

Developmental Mega Sample: Exploring Inter-Individual Variation

Saige Rutherford

Sripada Lab

Psychiatry, University of Michigan



**Philadelphia
Neurodevelopmental
Cohort**



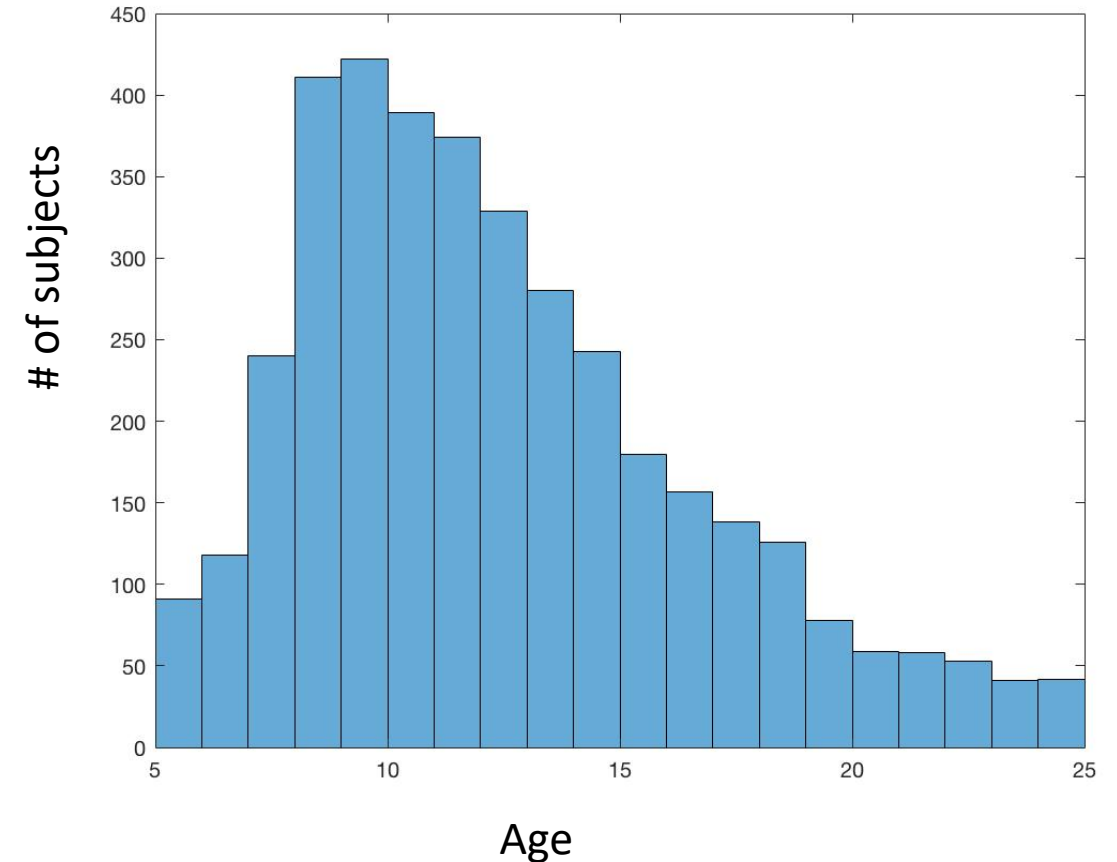
Adolescent Brain Cognitive Development

The Data

Inclusion criterion:

- Age range 5-25 years
- Resting-state + T1w images
- Site contributes > 30 subjects
- At least 4min of resting state data (after censoring)

| Study | # of Sites | # of Subjects |
|----------|------------|---------------|
| ABIDE 1 | 13 | 713 |
| ABIDE 2 | 11 | 648 |
| ADHD 200 | 6 | 630 |
| CMI | 2 | 451 |
| PNC | 1 | 724 |
| PING | 2 | 163 |
| NKI-RS | 1 | 90 |



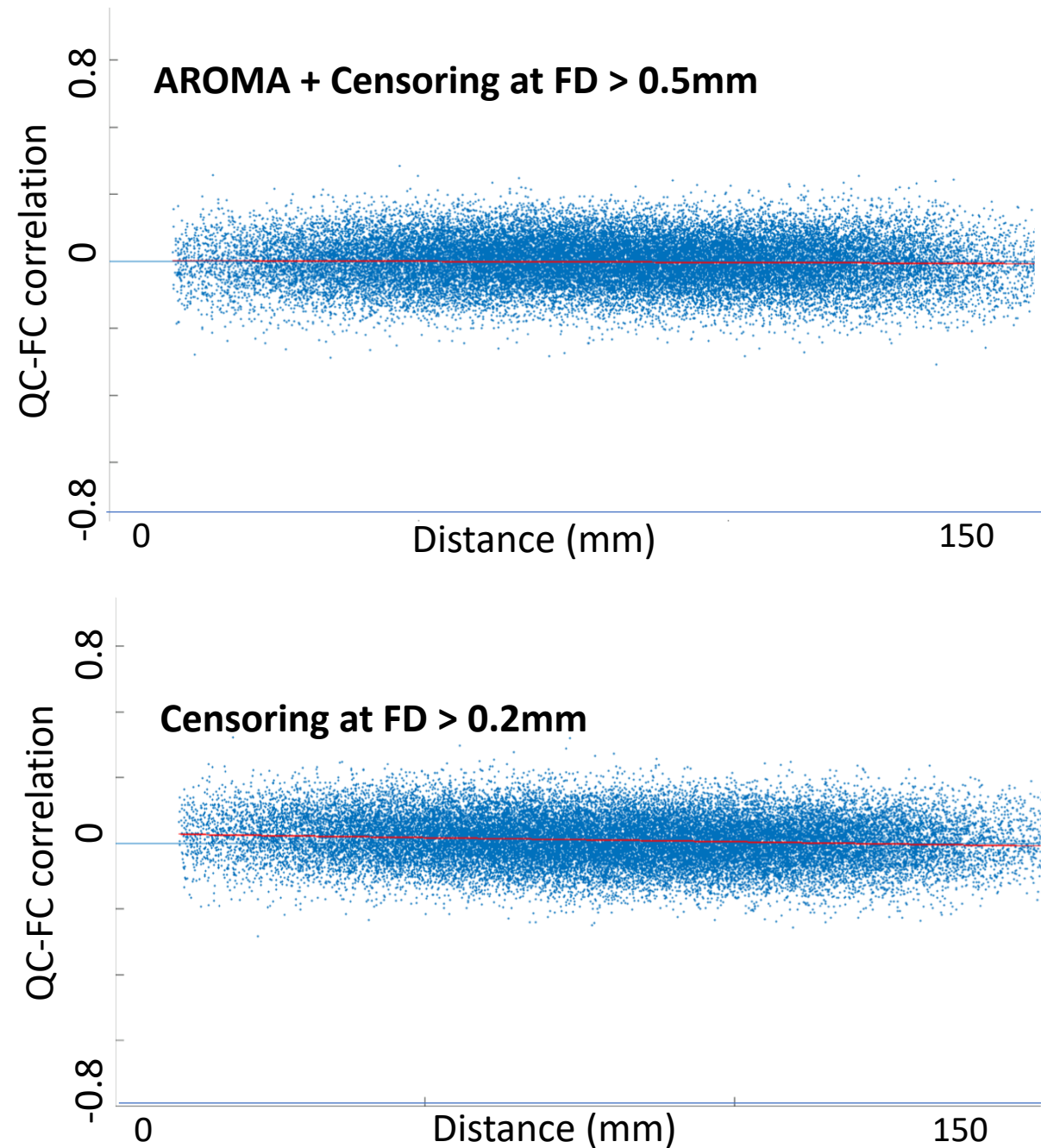
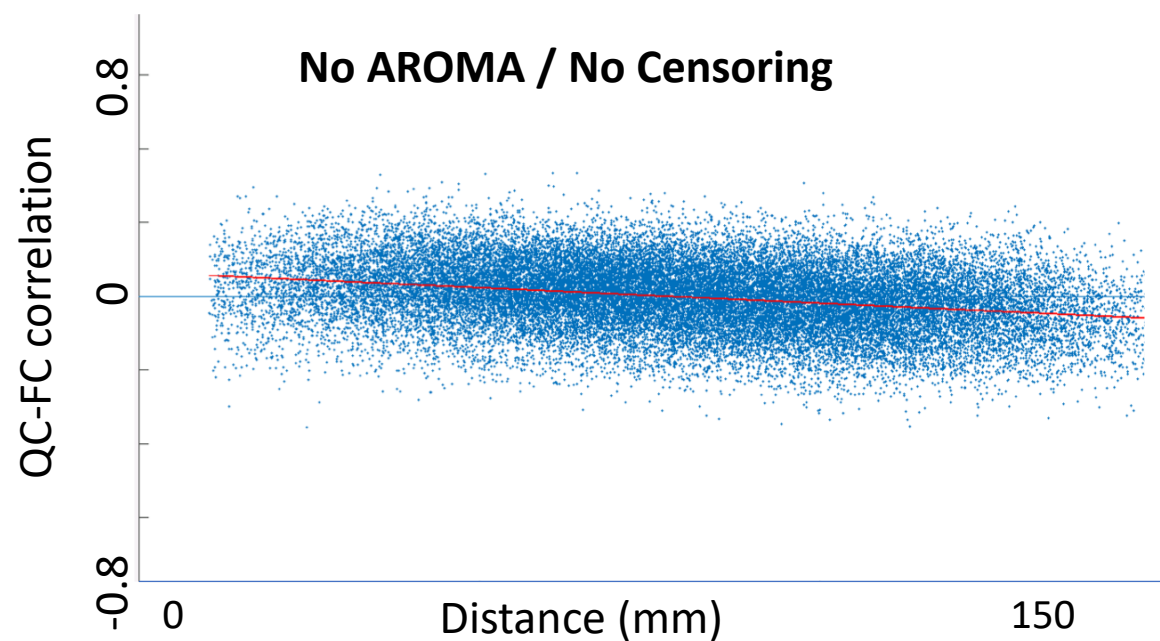
3,419 subjects!

How can we (properly) unite these data?

- Harmonize data
 - ✓ Consistent preprocessing pipeline for all studies/sites
- Quality control
 - ✓ Visual inspection of **all** data at multiple points
 - ✓ Create data quality score for each subject
- AROMA*
 - ✓ Robust ICA-based de-noising for reducing the effects of motion in the data
 - ✓ AROMA + censoring at $FD > 0.5\text{mm}$ is more effective than only censoring at $FD > 0.2\text{mm}$

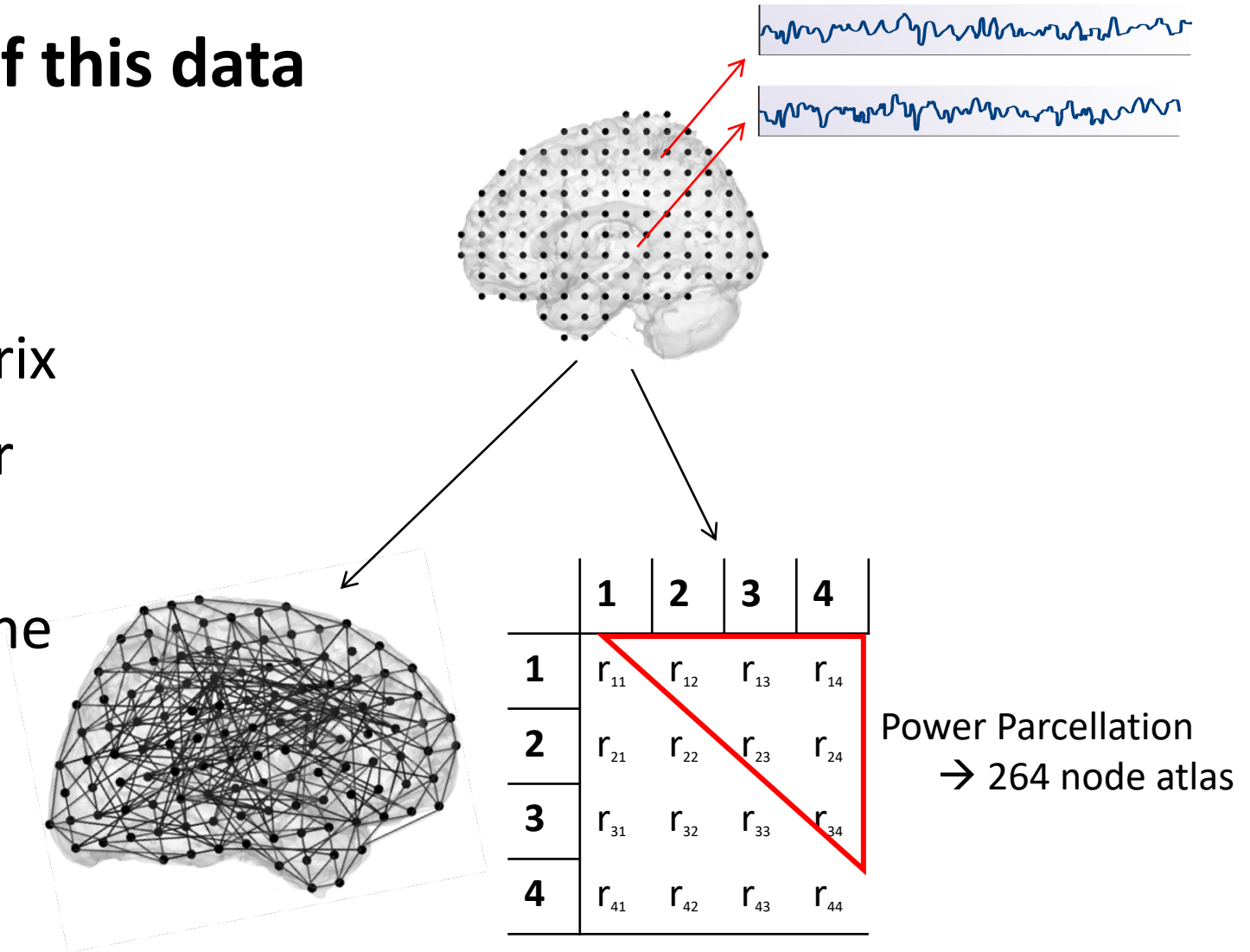
*Automatic Removal Of Motion Artifact

QC-FC Distance Dependence



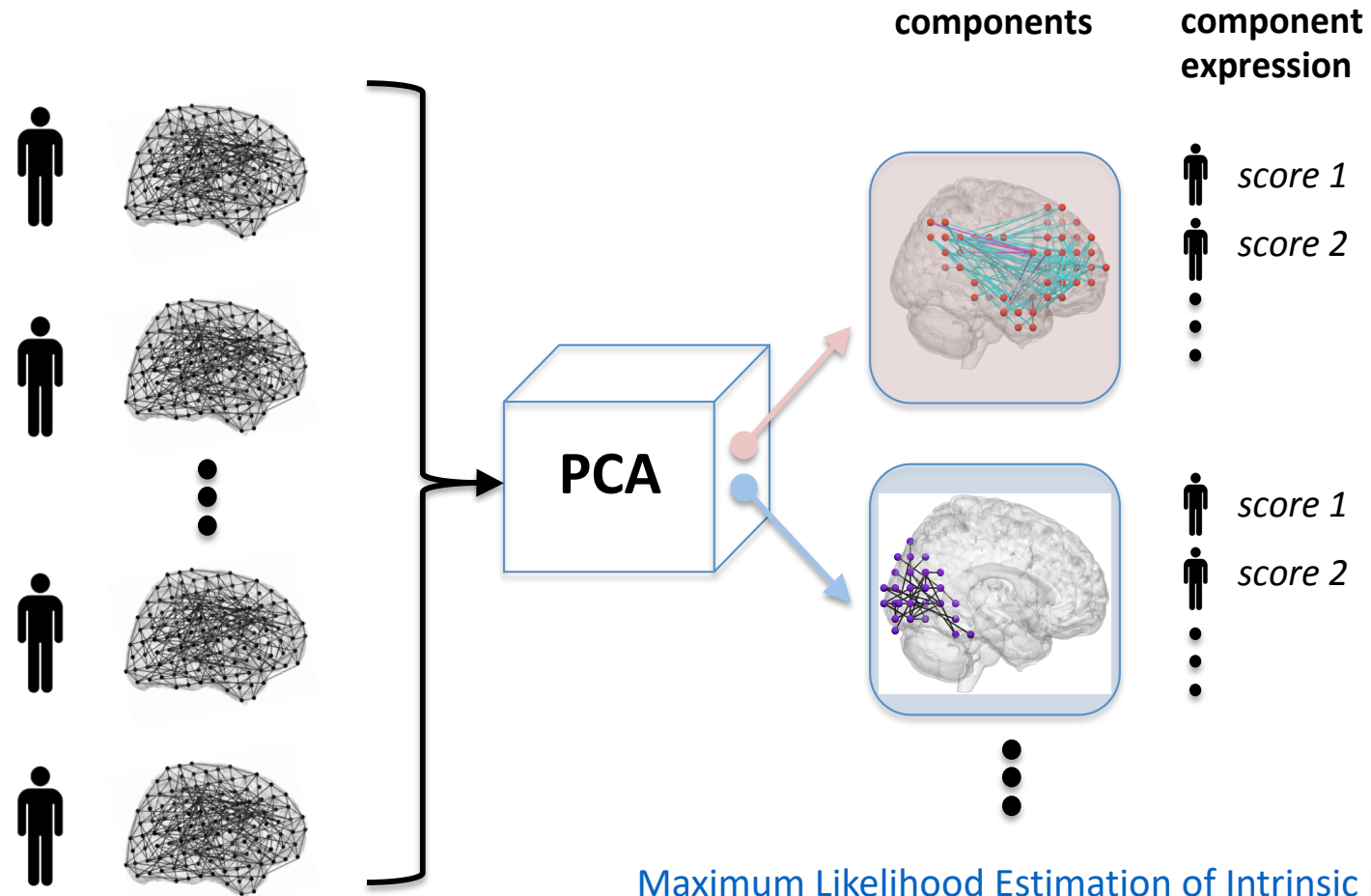
Time to think about some of the cool things we can do with all of this data

- ~3,400 Subjects
- Each subject contributes 264 x 264 connectivity matrix
- 34,584 unique variables per subject?
- 117,585,600 variables for the entire sample?!?!



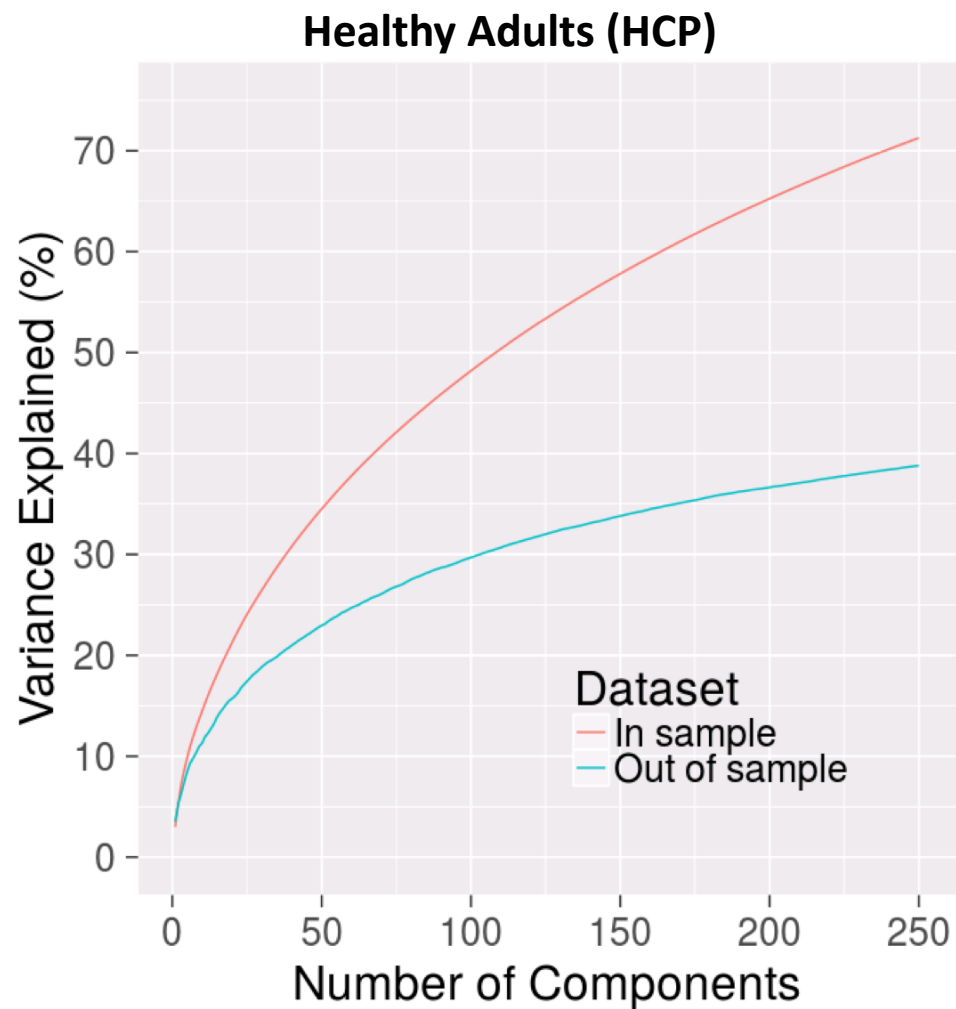
The Connectome-Basis Set

Functional Connectomes



→ Initial method to determine # of principal components to keep

Connectome Basis Set: Units of Inter-Individual Variation

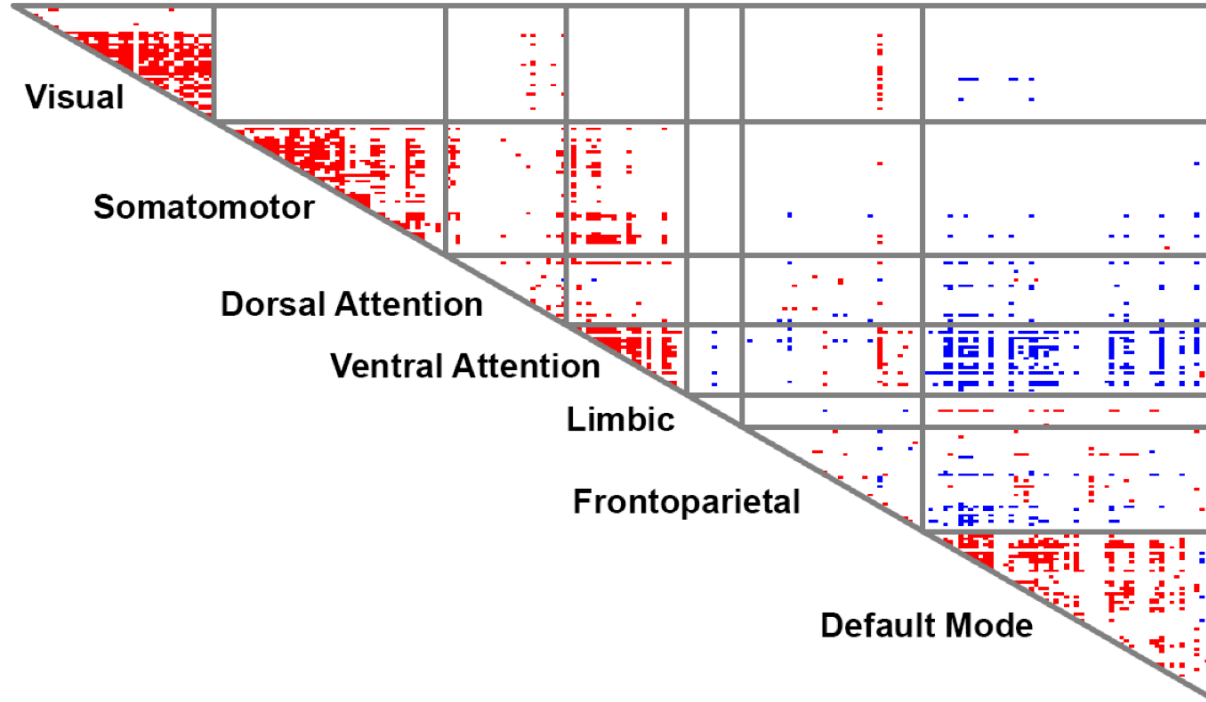


62 components → 25% out of sample variance explained

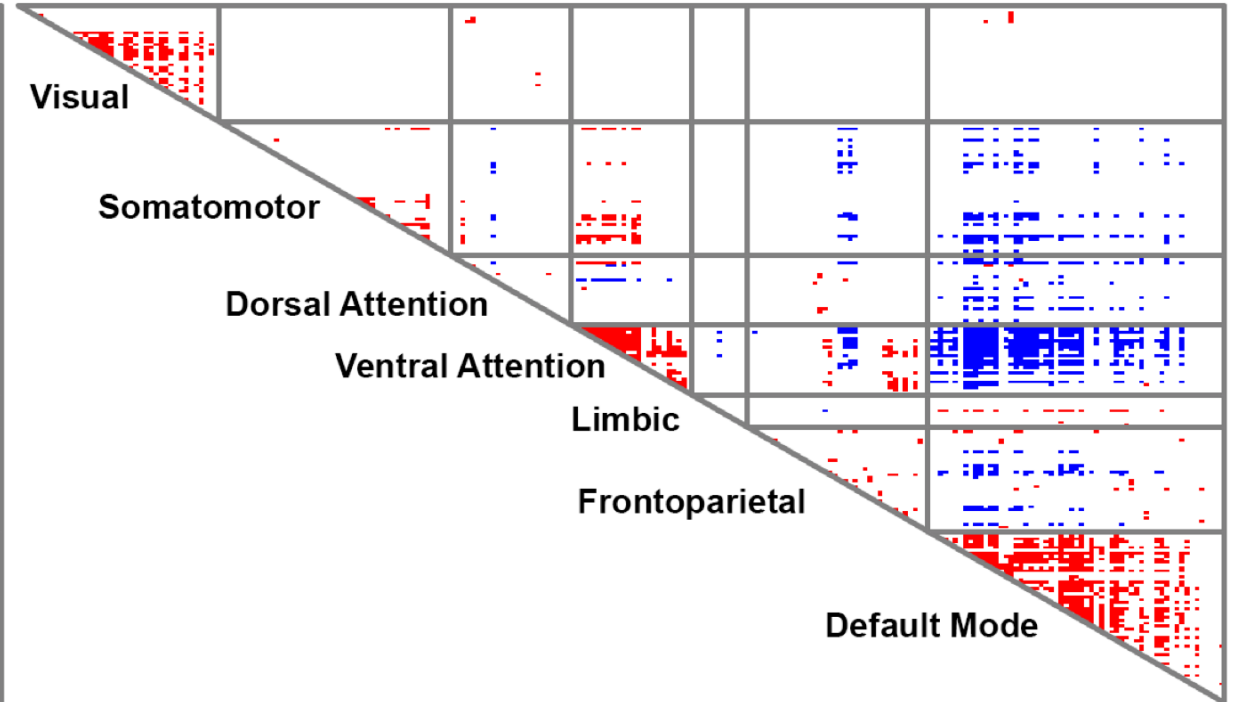


82 components → 25% out of sample variance explained

Adult Study Component 2



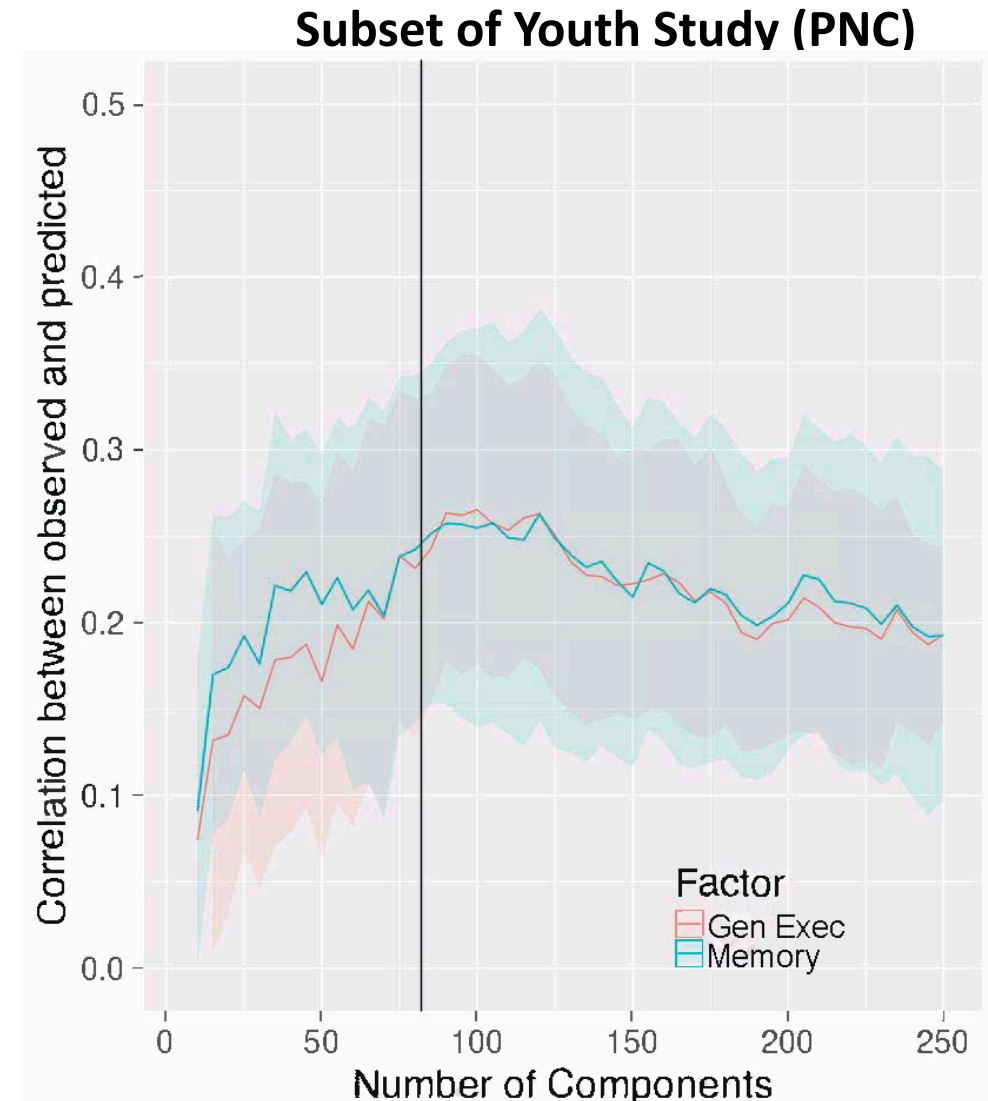
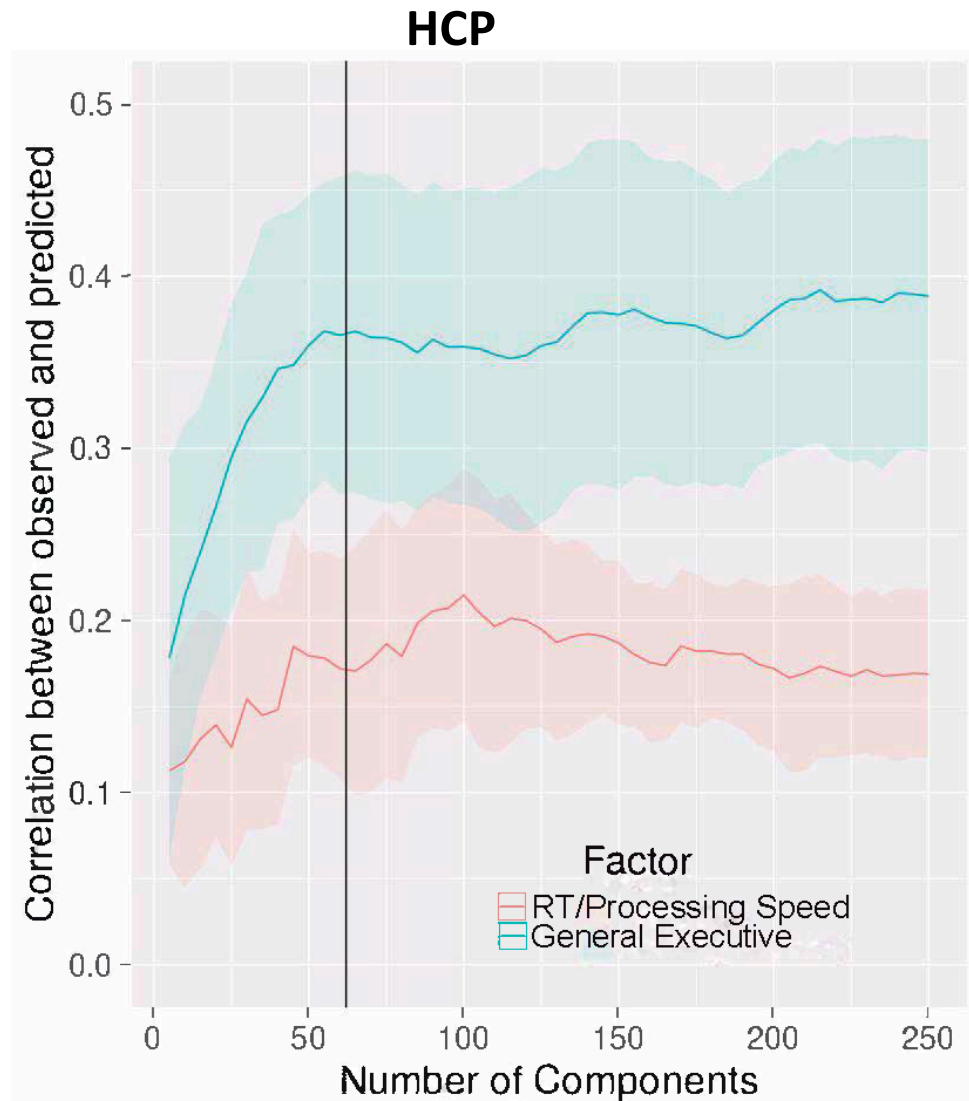
Youth Study Component 2



Integration
Segregation

External Validation of Number of Components:

How Many Components to Achieve Best Phenotype Prediction?



The Connectome Basis Set is

- Stable across brain parcellations of varying densities
 - tested 200-900 node atlases (Craddock PyCluster)
- Useful in phenotype prediction
- Has much higher test-retest reliability than individual edges
 - mean ICC of edges (day1/day2 in test set) = 0.4313
 - mean ICC of components = 0.6667
- Describes the nature and dimension of meaningful inter-individual connectomic variation

Thank You!

Twitter: **@being_saige**

Email: **saruther@med.umich.edu**

Thanks for support from:

NIAAA K23: AA-020297

NIMH BRAINS R01 MH107741

John Templeton Foundation

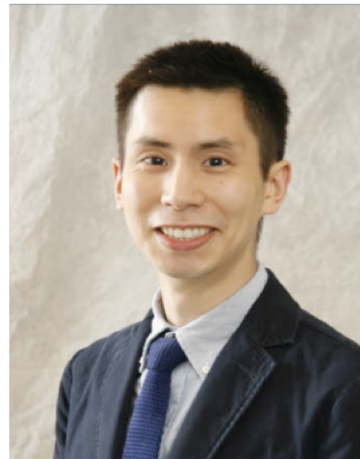
DANA Foundation



Chandra Sripada



Mike Angstadt



Mike Yee



Daniel Kessler