E., Deldin, P. J., & Jonides, J. (2011). Depression, rumination and the default network. *Social Cognitive and Affective Neuroscience*, *6*(5), 548–555. https://doi.org/10.1093/scan/nsq080

Bongers, I. L., Koot, H. M., Van Der Ende, J., & Verhulst, F. C. (2004). Developmental trajectories of externalizing behaviors in childhood and adolescence. *Child Development*, *75*(5), 1523–1537. https://doi.org/10.1111/j.1467-8624.2004.00755.x

Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*(3), 186–198. https://doi.org/10.1038/nrn2575

Bzdok, D., & Yeo, B. T. T. (2017). Inference in the age of big data: Future perspectives on neuroscience. *NeuroImage*, *155*, 549–564. https://doi.org/10.1016/j.neuroimage.2017.04.061

Calkins, M. E., Merikangas, K. R., Moore, T. M., Burstein, M., Behr, M. A., Satterthwaite, T. D., … Gur, R. E. (2015). The Philadelphia Neurodevelopmental Cohort : constructing a deep phenotyping collaborative. *Journal of Child Psychology and Psychiatry*, *12*, 1356–1369. https://doi.org/10.1111/jcpp.12416

Casey, B. J., Oliveri, M. E., & Insel, T. (2014). A neurodevelopmental perspective on the research domain criteria (RDoC) framework. *Biological Psychiatry*, *76*(5), 350–353. https://doi.org/10.1016/j.biopsych.2014.01.006

Chai, L. R., Khambhati, A. N., Ciric, R., Moore, T. M., Gur, R. C., Gur, R. E., … Bassett, D. S. (2017). Evolution of brain network dynamics in neurodevelopment. *Network Neuroscience*, *1*(1), 14–30. https://doi.org/10.1162/NETN_a_00001

Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., Ruparel, K., … Satterthwaite, T. D. (2017). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *NeuroImage*, *154*(March), 174–187. https://doi.org/10.1016/j.neuroimage.2017.03.020

Clementz, B. A., Sweeney, J. A., Hamm, J. P., Ivleva, E. I., Ethridge, L. E., Pearlson, G. D., … Tamminga, C. A. (2016). Identification of distinct psychosis biotypes using brain-based biomarkers. *American Journal of Psychiatry*, *173*(4), 373–384. https://doi.org/10.1176/appi.ajp.2015.14091200

Cuthbert, B. N., & Insel, T. R. (2010). Toward new approaches to psychotic disorders: The NIMH research domain criteria project. *Schizophrenia Bulletin*, *36*(6), 1061–1062. https://doi.org/10.1093/schbul/sbq108

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

Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., … Liston, C. (2016). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine*, *23*(1), 28–38. https://doi.org/10.1038/nm.4246

Fornito, A., Zalesky, A., & Breakspear, M. (2015). The connectomics of brain disorders. *Nature Reviews Neuroscience*, *16*(3), 159–172. https://doi.org/10.1038/nrn3901

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Essen, D. C. Van, & Raichle, M. E. (2005). From The Cover: The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, *102*(27), 9673–9678. https://doi.org/10.1073/pnas.0504136102

Gennatas, E. D., Avants, B. B., Wolf, D. H., Satterthwaite, T. D., Ruparel, K., Ciric, R., … Gur, R. C. (2017). 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*(20), 3550–16. https://doi.org/10.1523/JNEUROSCI.3550-16.2017

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

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

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Gu, S., Satterthwaite, T. D., Medaglia, J. D., Yang, M., Gur, R. E. R. C. R. E., Gur, R. E. R. C. R. E., & Bassett, D. S. (2015). Emergence of system roles in normative neurodevelopment. *Proceedings of the National Academy of Sciences*, *112*(44), 201502829. https://doi.org/10.1073/pnas.1502829112

Hallquist, M. N., Hwang, K., & Luna, B. (2013). 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. https://doi.org/10.1016/j.neuroimage.2013.05.116

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

Insel, B. T. R., & Cuthbert, B. N. (2015). Brain diorders? Precisely. *Science*, *348*(6234), 499–500.

Insel, T. R. (2014). Mental disorders in childhood: shifting the focus from behavioral symptoms to neurodevelopmental trajectories. *JAMA : The Journal of the American Medical Association*, *311*(17), 1727–8. https://doi.org/10.1001/jama.2014.1193

Jacobi, F., Wittchen, H.-U., Hölting, C., Höfler, M., Pfister, H., Müller, N., & Lieb, R. (2004). 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*(4), 597–611. https://doi.org/10.1017/S0033291703001399

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/NIMG.2002.1132

Kaczkurkin, A. N., Moore, T. M., Calkins, M. E., Ciric, R., Detre, J. A., Elliott, M. A., … Satterthwaite, T. D. (2017). Common and dissociable regional cerebral blood flow differences associate with dimensions of psychopathology across categorical diagnoses. *Molecular Psychiatry*. https://doi.org/10.1038/mp.2017.174

Kaczkurkin, A. N., Moore, T. M., Ruparel, K., Ciric, R., Calkins, M. E., Shinohara, R. T., … Satterthwaite, T. D. (2016). Elevated Amygdala Perfusion Mediates Developmental Sex Differences in Trait Anxiety. *Biological Psychiatry*, *80*(10), 775–785. https://doi.org/10.1016/j.biopsych.2016.04.021

Kaiser, R. H., Andrews-Hanna, J. R., Wager, T. D., & Pizzagalli, D. A. (2015). Large-Scale Network Dysfunction in Major Depressive Disorder. *JAMA Psychiatry*, *72*(6), 603. https://doi.org/10.1001/jamapsychiatry.2015.0071

Kessler, R. C. (2003). Epidemiology of women and depression. *Journal of Affective Disorders*, *74*(1), 5–13. https://doi.org/10.1016/S0165-0327(02)00426-3

Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., … Parsey, R. V. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, *46*(3), 786–802. https://doi.org/10.1016/j.neuroimage.2008.12.037

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

Lee, S. H., Ripke, S., Neale, B. M., Faraone, S. V, Purcell, S. M., Perlis, R. H., … Wray, N. R. (2013). Genetic relationship between five psychiatric disorders

estimated from genome-wide SNPs. *Nature Genetics*, *45*(9), 984–94. https://doi.org/10.1038/ng.2711

Lefebvre, S., Demeulemeester, M., Leroy, A., Delmaire, C., Lopes, R., Pins, D., … Jardri, R. (2016). Network dynamics during the different stages of hallucinations in schizophrenia. *Human Brain Mapping*, *37*(7), 2571–2586. https://doi.org/10.1002/hbm.23197

Lefort-Besnard, J., Bassett, D. S., Smallwood, J., Margulies, D. S., Derntl, B., Gruber, O., … Bzdok, D. (2018). Different shades of default mode disturbance in schizophrenia: Subnodal covariance estimation in structure and function. *Human Brain Mapping*, *39*(2), 644–661. https://doi.org/10.1002/hbm.23870

Li, W., Douglas Ward, B., Liu, X., Chen, G., Jones, J. L., Antuono, P. G., … Goveas, J. S. (2014). 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. https://doi.org/10.1136/jnnp-2014-309180

Lynall, M.-E., Bassett, D. S., Kerwin, R., McKenna, P. J., Kitzbichler, M., Muller, U., & Bullmore, E. (2010). Functional connectivity and brain networks in schizophrenia. *The Journal of Neuroscience*, *30*(28), 9477–87. https://doi.org/10.1523/jneurosci.0333-10.2010

McTeague, L. M., Huemer, J., Carreon, D. M., Jiang, Y., Eickhoff, S. B., & Etkin, A. (2017). Identification of Common Neural Circuit Disruptions in Cognitive Control Across Psychiatric Disorders. *American Journal of Psychiatry*, *174*(7), 676–685. https://doi.org/10.1176/appi.ajp.2017.16040400

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

Merikangas, K. R., He, J., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., … Swendsen, J. (2010). Lifetime Prevalence of Mental Disorders in U.S. Adolescents: Results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(10), 980–989.

https://doi.org/10.1016/j.jaac.2010.05.017

Mišić, B., Betzel, R. F., de Reus, M. A., van den Heuvel, M. P., Berman, M. G., McIntosh, A. R., & Sporns, O. (2016). Network-Level Structure-Function Relationships in Human Neocortex. *Cerebral Cortex*, (April), bhw089. https://doi.org/10.1093/cercor/bhw089

Muschelli, J., Nebel, M. B., Caffo, B. S., Barber, A. D., Pekar, J. J., & Mostofsky, S. H. (2014). Reduction of motion-related artifacts in resting state fMRI using aCompCor. *NeuroImage*, *96*, 22–35. https://doi.org/10.1016/j.neuroimage.2014.03.028

Nassar, R., Kaczkurkin, A. N., Xia, C. H., Sotiras, A., Pehlivanova, M., Moore, T. M., … Satterthwaite, T. D. (2018). Gestational Age is Dimensionally Associated with Structural Brain Network Abnormalities Across Development. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhy091

Pankow, A., Deserno, L., Walter, M., Fydrich, T., Bermpohl, F., Schlagenhauf, F., & Heinz, A. (2015). Reduced default mode network connectivity in schizophrenia patients. *Schizophrenia Research*, *165*(1), 90–3. https://doi.org/10.1016/j.schres.2015.03.027

Parkes, L., Fulcher, B., Yücel, M., & Fornito, A. (2018). An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. *NeuroImage*, *171*, 415–436. https://doi.org/10.1016/J.NEUROIMAGE.2017.12.073

Paus, T. (2005). Mapping brain maturation and cognitive development during adolescence. *Trends in Cognitive Sciences*, *9*(2), 60–68. https://doi.org/10.1016/j.tics.2004.12.008

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., … Petersen, S. E. (2011). Functional Network Organization of the Human Brain. *Neuron*, *72*(4), 665–678. https://doi.org/10.1016/J.NEURON.2011.09.006

Power, J. D., Fair, D. A., Schlaggar, B. L., & Petersen, S. E. (2010). The Development of Human Functional Brain Networks. *Neuron*, *67*(5), 735–748. https://doi.org/10.1016/j.neuron.2010.08.017

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., … Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, *460*(August). https://doi.org/10.1038/nature08185

Rapoport, J. L., Giedd, J. N., & Gogtay, N. (2012). Neurodevelopmental model of schizophrenia: update 2012. *Molecular Psychiatry*, *17*(12), 1228–1238. https://doi.org/10.1038/mp.2012.23

Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(4), 1118–1123.

Rubinov, M., & Sporns, O. (2009). Complex network measures of brain connectivity : Uses and interpretations. *NeuroImage*, *52*(3), 1059–1069. https://doi.org/10.1016/j.neuroimage.2009.10.003

Rudie, J. D., Brown, J. A., Beck-Pancer, D., Hernandez, L. M., Dennis, E. L., Thompson, P. M., … Dapretto, M. (2013). Altered functional and structural brain network organization in autism. *NeuroImage: Clinical*, *2*(1), 79–94. https://doi.org/10.1016/j.nicl.2012.11.006

Satterthwaite, T. D., Elliott, M. A., Ruparel, K., Loughead, J., Prabhakaran, K., Calkins, M. E., … Gur, R. E. (2014). Neuroimaging of the Philadelphia Neurodevelopmental Cohort. *NeuroImage*, *86*(2014), 544–553. https://doi.org/10.1016/j.neuroimage.2013.07.064

Satterthwaite, T. D., Kable, J. W., Vandekar, L., Katchmar, N., Bassett, D. S., Baldassano, C. F., … Wolf, D. H. (2015). 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*(9), 1–11. https://doi.org/10.1038/npp.2015.75

Satterthwaite, T. D., Vandekar, S. N., Wolf, D. H., Bassett, D. S., Ruparel, K., Shehzad, Z., … Gur, R. E. (2015). Connectome-wide network analysis of youth with Psychosis-Spectrum symptoms. *Molecular Psychiatry*, *20*(February), 1–8. https://doi.org/10.1038/mp.2015.66

Satterthwaite, T. D., Wolf, D. H., Roalf, D. R., Ruparel, K., Erus, G., Vandekar, S., … Gur, R. C. (2015). Linked Sex Differences in Cognition and Functional Connectivity in Youth. *Cerebral Cortex*, *25*(9), 2383–2394. https://doi.org/10.1093/cercor/bhu036

Satterthwaite, T. D., Wolf, D. H., Ruparel, K., Erus, G., Elliott, M. A., Eickhoff, S. B., … Gur, R. C. (2013). Heterogeneous impact of motion on fundamental patterns of developmental changes in functional connectivity during youth. *NeuroImage*, *83*(2013), 45–57. https://doi.org/10.1016/j.neuroimage.2013.06.045

Shanmugan, S., Wolf, D. H., Calkins, M. E., Moore, T. M., Ruparel, K., Hopson, R. D., … Satterthwaite, T. D. (2016). Common and Dissociable Mechanisms of Executive System Dysfunction Across Psychiatric Disorders in Youth. *The American Journal of Psychiatry*, *173*(5), 517–526. https://doi.org/10.1176/appi.ajp.2015.15060725

Sharma, A., Wolf, D. H., Ciric, R., Kable, J. W., Moore, T. M., Vandekar, S. N., … Satterthwaite, T. D. (2017). Common Dimensional Reward Deficits Across Mood and Psychotic Disorders: A Connectome-Wide Association Study. *American Journal of Psychiatry*, (6), appi.ajp.2016.1. https://doi.org/10.1176/appi.ajp.2016.16070774

Sharp, D. J., Scott, G., & Leech, R. (2014). Network dysfunction after traumatic brain injury. *Nat Rev Neurol*, *10*(3), 156–166. https://doi.org/10.1038/nrneurol.2014.15

Sheline, Y. I., Barch, D. M., Price, J. L., Rundle, M. M., Vaishnavi, S. N., Snyder, A. Z., … Raichle, M. E. (2009). The default mode network and self-referential processes in depression. *Proceedings of the National Academy of Sciences*, *106*(6), 1942–1947. https://doi.org/10.1073/pnas.0812686106

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

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

Sylvester, C. M., Corbetta, M., Raichle, M. E., Rodebaugh, T. L., Schlaggar, B. L., Sheline, Y. I., … Lenze, E. J. (2012). Functional network dysfunction in anxiety and anxiety disorders. *Trends in Neurosciences*, *35*(9), 527–535. https://doi.org/10.1016/j.tins.2012.04.012

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–20. https://doi.org/10.1109/TMI.2010.2046908

Tustison, N. J., Cook, P. A., Klein, A., Song, G., Das, S. R., Duda, J. T., … Avants, B. B. (2014). Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *NeuroImage*, *99*, 166–79. https://doi.org/10.1016/j.neuroimage.2014.05.044

Wang, J., Vachet, C., Rumple, A., Gouttard, S., Ouziel, C., Perrot, E., … Styner, M. (2014). Multi-atlas segmentation of subcortical brain structures via the AutoSeg software pipeline. *Frontiers in Neuroinformatics*, *8*, 7. https://doi.org/10.3389/fninf.2014.00007

Whitfield-Gabrieli, S., & Ford, J. M. (2012). Default Mode Network Activity and Connectivity in Psychopathology. *Annual Review of Clinical Psychology, Vol 8*, *8*, 49–75. https://doi.org/Doi 10.1146/Annurev-Clinpsy-032511-143049

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

Woodward, N. D., & Cascio, C. J. (2015). Resting-state functional connectivity in psychiatric disorders. *JAMA Psychiatry*, *72*(8), 743–744. https://doi.org/10.1016/B978-0-12-386043-9.00005-0.New

Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics. *PLoS ONE*, *8*(7). https://doi.org/10.1371/journal.pone.0068910

Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., … Buckner, R. L. (2011). The organization of the human

cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*, 1125–1165. https://doi.org/10.1152/jn.00338.2011.

Zhou, Y., Liang, M., Tian, L., Wang, K., Hao, Y., Liu, H., … Jiang, T. (2007). Functional disintegration in paranoid schizophrenia using resting-state fMRI. *Schizophrenia Research*, *97*(1–3), 194–205. https://doi.org/10.1016/j.schres.2007.05.029

CHAPTER 3

Multi-Scale Network Regression for Brain-Phenotype Associations

This chapter has been submitted for publication:

Xia C.H., Ma Z., Cui Z., Bzdok D., Bassett D.S., Satterthwaite T.D., Shinohara R.T.*, Witten D.*. 2019 “Multi-Scale Network Regression for Brain-Phenotype Associations”. *(under review)*

Abstract

Complex brain networks are increasingly characterized at different scales, including global summary statistics, community connectivity, and individual edges. While research relating brain networks to demographic and behavioral measurements have yielded many insights into brain-phenotype relationships, common analytical approaches only consider network information at a single scale, thus failing to incorporate rich information present at other scales. Here, we designed, implemented, and deployed Multi-Scale Network Regression (MSNR), a penalized multivariate approach for modeling brain networks that explicitly respects both edge- and community-level information by assuming a low rank and sparse structure, both encouraging less complex and more interpretable modeling. Capitalizing on a large neuroimaging cohort ($n = 1051$), we demonstrate that MSNR recapitulates interpretable and statistically significant connectivity patterns associated with brain development, sex differences, and motion-related artifacts. Notably, compared to single-scale methods, MSNR achieves a balance between out-of-sample prediction and model interpretability. Together, by jointly exploiting both edge- and community-level information, MSNR has the potential to yield novel insights into brain-behavior relationships.

Introduction

Studying brain-phenotype relationships in high-dimensional connectomics is an active area of research in the neuroscience community (Bassett & Sporns, 2017; Bullmore & Sporns, 2009). The advent of large neuroimaging datasets that have measures of brain connectivity for unprecedented numbers of subjects (Biswal et al., 2010; Bzdok & Yeo, 2017; Jernigan et al., 2016; Mennes, Biswal, Castellanos, & Milham, 2013; Satterthwaite et al., 2014; Schumann et al., 2010; Van Essen et al., 2012) have yielded novel insights into brain development (Fair et al., 2007; Power, Fair, Schlaggar, & Petersen, 2010; Satterthwaite et al., 2013), sex differences (Gur & Gur, 2016; Ingalhalikar et al., 2014; Dardo Tomasi & Volkow, 2012) neurological diseases (Buckner et al., 2009; Khambhati, Davis, Lucas, Litt, & Bassett, 2016) and psychiatric illnesses (Bassett, Xia, & Satterthwaite, 2018; Drysdale et al., 2016). As the availability of datasets with rich neural, genetic, and behavioral measurements from large numbers of subjects continues to increase, there is a growing need for statistical methods that are tailored for the discovery of complex relationships between brain networks and phenotypes (Craddock, Tungaraza, & Milham, 2015; Varoquaux & Craddock, 2013).

A typical brain network consists of hundreds of nodes, which denote anatomical brain regions and tens of thousands of edges, which indicate connections between pairs of nodes (Rubinov & Sporns, 2009).The network can be viewed on the *micro-scale*, *meso-scale*, or *macro-scale*. The set of edges that comprise the network make up the *micro-scale*. The *macro-scale* includes the

network's modularity, characteristic path, global efficiency, and other global summary statistics (Rubinov & Sporns, 2009). The *meso-scale* falls in between the micro-scale and macro-scale, and includes the communities that make up the network (Sporns & Betzel, 2016). A community refers to a collection of nodes that are highly connected to each other and have little connection to nodes in other communities. Prior work has demonstrated that brain network architecture present on these different scales is associated with development (Gu et al., 2015; Power et al., 2010), aging (Betzel et al., 2014; Damoiseaux et al., 2008; D Tomasi & Volkow, 2012), learning (Bassett, Yang, Wymbs, & Grafton, 2015; Jarosiewicz et al., 2008; Lewis, Baldassarre, Committeri, Romani, & Corbetta, 2009), cognition (Bressler & Menon, 2010; Crossley et al., 2013; Park & Friston, 2013), and neuropsychiatric diseases (Alexander-Bloch et al., 2010; Bassett et al., 2018; Fornito, Zalesky, & Breakspear, 2015; Xia et al., 2018).

Despite increased appreciation that multi-scale organization of the brain may be responsible for some of its major functions (Bassett & Siebenhühner, 2013; Betzel & Bassett, 2017), thus far, common strategies for studying the relationship between brain connectivity and phenotypes consider network features at a single scale (Craddock et al., 2015). For example, a popular single-scale strategy focuses on group-level comparisons of individual connections (i.e. edges) in brain networks (Craddock et al., 2015; Varoquaux & Craddock, 2013). This approach involves performing a statistical test on each edge. While this procedure is easy to implement, several drawbacks limit its effectiveness. Chief among these limitations are the need to account for multiple comparisons, and

a lack of interpretability (Craddock et al., 2015; Varoquaux & Craddock, 2013). To achieve high power while controlling false discovery, alternative edge-based methods have been developed, such as the network-based statistic (Zalesky, Fornito, & Bullmore, 2010) and multivariate distance matrix regression (Zapala & Schork, 2012). While these strategies have yielded important insights, they nonetheless focus exclusively on the micro-scale, often producing results that are difficult to interpret and that do not exploit the multi-scale information present in the brain networks.

Given the importance of community structure in brain networks and their readily interpretable characteristics (Betzel, Medaglia, & Bassett, 2018), it might be tempting to conduct a mass-univariate analysis at the meso-scale, considering *within*- and *between*-community connectivity as the input features. Such an approach dramatically reduces the dimensionality of the data, which in turn decreases the burden of multiple comparisons correction. A community-based approach also has the added benefit of not having to vectorize the connectivity matrix, as in an edge-based approach, which inevitably disrupts the innate structure in the data. However, summarizing hundreds or thousands of edges as one single number to represent the connection within or between brain communities can be problematic, especially for large communities such as the default mode network (Power et al., 2011), whose edges are spatially distributed across the anterior and posterior portions of the brain (Raichle, 2015). Stated another way, extracting the mean connectivity at the community level risks mixing disparate signals.

In this paper, we introduce *Multi-Scale Network Regression (MSNR)*, which simultaneously incorporates information across multiple scales in order to reveal associations between high-dimensional connectomic data and phenotypes of interest. We first describe the MSNR model and introduce an algorithm to fit it. Next, we capitalize on one of the largest neurodevelopmental imaging cohorts, the Philadelphia Neurodevelopmental Cohort (PNC), to empirically test MSNR's ability in delineating brain connectivity patterns that are associated with a wide variety of phenotypes. Importantly, we conduct head-to-head comparisons between MSNR and common single-scale strategies (both edge- and community-based), and show that MSNR achieves a balance between prediction performance and interpretability by considering information at multiple network scales.

Statistical Methodology

A STATISTICAL MODEL FOR MULTI-SCALE NETWORK REGRESSION

Given n subjects, let $A^1, \dots, A^n \in R^{p \times p}$ denote the symmetric adjacency matrices corresponding to their functional connectivity networks, where p is the number of nodes. For instance, $A^i_{jj'}$ could represent the Pearson correlation, a common measure of functional connectivity, between the jth and j'th nodes for the ith subject. Furthermore, we assume that the p nodes can be partitioned into K distinct communities $C_1, \dots, C_K$ that are known a priori: $\cup_{k=1}^{K} C_k = \{1, \dots, p\}, C_k \cap C_{k'} \neq \emptyset$ if $k \neq k'$. The notation $j \in C_k$ indicates that the jth node is in the kth community. If $K = p$, then the community structure is trivial, in the sense that each node belongs to its own community. Moreover, for each subject, $\boldsymbol{q}$ covariates have been measured, so that $\boldsymbol{X_i} = \left(\boldsymbol{X_i^1}\ \boldsymbol{X_i^2}\ \dots\ \boldsymbol{X_i^q}\right)^{\boldsymbol{T}} \in \mathbb{R}^{\boldsymbol{q}}$ is a covariate vector for the $\boldsymbol{i}$th subject, $\boldsymbol{i} = \boldsymbol{1}, \dots, \boldsymbol{n}$.

In what follows, we consider the model

$$A^i_{jj'} = \Theta_{jj'} + \sum_{f=1}^{q} X_i^f \cdot \left(\sum_{k=1}^{K} \sum_{k'=1}^{K'} \Gamma^f_{kk'} 1_{(j \in C_k, j' \in C_{k'})} \right) + \epsilon^i_{jj'}, i = 1, \dots, n, j, j' = 1, \dots, p, \qquad (1)$$

Where $\epsilon^i_{jj'}$ is a mean-zero noise term, and $\epsilon^i_{jj'} = \epsilon^i_{j'j}$. Θ is a symmetric $p \times p$ matrix that summarizes the mean connectivity, across all of the subjects, of each pair nodes, in the absence of covariates. Finally, for $f = 1, \dots, q$, Γ^f is a symmetric $K \times K$ matrix that quantifies the association between the fth feature and the functional connectivity between each pair of communities. For instance,

a one-unit increase in X_i^f is associated with a $\Gamma_{kk'}^f$ increase in the mean functional connectivity between nodes in the kth and k'th communities.

We now define a $p \times K$ matrix W for which $W_{jk} = 1_{(j \in C_k)}$, where $1_{(\cdot)}$ denotes an indicator variable. Then (1) can be re-written as

$$A^i = \Theta + \sum_{f=1}^{q} X_i^f \cdot (W \Gamma^f W^T) + \epsilon^i, \qquad i = 1, \dots, n. \tag{2}$$

In order to fit the model (2), we make two assumptions about the structures of the unknown parameter matrices Θ and $\Gamma^1, \dots, \Gamma^q$.

$Assumption\ 1$: Θ *has low rank* (Leonardi et al., 2013; K. Li, Guo, Nie, Li, & Liu, 2009; Smith et al., 2015). That is, $\Theta = VV^T$ where V is a $p \times d$ matrix, for a small positive constant d. This means that the p nodes effectively reside in a reduced subspace of d dimensions. The mean connectivity between any pair of nodes is simply given by their inner product in this low-dimensional subspace.

$Assumption\ 2$: $\Gamma^1, \dots, \Gamma^q$ *are sparse* (Meunier, Lambiotte, & Bullmore, 2010; Newman, 2006; Xia et al., 2018). That is, most of their elements are *exactly* equal to zero. If $\Gamma_{kk'}^f = 0$, then the value of the fth feature is unassociated with the mean connectivity between nodes in the kth and k'th communities. We note that *Assumption 1* is closely related to the random dot product graph model and related models (Durante & Dunson, 2018; Durante, Dunson, & Vogelstein, 2017; Fosdick & Hoff, 2015; Tang, Athreya, Sussman,

Lyzinski, & Priebe, 2017; Young & Scheinerman, 2007), whereas *Assumption 2* is a standard sparsity assumption for high-dimensional regression (Hastie, Tibshirani, & Friedman, 2008; Hastie, Tibshirani, Wainwright, Tibshirani, & Wainwright, 2015; Tibshirani, 1996). Under these two assumptions, a schematic of the model (2) can be seen in **Figure 3-1**.

Model (2) is closely related to both the stochastic block model (D. S. Choi, Wolfe, & Airoldi, 2012) and the random dot product graph model (Young & Scheinerman, 2007). In particular, if $\Theta = 0, q = 1$, and $X_i^1 = 1$ for $i = 1, \dots, n$, then (2) reduces to a stochastic block model with known communities $C_1, \dots, C_K$. And if $\Gamma^1 = \cdots = \Gamma^q = 0$ and *Assumption 1* holds, then (2) reduces to a random dot product graph model. However, unlike those two models (2) explicitly allows for the mean of the adjacency matrix to be a function of covariates, and effectively incorporates both edge- and community-level network information.

OPTIMIZATION PROBLEM

We now consider the task of fitting the model (2), under *Assumptions 1* and *2*. It is natural to consider the optimization problem

$$\left\{ \sum_{i=1}^{n} \left\| A^i - \left(\Theta + \sum_{f=1}^{q} X_i^f \cdot (W \Gamma^f W^T) \right) \right\|_F^2 + \lambda_1 rank(\Theta) + \lambda_2 \sum_{f=1}^{q} \left\| \Gamma^f \right\|_0 \right\}, \quad (3)$$

Where the notation $\| \cdot \|_F^2$ indicates the squared *Frobenius* norm of a matrix, i.e. $\|D\|_F^2 = \sum_{j=1}^{p} \sum_{j'}^{p} D_{jj'}^2$, and the notatiton $\| \cdot \|_0$ indicates the element-

wise cardinality (or l_0 norm) of a matrix, i.e. $\|D\|_0 = \sum_{j=1}^{p}\sum_{j'=1}^{p} 1_{(D_{jj'}\neq 0)}$. In (3), λ_1 and λ_2 are non-negative tuning parameter values that control the rank of Θ and the sparsity of $\Gamma^1 = \cdots = \Gamma^q$, respectively.

Unfortunately, due to the presence of the rank and l_0 penalties, the optimization problem (3) is highly non-convex, and no efficient algorithms are available to solve it. Therefore, in what follows, we will consider an alternative to (3), which results from replacing the non-convex rank and l_0 penalties in (3) with their convex relaxations. This leads to the optimization problem

$$\underset{\Theta,\,\Gamma^1,\ldots,\Gamma^q}{\text{minimize}}\left\{\sum_{i=1}^{n}\left\|A^i-\left(\Theta+\sum_{f=1}^{q}X_i^f\cdot(W\Gamma^fW^T)\right)\right\|_F^2+\lambda_1\|\Theta\|_*+\lambda_2\sum_{f=1}^{q}\left\|\Gamma^f\right\|_1\right\}. \quad (4)$$

In (4), the notation $\|\cdot\|_*$ indicates the *nuclear norm* of a matrix, i.e. the sum of its singular values (Bien & Witten, 2016; Fazel, 2002; Recht, Fazel, & Parrilo, 2010). The nuclear norm is a convex surrogate for the rank of a matrix. The notation $\|\cdot\|_1$ indicates the element-wise l_1 (or *lasso)* norm of a matrix, i.e. $\|D\|_1 = \sum_{j=1}^{p}\sum_{j'=1}^{p}\left|D_{jj'}\right|$; this is a convex relaxation of the l_0 norm (Hastie, 2015; Hastie et al., 2008; Tibshirani, 1996). In (4), the non-negative tuning parameters λ_1 and λ_2 encourage Θ and $\Gamma^1, \ldots, \Gamma^q$ to be low-rank and sparse, respectively.

Importantly, the optimization problem (4) is convex, and so fast algorithms are available to solve it for the global optimum. In the next section, we derive a block coordinate descent algorithm for solving (4). Simulation studies

indicated that MSNR behaved in the manner that was dependent on the signal-to-noise ratio and the observation-to-feature ratio, particularly in its ability to model underlying connectivity patterns (see Supplementary Information).

BLOCK COORDINATE DESCENT ALGORITHM TO SOLVE (4)

We now derive a block coordinate descent algorithm for solving (4) (Bien & Witten, 2016; Friedman, Hastie, Höfling, & Tibshirani, 2007; Hastie et al., 2008; Tseng, 2001). Roughly speaking, we will cycle through the parameters Θ, $\Gamma^1 = \cdots = \Gamma^q$, and minimize the objective (4) with respect to each one in turn, holding all others fixed. Because the loss function is differentiable and the penalties are separable with respect to each block of parameters, this approach is guaranteed to yield the global optimum. The algorithm is as follows:

1. Initialize a $p \times p$ matrix $\widehat{\Theta}$, and $K \times K$ matrices $\hat{\Gamma}^1, \ldots, \hat{\Gamma}^q$.
2. Iterate until convergence:
   a. Update Θ by minimizing (4) with respect to Θ, holding $\hat{\Gamma}^1, \ldots, \hat{\Gamma}^q$ fixed:

$$\widehat{\Theta} \leftarrow \left\{ \sum_{i=1}^{n} \left\| A^i - \left(\Theta + \sum_{f=1}^{q} X_i^f \cdot \left(W \hat{\Gamma}^f W^T \right) \right) \right\|_F^2 + \lambda_1 \|\Theta\|_* \right\}. \quad (5)$$

   b. For $f = 1, \ldots, q$, update Γ^f by minimizing (4) with respect to Γ^f, holding $\widehat{\Theta}$ and $\hat{\Gamma}^1, \ldots, \hat{\Gamma}^{f-1}, \hat{\Gamma}^{f+1}, \ldots, \hat{\Gamma}^q$ fixed:

$$\hat{\Gamma}^f \leftarrow \left\{ \sum_{i=1}^{n} \left\| A^i - \left(\widehat{\Theta} + \sum_{f' \neq f} X_i^{f'} \cdot \left(W \hat{\Gamma}^{f'} W^T \right) + X_i^f \cdot \left(W \Gamma^f W^T \right) \right) \right\|_F^2 + \lambda_2 \left\| \Gamma^f \right\|_1 \right\}. \quad (6)$$

Both (5) and (6) are convex optimization problems, for which closed form solutions are available, as detailed in the following propositions. These propositions make use of the soft-thresholding operator, defined as

$$S(a, b) = (|a| - b, 0)\, sign(a), \quad (7)$$

and applied element-wise to the elements of a matrix.

Proposition 1. Define

$$\widetilde{\mathrm{A}}^{\mathrm{i}} \equiv \mathrm{A}^{\mathrm{i}} - \sum_{\mathrm{f}=1}^{\mathrm{q}} \mathrm{X}_{\mathrm{i}}^{\mathrm{f}} \cdot \left(\mathrm{W} \hat{\Gamma}^{\mathrm{f}} \mathrm{W}^{\mathrm{T}} \right),$$

and let $\mathrm{UDV}^{\mathrm{T}}$denote the singular value decomposition of $\frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \widetilde{\mathrm{A}}^{\mathrm{i}}$: that is, $\frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \widetilde{\mathrm{A}}^{\mathrm{i}} = \mathrm{UDV}^{\mathrm{T}}$, where U and V are $\mathrm{p} \times \mathrm{p}$ matrices, $\mathrm{U}^{\mathrm{T}}\mathrm{U} = \mathrm{UU}^{\mathrm{T}} = \mathrm{V}^{\mathrm{T}}\mathrm{V} = \mathrm{VV}^{\mathrm{T}} = \mathrm{I}$, and D is a diagonal matrix with non-negative elements on the diagonal elements on the diagonal. Then, the solution to the optimization problem (5) is

$$\widehat{\Theta} = \mathrm{US}\left(\mathrm{D}, \frac{\lambda_1}{2\mathrm{n}} \right) \mathrm{V}^{\mathrm{T}},$$

where the soft-thresholding operator defined in (7) is applied element-wise.

Let $\mathrm{p_k} \equiv |\mathrm{C_k}|$, the cardinality of the kth community; note that $\sum_{\mathrm{k}=1}^{\mathrm{K}} \mathrm{p_k} = \mathrm{p}$.

Proposition 2. Let W_j denote the jth row of the matrix W. For $f = 1, \ldots, q$, define

$$\overline{A}_{jj'} \equiv A^{i}_{jj'} - \overline{\Theta}_{jj'} - \sum_{f' \neq f} X_i^{f'} \cdot \left(W_j^T \hat{\Gamma}^{f'} W_{j'} \right),$$

$$\tilde{y}^{f}_{kk'} \equiv \frac{\sum_{j \in C_k} \sum_{j' \in C_k} \sum_{i=1}^{n} \overline{A}_{jj'} \cdot X_i^f}{\sum_{i=1}^{n} \left(X_i^f \right)^2 p_k p_{k'}},$$

and

$$\tilde{\lambda}^{f}_{kk'} \equiv \frac{\lambda_2}{\sum_{i=1}^{n} \left(X_i^f \right)^2 p_k p_{k'}} .$$

Then, the solution to the optimization problem (6) is of the form

$$\tilde{\Gamma}^{f}_{kk'} \equiv S\left(\tilde{y}^{f}_{kk'}, \frac{\tilde{\lambda}^{f}_{kk'}}{2} \right).$$

Proofs of Propositions 1 and 2 are provided in the Appendix.

CODE AVAILABILITY

An implementation of the algorithm described above is available in R at `bitbucker.org/rshinohara/networkregression`.

Methods

PHILADELPHIA NEURODEVELOPMENTAL COHORT

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 (Satterthwaite et al., 2016, 2014). In total, 1601 participants completed the cross-sectional neuroimaging protocol. Of these participants, 154 were excluded for meeting any of the following criteria: gross radiological abnormalities, history of medical problems that might affect brain function, history of inpatient psychiatric hospitalization, use of psychoactive medications at the time of data acquisition. Of the remaining 1447 participants, 51 were excluded for low quality or incomplete FreeSurfer reconstruction of T1-weighted images. Of the remaining 1396 participants, 381 were excluded for missing rs-fMRI, voxelwise coverage or excessive motion, which is defined as having an average framewise motion more than 0.20mm and more than 20 frames exhibiting over 0.25mm movement (using calculation from Jenkinson, Bannister, Brady, & Smith, 2002). These exclusions produced a final sample consisting of 1015 youths (mean age 15.78, SD = 3.34; 461 males and 554 females).

IMAGING 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 (Satterthwaite et al., 2016, 2014). 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 $3\ mm$ isotropic; 124 volumes). Prior to scanning, in order to acclimatize 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 PRE-PROCESSING

A study-specific template was generated from a sample of 120 PNC subjects balanced across sex, race, and age using the `buildTemplateParallel` procedure in ANTs (Avants, Tustison, Song, et al., 2011). Study-specific tissue priors were created using a multi-atlas segmentation procedure (Wang et al., 2013). Next, each subject's high-resolution structural image was processed using the ANTs Cortical Thickness Pipeline (Tustison et al., 2014). Following bias field correction (Tustison et al., 2010), each structural image was diffeomorphically registered to the study-specific PNC template using the top-performing `SyN` deformation (Klein et al., 2009). Study-specific tissue priors were used to guide brain extraction and segmentation of the subject's structural image (Avants, Tustison, Wu, Cook, & Gee, 2011).

FUNCTIONAL PRE-PROCESSING

Task-free functional images were processed using the XCP Engine (Ciric et al., 2018, 2017a), which was configured to execute a top-performing pipeline for removal of motion-related variance (Ciric et al., 2018). 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` (Jenkinson et al., 2002) (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

(Greve & Fischl, 2009). Confounding signals in the data were 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 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 (Hallquist, Hwang, & Luna, 2013).

NETWORK CONSTRUCTION

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. Connectomes were computed across all regions within a common parcellation with 264 nodes and 13 communities (Power et al., 2011). We excluded 28 nodes that were not sorted into any community, therefore resulting in the final $p = 236$ and $K = 13$ (**Figure 3-1a**). The *a priori* community structure for this set of nodes was delineated using the Infomap algorithm (Rosvall & Bergstrom, 2008) and were replicated in an independent sample. This parcellation was selected for our analysis as it has been previously used for studying individual differences in brain connectivity, including those related to brain development (Gu et al., 2015; Satterthwaite et al., 2013), sex differences (Satterthwaite et al., 2015), and in-scanner motion (Ciric et al., 2018).

CROSS-VALIDATION

We first randomly selected 20% of the total sample ($n = 1015$) to serve as the left-out validation set ($n = 202$). We then performed five-fold cross validation on the remaining 80% of the sample ($n = 813$) in order to select the values of the tuning parameters λ_1 and λ_2 for MSNR (James, Witten, Hastie, & Tibshirani, 2013, **Figure 3-2b**). In each fold, the independent variables ($X_{n\times q}$) were centered to a mean of zero and scaled by each column's standard deviation.

The prediction error used in cross-validation was the *Frobenius* norm of the difference between estimated and true connectivity matrices in the test set, $\|A^i - \hat{A}^i\|_F^2$ (**Figure 3-2c**). We ensured the prediction error was sample size independent by using the average prediction error over all subjects in the test set.

PERMUTATION PROCEDURE

To estimate the distribution of prediction error under the null hypothesis of no association between functional connectivity and phenotype, we permuted the rows of the covariate matrix $X_{n\times q}$. For each permutation, we tuned λ_1 and λ_2 using cross-validation, and calculated the prediction error in the left-out validation set. The p-value was defined to be the proportion of prediction errors among the 1,000 permuted datasets that are smaller than the prediction error on the observed data,

$$p_{permutation} = \frac{\sum_1^{1000} 1_{e_1 \leq \bar{e}}}{1000}, \quad (8)$$

where $e_1, \ldots, e_{1000}$ denote the prediction errors on the 1,000 permuted data sets, and $\bar{e}$ denotes the prediction error on the original data. Here, $1_{(A)}$ is an indicator variable that equals 1 if the event A, and 0 otherwise.

COMPARISON TO SINGLE-SCALE APPROACHES

We compared the performance of MSNR to two single-scale network regression strategies, namely individual edge model (Grillon et al., 2013; Lewis et al., 2009) and community mean model (Betzel et al., 2014; King et al., 2018; Yan et al., 2019; Yu et al., 2019). These two approaches have been commonly used to study connectivity-phenotype relationships (Craddock et al., 2015; Varoquaux & Craddock, 2013) and differ primarily in terms of the scale of brain network examined (**Figure 3-3**). Details are as follows:

Individual edge model. We vectorized the upper triangle of the adjacency matrix A^i for the ith subject, $i = 1, \ldots, n$, in order to create a $n \times p(p-1)/2$ matrix. For each of the $p(p-1)/2$ columns of this matrix, we fit a linear regression to model that column using three covariates: age, sex, and in-scanner motion (**Figure 3-3a**). Specifically, we built a linear model for each edge using `mgcv` package in R, with the formula `edge ~ age + sex + motion` (Wood, 2017, **Figure 3-3b**). This included a penalization on roughness, and we estimated the penalty parameter by recasting the problem as a mixed effect model and estimating this via restricted maximum likelihood or REML (Wood, 2011; Wood,

Pya, & Säfken, 2016). We corrected the results for multiple comparisons using the false discovery rate (FDR, $q < 0.05$, Storey, 2002) and reshaped the $p(p-1)/2$ columns to a $p \times p$ matrix for visualizing significant coefficients. For calculating out-of-sample prediction error, we used linear models fit for all edges. The prediction error was calculated in the same way as in MSNR.

Community mean model. Community-based linear models were built with mean *within*- and *between*-community connectivity as the dependent variables. The within-community connectivity is defined as

$$\frac{\sum_{j,j' \in C_k} A^i_{jj'}}{|C_k| \times |C_k - 1|}, \tag{9}$$

where $A^i_{jj'}$ is the weighted edge strength between the node j and node j' both of which belong to the same community C_k, for the ith subject. The cardinality of the community assignment vector, $|C_k|$, represents the number of nodes in the kth community. The between-community connectivity is defined as

$$\frac{\sum_{j \in C_k, j' \in C_{k'}} A^i_{jj'}}{|C_k| \times |C_{k'}|}. \tag{10}$$

Here, C_k and $C_{k'}$ represent two different communities, and $|C_k|$ and $\left|C_{k'}\right|$ are the number of nodes in each community, respectively.

By applying (9) and (10) to each subject, we created a $n \times [\frac{K(K-1)}{2} + K]$ matrix. For each of the $\frac{K(K-1)}{2} + K$ columns of this matric, we fit a linear model to predict that column using three covariates: age, sex, and in-scanner motion.

Similar to the edge-based model, we built a linear model for each edge using `mgcv` package in R, with the formula `community ~ age + sex + motion` (Wood, 2017) and roughness penalty estimation by REML (Wood, 2011; Wood et al., 2016, **Figure 3-3b**). We corrected the results for multiple comparisons using the false discovery rate (FDR, $q < 0.05$, Storey, 2002) and reshaped the $\frac{K(K-1)}{2} + K$ columns to a $K \times K$ matrix for visualizing significant coefficients. For calculating out-of-sample prediction eorr, we used linear models fit for all communities. The prediction error was calculated in the same way as in MSNR.

Results

MSNR SHOWS HIGH ACCURACY IN A LARGE DEVELOPMENTAL SAMPLE

We applied MSNR to data from the Philadelphia Neurodevelopmental Cohort (PNC) (Satterthwaite et al., 2016, 2014) in order to uncover meaningful brain-phenotype relationships. In total, we studied $n = 1015$ participants aged 8-22, who completed resting state functional neuroimaging as part of the PNC. We constructed functional connectivity matrices from a commonly-used parcellation scheme ($p = 236$ nodes) and community membership assignment ($K = 13$ communities) (Power et al., 2011, **Figure 3-2a**). We first randomly selected 20% of the total sample as the left-out validation set ($n = 202$), with which we assessed the prediction performance of all subsequent models **Figure 3-2b**). The prediction performance was defined as the *Frobenius* norm of the difference between the observed and estimated adjacency matrices in the validation set (**Figure 3-2c**). For this proof-of-concept empirical study, we examined the association of functional connectivity with age, sex, and in-scanner motion. On the remaining 80% of the observations, we selected tuning parameters, λ_1 and λ_2, through five-fold cross-validation (**Figure 3-2b**). We iteratively refined the cross-validation grid (**Figure 3-4a-c**) in order to obtain the best possible tuning parameter values. Importantly, no boundary effect was observed in any of the iterations during successive grid searches, revealing a smooth convex landscape for the objective (**Figure 3-4d**).

We subsequently evaluated the model's out-of-sample prediction error on the 20% of observations that made up the left-out validation set. The prediction error on the validation set was comparable to the average error in the cross-validation procedure (**Figure 3-4e**). In addition, we performed a permutation test to compare the model's prediction error to the distribution of prediction error under the null hypothesis of no association between brain networks and the predictors (**Figure 3-4e**), which we estimated by permuting the rows of the covariate data matrix. This procedure disrupted the linkage between functional connectivity and phenotypes, while preserving the covariance structure of the covariates. For each permutation, we repeated the process of selecting tuning parameter values by cross-validation, fitting an MSNR model on the training set, and calculating its prediction error on the validation set. Out of 1,000 permutations, no out-of-sample prediction error was lower than that of the MSNR model built using the original data, indicating that the multivariate model had significantly better prediction performance ($p < 0.001$).

MSNR RECAPITULATES KNOWN INDIVIDUAL DIFFERENCES IN FUNCTIONAL CONNECTIVITY

The connectivity-phenotype relationships are summarized in the matrices Γ^1, Γ^2, and Γ^3 in the MSNR model. We counted the number of positive and negative coefficients within each estimated matrix; these represent, respectively, positive and negative associations between community membership and age, sex, and in-scanner motion (**Figure 3-5**). Consistent with the previous literature

(Satterthwaite et al., 2013), we found that as age increased, there were more within-community, rather than between-community connectivity, that strengthened over age. (**Figure 3-5a**). Conversely, as age increased, there were more between-community, rather than within-community connectivity, that weakened over age. This pattern suggests that functional brain networks tend to segregate during normative brain development. Replicating findings from a previous report which evaluated a different subsample of PNC data using mass-univariate analyses (Satterthwaite et al., 2015), here we observed that stronger within-community connectivity, rather than between-community, was more representative of functional brain networks in males; whereas stronger between-community connectivity, rather than within-community, was more representative of functional brain networks in females (**Figure 3-5b**). Finally, following on prior studies, we evaluated the degree to which the association between in-scanner motion and connectivity varies by inter-node distance, defined as the Euclidean distance between two spherical brain parcellations in the MNI space (Brett, Johnsrude, & Owen, 2002, **Figure 3-5c**). As expected, the MSNR coefficients for in-scanner motion in relation to functional connectivity were negatively correlated with the distances between pairs of communities. In other words, when two brain regions are close together, the presence of in-scanner motion typically is associated with an increase in their connectivity. This is consistent with prior reports that in-scanner motion induces a distance-dependent bias in estimation of functional connectivity (Ciric et al., 2017b; Satterthwaite et al., 2012).

COMPARISON WITH TYPICAL MASS-UNIVARIATE SINGLE-SCALE STRATEGIES

Next, we compared MSNR to common single-scale mass-univariate approaches that make use of linear models at the edge-level or the community-level (**Figure 3-6**). We computed the out-of-sample performances of the two single-scale approaches using the left-out validation set. The prediction error of the community-based model on the left-out validation set was poor, whereas the prediction error of the edge-based model was similar to that of MSNR (**Figure 3-6a**). Given that not all models built in the mass-univariate analyses were significant, our estimation of prediction error for edge- and community-based models were likely to be overly optimistic since we used all fitted models for the purpose of out-of-sample prediction.

Next, we examined the interpretability of coefficients obtained in each model after applying FDR correction to control for multiple comparisons in single-scale approaches (Storey, 2002). We found that while the edge-based model and MSNR achieved similar out-of-sample prediction, coefficients estimated in MSNR (**Figure 3-6b**) were more interpretable than those from edge-based models (**Figure 3-6c**). The number of coefficients in edge-based models for each covariate exceeded that of MSNR by three orders of magnitude. On the other hand, at the expense of low out-of-sample prediction performance, community-based models exhibited similar interpretability as MSNR (**Figure 3-6d**).

Discussion

In the past decade, the neuroscience community has shifted away from studying localized regions of the brain towards studying inter-regional relationships, or connectivity (Bassett & Sporns, 2017; Rubinov & Sporns, 2009). The association of connectivity network architecture with learning and memory (Solomon et al., 2017), decision-making (Neubert, Mars, Sallet, & Rushworth, 2015), development and aging throughout the lifespan (Baum et al., 2016; Betzel et al., 2014; Fair et al., 2007; Power et al., 2010), and neuropsychiatric disorders (Bassett et al., 2018; van den Heuvel & Fornito, 2014; Xia et al., 2018) is of profound interest to the burgeoning network neuroscience literature, and can be studied on the scale of individual edges (*micro*-scale), communities (*meso*-scale), or the network as a whole (*macro*-scale) (Betzel & Bassett, 2017). Most existing approaches for analyzing networks, such as mass-univariate analyses, operate on a single scale (Betzel et al., 2014; Grillon et al., 2013; King et al., 2018; Lewis et al., 2009; Yan et al., 2019; Yu et al., 2019).

In recent years, interest has centered on multi-scale modeling approaches (Breakspear & Stam, 2005; Jenatton et al., 2012; Y. Li et al., 2013, 2011), which aim to integrate information across homogeneous regions in the brain while still modeling data on finer scales. These methods have mainly focused on the problem of smoothing without prior knowledge of anatomical or functional parcellations of the brain, and have been adapted for both classification (H. Choi & Baraniuk, 2001; Romberg, Hyeokho Choi, Baraniuk, & Kingbury, 2000) and

regression (Y. Li et al., 2011) as well as in longitudinal settings (Y. Li et al., 2013).

Building upon this recent work, we developed multi-scale network regression (MSNR) to study relationships between high-dimensional brain networks and variables of interest. Specifically, our proposal models the adjacency matrix for each observation by integrating both *micro*- and *meso*-scale information. By applying a low-rank assumption to the mean functional connectivity network (Leonardi et al., 2013; K. Li et al., 2009; Smith et al., 2015) and a sparsity assumption to the community-level network (Crossley et al., 2013; Meunier et al., 2010; Newman, 2006), we substantially decrease the number of parameters and encourage interpretable brain-phenotype relationships.

Leveraging a large neuroimaging dataset of over one thousand youths, we demonstrated that MSNR recapitulates known individual differences in functional connectivity, including those related to development (Satterthwaite et al., 2013), sex differences (Satterthwaite et al., 2015), and in-scanner motion (Satterthwaite et al., 2012). Additionally, compared to common single-scale mass-univariate regression methods, MSNR achieved a balance between prediction performance and model complexity, with improved interpretability. All told, MSNR represents a new method for identifying individual differences in high-dimensional brain networks.

Several limitations of the MSNR approach should be noted. First, “scale” does not have a single definition. In fact, as pointed out by (Betzel & Bassett, 2017), scale can represent at least three different entities depending on the context: multi-scale topological structure, multi-scale temporal structure, and multi-scale spatial structure. In MSNR, we only considered multi-scale topological structure. Incorporating additional information from multiple scales beyond network topology will likely generate more nuanced and richer models for brain networks. Second, while we carefully conducted a permutation test to assess the statistical significance of the entire model, we did not provide an inferential procedure for determining the association between brain networks and each variable of interest. In particular, MSNR makes no claim of statistical significance for the coefficients in the matrices $\Gamma^1, \ldots, \Gamma^q$, which describe the community-level relationships with the covariates. Due to the inclusion of penalty terms in the MSNR framework, making such inferential statements is a challenging open problem.

In summary, by explicitly modeling variability at the edge and community levels, we developed a multi-scale network regression approach that achieved a balance between the trade-off of prediction and model complexity, potentially offering enhanced interpretability. Empirically, we demonstrated its advantages over alternative methods and illustrated its ability to uncover meaningful signals in a large neuroimaging dataset. Approaches such as MSNR have the potential

to yield novel insights into brain-behavior relationships that incorporate realistic multi-scale network architecture.

Supplementary Information

SIMULATION STUDY

We used the Brain Connectivity Toolbox (Rubinov & Sporns, 2009) to create random modular small-world adjacency matrices of dimension $p \times p$ with specified community assignments ($K = 4$) representing the edge-level information. There adjacency matrices were then used as the ground truth mean connectivity in stimulated data, Θ_0. We also created sparse $K \times K$ matrices $\Gamma_0^1, \ldots, \Gamma_0^q$, representing ground truth community-level brain-phenotype relationships. We constructed the ground truth adjacency matrix for the ith observation as $A_0^i = \Theta_0 + \gamma \sum_{f=1}^{q} X_i^f \cdot (W \Gamma_0^f W^T)$, where the elements X_i^f were independently generated from a normal distribution, scaled by a factor of γ to represent the effect size. Then, we generated the observed connectivity matrix $A^i = A_0^i + \epsilon_i$ for a noise matrix ϵ_i.

We created synthetic data with varying characteristics, such as different numbers of nodes ($p \in \{32, 64, 128\}$), sample sizes ($n \in \{50, 100, 150\}$), effect sizes ($\gamma \in \{0, 0.1, 0.5, 1\}$), and noise levels ($\epsilon \in \{0, 0.1, 0.5, 1\}$), for a total of 108 combinations of these parameters. For each combination, we generated three equally-size sets, for training, testing, and validation. Tuning parameters λ_1 and λ_2 were selected using the training and testing sets, and the out-of-sample prediction error was computed on the validation set.

We found that MSNR achieved the lowest out-of-sample prediction error when the ratio between the number of subjects and the number of nodes was the largest ($n = 150$, $p = 32$) (**Supplementary Figure 1**). In addition, the amount of noise impacted MSNR's prediction performance in a graded fashion, with a three-fold difference between the lowest noise level (0.1) and the highest noise level (1). In contrast, MSNR was less sensitive to the varying levels of γ, which represents the effect size of the community level relationship of the covariates. These results were to be expected.

Proof of Proposition 1.

Given the definition of $\tilde{A}_i$, (5) reduces to the optimization problem

$$\underset{\Theta}{\text{minimize}} \left\{ \sum_{i=1}^{n} \left\| \Theta - \tilde{A}_i \right\|_F^2 + \lambda_1 \|\Theta\|_* \right\}. \tag{11}$$

We notice that

$$\begin{aligned} \sum_{i=1}^{n} \left\| \Theta - \tilde{A}^i \right\|_F^2 &= n \left(\|\Theta\|_F^2 - 2trace \left[\Theta \left(\sum_{i=1}^{n} \frac{\tilde{A}^i}{n} \right) \right] \right) + C \\ &= n \left\| \Theta - \sum_{i=1}^{n} \frac{\tilde{A}^i}{n} \right\|_F^2 + C', \end{aligned}$$

where C and C' are not a function of Θ. Therefore, (11) can be re-written as

$$\underset{\Theta}{\text{minimize}} \left\{ \left\| \Theta - \sum_{i=1}^{n} \frac{\tilde{A}^i}{n} \right\|_F^2 + \lambda_1 \|\Theta\|_* \right\}. \tag{12}$$

The result follows directly from Lemma 1 of Mazumder et al. (2010).

Proof of Proposition 2.

We wish to solve the problem

$$\underset{\Gamma^f}{\text{minimize}} \left\{ \sum_{i=1}^{n} \left\| A^i - \left(\widehat{\Theta} + \sum_{f' \neq f} X_i^{f'} \cdot \left(W \hat{\Gamma}^{f'} W^T \right) + X_i^f \cdot \left(W \Gamma^f W^T \right) \right) \right\|_F^2 + \lambda_2 \left\| \Gamma^f \right\|_1 \right\}.$$

Given the definition of $\bar{A}^i$, this amounts to solving

$$\underset{\Gamma^f}{\text{minimize}} \left\{ \sum_{i=1}^{n} \left\| \bar{A}^i - X_i^f \cdot \left(W \Gamma^f W^T \right) \right\|_F^2 + \lambda_2 \left\| \Gamma^f \right\|_1 \right\}. \tag{13}$$

So, for $k = 1, \ldots, K$ and $k' = 1, \ldots, K$, we must solve the problem

$$\underset{\Gamma^f_{kk'}}{\text{minimize}} \left\{ \sum_{i=1}^{n} \sum_{j \in C_k} \sum_{j' \in C_{k'}} \left(\bar{A}^i_{jj'} - X_i^f \Gamma^f_{kk'} \right)^2 + \lambda_2 \left| \Gamma^f_{kk'} \right| \right\}. \tag{14}$$

And note that

$$\sum_{i=1}^{n} \sum_{j \in C_k} \sum_{j' \in C_{k'}} \left(\bar{A}^i_{jj'} - X_i^f \Gamma^f_{kk'} \right)^2 + \lambda_2 \left| \Gamma^f_{kk'} \right|$$

$$= C - 2 \left(\sum_{i=1}^{n} \sum_{j \in C_k} \sum_{j' \in C_{k'}} \bar{A}^i_{jj'} X_i^f \right) \Gamma^f_{kk'} + p_k p_{k'} \sum_{i=1}^{n} \left(X_i^f \right)^2 \left(\Gamma^f_{kk'} \right)^2$$

where C is not a function of Γ^f. So the problem of interest amounts to minimizing

$$-2\frac{\left(\sum_{i=1}^{n}\sum_{j\in C_k}\sum_{j'\in C_{k'}}\bar{A}_{jj'}^{i}X_i^f\right)}{p_k p_{k'}\sum_{i=1}^{n}\left(X_i^f\right)^2}\Gamma_{kk'}^f+\left(\Gamma_{kk'}^f\right)^2+\frac{\lambda_2}{p_k p_{k'}\sum_{i=1}^{n}\left(X_i^f\right)^2}\left|\Gamma_{kk'}^f\right|$$

with respect to $\Gamma_{kk'}^f$. Thus, the minimizer is

$$S\left(\frac{\sum_{i=1}^{n}X_i^f\sum_{j\in C_k}\sum_{j'\in C_{k'}}\bar{A}_{jj'}^{i}}{p_k p_{k'}\sum_{i=1}^{n}\left(X_i^f\right)^2},\frac{\lambda_2}{2p_k p_{k'}\sum_{i=1}^{n}\left(X_i^f\right)^2}\right).$$

Figures

Figure 3-1

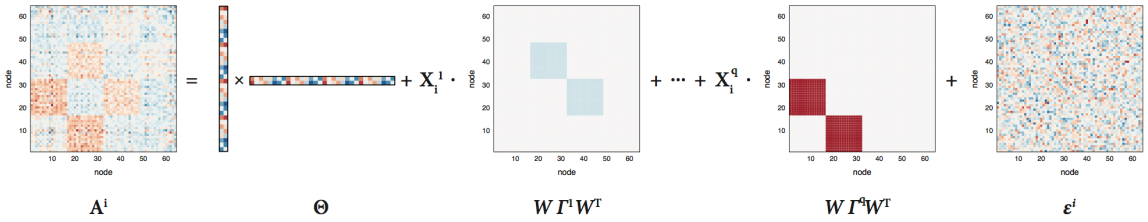


Figure 3-1 A schematic for Multi-Scale Network Regression. Under model (2), A^i is the adjacency matrix for the ith subject, Θ is a low-rank matrix representing the mean connectivity across all subjects, $\Gamma^1, \ldots, \Gamma^q$ are sparse matrices representing the community-level connectivity associated with the covariates $\left(X_i^1, \ldots, X_i^q\right)$, and ϵ^i is the noise.

Figure 3-2

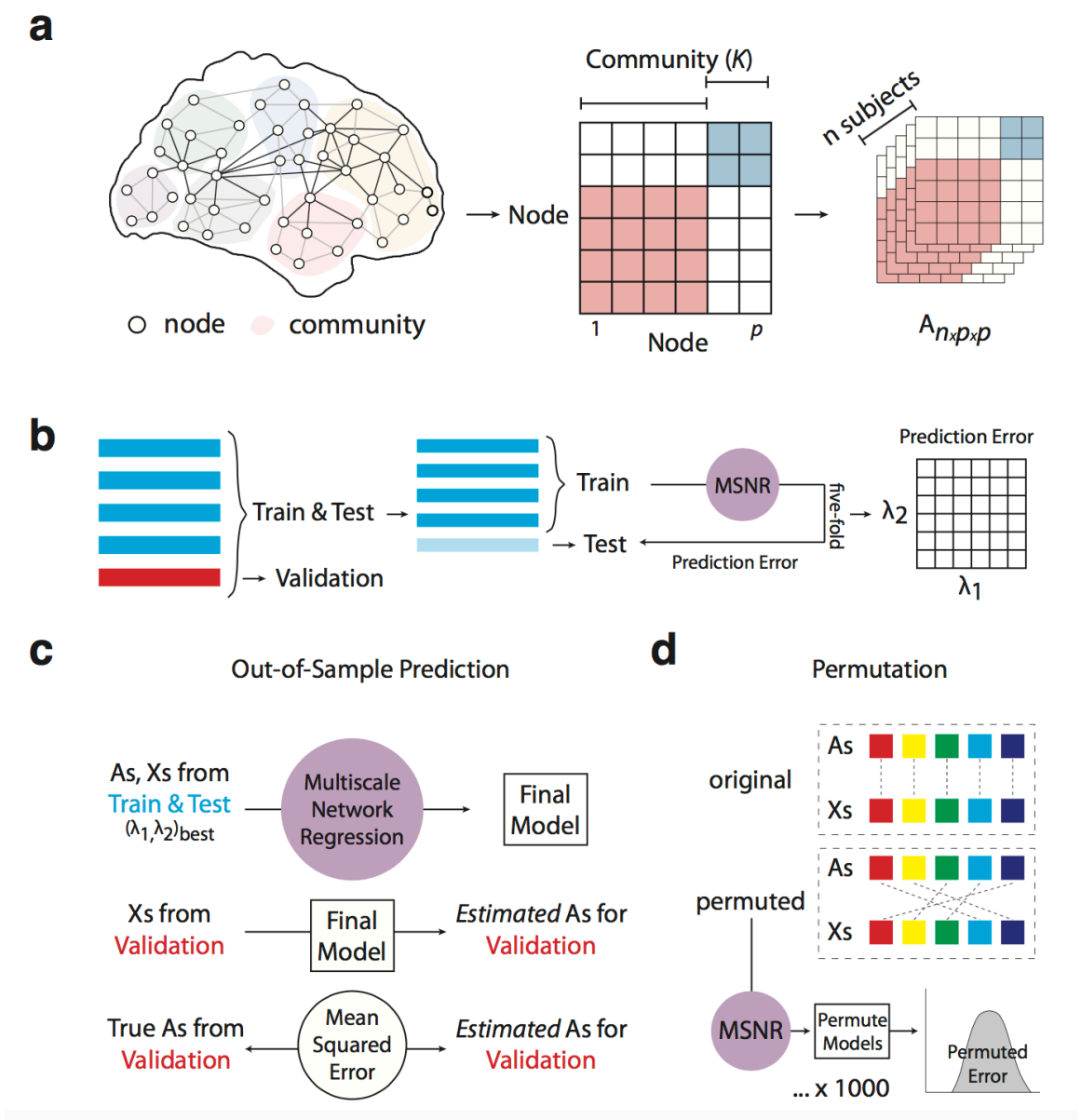


Figure 3-2: A schematic for MSNR model training and evaluation. a) MSNR is designed to study the brain connectivity-phenotype relationship by taking into account both edge- and community-level information. The model takes in a

$n \times p \times p$ matrix, where n is the number of subjects and where p is the number of nodes in each symmetric adjacency matrix. The nodes belong to K communities, determined a priori. **b**) 20% ($n = 202$) of the total sample ($n = 1,015$) were randomly selected as the left-out validation data. We conducted five-fold cross-validation to select the values of the tuning parameters λ_1 and λ_2, which were applied to the nuclear norm penalty on the mean connectivity matrix (Θ) and the l_1 norm of the community-level connectivity-covariate relationship matrices ($\Gamma^1, \ldots, \Gamma^q$), respectively. **c**) The model was then trained using the tuning parameters determined in b) on the 80% ($n = 813$) of the total data not in the left-out validation set. Out-of-sample prediction error was then calculated as the *Frobenius* norm of the difference between the known and estimated connectivity matrices on the validation set. **d**) We also evaluated the final model through a permutation procedure, where we broke the linkage between brain connectivity and covariate data to generate a null distribution of out-of-sample prediction error.

Figure 3-3

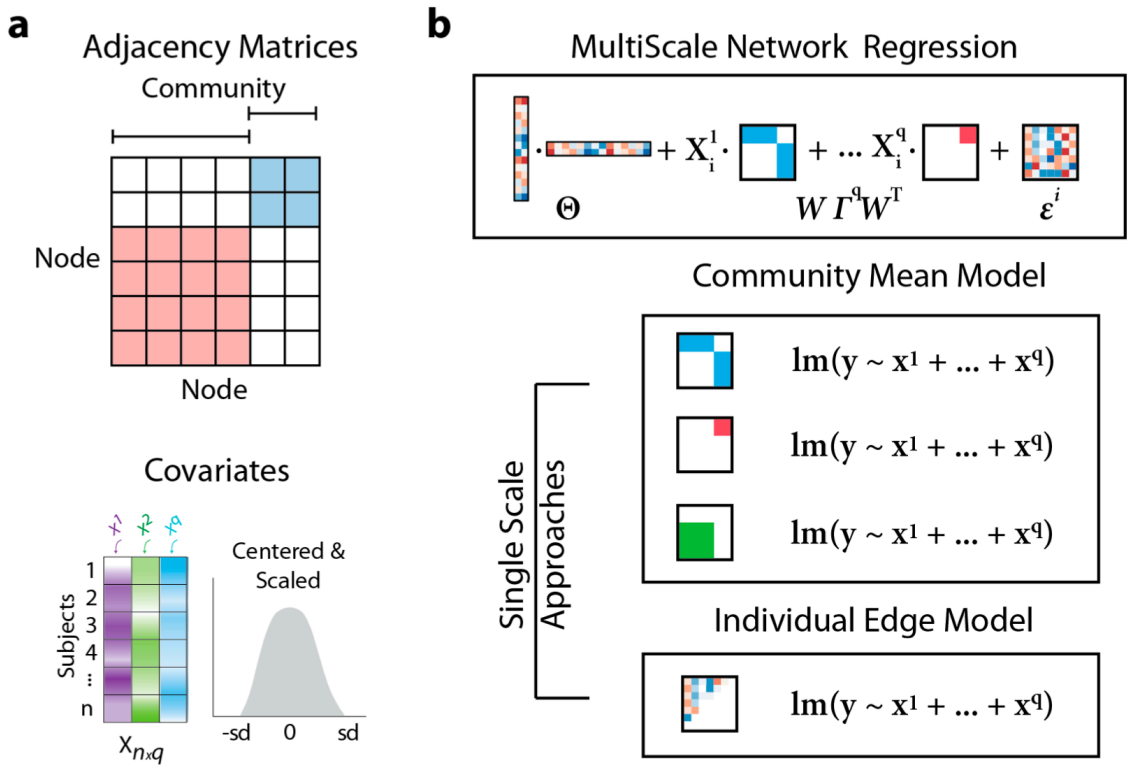


Figure 3-3: Benchmarking MSNR against common single-scale approaches. **a**) On the PNC data, we considered prediction of out-of-sample connectivity matrices from age, sex, and in-scanner motion. Specifically, input network data were $n \times p \times p$ connectivity matrices of n subjects with p nodes sorted a priori into K communities. Additionally, covariate data were a $n \times q$ matrix of q measurements, with each column centered with zero mean and scaled by its standard deviation. **b**) Specifically, we compared MSNR to two common network analysis approaches that only consider information present on a single scale. Linear models were fit for each edge or community connectivity for the individual edge and community mean model, respectively.

Figure 3-4

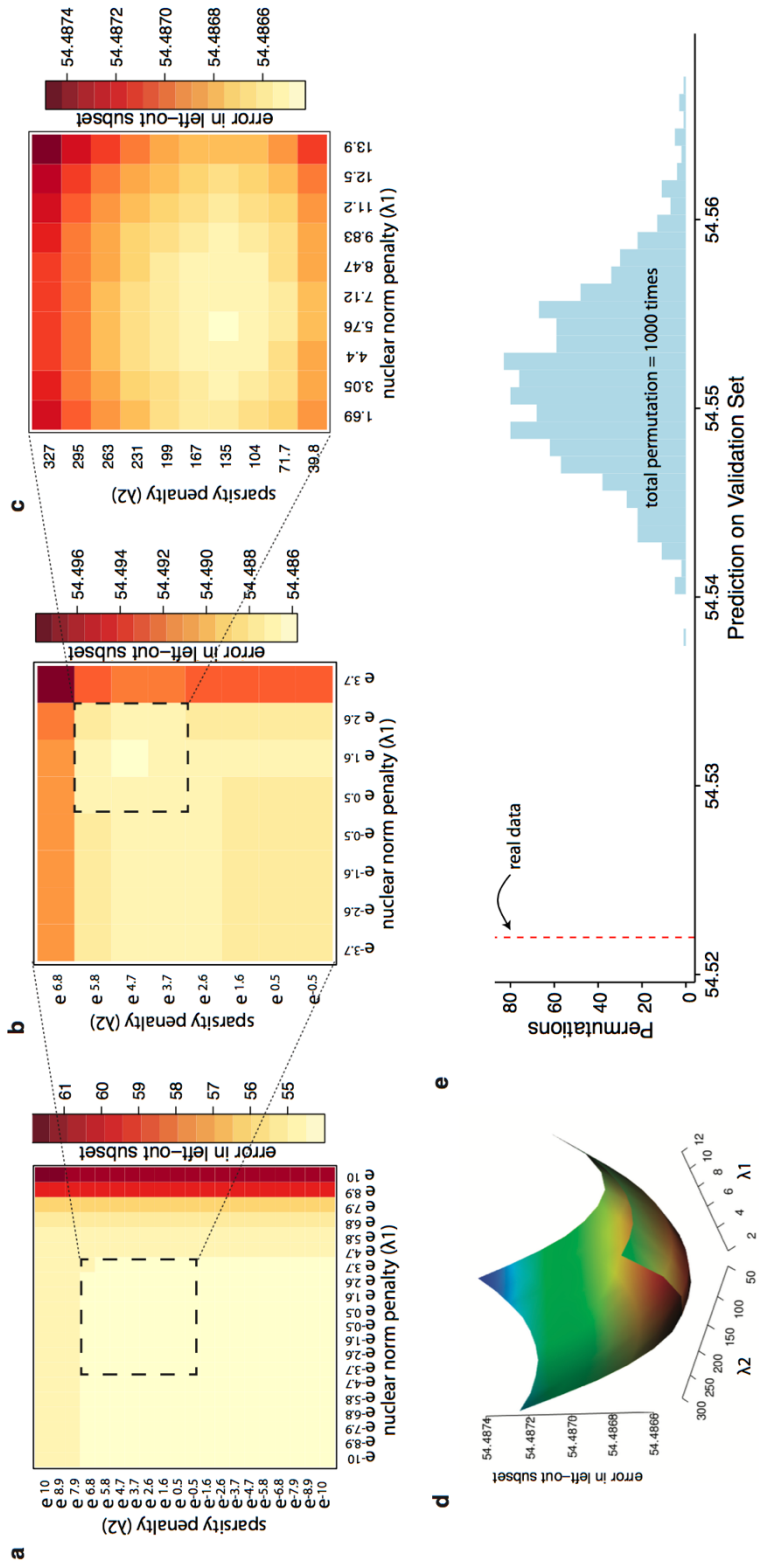

a
b
c
d
e
sparsity penalty (λ2)
nuclear norm penalty (λ1)
error in left−out subset
Permutations
real data
total permutation = 1000 times
Prediction on Validation Set
λ1
λ2

Figure 3-4 Tuning parameter selection and model evaluation of MSNR in a real-world large neuroimaging dataset. a) We used five-fold cross-validation to estimate the test prediction error associated with various values of λ_1 and λ_2. **b**) After the initial search, we conducted another search on a finer scale, focusing on the range of λ_1 and λ_2 indicated by the dashed-line box. **c**) The optimal tuning parameter values were found to be $\lambda_1 = 5.76$ and $\lambda_2 = 135$. No boundary effect was observed in the grid search, revealing a smooth convex landscape for the objective, also visualized in **d**), with warmer color indicating lower prediction error. **e**) The permutation procedure indicated that MSNR fit to the original data significantly outperformed MSNR fit to permuted data, with an out-of-sample prediction error about six standard deviations below the mean of the null distribution ($p < 0.001$).

Figure 3-5

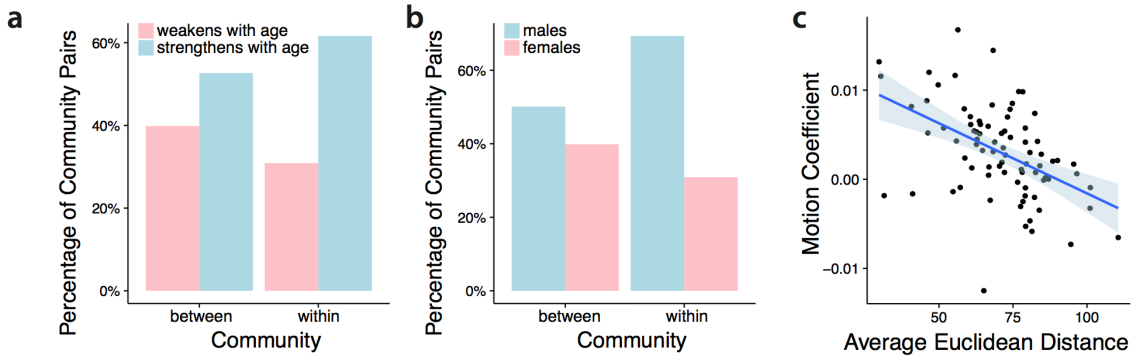


Figure 3-5: MSNR describes meaningful individual differences in brain connectivity. a) More within-community, rather than between-community, connectivity strengthened as the age increased. Conversely, more between-community, rather than within-community, connectivity weakened over age. **b**) Stronger within-community than between-community connectivity was more representative of male functional brain networks, whereas stronger between-community than within-community connectivity was more representative of female functional brain networks. **c**) Coefficient for in-scanner motion was negatively correlated with the average Euclidean distance between communities ($p < 0.001$).

Figure 3-6

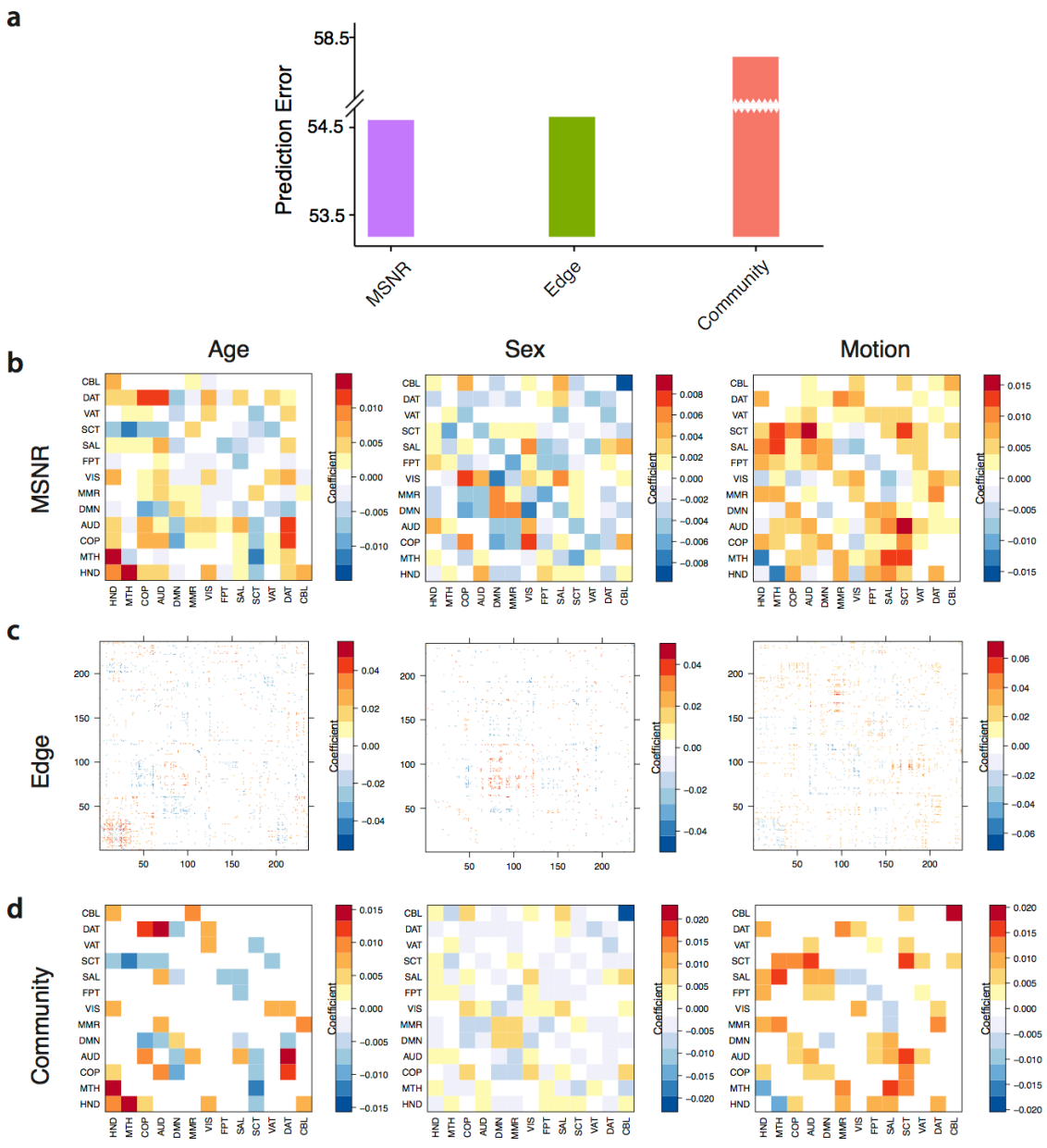
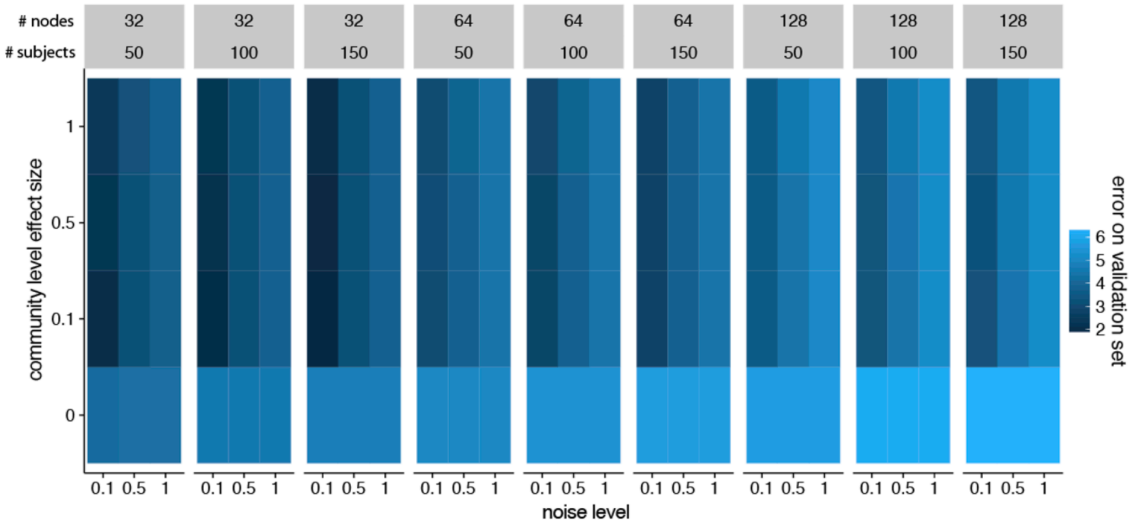


Figure 3-6: MSNR achieves a balance between out-of-sample prediction performance and model interpretability compared to common single-scale mass-univariate approaches. a) We compared out-of-sample prediction performance of MSNR to edge- and community-based single-scale approaches.

The community-based approach performed poorly, while the edge-based approach and MSNR had similar out-of-sample prediction error. All models fitted in mass-univariate approaches were used to calculate prediction error. **b**) MSNR coefficients in $\Gamma^1, \Gamma^2, \Gamma^3$, correspond to age, sex, and in-scanner motion, respectively. Warm colors indicate increased connectivity and cold colors indicate decreased connectivity as the covariate increased. White color indicates zero values. Results from single-scale models were visualized in **c**) for edge-based and in **d**) for community-based approaches. Multiple comparisons were corrected using FDR.

Figure 3-7



Supplementary Figure 1: Performance of MSNR in a simulation study. We simulated data with varying numbers of observations (n) and nodes (p), effect size (γ) of $\Gamma^1, \ldots \Gamma^q$, and noise levels ($\varepsilon$). As expected, the performance of MSNR improved as the ratio of n to p increased, and as the signal-to-noise ratio increased. In contrast, MSNR was less sensitive to the varying levels of γ, which represents the effect size of the community level relationship of the covariates.

References

Alexander-Bloch, A. F., Gogtay, N., Meunier, D., Birn, R., Clasen, L., Lalonde, F., … Bullmore, E. T. (2010). Disrupted Modularity and Local Connectivity of Brain Functional Networks in Childhood-Onset Schizophrenia. *Frontiers in Systems Neuroscience*, *4*, 147. https://doi.org/10.3389/fnsys.2010.00147

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*, *54*(3), 2033–2044. https://doi.org/10.1016/J.NEUROIMAGE.2010.09.025

Avants, B. B., Tustison, N. J., Wu, J., Cook, P. A., & Gee, J. C. (2011). An open source multivariate framework for n-tissue segmentation with evaluation on public data. *Neuroinformatics*, *9*(4), 381–400. https://doi.org/10.1007/s12021-011-9109-y

Bassett, D. S., & Siebenhühner, F. (2013). Multiscale Network Organization in the Human Brain. In *Multiscale Analysis and Nonlinear Dynamics* (pp. 179–204). Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. https://doi.org/10.1002/9783527671632.ch07

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, *20*(3), 353–364. https://doi.org/10.1038/nn.4502

Bassett, D. S., Xia, C. H., & Satterthwaite, T. D. (2018). Understanding the Emergence of Neuropsychiatric Disorders With Network Neuroscience. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. https://doi.org/10.1016/J.BPSC.2018.03.015

Bassett, D. S., Yang, M., Wymbs, N. F., & Grafton, S. T. (2015). Learning-induced autonomy of sensorimotor systems. *Nature Neuroscience*, *18*(5), 744–751. https://doi.org/10.1038/nn.3993

Baum, G. L., Ciric, R., Roalf, D. R., Betzel, R. F., Moore, T. M., Shinohara, R. T., … Satterthwaite, T. D. (2016). Modular Segregation of Structural Brain Networks Supports the Development of Executive Function in Youth.

Betzel, R. F., & Bassett, D. S. (2017). Multi-scale brain networks. *NeuroImage*,

160, 73–83. https://doi.org/10.1016/J.NEUROIMAGE.2016.11.006

Betzel, R. F., Byrge, L., He, Y., Goñi, J., Zuo, X.-N., & Sporns, O. (2014). Changes in structural and functional connectivity among resting-state networks across the human lifespan. *NeuroImage*, *102*, 345–357. https://doi.org/10.1016/J.NEUROIMAGE.2014.07.067

Betzel, R. F., Medaglia, J. D., & Bassett, D. S. (2018). Diversity of meso-scale architecture in human and non-human connectomes. *Nature Communications*, *9*(1), 346. https://doi.org/10.1038/s41467-017-02681-z

Bien, J., & Witten, D. (2016). Penalized estimation in complex models. In *Handbook of Big Data* (pp. 285–299).

Biswal, B. B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C., Smith, S. M., … Milham, M. P. (2010). Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(10), 4734–4739. https://doi.org/10.1073/pnas.0911855107

Breakspear, M., & Stam, C. J. (2005). Dynamics of a neural system with a multiscale architecture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1457), 1051–1074. https://doi.org/10.1098/rstb.2005.1643

Bressler, S. L., & Menon, V. (2010). Large-scale brain networks in cognition: emerging methods and principles. *Trends in Cognitive Sciences*, *14*(6), 277–290. https://doi.org/10.1016/J.TICS.2010.04.004

Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nature Reviews Neuroscience*, *3*(3), 243–249. https://doi.org/10.1038/nrn756

Buckner, R. L., Sepulcre, J., Talukdar, T., Krienen, F. M., Liu, H., Hedden, T., … Johnson, K. A. (2009). Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(6), 1860–1873. https://doi.org/10.1523/JNEUROSCI.5062-08.2009

Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*(3), 186–198. https://doi.org/10.1038/nrn2575

Bzdok, D., & Yeo, B. T. T. (2017). Inference in the age of big data: Future perspectives on neuroscience. *NeuroImage*, *155*, 549–564. https://doi.org/10.1016/j.neuroimage.2017.04.061

Choi, D. S., Wolfe, P. J., & Airoldi, E. M. (2012). Stochastic blockmodels with a growing number of classes. *Biometrika*, *99*(2), 273–284. https://doi.org/10.1093/biomet/asr053

Choi, H., & Baraniuk, R. G. (2001). Multiscale image segmentation using wavelet-domain hidden Markov models. *IEEE Transactions on Image Processing*, *10*(9), 1309–1321. https://doi.org/10.1109/83.941855

Ciric, R., Rosen, A. F. G., Erus, G., Cieslak, M., Adebimpe, A., Cook, P. A., … Satterthwaite, T. D. (2018). Mitigating head motion artifact in functional connectivity MRI. *Nature Protocols*, *13*(12), 2801–2826. https://doi.org/10.1038/s41596-018-0065-y

Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., Ruparel, K., … Satterthwaite, T. D. (2017a). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *NeuroImage*, *154*, 174–187. https://doi.org/10.1016/J.NEUROIMAGE.2017.03.020

Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., Ruparel, K., … Satterthwaite, T. D. (2017b). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *NeuroImage*, *154*(March), 174–187. https://doi.org/10.1016/j.neuroimage.2017.03.020

Craddock, R. C., Tungaraza, R. L., & Milham, M. P. (2015). Connectomics and new approaches for analyzing human brain functional connectivity. *GigaScience*, *4*(1), 13. https://doi.org/10.1186/s13742-015-0045-x

Crossley, N. A., Mechelli, A., Vértes, P. E., Winton-Brown, T. T., Patel, A. X., Ginestet, C. E., … Bullmore, E. T. (2013). Cognitive relevance of the

community structure of the human brain functional coactivation network. *Proceedings of the National Academy of Sciences*, *110*(28), 11583–11588. https://doi.org/10.1073/PNAS.1220826110

Damoiseaux, J. S., Beckmann, C. F., Arigita, E. J. S., Barkhof, F., Scheltens, P., Stam, C. J., … Rombouts, S. A. R. B. (2008). Reduced resting-state brain activity in the “default network” in normal aging. *Cerebral Cortex*, *18*(8), 1856–1864. https://doi.org/10.1093/cercor/bhm207

Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., … Liston, C. (2016). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine*, *23*(1), 28–38. https://doi.org/10.1038/nm.4246

Durante, D., & Dunson, D. B. (2018). Bayesian Inference and Testing of Group Differences in Brain Networks. *Bayesian Analysis*, *13*(1), 29–58. https://doi.org/10.1214/16-BA1030

Durante, D., Dunson, D. B., & Vogelstein, J. T. (2017). Nonparametric Bayes Modeling of Populations of Networks. *Journal of the American Statistical Association*, *112*(520), 1516–1530. https://doi.org/10.1080/01621459.2016.1219260

Fair, D. A., Dosenbach, N. U. F., Church, J. a, Cohen, A. L., Brahmbhatt, S., Miezin, F. M., … Schlaggar, B. L. (2007). 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. https://doi.org/10.1073/pnas.0705843104

Fazel, M. (2002). *Matrix rank minimization with applications*. Stanford University.

Fornito, A., Zalesky, A., & Breakspear, M. (2015). The connectomics of brain disorders. *Nature Reviews Neuroscience*, *16*(3), 159–172. https://doi.org/10.1038/nrn3901

Fosdick, B. K., & Hoff, P. D. (2015). Testing and Modeling Dependencies Between a Network and Nodal Attributes. *Journal of the American Statistical Association*, *110*(511), 1047–1056. https://doi.org/10.1080/01621459.2015.1008697

Friedman, J., Hastie, T., Höfling, H., & Tibshirani, R. (2007). Pathwise coordinate optimization. *The Annals of Applied Statistics*, *1*(2), 302–332. https://doi.org/10.1214/07-AOAS131

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Grillon, M.-L., Oppenheim, C., Varoquaux, G., Charbonneau, F., Devauchelle, A.-D., Krebs, M.-O., … Huron, C. (2013). Hyperfrontality and hypoconnectivity during refreshing in schizophrenia. *Psychiatry Research: Neuroimaging*, *211*(3), 226–233. https://doi.org/10.1016/J.PSCYCHRESNS.2012.09.001

Gu, S., Satterthwaite, T. D., Medaglia, J. D., Yang, M., Gur, R. E., Gur, R. C., & Bassett, D. S. (2015). Emergence of system roles in normative neurodevelopment. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(44), 13681–13686. https://doi.org/10.1073/pnas.1502829112

Gur, R. E., & Gur, R. C. (2016). Sex differences in brain and behavior in adolescence: Findings from the Philadelphia Neurodevelopmental Cohort. *Neuroscience & Biobehavioral Reviews*, *70*, 159–170. https://doi.org/10.1016/J.NEUBIOREV.2016.07.035

Hallquist, M. N., Hwang, K., & Luna, B. (2013). 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. https://doi.org/10.1016/J.NEUROIMAGE.2013.05.116

Hastie, T. (2015). *Statistical Learning with Sparsity*. Chapman and Hall/CRC. https://doi.org/10.1201/b18401

Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning* (2nd ed.). Stanford, CA: Springer.

Hastie, T., Tibshirani, R., Wainwright, M., Tibshirani, R., & Wainwright, M. (2015). *Statistical Learning with Sparsity*. Chapman and Hall/CRC. https://doi.org/10.1201/b18401

Ingalhalikar, M., Smith, A., Parker, D., Satterthwaite, T. D., Elliott, M. A., Ruparel, K., … Verma, R. (2014). Sex differences in the structural connectome of the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(2), 823–828. https://doi.org/10.1073/pnas.1316909110

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (8th ed.). Springer. https://doi.org/10.1007/978-1-4614-7138-7

Jarosiewicz, B., Chase, S. M., Fraser, G. W., Velliste, M., Kass, R. E., & Schwartz, A. B. (2008). Functional network reorganization during learning in a brain-computer interface paradigm. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(49), 19486–19491. https://doi.org/10.1073/pnas.0808113105

Jenatton, R., Gramfort, A., Michel, V., Obozinski, G., Eger, E., Bach, F., & Thirion, B. (2012). Multiscale Mining of fMRI Data with Hierarchical Structured Sparsity. *SIAM Journal on Imaging Sciences*, *5*(3), 835–856. https://doi.org/10.1137/110832380

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/NIMG.2002.1132

Jernigan, T. L., Brown, T. T., Hagler, D. J., Akshoomoff, N., Bartsch, H., Newman, E., … Pediatric Imaging, Neurocognition and Genetics Study. (2016). The Pediatric Imaging, Neurocognition, and Genetics (PING) Data Repository. *NeuroImage*, *124*(Pt B), 1149–1154. https://doi.org/10.1016/j.neuroimage.2015.04.057

Khambhati, A. N., Davis, K. A., Lucas, T. H., Litt, B., & Bassett, D. S. (2016). Virtual Cortical Resection Reveals Push-Pull Network Control Preceding Seizure Evolution. *Neuron*, *91*(5), 1170–1182. https://doi.org/10.1016/J.NEURON.2016.07.039

King, J. B., Prigge, M. B. D., King, C. K., Morgan, J., Dean, D. C., Freeman, A., … Anderson, J. S. (2018). Evaluation of Differences in Temporal Synchrony

Between Brain Regions in Individuals With Autism and Typical Development. *JAMA Network Open*, *1*(7), e184777. https://doi.org/10.1001/jamanetworkopen.2018.4777

Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., … Parsey, R. V. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, *46*(3), 786–802. https://doi.org/10.1016/J.NEUROIMAGE.2008.12.037

Leonardi, N., Richiardi, J., Gschwind, M., Simioni, S., Annoni, J.-M., Schluep, M., … Van De Ville, D. (2013). Principal components of functional connectivity: A new approach to study dynamic brain connectivity during rest. *NeuroImage*, *83*, 937–950. https://doi.org/10.1016/J.NEUROIMAGE.2013.07.019

Lewis, C. M., Baldassarre, A., Committeri, G., Romani, G. L., & Corbetta, M. (2009). Learning sculpts the spontaneous activity of the resting human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(41), 17558–17563. https://doi.org/10.1073/pnas.0902455106

Li, K., Guo, L., Nie, J., Li, G., & Liu, T. (2009). Review of methods for functional brain connectivity detection using fMRI. *Computerized Medical Imaging and Graphics*, *33*(2), 131–139. https://doi.org/10.1016/J.COMPMEDIMAG.2008.10.011

Li, Y., Gilmore, J. H., Shen, D., Styner, M., Lin, W., & Zhu, H. (2013). Multiscale adaptive generalized estimating equations for longitudinal neuroimaging data. *NeuroImage*, *72*, 91–105. https://doi.org/10.1016/J.NEUROIMAGE.2013.01.034

Li, Y., Zhu, H., Shen, D., Lin, W., Gilmore, J. H., & Ibrahim, J. G. (2011). Multiscale Adaptive Regression Models for Neuroimaging Data. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *73*(4), 559–578. https://doi.org/10.1111/j.1467-9868.2010.00767.x

Mazumder, R., Hastie, T., Edu, H., Tibshirani, R., Edu, T., & Jaakkola, T. (2010). *Spectral Regularization Algorithms for Learning Large Incomplete Matrices*. *Journal of Machine Learning Research* (Vol. 11).

Mennes, M., Biswal, B. B., Castellanos, F. X., & Milham, M. P. (2013). Making data sharing work: The FCP/INDI experience. *NeuroImage*, *82*, 683–691. https://doi.org/10.1016/J.NEUROIMAGE.2012.10.064

Meunier, D., Lambiotte, R., & Bullmore, E. T. (2010). Modular and Hierarchically Modular Organization of Brain Networks. *Frontiers in Neuroscience*, *4*, 200. https://doi.org/10.3389/fnins.2010.00200

Neubert, F.-X., Mars, R. B., Sallet, J., & Rushworth, M. F. S. (2015). Connectivity reveals relationship of brain areas for reward-guided learning and decision making in human and monkey frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(20), E2695-704. https://doi.org/10.1073/pnas.1410767112

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, *103*(23), 8577–8582. https://doi.org/10.1073/pnas.0601602103

Park, H. J., & Friston, K. (2013). Structural and functional brain networks: from connections to cognition. *Science*, *342*(November), 579–588. https://doi.org/10.1126/science.1238411

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., … Petersen, S. E. (2011). Functional Network Organization of the Human Brain. *Neuron*, *72*(4), 665–678. https://doi.org/10.1016/J.NEURON.2011.09.006

Power, J. D., Fair, D. A., Schlaggar, B. L., & Petersen, S. E. (2010). The Development of Human Functional Brain Networks. *Neuron*, *67*(5), 735–748. https://doi.org/10.1016/J.NEURON.2010.08.017

Raichle, M. E. (2015). The Brain's Default Mode Network. *Annual Review of Neuroscience*, *38*(1), 433–447. https://doi.org/10.1146/annurev-neuro-071013-014030

Recht, B., Fazel, M., & Parrilo, P. A. (2010). Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization. *SIAM Review*, *52*(3), 471–501. https://doi.org/10.1137/070697835

Romberg, J., Hyeokho Choi, Baraniuk, R., & Kingbury, N. (2000). Multiscale classification using complex wavelets and hidden Markov tree models. In *Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101)* (pp. 371–374 vol.2). IEEE. https://doi.org/10.1109/ICIP.2000.899396

Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(4), 1118–1123.

Rubinov, M., & Sporns, O. (2009). Complex network measures of brain connectivity : Uses and interpretations. *NeuroImage*, *52*(3), 1059–1069. https://doi.org/10.1016/j.neuroimage.2009.10.003

Satterthwaite, T. D., Connolly, J. J., Ruparel, K., Calkins, M. E., Jackson, C., Elliott, M. A., … Gur, R. E. (2016). The Philadelphia Neurodevelopmental Cohort: A publicly available resource for the study of normal and abnormal brain development in youth. *NeuroImage*, *124*, 1115–1119. https://doi.org/10.1016/J.NEUROIMAGE.2015.03.056

Satterthwaite, T. D., Elliott, M. A., Ruparel, K., Loughead, J., Prabhakaran, K., Calkins, M. E., … Gur, R. E. (2014). Neuroimaging of the Philadelphia Neurodevelopmental Cohort. *NeuroImage*, *86*(2014), 544–553. https://doi.org/10.1016/j.neuroimage.2013.07.064

Satterthwaite, T. D., Wolf, D. H., Loughead, J., Ruparel, K., Elliott, M. A., Hakonarson, H., … Gur, R. E. (2012). Impact of in-scanner head motion on multiple measures of functional connectivity: Relevance for studies of neurodevelopment in youth. *NeuroImage*, *60*(1), 623–632. https://doi.org/10.1016/j.neuroimage.2011.12.063

Satterthwaite, T. D., Wolf, D. H., Roalf, D. R., Ruparel, K., Erus, G., Vandekar, S., … Gur, R. C. (2015). Linked Sex Differences in Cognition and Functional Connectivity in Youth. *Cerebral Cortex*, *25*(9), 2383–2394. https://doi.org/10.1093/cercor/bhu036

Satterthwaite, T. D., Wolf, D. H., Ruparel, K., Erus, G., Elliott, M. A., Eickhoff, S. B., … Gur, R. C. (2013). Heterogeneous impact of motion on fundamental patterns of developmental changes in functional connectivity during youth.

NeuroImage, *83*(2013), 45–57. https://doi.org/10.1016/j.neuroimage.2013.06.045

Schumann, G., Loth, E., Banaschewski, T., Barbot, A., Barker, G., Büchel, C., … Struve, M. (2010). The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Molecular Psychiatry*, *15*(12), 1128–1139. https://doi.org/10.1038/mp.2010.4

Smith, S. M., Nichols, T. E., Vidaurre, D., Winkler, A. M., J Behrens, T. E., Glasser, M. F., … Miller, K. L. (2015). A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature Neuroscience*, *18*(September), 1–7. https://doi.org/10.1038/nn.4125

Solomon, E. A., Kragel, J. E., Sperling, M. R., Sharan, A., Worrell, G., Kucewicz, M., … Kahana, M. J. (2017). Widespread theta synchrony and high-frequency desynchronization underlies enhanced cognition. *Nature Communications*, *8*(1), 1704. https://doi.org/10.1038/s41467-017-01763-2

Sporns, O., & Betzel, R. F. (2016). Modular Brain Networks. *Annual Review of Psychology*, *67*(1), 613–640. https://doi.org/10.1146/annurev-psych-122414-033634

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(3), 479–498. https://doi.org/10.1111/1467-9868.00346

Tang, M., Athreya, A., Sussman, D. L., Lyzinski, V., & Priebe, C. E. (2017). A nonparametric two-sample hypothesis testing problem for random graphs. *Bernoulli*, *23*(3), 1599–1630. https://doi.org/10.3150/15-BEJ789

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tomasi, D, & Volkow, N. D. (2012). Aging and functional brain networks. *Molecular Psychiatry*, *17*(5), 549–558. https://doi.org/10.1038/mp.2011.81

Tomasi, Dardo, & Volkow, N. D. (2012). Gender differences in brain functional

connectivity density. *Human Brain Mapping*, *33*(4), 849–860. https://doi.org/10.1002/hbm.21252

Tseng, P. (2001). Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization. *Journal of Optimization Theory and Applications*, *109*(3), 475–494. https://doi.org/10.1023/A:1017501703105

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/TMI.2010.2046908

Tustison, N. J., Cook, P. A., Klein, A., Song, G., Das, S. R., Duda, J. T., … Avants, B. B. (2014). Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *NeuroImage*, *99*, 166–179. https://doi.org/10.1016/j.neuroimage.2014.05.044

van den Heuvel, M. P., & Fornito, A. (2014). Brain Networks in Schizophrenia. *Neuropsychology Review*, *24*(1), 32–48. https://doi.org/10.1007/s11065-014-9248-7

Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E. J., Bucholz, R., … Yacoub, E. (2012). The Human Connectome Project: A data acquisition perspective. *NeuroImage*, *62*(4), 2222–2231. https://doi.org/10.1016/J.NEUROIMAGE.2012.02.018

Varoquaux, G., & Craddock, R. C. (2013). Learning and comparing functional connectomes across subjects. *NeuroImage*, *80*, 405–415. https://doi.org/10.1016/J.NEUROIMAGE.2013.04.007

Wang, H., Suh, J. W., Das, S. R., Pluta, J. B., Craige, C., & Yushkevich, P. A. (2013). Multi-Atlas Segmentation with Joint Label Fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(3), 611–623. https://doi.org/10.1109/TPAMI.2012.143

Wood, S. N. (2011). Fast stable REML and ML estimation of semiparametric GLMs. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *73*(1), 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x

Wood, S. N. (2017). *Generalized additive models : an introduction with R*. Chapman and Hall/CRC.

Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing Parameter and Model Selection for General Smooth Models. *Journal of the American Statistical Association*, *111*(516), 1548–1563. https://doi.org/10.1080/01621459.2016.1180986

Xia, C. H., Ma, Z., Ciric, R., Gu, S., Betzel, R. F., Kaczkurkin, A. N., … Satterthwaite, T. D. (2018). Linked dimensions of psychopathology and connectivity in functional brain networks. *Nature Communications*, *9*(1), 3003. https://doi.org/10.1038/s41467-018-05317-y

Yan, C.-G., Chen, X., Li, L., Castellanos, F. X., Bai, T.-J., Bo, Q.-J., … Zang, Y.-F. (2019). Reduced default mode network functional connectivity in patients with recurrent major depressive disorder. *Proceedings of the National Academy of Sciences of the United States of America*, 201900390. https://doi.org/10.1073/pnas.1900390116

Young, S. J., & Scheinerman, E. R. (2007). Random Dot Product Graph Models for Social Networks. In *Algorithms and Models for the Web-Graph* (pp. 138–149). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-77004-6_11

Yu, M., Linn, K. A., Shinohara, R. T., Oathes, D. J., Cook, P. A., Duprat, R., … Sheline, Y. I. (2019). Childhood trauma history is linked to abnormal brain connectivity in major depression. https://doi.org/10.1073/pnas.1900801116

Zalesky, A., Fornito, A., & Bullmore, E. T. (2010). Network-based statistic: Identifying differences in brain networks. *NeuroImage*, *53*(4), 1197–1207. https://doi.org/10.1016/J.NEUROIMAGE.2010.06.041

Zapala, M. A., & Schork, N. J. (2012). Statistical Properties of Multivariate Distance Matrix Regression for High-Dimensional Data Analysis. *Frontiers in Genetics*, *3*, 190. https://doi.org/10.3389/fgene.2012.00190

CHAPTER 4

General Discussion

Synthesis of results and overall discussion

The marked level of co-morbidity across psychiatric diagnoses and heterogeneity within individual diagnostic categories suggest that current symptomology-based diagnostic criteria do not “carve nature at its joints.”(Cuthbert & Insel, 2010; B. T. R. Insel & Cuthbert, 2015) Brain reconfiguration during adolescence is a complex neurodevelopmental process, deviations from which may underlie many mental illnesses that arise in young adulthood (Bassett, Xia, & Satterthwaite, 2018; T. R. Insel, 2014). Circuit-level abnormalities, theorized as a result of brain network misconfiguration during development, do not neatly respect clinical diagnostic boundaries, suggesting common mechanisms that cut across clinically diagnosed psychiatric disorders (Kapur, Phillips, & Insel, 2012). However, a fundamental understanding of how specific deviations from the normal remodeling in the developing brain are associated with a diverse range of psychiatric symptoms has remained elusive.

In chapter 2, we delineated linked dimensions of psychopathology that were highly associated with complex patterns of functional brain connectivity. Specifically, we leveraged a large cohort of youth as part of the Philadelphia Neurodevelopmental Cohort (PNC), who have completed functional MRI imaging and comprehensive psychiatric symptom evaluation (Calkins et al., 2015; Satterthwaite et al., 2016). Explicitly agnostic to any specific diagnosis, we applied sparse Canonical Correlation Analysis (Witten, Tibshirani, & Hastie,

2009), an unsupervised learning method, to extract latent representations of symptom-connectivity relationships in a multivariate fashion. As a result, we discovered four linked dimensions of psychopathology and functional brain connectivity patterns – mood, psychosis, fear, and externalizing behavior. These brain-guided psychopathological dimensions crossed traditional categorical boundaries, while concurring with clinical experience. Each linked dimension exhibited unique connectivity patterns; however, across all psychopathology, loss of normative segregation between the default mode and executive networks emerged as a common feature of connectivity dysfunction. Moreover, significant development effect was present for mood and psychosis dimensions, and sex differences were present for dimensions of mood and fear.

In chapter 3, we built upon the momentum in the neuroscience community of investigating complex functional connectivity patterns that are associated with a wide range of measurements, including psychopathology that we examined in the previous chapter. Specifically, we recognized both the intense need and relative deficiency in proper methods to study brain-phenotype relationships, especially in high-dimensional brain networks, where number of features often far exceeds the number of observations available. To this end, we designed, implemented, and deployed a new penalized multivariate analytical tool to study brain-phenotype relationship based on a multi-scale perspective of brain networks, which we call Multi-Scale Network Regression (MSNR). In particular, compared to common single-scale networks that only consider the edge or the

community level information of networks alone, MSNR achieved a balance between out-of-sample prediction performance and model interpretability. To do this, MSNR imposed a low rank and a sparse structure on the edge and community features, respectively. In an empirical study where we deployed MSNR to the PNC dataset, MSNR recapitulated known multivariate relationships between functional brain networks and age, sex, as well as in-scanner motion.

In sum, this dissertation uncovered latent representations of psychopathological dimensions that are linked to common and dissociable connectivity patterns, which cut across existing diagnostic categories. In addition, we extended the approach to incorporate information present at multiple scales of brain networks that can be used to model a variety of phenotype measurements. By considering more realistic network architecture, the new method, named *Multi-Scale Network Regression*, could yield novel insights to brain-phenotype relationships with improved generalizability and interpretability.

Future directions

In the preceding chapters, leveraging a large neuroimaging dataset of youth and recent advance in machine learning, we provided evidence of common and dissociable brain connectivity patterns that are correlated with dimensions of psychopathology across diagnostic criteria, and offered a new statistical learning tool to investigate multivariate connectivity patterns with diverse range of

measurement, beyond psychopathology, in a generalizable and interpretable fashion. However, there are a few limitations in both studies that would restrict the potential impact of the findings and tools, but at the same time offer exciting opportunities for follow up investigations.

In chapter 2, using sparse Canonical Correlation Analysis and functional brain networks of nearly 1,000 youth, we demonstrated that complex psychiatric symptoms are associated with specific patterns of abnormal connectivity during brain development. Although this study benefited from a large sample, advanced multivariate methods, and replication of results in a left-out-sample, several limitations should be noted. First, this approach of linking symptoms across diagnostic categories to aberrations in functional connectivity is limited by the item-level clinical data used. In particular, while we were agnostic to subjects' exact diagnosis, the individual symptoms were from a structured clinical interview, legacy from categorical conceptualization of psychopathology (Calkins et al., 2015). Second, the generalizability of the current is impaired by the fact we could not use a truly independent dataset to validate our findings (James, Witten, Hastie, & Tibshirani, 2013). The use of a left-out one-third of the total data was a reasonable, but far from perfect, proxy to independent data acquired in a different setting. Third, the cross-sectional nature of the data further limited our ability to answer question how deviations from normative process of brain development underlie vulnerability to psychopathology. Finally, functional connectivity was only a small set of the richness present in the PNC dataset (Satterthwaite et al.,

2016). Inevitably, we were only able to capture a small set of variation and potential signals pertaining to individual differences in their biology.

Given these limitations of the study in chapter 2, future follow up studies should focus on: 1) Incorporating additional datatypes including digital phenotyping and genomics to capture different sources of important biological heterogeneity (T. R. Insel, 2017); 2) harmonizing datasets across clinical and imaging methodologies so that findings in this study could be validated in a truly independent dataset acquired in different settings (Fortin et al., 2018, 2017; Yu et al., 2018); 3) taking advance of the longitudinal component of the PNC to more robustly test how individual development of their brain associate with their psychiatric and behavioral changes (Satterthwaite et al., 2016); 4) incorporating multi-modal imaging data, beyond functional connectivity, to examine dimensions of structural as well as function-structure coupling that are associated with psychopathology in the developing brain.

In chapter 3, we proposed a new tool to extract brain-phenotype relationship in high-dimensional connectomics. By integrating information present on multiple levels of brain networks, we designed a multi-scale approach, MSNR, to study complex connectivity patterns underlying phenotype-of-interests. With its ability to achieve a balance between out-of-sample prediction and model interpretability, this multivariate analysis tool has the potential to yield novel insights into brain-phenotype associations. Several limitations should be noted.

First, the current only addressed the scale in a narrow sense. Specifically, here scale referred to the topological scale of brain networks under investigation, including microscale edges, mesoscale community structure, and macroscale global summary statistics. However, scale of the network could also include temporal scale and spatial structures, not just limited to topology (Bassett & Siebenhühner, 2013; Betzel & Bassett, 2017). Second, the ability to make inference on the resulting model from MSNR is limited to the multivariate patterns associated with all covariates included in the model. Due to low rank and sparsity constraints present, the current study did not address the potentially more useful question of how one would make inference on one individual variable tested in the model.

Given these limitations of the study in chapter 3, future follow up study should focus on: 1) extending MSNR to incorporate information present on other definitions of scales, such as time-varying dynamic networks and spatial networks that acknowledge unique brain anatomy. 2) implementing a practical way to test the statistical significance of each of the community-level coefficients, namely, $\Gamma^1 \ldots \Gamma^q$. To the potential end-users of MSNR in the broad neuroscience community, this tool would be much more useful if one could make inference on one variable in the context of multivariate analysis.

Conclusions

In summary, this body of work fits into the broad context of computational psychiatry, where there is intense interest in the quest for brain-based biomarkers for psychopathology to overcome the barriers of heterogeneity and co-morbidity in current categorical diagnostic framework. Integrating recent advances in multiple disciplines, across machine learning, network science, developmental neuroscience, and psychiatry, this work delineated common and dissociable functional brain connectivity patterns that are linked to dimensions of psychopathology across clinical boundaries. We also offered a new tool to extend such multivariate method to extract brain-phenotype relationships beyond psychopathology to a wide range of measurements. Going forward, marrying the appropriate hammer to clinically critical questions would be the key to evaluating the suitability of these brain-derived dimensions of psychopathology for charting developmental trajectories and prediction of clinical outcome.

Reference

Bassett, D. S., & Siebenhühner, F. (2013). Multiscale Network Organization in the Human Brain. In *Multiscale Analysis and Nonlinear Dynamics* (pp. 179–204). Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA. https://doi.org/10.1002/9783527671632.ch07

Bassett, D. S., Xia, C. H., & Satterthwaite, T. D. (2018). Understanding the Emergence of Neuropsychiatric Disorders With Network Neuroscience. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*(9), 742–753. https://doi.org/10.1016/J.BPSC.2018.03.015

Betzel, R. F., & Bassett, D. S. (2017). Multi-scale brain networks. *NeuroImage*, *160*, 73–83. https://doi.org/10.1016/J.NEUROIMAGE.2016.11.006

Calkins, M. E., Merikangas, K. R., Moore, T. M., Burstein, M., Behr, M. A., Satterthwaite, T. D., … Gur, R. E. (2015). The Philadelphia Neurodevelopmental Cohort: Constructing a deep phenotyping collaborative. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *56*(12), 1356–1369. https://doi.org/10.1111/jcpp.12416

Cuthbert, B. N., & Insel, T. R. (2010). Toward new approaches to psychotic disorders: The NIMH research domain criteria project. *Schizophrenia Bulletin*, *36*(6), 1061–1062. https://doi.org/10.1093/schbul/sbq108

Fortin, J.-P., Cullen, N., Sheline, Y. I., Taylor, W. D., Aselcioglu, I., Cook, P. A., … Shinohara, R. T. (2018). Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*, *167*, 104–120. https://doi.org/10.1016/J.NEUROIMAGE.2017.11.024

Fortin, J.-P., Parker, D., Tunç, B., Watanabe, T., Elliott, M. A., Ruparel, K., … Shinohara, R. T. (2017). Harmonization of multi-site diffusion tensor imaging data. *NeuroImage*, *161*, 149–170. https://doi.org/10.1016/J.NEUROIMAGE.2017.08.047

Insel, B. T. R., & Cuthbert, B. N. (2015). Brain diorders? Precisely. *Science*, *348*(6234), 499–500.

Insel, T. R. (2014). Mental disorders in childhood: shifting the focus from behavioral symptoms to neurodevelopmental trajectories. *JAMA : The Journal of the American Medical Association*, *311*(17), 1727–1728. https://doi.org/10.1001/jama.2014.1193

Insel, T. R. (2017). Digital Phenotyping. *JAMA*, *318*(13), 1215. https://doi.org/10.1001/jama.2017.11295

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (8th ed.). Springer. https://doi.org/10.1007/978-1-4614-7138-7

Kapur, S., Phillips, A. G., & Insel, T. R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it ? *Molecular Psychiatry*, *17*(12), 1174–1179. https://doi.org/10.1038/mp.2012.105

Satterthwaite, T. D., Connolly, J. J., Ruparel, K., Calkins, M. E., Jackson, C., Elliott, M. A., … Gur, R. E. (2016). The Philadelphia Neurodevelopmental Cohort: A publicly available resource for the study of normal and abnormal brain development in youth. *NeuroImage*, *124*, 1115–1119. https://doi.org/10.1016/J.NEUROIMAGE.2015.03.056

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

Yu, M., Linn, K. A., Cook, P. A., Phillips, M. L., McInnis, M., Fava, M., … Sheline, Y. I. (2018). Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fMRI data. *Human Brain Mapping*, *39*(11), 4213–4227. https://doi.org/10.1002/hbm.24241

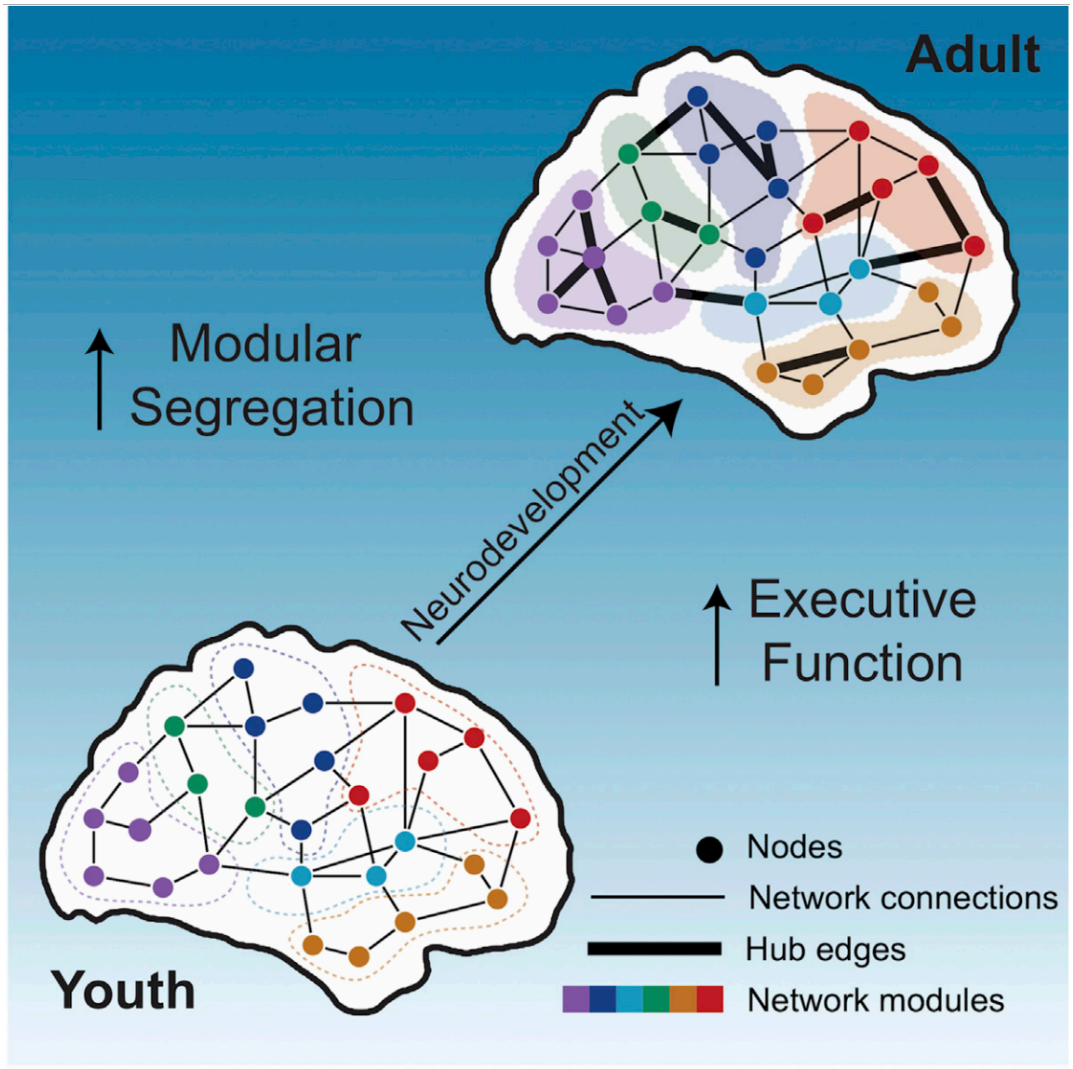
Appendix

A Figure Gallery

Appendix

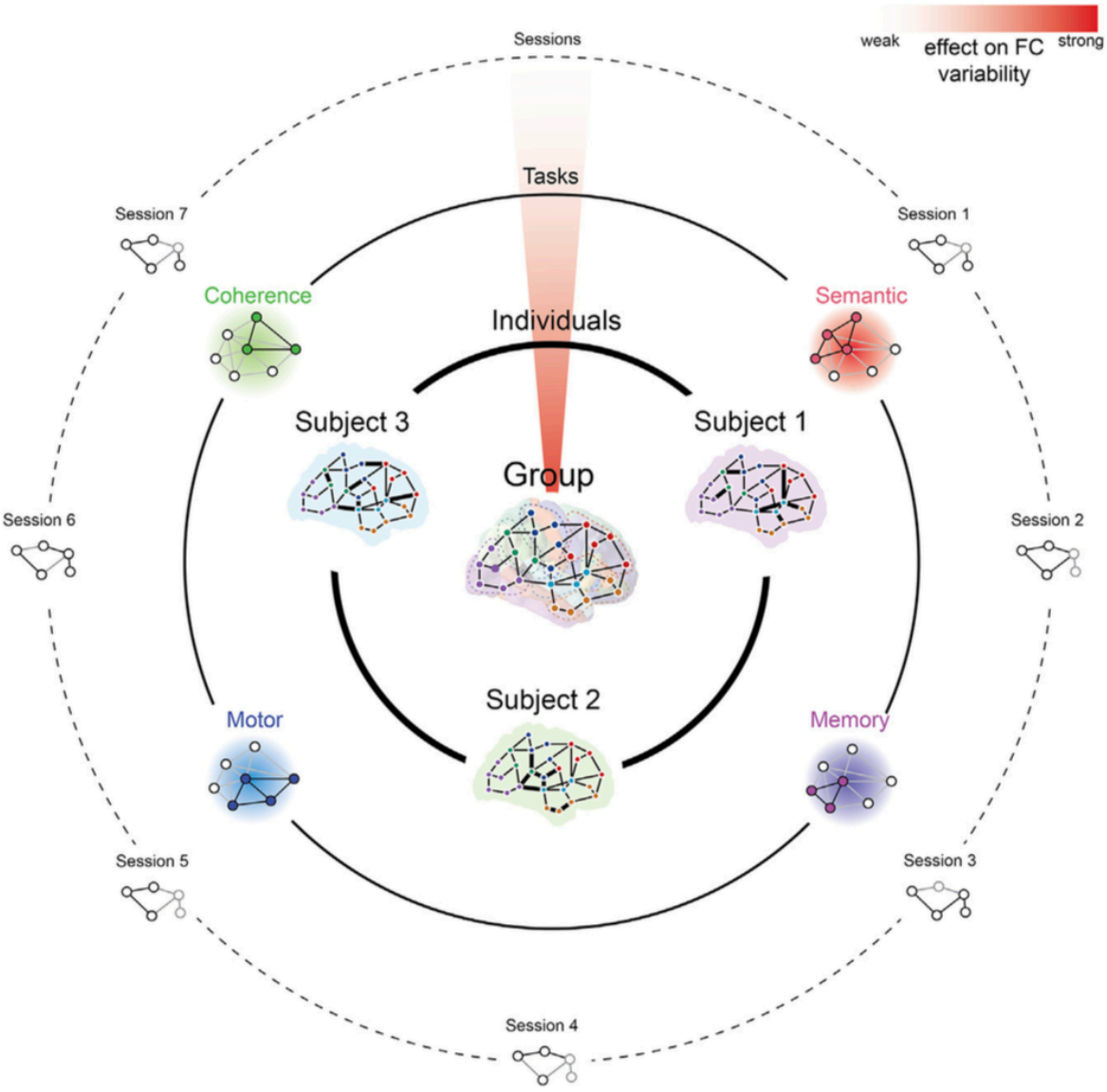
One of my passions in science is the communication of complex data through appealing visualization. In addition to the figures in Chapter 2 and 3 that are part of my first-author work, I also contributed visually to various other projects throughout my graduate work. Below is a select collection of the illustrations I helped create.

Appendix Figure 1



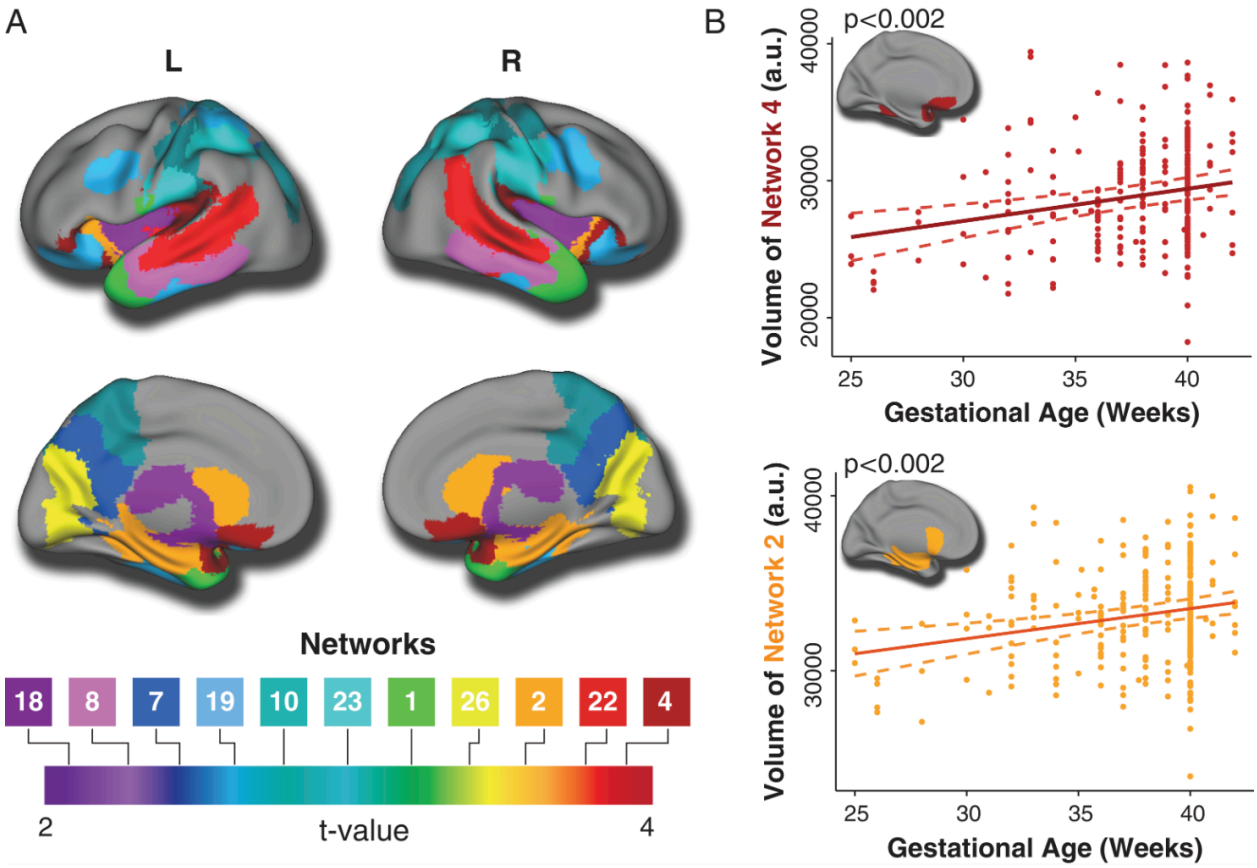
Published as the Graphic Abstract in Baum, G. L., Ciric, R., Roalf, D. R., Betzel, R. F., Moore, T. M., Shinohara, R. T., Kahn, A.E., Vandekar, S.N., Rupert. P.E., Quarmley, M., Cook, P.A., Elliott, M.A., Ruparel, K., Gur, R.E., Gur, R.C., Bassett, D.S., Satterthwaite, T. D. (2017). Modular Segregation of Structural Brain Networks Supports the Development of Executive Function in Youth. Current Biology, 27(11), 1561-1572.e8.

Appendix Figure 2



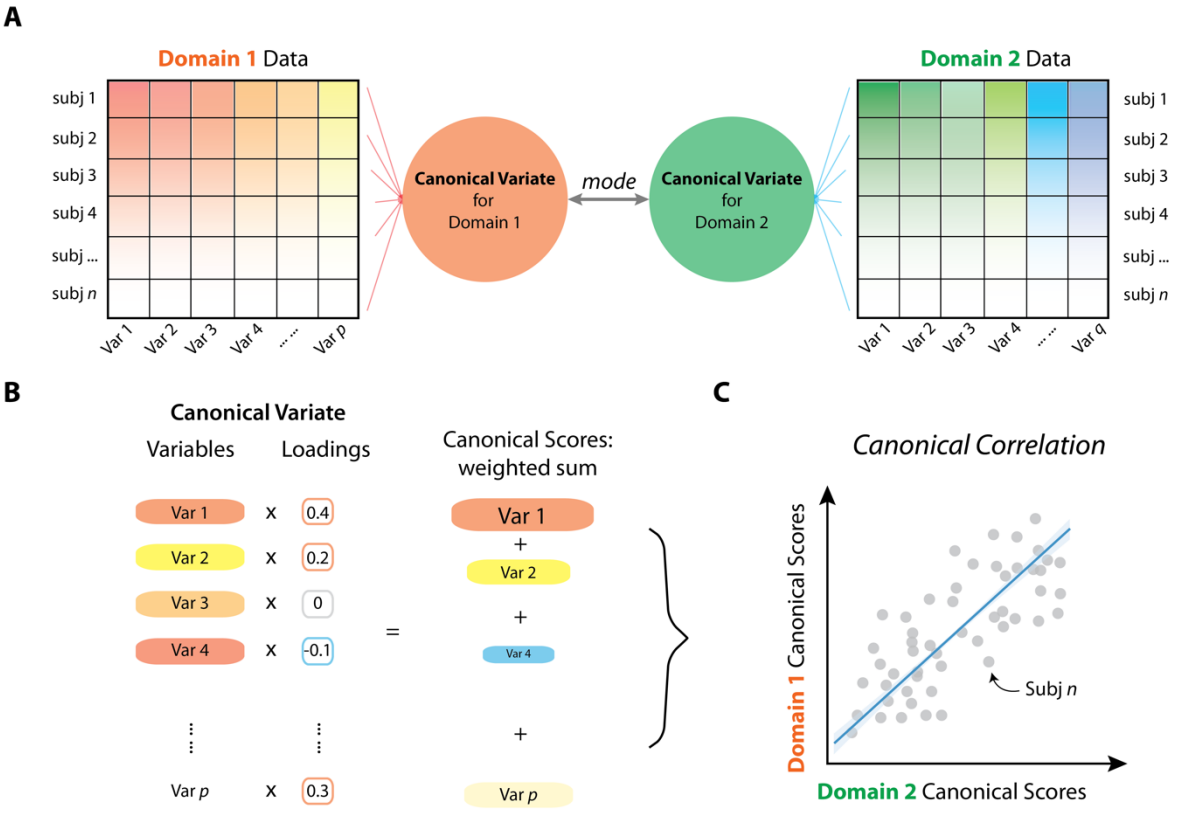
Published as Figure 1 in Satterthwaite, T. D., Xia., C.H., Bassett, D.S., (2018). Personalized Neuroscience: Common and Individual-Specific Features in Functional Brain Networks. Neuron, 98, 243-245

Appendix Figure 3



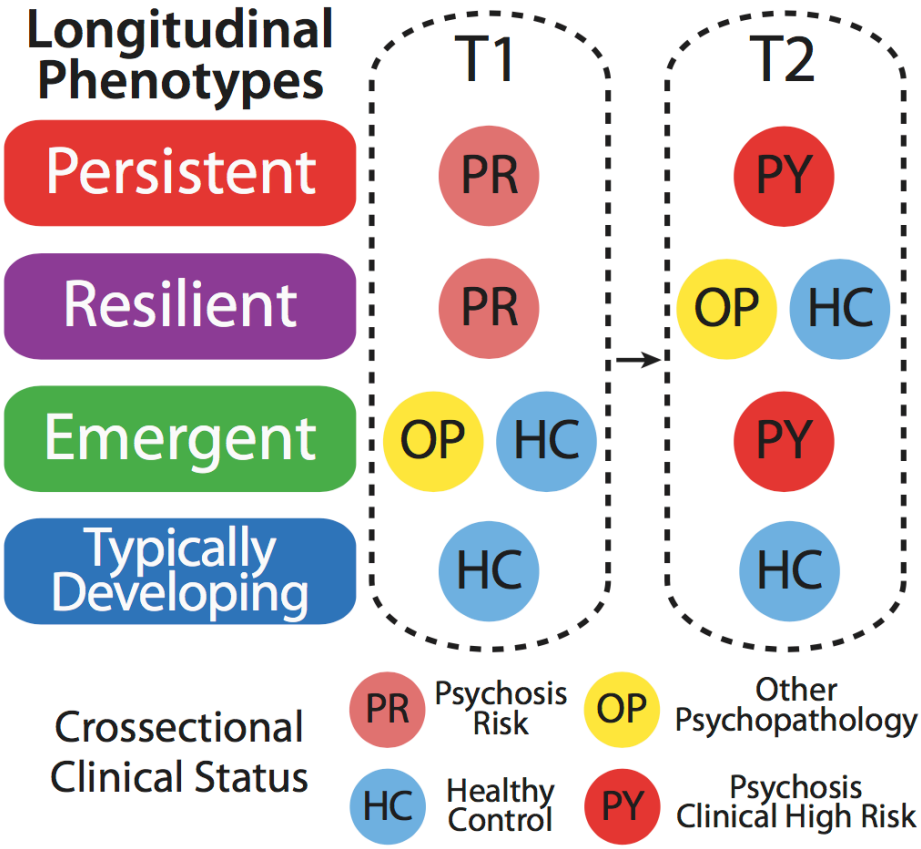
Published as Figure 3 in Nassar, R., Kaczkurkin, A.N., Xia, C.H., Sotiras, A., Pehlivanova, M., Moore, T.M., Garcia de la Garza, A., Roalf, D.R., Rosen, A.F.G., Lorch, S.A., Ruparel, K., Shinohara, R.T., Davatzikos, C., Gur, R.C., Gur, R.E., Satterthwaite, T.D. (2019). Gestational age is dimensionally associated with structural brain network abnormalities across development. Cerebal Cortex, 98(5), 2102-2114

Appendix Figure 4



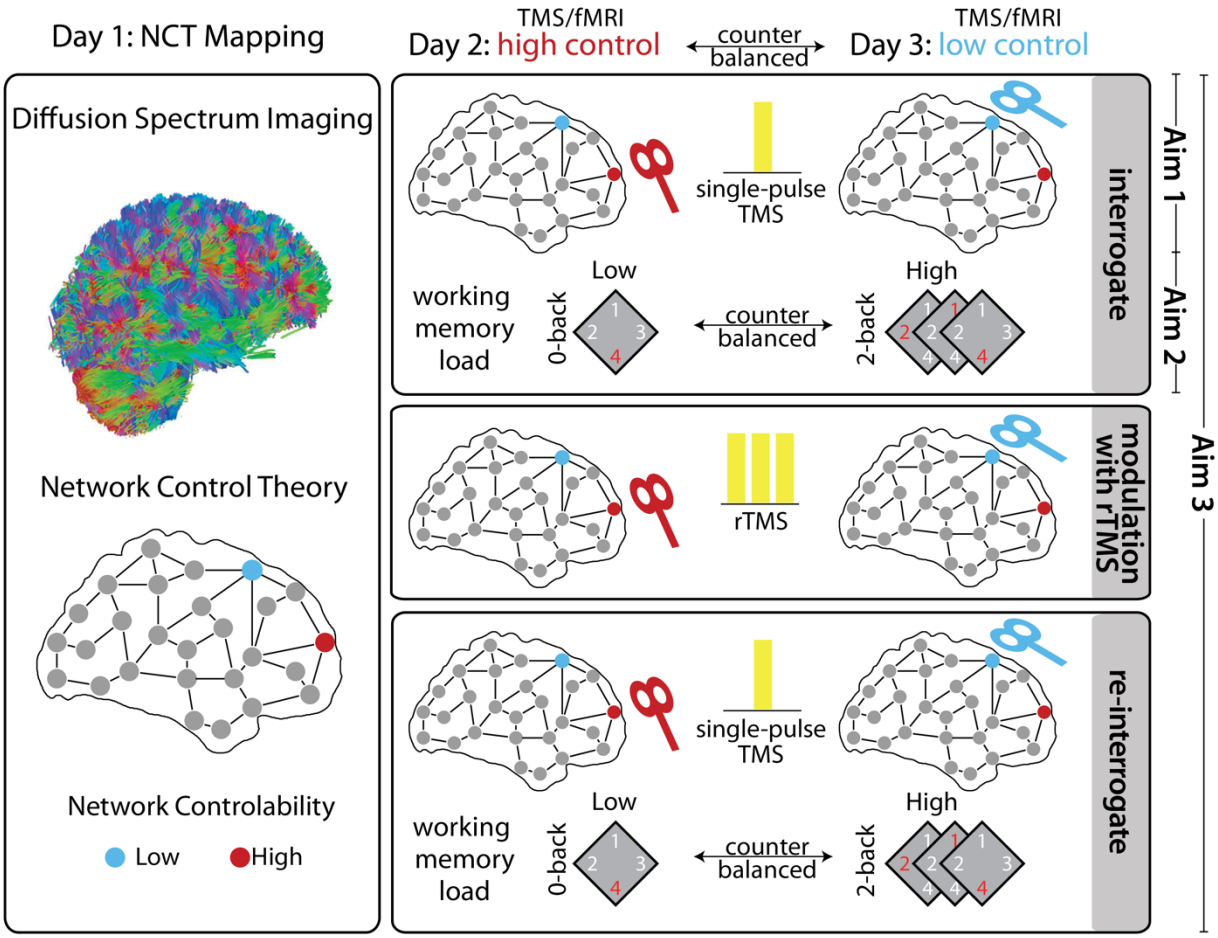
Published as Figure 1 in Wang, H.T., Smallwood, J., Mourao-Miranda, J., Xia, C.H., Satterthwaite, T.D., Bassett, D.S., Bzdok, D. (2019). Finding the needle in high-dimensional haystack: a tutorial on canonical correlation analysis. arXiv:1812.02598

Appendix Figure 5



Prepared originally for Roalf, D.R., Garcia de la Garza, A., Rosen, A., Calkins, M.E., Moore, T.M., Quarmley, M., Ruparel, Ko., Xia, C.H., Rupert, P.E., Satterthwaite, T.D., Shinohara, R.T., Elliott, M.A., Gur, R.C., Gur, R.E. (2019) Alterations in white matter microstructure in individuals at persistent risk for psychosis. Molecular Psychiatry

Appendix Figure 6



Prepared originally for a now funded R01 grant by Desmond Oathes, Danielle Bassett, and Ted Satterthwaite, named “network control and functional context: mechanisms for TMS response”.