Connectome-Wide Association Studies (CWAS)

Zarrar Shehzad[1], Philip T Reiss[2,3], Jonathan Adelstein[2], John W Emerson[4], Camille Chabernaud[2], Maarten Mennes[2], Adriana DiMartino[2], Katie McMahon[5], David Copland[6], F Xavier Castellanos[2,3], Clare Kelly[2], Michael P Milham[2,3]

1 Department of Psychology, Yale University, New Haven, CT, USA; **2** Department of Child and Adolescent Psychiatry, New York University, New York, NY, USA; **3** Nathan Kline Institute for Psychiatric Research, Orangeburg , NY, USA; **4** Department of Statistics, Yale University, New Haven, CT, USA; **5** Center for Advanced Imaging, University of Queensland, Brisbane, QLD, Australia; **6** Language Neuroscience Lab, University of Queensland, Brisbane, QLD, Australia

zarrar.shehzad@yale.edu

Introduction

- A challenge for neuroscience is to understand brain-behavior relationships in the functional connectome
- Comprehensive analyses have been lacking due to statistical approaches that are computationally intensive or dependent on *a-priori* hypotheses
- To overcome these limitations, we
  - Probe the entire connectome using a novel data-mining technique
  - Relate a phenotype to every voxel's pattern of whole-brain functional connectivity instead of every individual connection

Methods

Tested on 4 Resting State fMRI Datasets (R-fMRI):

- ***Age***[1,2]: 96 healthy children to young adults (aged 7-25; 53 F)
- ***Childhood ADHD***[1,3]: 37 with (15 F) and 37 without (21 F) ADHD (aged 7-13)
- ***IQ***[1]: 58 healthy adults (aged 18-48; 25 F) with full-scale IQ using WASI
- ***L-dopa Drug Manipulation***[4]: 19 healthy adults (mean age 26.2) administered 100mg L-dopa or placebo on separate scan visits

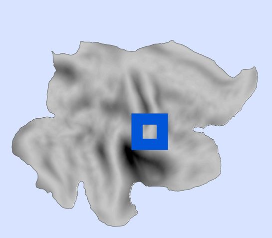
Preparation of Functional Connectivity Data:

- Preprocessing was in line with previous R-fMRI studies[2,4,5]
- Preprocessed time-series were transformed to MNI152 space ($4mm^3$)

Connectome-Wide Association Studies (CWAS):

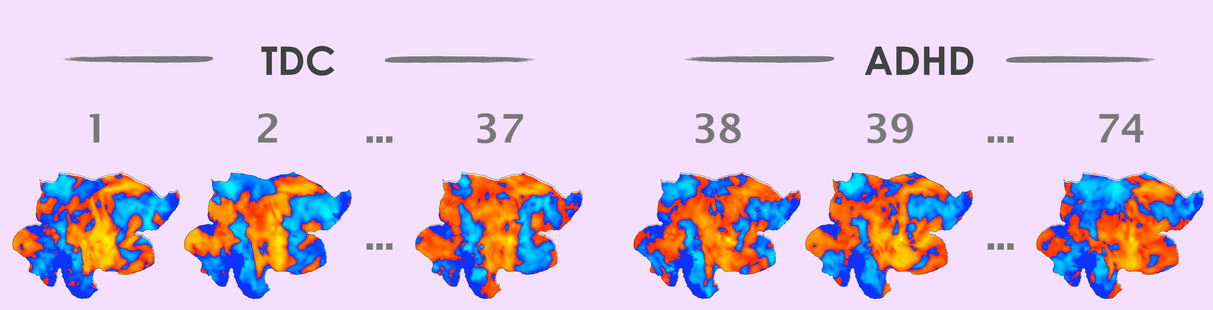
- Analyses were performed in R (version 2.11) using 4 2.66-GHz Intel Xeon processors and a maximum of 12 GB of RAM
- Code is publicly available through the R package *connectir*[6]
- Voxels with $p < 0.05$ were considered significant; for multiple comparison correction, a minimum cluster size ($p < 0.05$) was set based on Monte Carlo simulations using *3dClustSim*[7]
- To illustrate our approach, a **sample workflow** is presented below
  - *Goal*: Find voxels whose whole-brain connectivity patterns significantly differentiate typically developing children (TDC) from children with ADHD
  - *Connectivity Map/Pattern:* A temporal Pearson correlation (r) between a given voxel and every other grey-matter voxel in the brain
  - *Abbreviation*: MDMR = Multivariate Distance Matrix Regression

1. Voxel

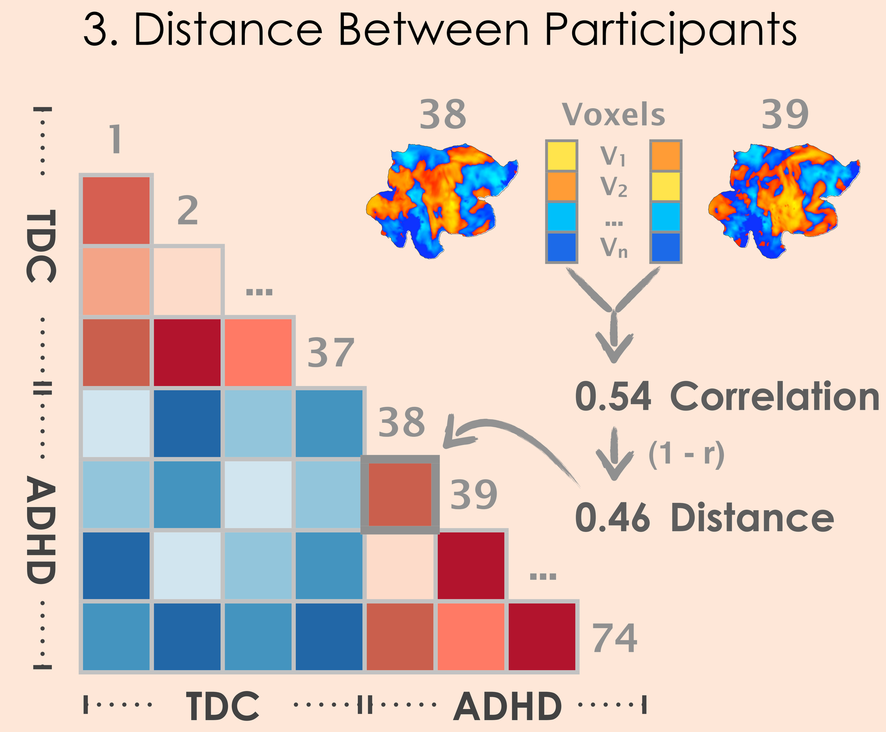


Note: Repeat for every voxel

2. Participant Connectivity Maps



3. Distance Between Participants



4. MDMR[5,8,9]

	df	SS	MS	*F*
Group	1	0.98	0.98	2.9
Error	72	24.6	0.34	

Estimate how well group membership explains the distances between participants, using a pseudo-*F* statistic.

5. Significance

Permutation test: $p < 0.05$.

Shuffle group membership and recompute pseudo-*F* statistic, 15,000 times.

Details for MDMR[5]

The distance matrix must be transformed and centered before #4.

$$A = (a_{ij}) = \left(-\frac{1}{2}d_{ij}^2\right)_{1 \le i, j \le n}$$

note: ***d*** = 1-r, **G** = Gower's centered matrix, **I** = identity matrix, **1** = vector of *n* 1's.

$$\mathbf{G} = \left(\mathbf{I} - \frac{1}{n}11^T\right)A\left(\mathbf{I} - \frac{1}{n}11^T\right)$$

$$F = \frac{tr(\mathbf{HGH})/(m-1)}{tr[(\mathbf{I}-\mathbf{H})\mathbf{G}(I-\mathbf{H})]/(n-m)}$$

A pseudo-F statistic is computed for each explanatory variable.

note: **H** = hat matrix, ***m*** = # of predictors, & ***n*** = # of participants.

Figure 1: CWAS

Shown are **voxels** whose **whole-brain functional connectivity maps significantly vary according to age (7-25 years)**, **childhood ADHD diagnosis**, **IQ (range 86-129)**, or **L-dopa drug manipulation**.

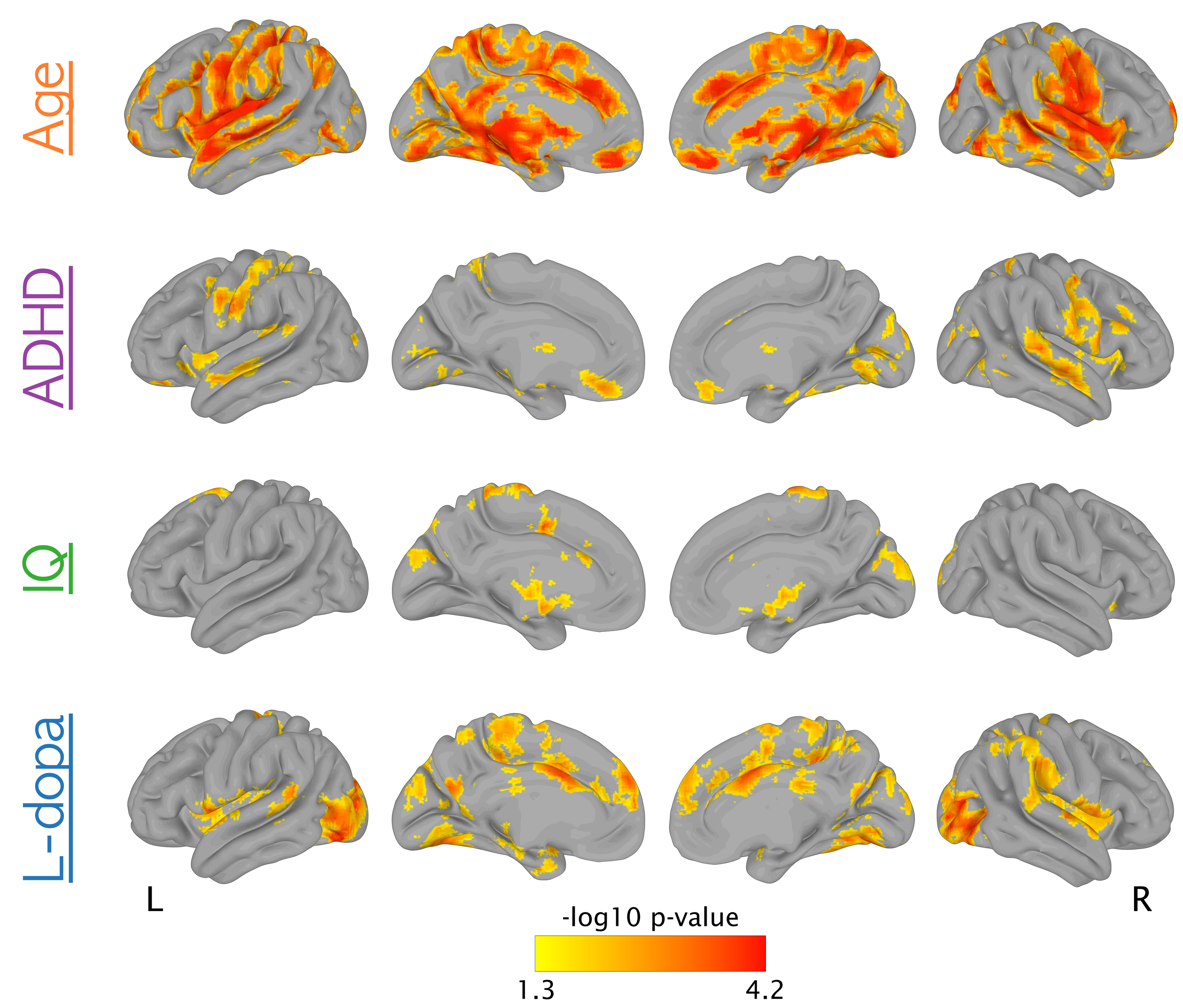
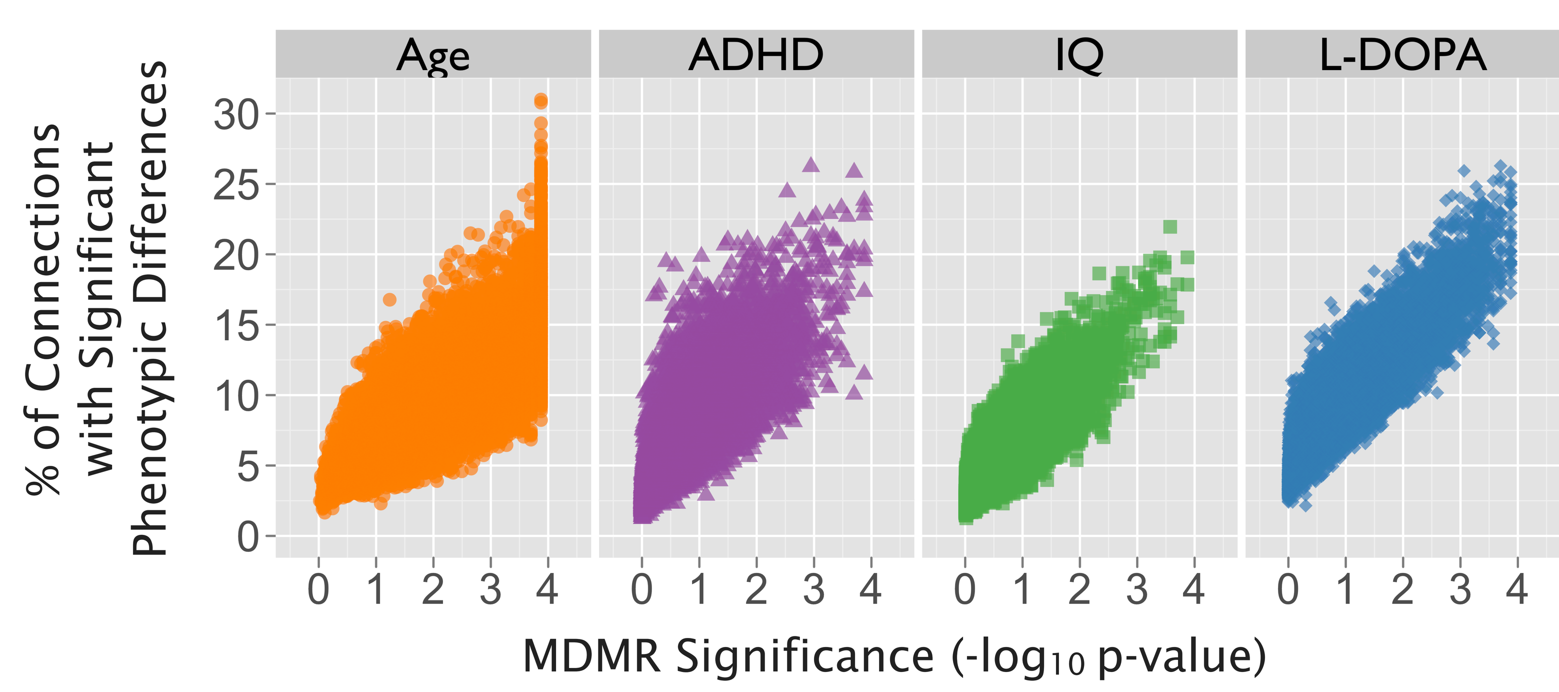


Figure 2: Validation

Voxels with high MDMR significance exhibit a large number of individual connections with significant connectivity-phenotype relationships. Thus, **CWAS constitutes a screening tool that can inform follow-up seed-based correlation analyses**.



Discussion

- Here we have:
  - Demonstrated the feasibility of CWAS
  - Captured the complete set of connectivity-phenotype relationships that previously required several traditional connectivity studies
- Our method is ideal for mining neuroscience data:
  - *data-driven:* minimal *a-priori* hypotheses or analytic parameters
  - *computationally efficient*: each CWAS took 1-3 hours
  - *extensible*: can be applied to any brain-based measure
  - *scalable*: can handle large-scale public datasets[1]
  - *discovery rich*: can provide new hypotheses for follow-up analyses

References

1. http://fcon_1000.projects.nitrc.org
2. Zuo X (2010). *J Neuro*. 30: 15034-43.
3. Also see posters #129 & #158
4. Kelly C (2009). *J Neuro*. 29: 7364:78.
5. Reiss PT (2010). *Biometrics*. 66: 636-43.
6. http://scanlab.psych.yale.edu/connectir
7. http://afni.nimh.nih.gov/afni
8. McArdle BH (2001). *Ecology*. 82: 290-7.
9. Zapala M (2006). *PNAS*. 103: 19430-5.

Supported by NSERC Post-Graduate Fellowship (ZS) & grants from Australian Research Council [DP0452264 (KM, DC)], Leon Levy Foundation (MPM, ADM), NIMH [R01MH081218 (FXC), K23MH087770 (ADM)], and Stavros Niarchos Foundation (FXC).