NeuroImage 93 (2014) 74–94

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

A multivariate distance-based analytic framework for connectome-wide association studies

Zarrar Shehzad [a,b,c,*], Clare Kelly [d], Philip T. Reiss [e,b], R. Cameron Craddock [b,c], John W. Emerson [f], Katie McMahon [g], David A. Copland [h], F. Xavier Castellanos [d,b], Michael P. Milham [c,b,**]

[a] Department of Psychology, Yale University, New Haven, CT, USA
[b] Nathan Kline Institute for Psychiatric Research, Orangeburg, NY, USA
[c] Center for the Developing Brain, Child Mind Institute, New York, NY, USA
[d] Phyllis Green and Randolph Cōwen Institute for Pediatric Neuroscience, Department of Child and Adolescent Psychiatry, New York University, New York, NY, USA
[e] Division of Biostatistics, Department of Child and Adolescent Psychiatry, New York University, New York, NY, USA
[f] Department of Statistics, Yale University, New Haven, CT, USA
[g] Centre for Advanced Imaging, University of Queensland, Brisbane, Queensland, Australia
[h] Centre for Clinical Research and School of Health and Rehabilitation Sciences, University of Queensland, Brisbane, Queensland, Australia

ARTICLE INFO

Article history:
Accepted 19 February 2014
Available online 28 February 2014

Keywords:
Brain–behavior relationships
Connectome
Discovery
Functional connectivity
Phenotype
Resting-state

ABSTRACT

The identification of phenotypic associations in high-dimensional brain connectivity data represents the next frontier in the neuroimaging connectomics era. Exploration of brain–phenotype relationships remains limited by statistical approaches that are computationally intensive, depend on a priori hypotheses, or require stringent correction for multiple comparisons. Here, we propose a computationally efficient, data-driven technique for connectome-wide association studies (CWAS) that provides a comprehensive voxel-wise survey of brain–behavior relationships across the connectome; the approach identifies voxels whose whole-brain connectivity patterns vary significantly with a phenotypic variable. Using resting state fMRI data, we demonstrate the utility of our analytic framework by identifying significant connectivity–phenotype relationships for full-scale IQ and assessing their overlap with existent neuroimaging findings, as synthesized by openly available automated meta-analysis (www.neurosynth.org). The results appeared to be robust to the removal of nuisance covariates (i.e., mean connectivity, global signal, and motion) and varying brain resolution (i.e., voxelwise results are highly similar to results using 800 parcellations). We show that CWAS findings can be used to guide subsequent seed-based correlation analyses. Finally, we demonstrate the applicability of the approach by examining CWAS for three additional datasets, each encompassing a distinct phenotypic variable: neurotypical development, Attention-Deficit/Hyperactivity Disorder diagnostic status, and L-DOPA pharmacological manipulation. For each phenotype, our approach to CWAS identified distinct connectome-wide association profiles, not previously attainable in a single study utilizing traditional univariate approaches. As a computationally efficient, extensible, and scalable method, our CWAS framework can accelerate the discovery of brain–behavior relationships in the connectome.

Introduction

The human connectome, comprising the complete set of neural interactions in the brain, provides the framework for behavior and cognition (Craddock et al., 2013; Sporns, 2011; Sporns et al., 2005). A key challenge for neuroscience is to understand the relationship between inter-individual variations in the organization of functional systems within the connectome, and environmental and phenotypic factors (Akil et al., 2011; Biswal et al., 2010). Phenotypic variables such as task performance, psychological traits, and disease states have been found to be associated with variation within and between specific functional brain circuits (Andrews-Hanna et al., 2010; Fornito and Bullmore, 2010; Greicius, 2008; Kelly et al., 2008, 2012; Zhu et al., 2008). However, connectome-wide association studies (CWAS) permitting the exploration of brain–behavior relationships across the entire connectome remain a challenge as they entail a massive number of comparisons (Milham, 2012). For example, an investigation of connectivity for 25,000 voxels requires considering more than 300 million voxel pairings.

As in genome-wide association studies (Burton et al., 2007; McCarthy et al., 2008), investigations of connectome-wide associations typically employ mass-univariate statistical analyses. In the univariate approach,

* Correspondence to: Z. Shehzad, Department of Psychology, Yale University, 2 Hillhouse Ave., New Haven, CT, USA.
** Correspondence to: M. P. Milham, Center for the Developing Brain, Child Mind Institute, 445 Park Avenue, New York, NY 10022 New York, NY, USA. Fax: +1 646 625 4348.
E-mail addresses: zarrar.shehzad@gmail.com (Z. Shehzad), michael.milham@childmind.org (M.P. Milham).

http://dx.doi.org/10.1016/j.neuroimage.2014.02.024
1053-8119/

a phenotypic measure is related to only one functional connection at a time (e.g., between regions of interest [ROIs], or between voxels in a whole-brain analysis); concurrent contributions from other connections are necessarily ignored (Cole et al., 2010). The large number of statistical tests entailed by this approach (thousands to millions) increases the potential for false positives, requiring stringent correction for multiple comparisons (Benjamini and Hochberg, 1995; Chumbley and Friston, 2009; Genovese et al., 2002; Hu et al., 2010; Worsley et al., 2005; Zalesky et al., 2010a). In addition, visualization and interpretation of results from such a massive number of univariate analyses in three dimensions represent major challenges (e.g., visualization of the voxel-wise connectome) (Margulies et al., 2013).

Multivariate learning methods have been advocated as an alternative approach for exploring connectivity-phenotype associations (Margulies et al., 2010). In the multivariate framework, the *simultaneous* contribution of *entire sets* of functional connections to a phenotypic variable (e.g., age, clinical diagnosis, behavioral performance) is evaluated. This reduces the number of connectivity–phenotype evaluations to one per set and thus reduces the scale of the multiple comparison problem. Given that cognitive and perceptual processes are driven by patterns of concurrent activity across distributed brain networks rather than individual regions (Haynes and Rees, 2006; Norman et al., 2006), simultaneous assessment of multiple connections may capture connectivity–phenotype relationships more accurately (Varoquaux and Craddock, 2013). While a variety of multivariate approaches can assess phenotypic associations in the connectome (Varoquaux and Craddock, 2013), several factors led us to pursue multivariate distance matrix regression (MDMR) (Anderson, 2001; McArdle and Anderson, 2001; Reiss et al., 2010; Zapala and Schork, 2006). These include: 1) the ability to examine more than one predictor variable at a time (i.e., covariates can be incorporated), 2) applicability for categorical and/or continuous variables, 3) minimal requirements for parameter-specific or analytic decision-making (e.g., a user only needs to choose the distance measure), and 4) ease of interpretability due to regression-like analytic structure. MDMR has also been shown to have excellent test level accuracy and good statistical power (Schork et al., 2008; Zapala and Schork, 2006, 2012; Lin and Schaid, 2009).

Here, we provide a whole-brain framework for identifying phenotypic associations in the connectome. While the approach is illustrated using several examinations of functional connectivity, it can also be applied to structural connectivity. At each node in the connectome (defined by voxels, brain areas, or parcellation units), we test whether inter-individual whole-brain connectivity patterns are related to differences in one or more phenotypic variables of interest. This is accomplished using a two-step approach. First, for each node in the connectome, we calculate a whole brain functional connectivity map, and then calculate the similarity between the connectivity maps of all possible pairings of participants using spatial correlation, yielding an $n \times n$ matrix (n = number of participants). Then, at each node, we use MDMR to test whether a variable of interest (e.g., a clinical diagnosis) is associated with the between-subject distances: significance is determined using permutation testing (Fig. 1). The end result is a statistic for each node that indicates the strength of the relationship between a phenotypic measure and variations in its connectivity patterns across participants.

Our approach attempts to lower common barriers to effective discovery science and full-brain exploration of the connectome (Biswal et al., 2010; Van Horn and Gazzaniga, 2002); in particular, the high computational demands and resulting necessity to incorporate a priori information to constrain the problem. The proposed MDMR-based framework does not require the user to pre-specify or estimate the dimensionality of the data (as in independent component analysis or clustering; Beckmann, 2012; Bellec et al., 2010; Damoiseaux et al., 2006; Hartigan, 1975; Mckeown et al., 1998) or select parameters for graph construction (as in network centrality-based approaches; Buckner et al., 2009; Bullmore and Sporns, 2009; Bullmore and Bassett, 2011; Zuo et al., 2012). The resolution of brain representations (e.g., voxels) does not need to be reduced to facilitate computation, as is common with graph theoretic analyses (Buckner et al., 2009; Cole et al., 2010; Zalesky et al., 2010b). Finally, there is no need to select particular seeds or networks, as in seed-based correlation analyses (Cole et al., 2010).

Our primary demonstration and evaluation of the MDMR-based CWAS approach focuses on the identification of connectome-wide associations for IQ using resting state data from the publicly available Enhanced Nathan Kline Institute Rockland Sample (NKI-RS; http://fcon_1000.projects.nitrc.org/indi/enhanced). Advantages of this dataset include moderate sample size (n = 104) and multiple resting-state scans (albeit with different imaging sequence parameters). This enabled us to evaluate the robustness of our results, their overlap with existing meta-analyses of IQ–brain relationships, and their utility in guiding subsequent seed-based correlation analyses. The impact of potential confound signals (e.g., motion), preprocessing strategies (e.g., global signal regression), and brain parcellation strategies were considered, as well as overall differences in measured brain connectivity from one individual to the next. In addition, we provide three other examples of MDMR-based CWAS applications using a range of phenotypes and experimental designs (Attention-Deficit/Hyperactivity Disorder [ADHD] vs. controls, age-related developmental effects, and administration of L-DOPA vs. placebo). A summary of our approach is provided in Fig. 1, and an overview of analyses is provided in Fig. 2. Code for all analyses and figures in the paper are available online (https://github.com/czarrar/cwas-paper).

Methods

Participants

We examined resting-state fMRI scans from four community-based datasets (see Table 1 for demographics). The four datasets included: 1) *IQ*: healthy adults with Full Scale IQ estimated with the Wechsler Abbreviated Scale of Intelligence (WASI) from the NKI-RS with ages 18 to 65 (http://fcon_1000.projects.nitrc.org/indi/enhanced; Nooner et al., 2012); 2) *Development*: healthy individuals ranging from children to young adults (http://fcon_1000.projects.nitrc.org/indi/retro/Power2012.html; Power et al., 2012); 3) *ADHD*: typically developing children and children meeting DSM-IV criteria for ADHD sampled from the NYU site of the ADHD200 dataset that were matched for age and sex (http://fcon_1000.projects.nitrc.org/indi/adhd200; Chabernaud et al., 2012; Di Martino et al., 2011; Kelly et al., 2009; Koyama et al., 2011; Mennes et al., 2012; Zuo et al., 2010); and 4) *L-DOPA*: healthy adults administered 100 mg of L-DOPA or placebo double-blind on two separate scan visits (Kelly et al., 2009). Datasets 1–3 and the placebo scans from Dataset 4 are publicly available for download from the International Neuroimaging Data-sharing Initiative (INDI) at http://fcon_1000.projects.nitrc.org.

Data acquisition

For the first dataset (IQ), imaging data were acquired using a Siemens Tim Trio 3 T scanner with a 32-channel head coil at the Center for Advanced Brain Imaging, NKI. Three different resting-state fMRI scans were collected in the following order: (i) Scan 1, a multiband echoplanar imaging (EPI) sequence (Moeller et al., 2010; Xu et al., 2013) (900 time points, repetition time [TR] = 645 ms, echo time [TE] = 30 ms, flip angle = 60°, 40 slices, voxel size = 3 × 3 × 3 mm), (ii) Scan 2, a higher spatial resolution multiband EPI sequence (404 time points, TR = 1400 ms, TE = 30 ms, flip angle = 60°, 64 slices, voxel size = 2 × 2 × 2 mm), and (iii) Scan 3, a standard EPI sequence (120 time points, TR = 2500 ms, TE = 30 ms, flip angle = 80°, 38 slices, voxel size = 3 × 3 × 3.33 mm). Scan 3 was not included in the present study because its duration is half that of Scans 1 and 2, decreasing