Brain-phenotype models fail for individuals who defy sample stereotypes

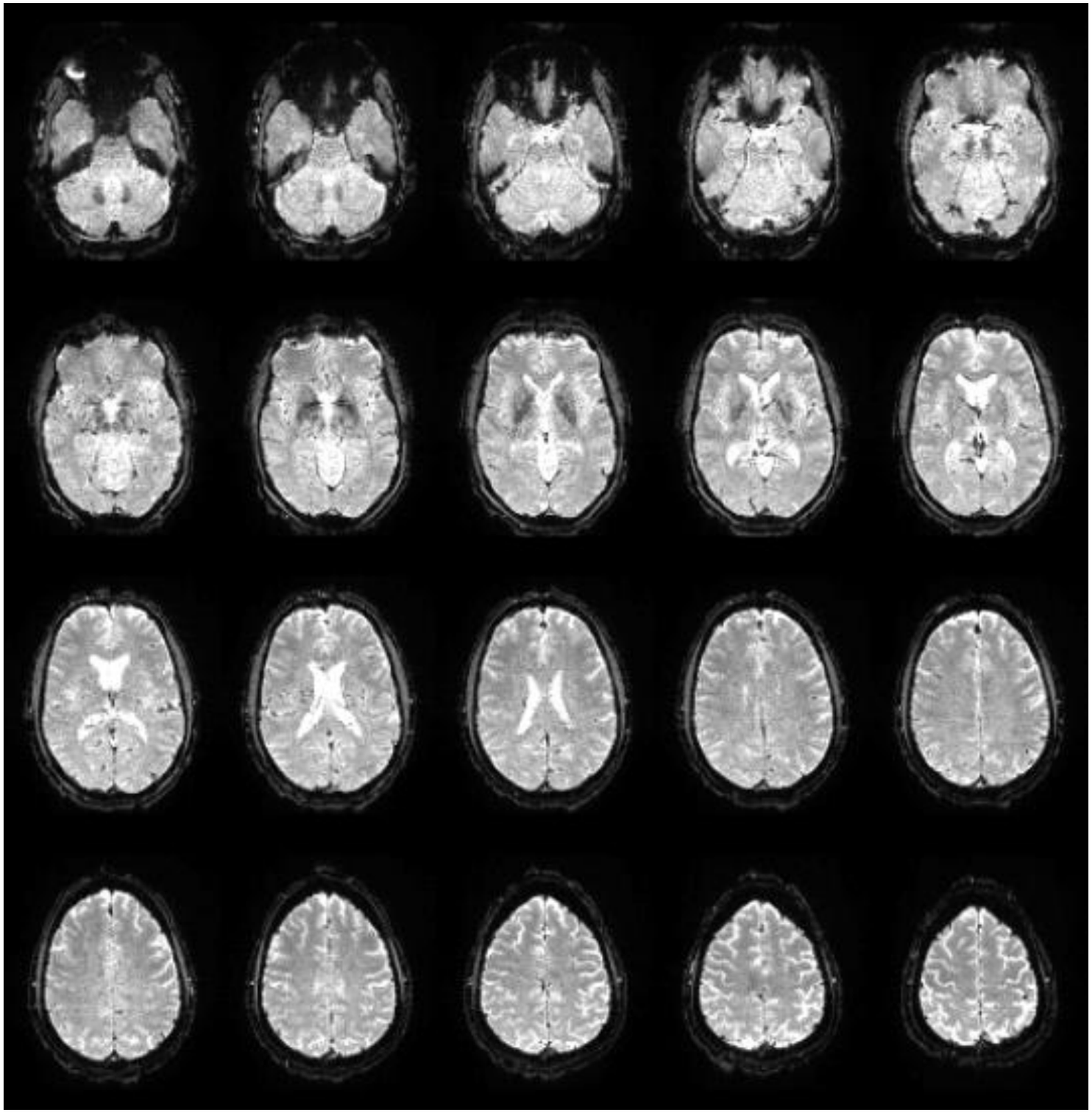
Abby Greene

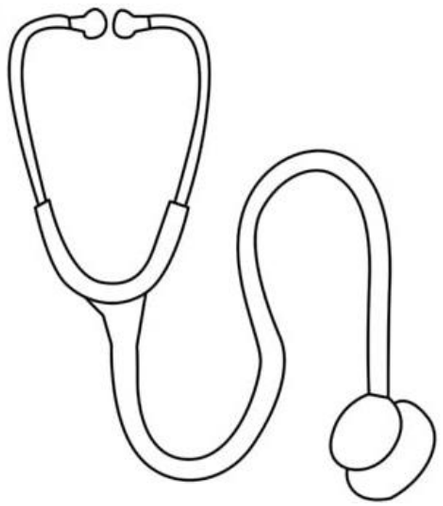
March 4, 2023

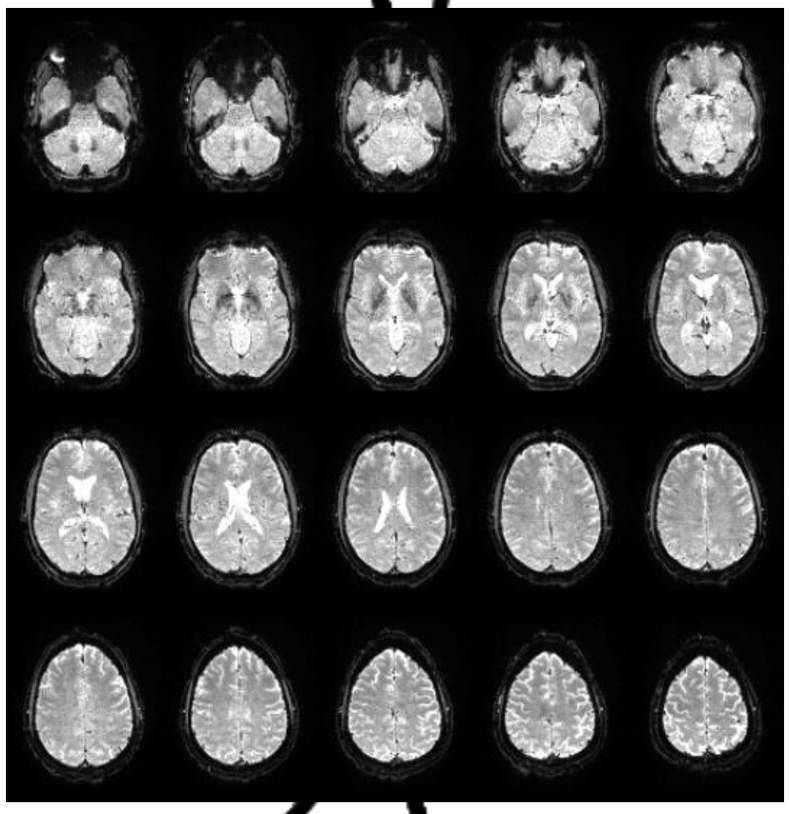
APPA annual meeting

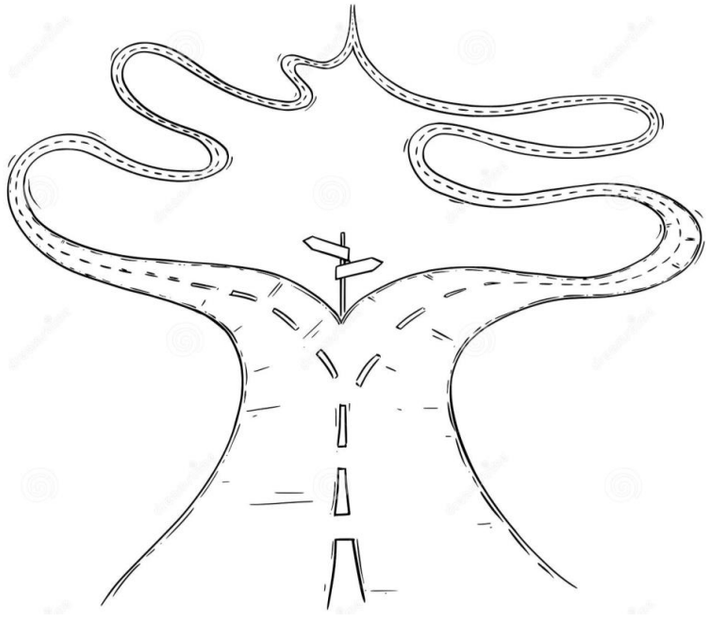
Disclosure

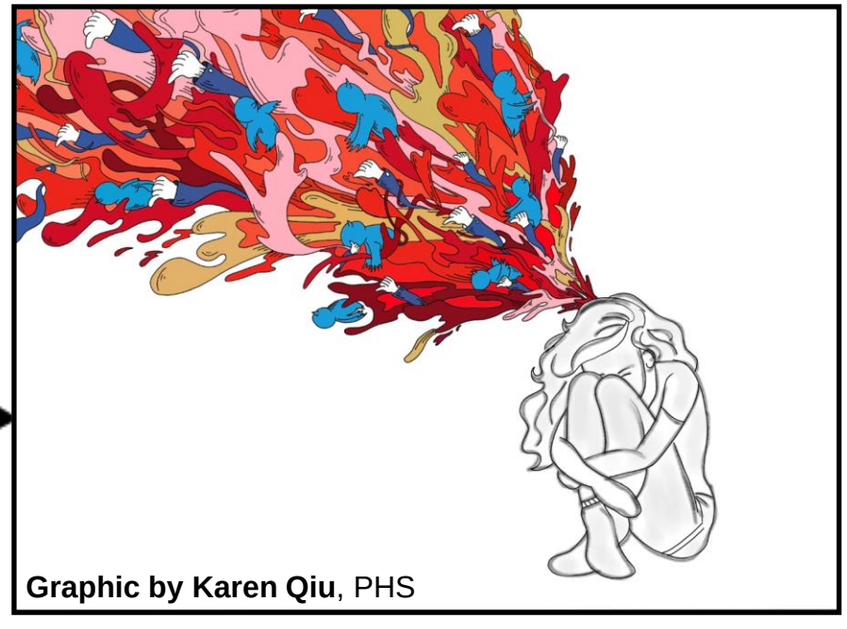
No financial or other conflicts of interest to disclose

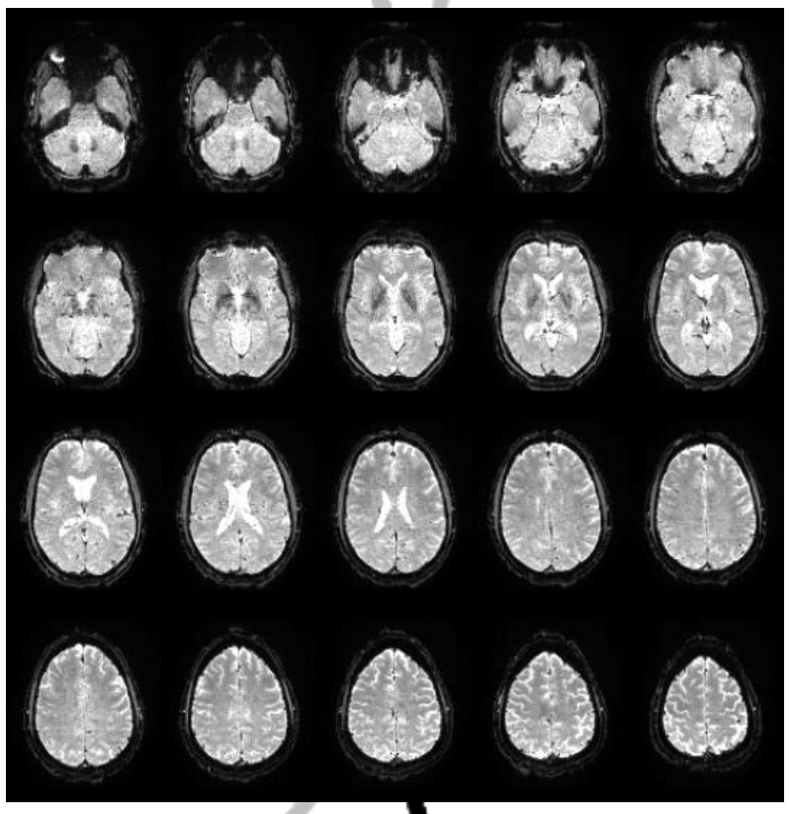


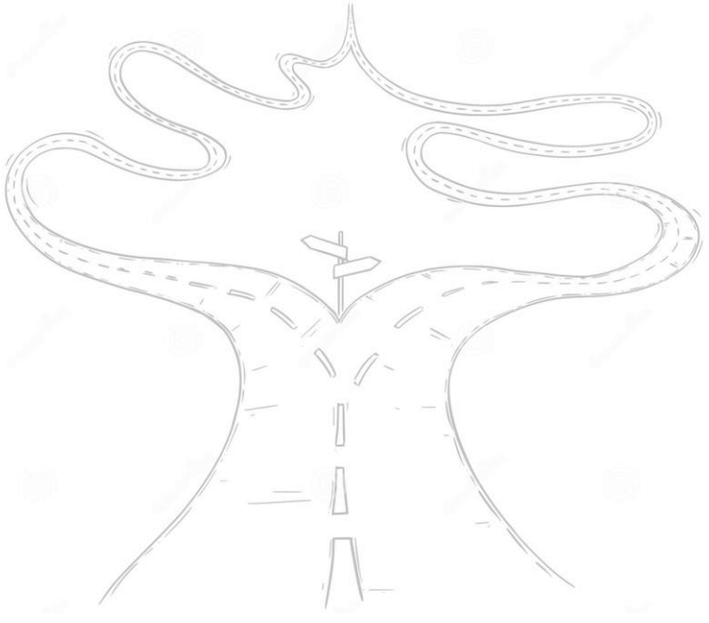


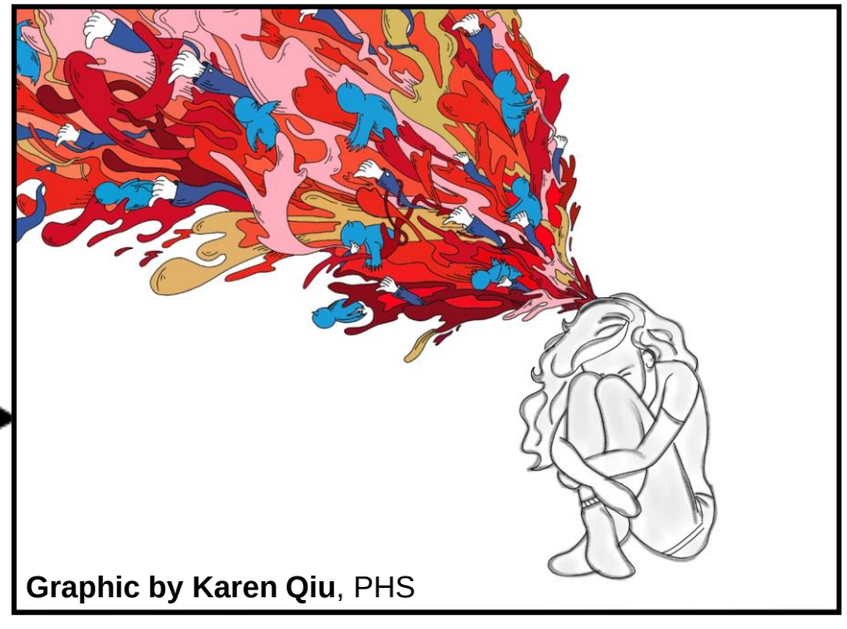




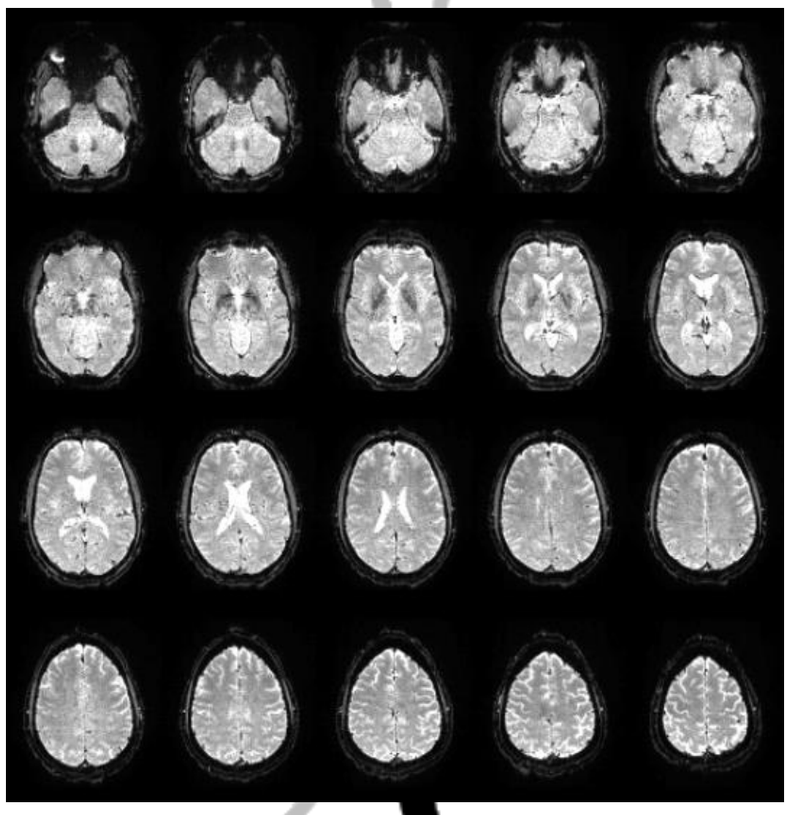

Graphic by Karen Qiu, PHS

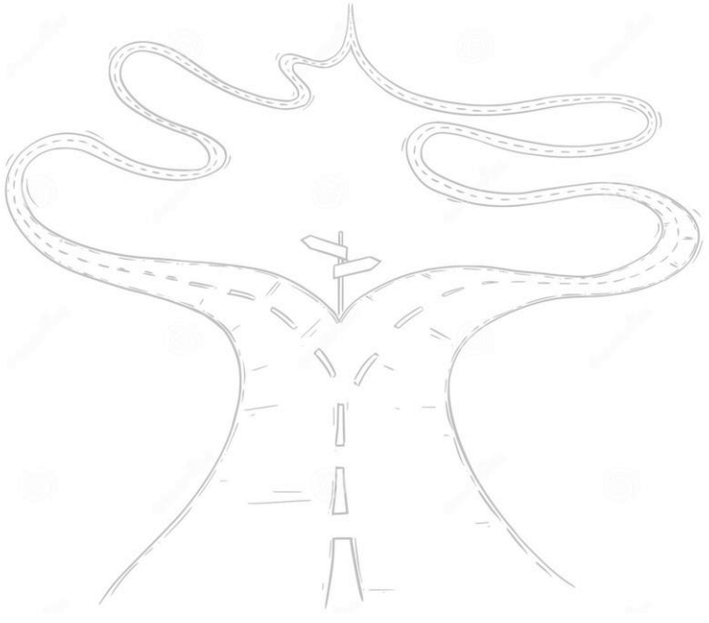


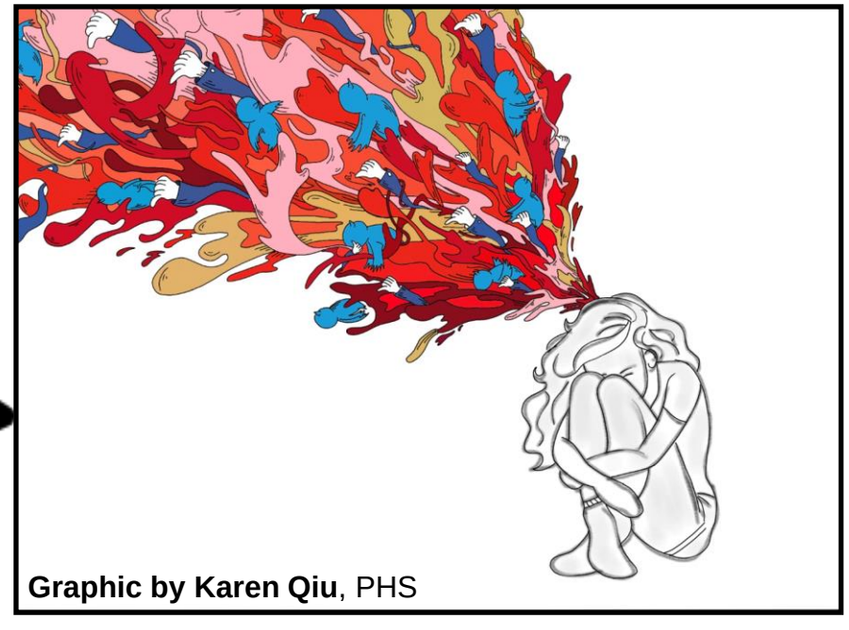




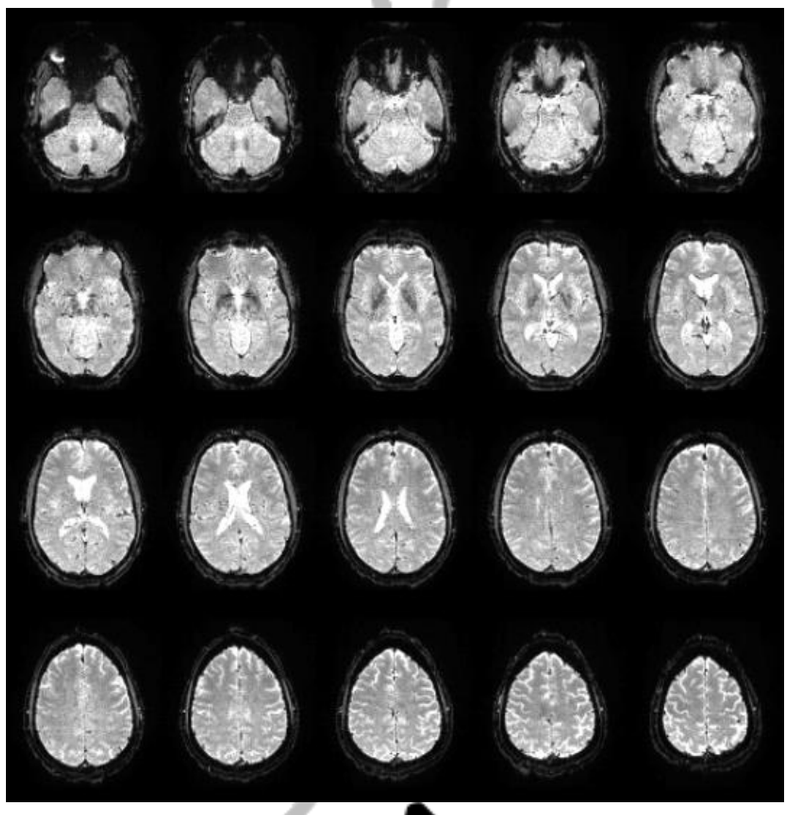
Graphic by Karen Qiu, PHS

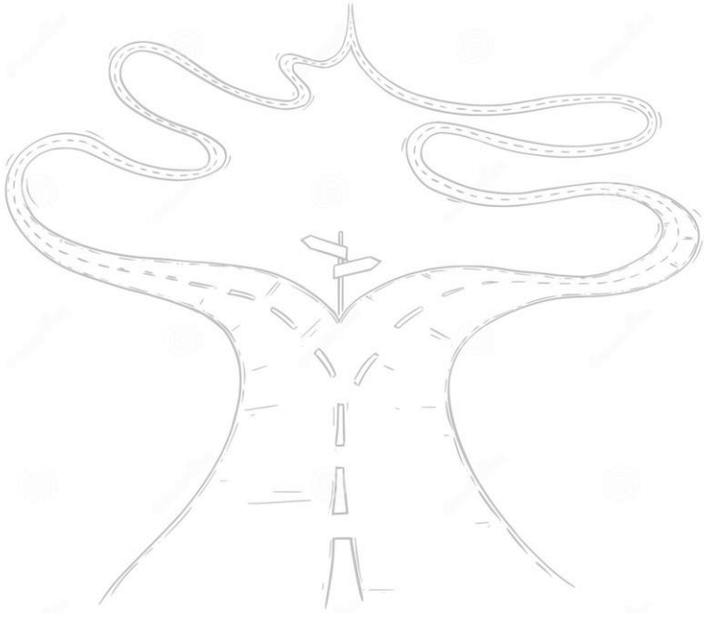


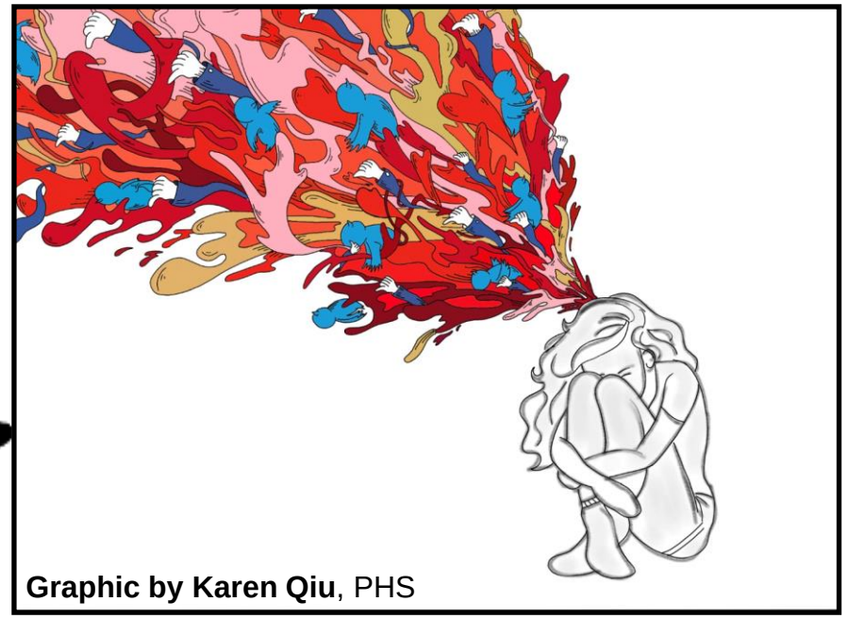




Graphic by Karen Qiu, PHS







Graphic by Karen Qiu, PHS

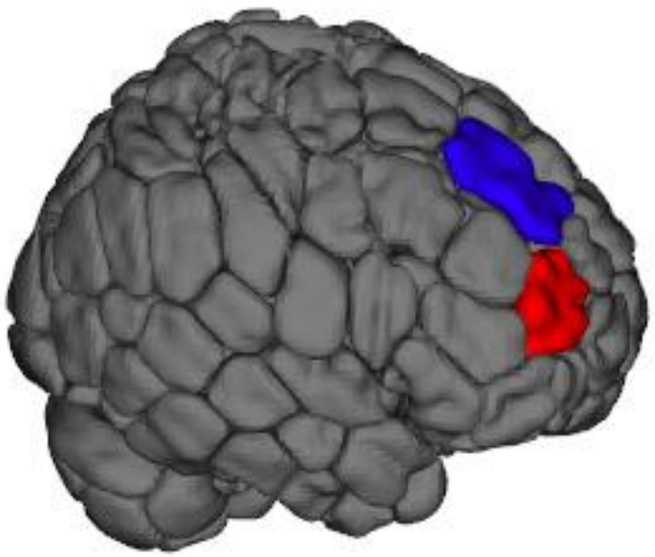
Brain-phenotype models fail for individuals who defy sample stereotypes

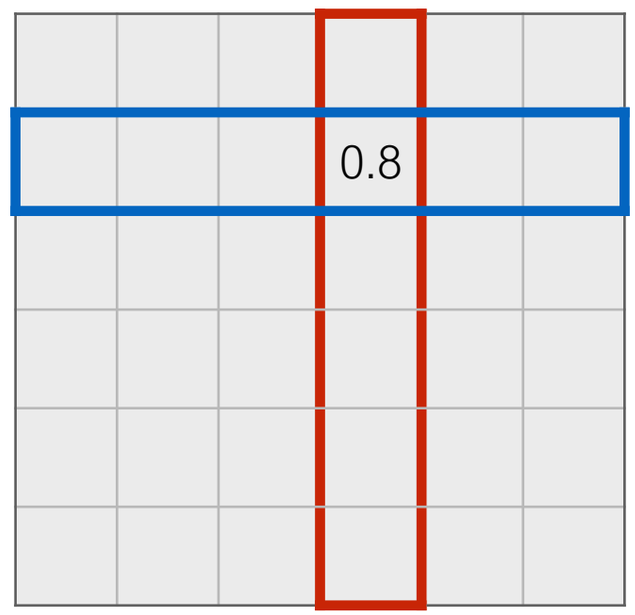
- Predictive modeling to relate brain to phenotype
- Questioning the generalizability assumption
- In whom do models fail, and why?

Brain-phenotype models fail for individuals who defy sample stereotypes

- **Predictive modeling to relate brain to phenotype**
- Questioning the generalizability assumption
- In whom do models fail, and why?

Functional connectivity





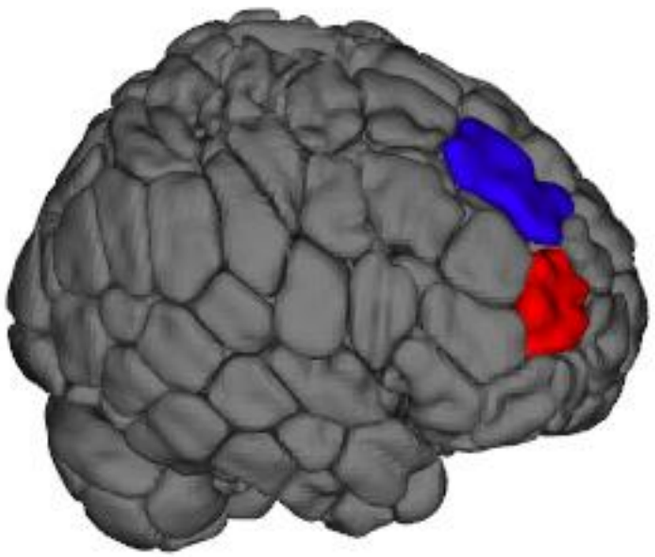
Parcellate

Correlate node time courses

Compute FC matrix

Shen et al., 2013

Functional connectivity



	0.9	-0.2	0.3	0.7	-0.6
		-0.7	0.8	-0.4	0.1
			0.1	-0.1	0.6
				0.8	-0.2
					0.3

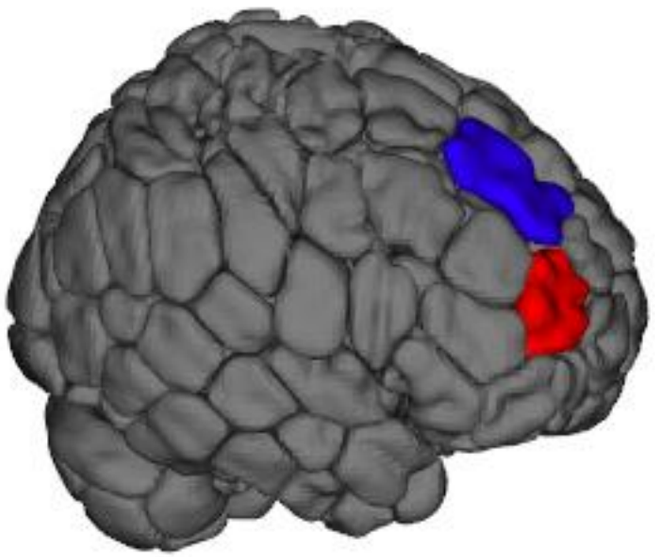
Parcellate

Correlate node time courses

Compute FC matrix

Shen et al., 2013

Functional connectivity



	0.9	-0.2	0.3	0.7	-0.6
0.9		-0.7	0.8	-0.4	0.1
-0.2	-0.7		0.1	-0.1	0.6
0.3	0.8	0.1		0.8	-0.2
0.7	-0.4	-0.1	0.8		0.3
-0.6	0.1	0.6	-0.2	0.3	

Parcellate

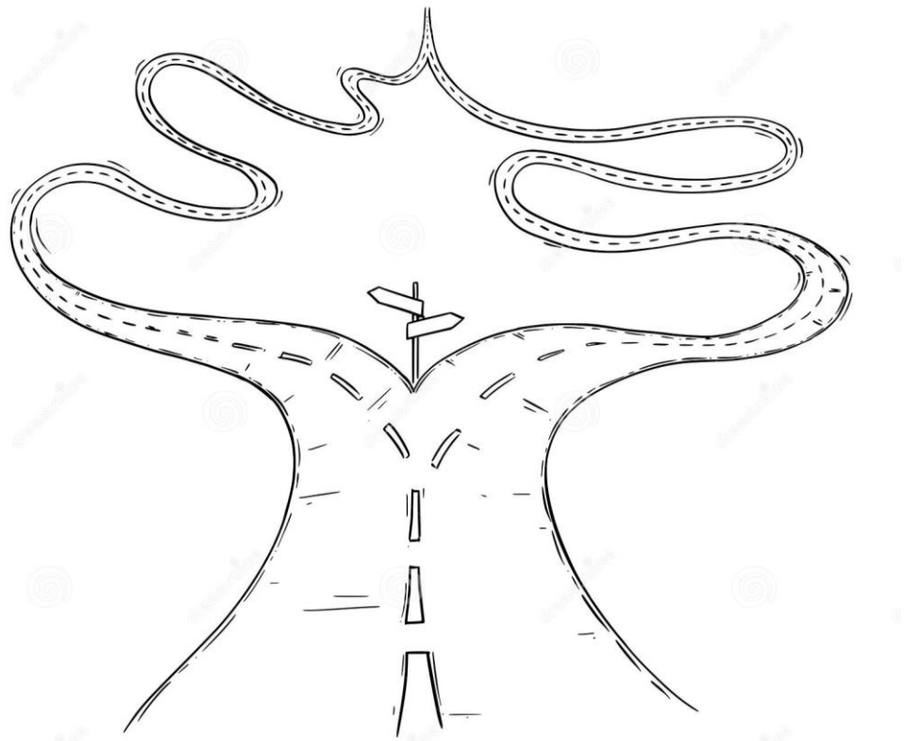
Correlate node time courses

Compute FC matrix

Shen et al., 2013

How to relate FC to cognition and behavior?

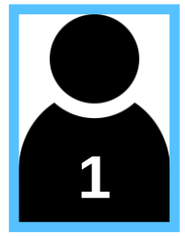
	0.9	-0.2	0.3	0.7	-0.6
0.9		-0.7	0.8	-0.4	0.1
-0.2	-0.7		0.1	-0.1	0.6
0.3	0.8	0.1		0.8	-0.2
0.7	-0.4	-0.1	0.8		0.3
-0.6	0.1	0.6	-0.2	0.3	



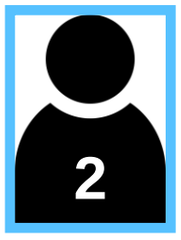
Connectome-based predictive modeling can predict a wide range of phenotypes from FC

Participant FC gF

Train



	0.9	-0.2	0.3	0.7	-0.6
0.9		-0.7	0.8	-0.4	0.1
-0.2	-0.7		0.1	-0.1	0.6
0.3	0.8	0.1		0.8	-0.2
0.7	-0.4	-0.1	0.8		0.3
-0.6	0.1	0.6	-0.2	0.3	



	0.8	-0.4	0.7	0.2	-0.3
0.8		0.1	0.6	-0.2	0.3
-0.4	0.1		0.2	0.3	0.9
0.7	0.6	0.2		0.3	-0.6
0.2	-0.2	0.3	0.3		0.1
-0.3	0.3	0.9	-0.6	0.1	

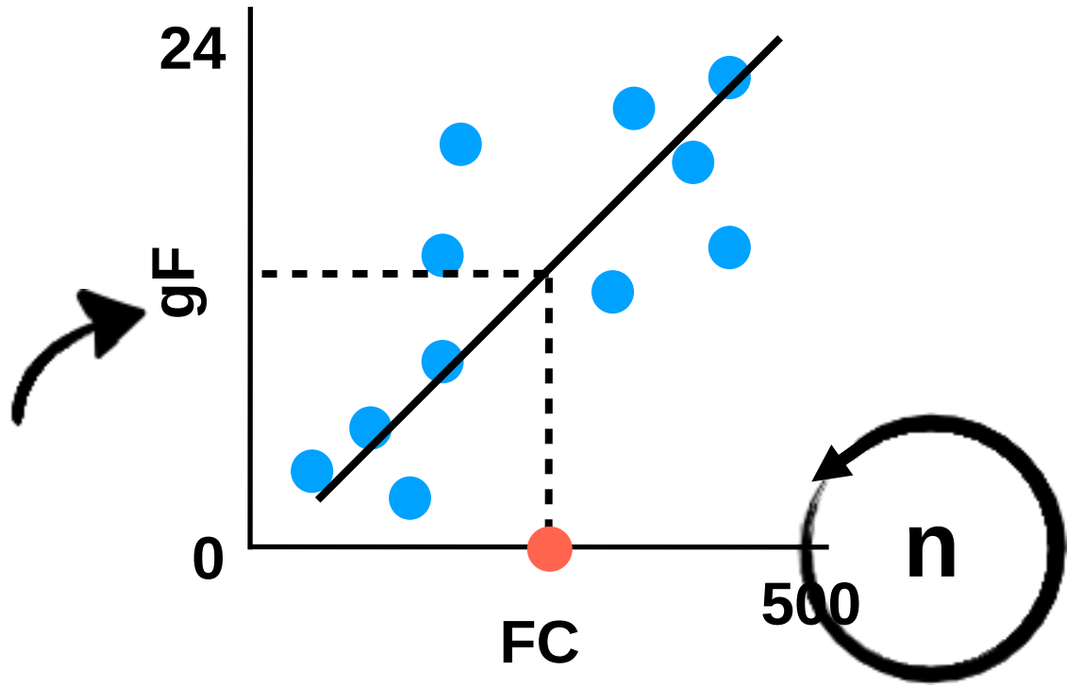
	0.7	-0.5	0.1	0.6	-0.4
0.7		0.2	0.3	-0.7	0.3
-0.5	0.2		0.2	0.2	0.3
0.1	0.3	0.2		0.6	-0.4
0.6	-0.7	0.2	0.6		0.2
-0.4	0.3	0.3	-0.4	0.2	

gF

15

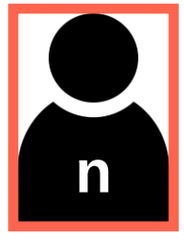
13

7



Test

Participant FC gF



	0.6	-0.6	0.1	0.5	-0.4
0.6		0.2	0.2	-0.9	0.3
-0.6	0.2		0.2	0.1	0.3
0.1	0.2	0.2		0.6	-0.4
0.5	-0.9	0.1	0.6		0.2
-0.4	0.3	0.3	-0.4	0.2	

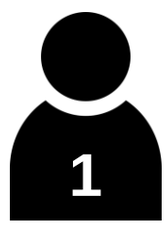
12?

Shen et al., 2017; Finn, Shen, et al., 2015

CPM can predict a wide range of phenotypes

Participant FC Attention
(d', ADHD symptoms)

gF
- Finn, Shen et al. (2015)
- Noble et al. (2017)
- Dubois et al. (2017)
- Greene et al. (2018)



	0.9	-0.2	0.3	0.7	-0.6
0.9		-0.7	0.8	-0.4	0.1
-0.2	-0.7		0.1	-0.1	0.6
0.3	0.8	0.1		0.8	-0.2
0.7	-0.4	-0.1	0.8		0.3
-0.6	0.1	0.6	-0.2	0.3	

1.2

	0.8	-0.4	0.7	0.2	-0.3
0.8		0.1	0.6	-0.2	0.3
-0.4	0.1		0.2	0.3	0.9
0.7	0.6	0.2		0.3	-0.6
0.2	-0.2	0.3	0.3		0.1
-0.3	0.3	0.9	-0.6	0.1	

3.4

	0.7	-0.5	0.1	0.6	-0.4
0.7		0.2	0.3	-0.7	0.3
-0.5	0.2		0.2	0.2	0.3
0.1	0.3	0.2		0.6	-0.4
0.6	-0.7	0.2	0.6		0.2
-0.4	0.3	0.3	-0.4	0.2	

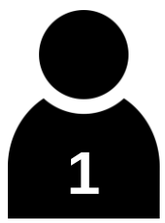
2.7

CPM can predict a wide range of phenotypes

Participant FC Autism symptoms
(SRS, ADOS scores)

gF
- Finn, Shen et al. (2015)
- Noble et al. (2017)
- Dubois et al. (2017)
- Greene et al. (2018)

Attention
(d', ADHD symptoms)
- Rosenberg et al. (2015)



	0.9	-0.2	0.3	0.7	-0.6
0.9		-0.7	0.8	-0.4	0.1
-0.2	-0.7		0.1	-0.1	0.6
0.3	0.8	0.1		0.8	-0.2
0.7	-0.4	-0.1	0.8		0.3
-0.6	0.1	0.6	-0.2	0.3	

27

	0.8	-0.4	0.7	0.2	-0.3
0.8		0.1	0.6	-0.2	0.3
-0.4	0.1		0.2	0.3	0.9
0.7	0.6	0.2		0.3	-0.6
0.2	-0.2	0.3	0.3		0.1
-0.3	0.3	0.9	-0.6	0.1	

63

	0.7	-0.5	0.1	0.6	-0.4
0.7		0.2	0.3	-0.7	0.3
-0.5	0.2		0.2	0.2	0.3
0.1	0.3	0.2		0.6	-0.4
0.6	-0.7	0.2	0.6		0.2
-0.4	0.3	0.3	-0.4	0.2	

82

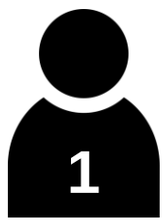
CPM can predict a wide range of phenotypes

Participant FC Personality
(Neuroticism, extraversion)

gF
- Finn, Shen et al. (2015)
- Noble et al. (2017)
- Dubois et al. (2017)
- Greene et al. (2018)

Attention
(d', ADHD symptoms)
- Rosenberg et al. (2015)

Autism symptoms
(SRS, ADOS scores)
- Lake et al. (2019)



	0.9	-0.2	0.3	0.7	-0.6
0.9		-0.7	0.8	-0.4	0.1
-0.2	-0.7		0.1	-0.1	0.6
0.3	0.8	0.1		0.8	-0.2
0.7	-0.4	-0.1	0.8		0.3
-0.6	0.1	0.6	-0.2	0.3	

34

	0.8	-0.4	0.7	0.2	-0.3
0.8		0.1	0.6	-0.2	0.3
-0.4	0.1		0.2	0.3	0.9
0.7	0.6	0.2		0.3	-0.6
0.2	-0.2	0.3	0.3		0.1
-0.3	0.3	0.9	-0.6	0.1	

26

	0.7	-0.5	0.1	0.6	-0.4
0.7		0.2	0.3	-0.7	0.3
-0.5	0.2		0.2	0.2	0.3
0.1	0.3	0.2		0.6	-0.4
0.6	-0.7	0.2	0.6		0.2
-0.4	0.3	0.3	-0.4	0.2	

17

CPM can predict a wide range of phenotypes

Participant FC Creative ability
(Divergent thinking)

gF
- Finn, Shen et al. (2015)
- Noble et al. (2017)
- Dubois et al. (2017)
- Greene et al. (2018)

Attention
(d', ADHD symptoms)
- Rosenberg et al. (2015)

Autism symptoms
(SRS, ADOS scores)
- Lake et al. (2019)

	0.9	-0.2	0.3	0.7	-0.6
0.9		-0.7	0.8	-0.4	0.1
-0.2	-0.7		0.1	-0.1	0.6
0.3	0.8	0.1		0.8	-0.2
0.7	-0.4	-0.1	0.8		0.3
-0.6	0.1	0.6	-0.2	0.3	

1.2

	0.8	-0.4	0.7	0.2	-0.3
0.8		0.1	0.6	-0.2	0.3
-0.4	0.1		0.2	0.3	0.9
0.7	0.6	0.2		0.3	-0.6
0.2	-0.2	0.3	0.3		0.1
-0.3	0.3	0.9	-0.6	0.1	

-0.3

	0.7	-0.5	0.1	0.6	-0.4
0.7		0.2	0.3	-0.7	0.3
-0.5	0.2		0.2	0.2	0.3
0.1	0.3	0.2		0.6	-0.4
0.6	-0.7	0.2	0.6		0.2
-0.4	0.3	0.3	-0.4	0.2	

2.7

Personality
(Neuroticism, extraversion)
- Hsu et al. (2017)

CPM can predict a wide range of phenotypes

Participant FC Reading recall

gF
- Finn, Shen et al. (2015)
- Noble et al. (2017)
- Dubois et al. (2017)
- Greene et al. (2018)

Attention
(d', ADHD symptoms)
- Rosenberg et al. (2015)

Autism symptoms
(SRS, ADOS scores)
- Lake et al. (2019)

	0.9	-0.2	0.3	0.7	-0.6
0.9		-0.7	0.8	-0.4	0.1
-0.2	-0.7		0.1	-0.1	0.6
0.3	0.8	0.1		0.8	-0.2
0.7	-0.4	-0.1	0.8		0.3
-0.6	0.1	0.6	-0.2	0.3	

57

	0.8	-0.4	0.7	0.2	-0.3
0.8		0.1	0.6	-0.2	0.3
-0.4	0.1		0.2	0.3	0.9
0.7	0.6	0.2		0.3	-0.6
0.2	-0.2	0.3	0.3		0.1
-0.3	0.3	0.9	-0.6	0.1	

86

	0.7	-0.5	0.1	0.6	-0.4
0.7		0.2	0.3	-0.7	0.3
-0.5	0.2		0.2	0.2	0.3
0.1	0.3	0.2		0.6	-0.4
0.6	-0.7	0.2	0.6		0.2
-0.4	0.3	0.3	-0.4	0.2	

63

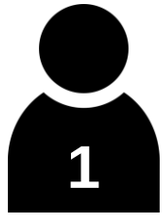
Personality
(Neuroticism, extraversion)
- Hsu et al. (2017)

Creative ability
(Divergent thinking)
- Beaty et al. (2018)

CPM can predict a wide range of phenotypes

Participant

FC



	0.9	-0.2	0.3	0.7	-0.6
0.9		-0.7	0.8	-0.4	0.1
-0.2	-0.7		0.1	-0.1	0.6
0.3	0.8	0.1		0.8	-0.2
0.7	-0.4	-0.1	0.8		0.3
-0.6	0.1	0.6	-0.2	0.3	

gF
- Finn, Shen et al. (2015)
- Noble et al. (2017)
- Dubois et al. (2017)
- Greene et al. (2018)

	0.8	-0.4	0.7	0.2	-0.3
0.8		0.1	0.6	-0.2	0.3
-0.4	0.1		0.2	0.3	0.9
0.7	0.6	0.2		0.3	-0.6
0.2	-0.2	0.3	0.3		0.1
-0.3	0.3	0.9	-0.6	0.1	

Attention
(d', ADHD symptoms)
- Rosenberg et al. (2015)

	0.7	-0.5	0.1	0.6	-0.4
0.7		0.2	0.3	-0.7	0.3
-0.5	0.2		0.2	0.2	0.3
0.1	0.3	0.2		0.6	-0.4
0.6	-0.7	0.2	0.6		0.2
-0.4	0.3	0.3	-0.4	0.2	

Autism symptoms
(SRS, ADOS scores)
- Lake et al. (2019)

Personality
(Neuroticism, extraversion)
- Hsu et al. (2017)

Creative ability
(Divergent thinking)
- Beaty et al. (2018)

Reading recall
- Jangraw et al. (2018)

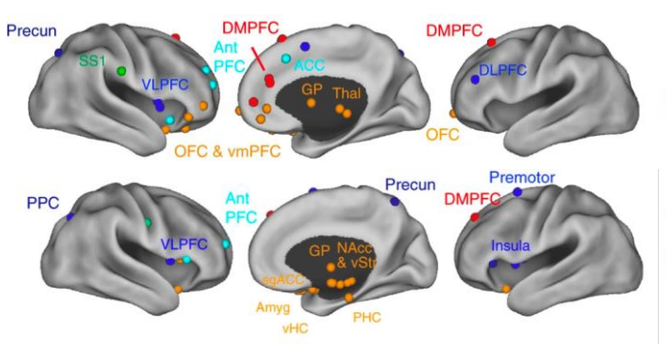
Growing interest in predictive modeling of brain-phenotype relationships

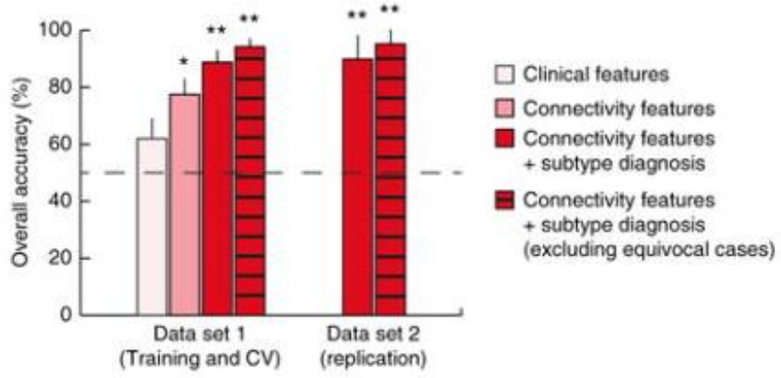
Anxiety Anhedonia

nature medicine

Resting-state connectivity biomarkers define neurophysiological subtypes of depression

Andrew T Drysdale[1–3], Logan Grosenick[4,5], Jonathan Downar[6], Katharine Dunlop[6], Farrokh Mansouri[6], Yue Meng[1], Robert N Fetcho[1], Benjamin Zebley[7], Desmond J Oathes[8], Amit Etkin[9,10], Alan F Schatzberg[9], Keith Sudheimer[9], Jennifer Keller[9], Helen S Mayberg[11], Faith M Gunning[2,12], George S Alexopoulos[2,12], Michael D Fox[13], Alvaro Pascual-Leone[13], Henning U Voss[14], BJ Casey[15], Marc J Dubin[1,2] & Conor Liston[1–3]

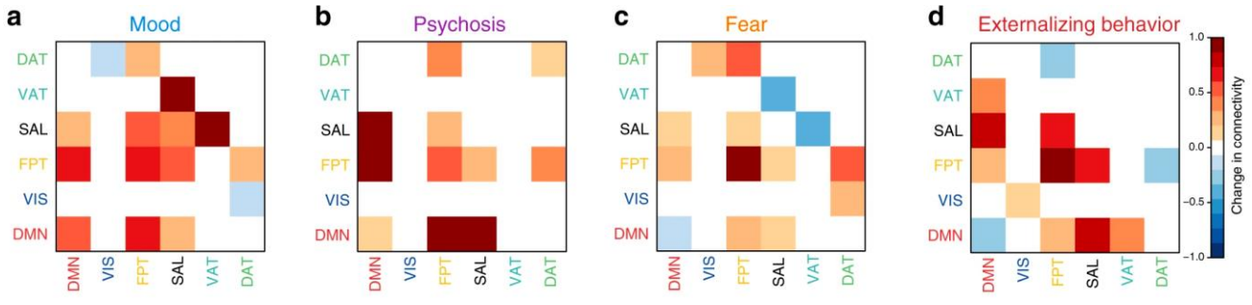




nature COMMUNICATIONS

Linked dimensions of psychopathology and connectivity in functional brain networks

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



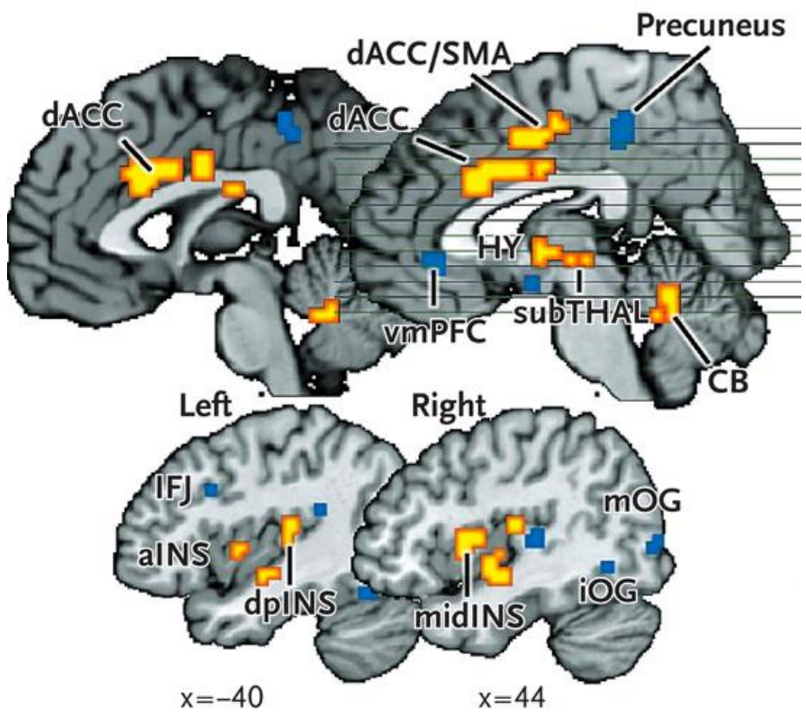
The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

An fMRI-Based Neurologic Signature of Physical Pain

Tor D. Wager, Ph.D., Lauren Y. Atlas, Ph.D., Martin A. Lindquist, Ph.D., Mathieu Roy, Ph.D., Choong-Wan Woo, M.A., and Ethan Kross, Ph.D.

Pain-Predictive Signature Pattern



Brain-phenotype models fail for individuals who defy sample stereotypes

- **Predictive modeling to relate brain to phenotype**
- Questioning the generalizability assumption
- In whom do models fail, and why?

Brain-phenotype models fail for individuals who defy sample stereotypes

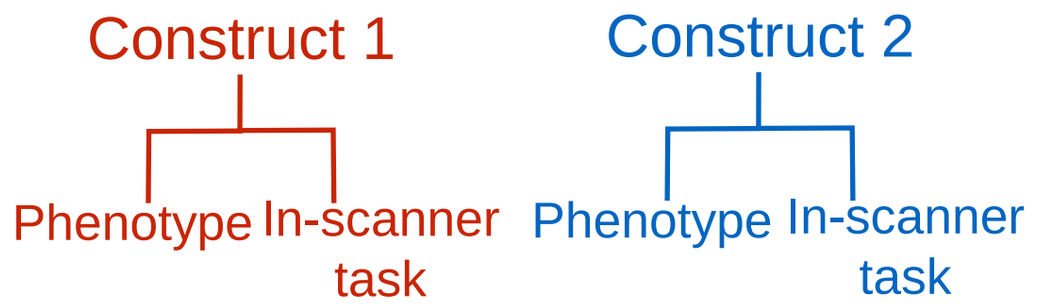
- Predictive modeling to relate brain to phenotype
- **Questioning the generalizability assumption**
- In whom do models fail, and why?

A sample to study failures of generalizability

- **Aim:** To build task FC-based predictive models of cognitive and behavioral phenotypes across a wide range of function

- **Approach:**

  - Large (*n* = 194 and growing…), heterogeneous sample
  - In-scanner task battery
  - Out-of-scanner battery of phenotype measures



RDoC
Research Domain Criteria Initiative

In-scanner task battery Post-scan battery

A sample to study failures of generalizability

Task	RDoC Domain	RDoC Construct
Card Guessing Task	Positive valence	Reward responsiveness
N-Back Task	Cognitive systems	Working memory
Stop-Signal Task	Cognitive systems	Cognitive control
Reading the Mind in the Eyes Task	Social processes	Perception and understanding of others
Perceptual Task (movie watching)	Cognitive systems	Perception
Sustained Attention Task	Cognitive systems	Attention

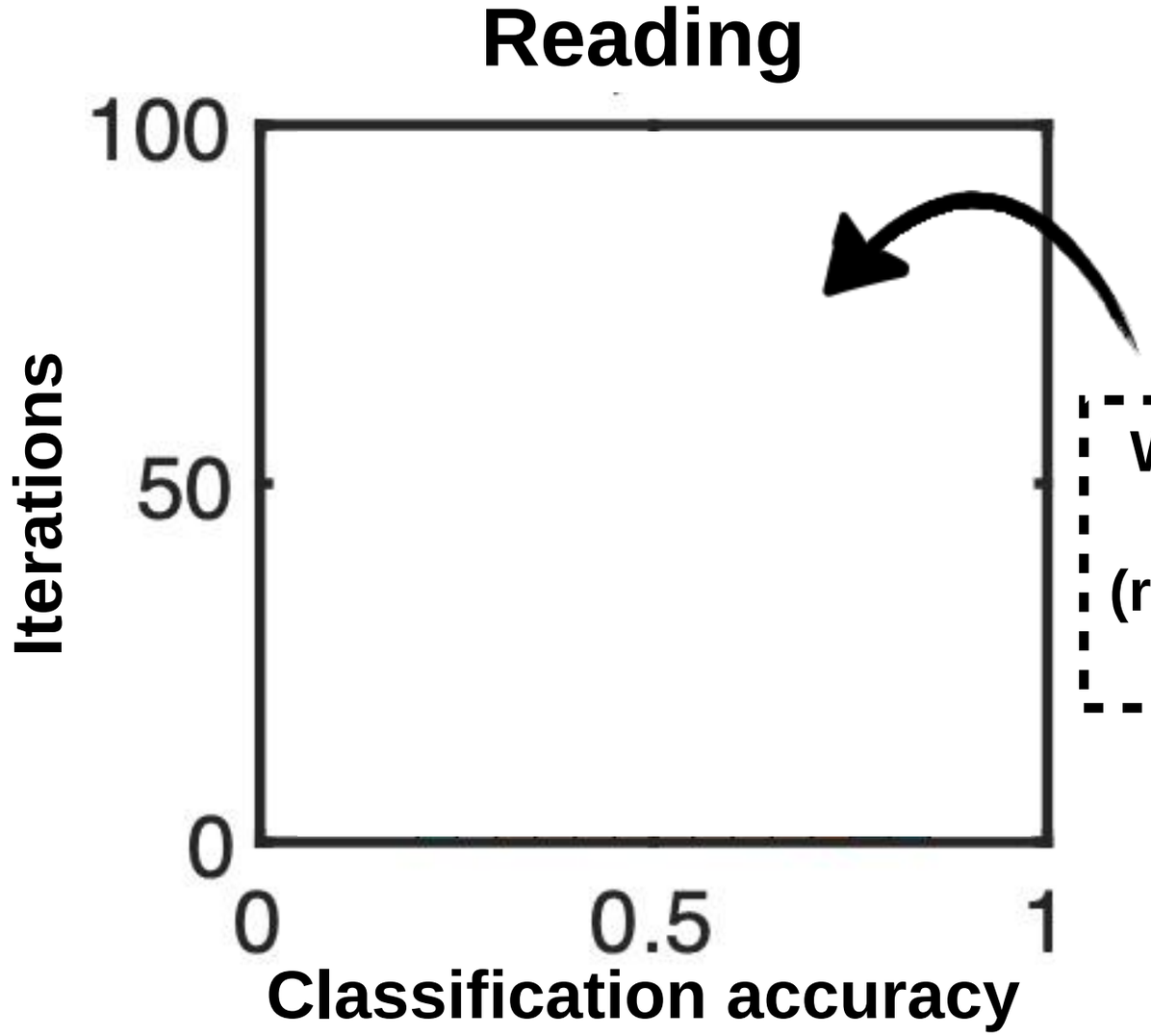
In-scanner task battery **Post-scan battery**

A sample to study failures of generalizability

Measure	Subtests	RDoC Domain	RDoC Construct
Demographics questionnaire	N/A	N/A	N/A
Handedness inventory	N/A	N/A	N/A
Interpersonal Reactivity Index	N/A	Social processes	Perception and understanding of others
Perceived Stress Scale	N/A	N/A	N/A
Positive and Negative Affect Schedule	N/A	Positive valence	Reward responsiveness
Pittsburgh Sleep Quality Index	N/A	Arousal and regulatory systems	Sleep-wakefulness
Temperament questionnaire	N/A	N/A	N/A
Task strategy/difficulty questionnaire	N/A	N/A	N/A
Delis-Kaplan Executive Function Scale	Verbal Fluency	Cognitive systems	Language
	Trail Making	Cognitive systems	Attention, Perception
	Color-Word Interference	Cognitive systems	Cognitive control
	Twenty Questions	Social processes	Social communication

Measure	Subtests	RDoC Domain	RDoC Construct
Wechsler Intelligence Scales (WISC/WAIS)	Letter-Number Sequencing	Cognitive systems	Working memory
	Symbol Search	Cognitive systems	Attention, Perception
	Cancellation	Cognitive systems	Attention, Perception
	Coding	Cognitive systems	Attention, Perception
Wechsler Abbreviated Scale of Intelligence	Vocabulary	Cognitive systems	Language
	Matrix Reasoning	N/A	N/A
Wide Range Assessment of Memory and Learning	Finger Windows	Cognitive systems	Working memory
	List Learning	Cognitive systems	Working memory, Declarative memory
Wide Range Achievement Test	Reading	N/A	N/A
Behavior Rating Inventory of Executive Function	N/A	Cognitive systems	Cognitive control
Boston Naming Test	N/A	Cognitive systems	Language
Brief Symptom Inventory	N/A	N/A	N/A
Mini-International Neuropsychiatric Interview	N/A	N/A	N/A

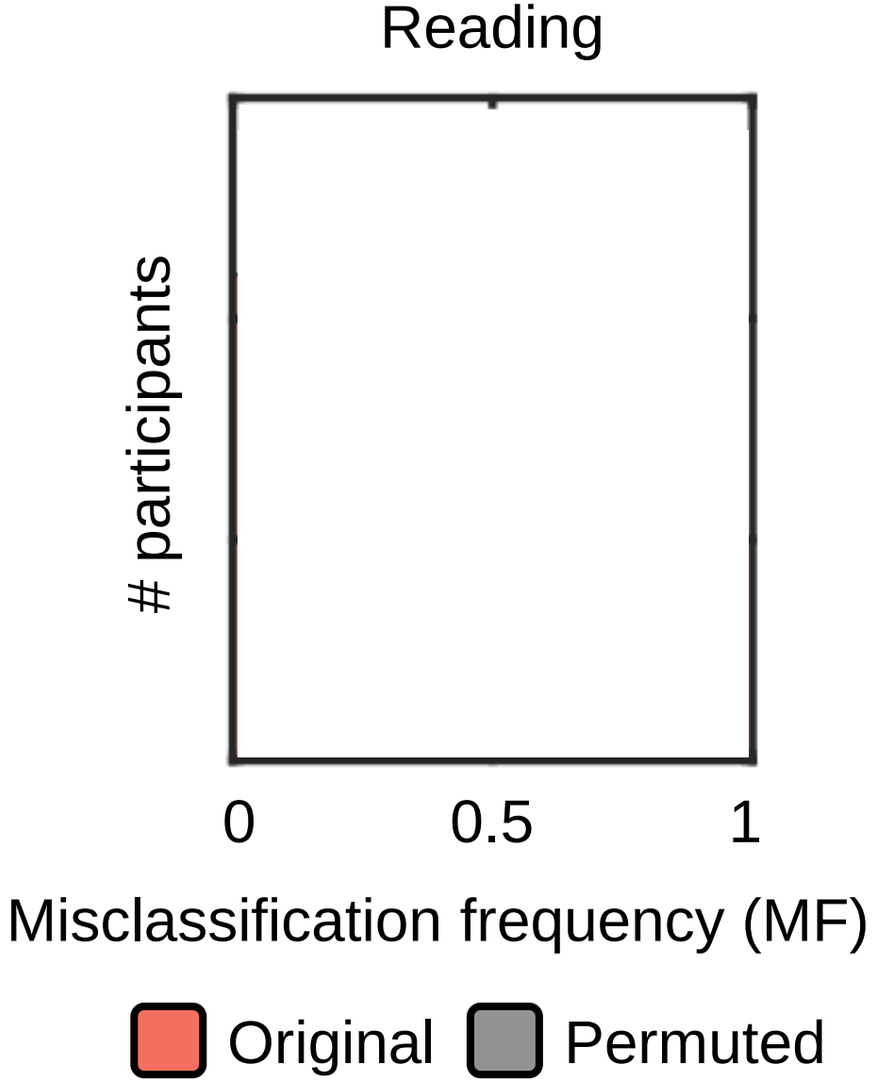
Can we predict phenotype in these data, and in whom?



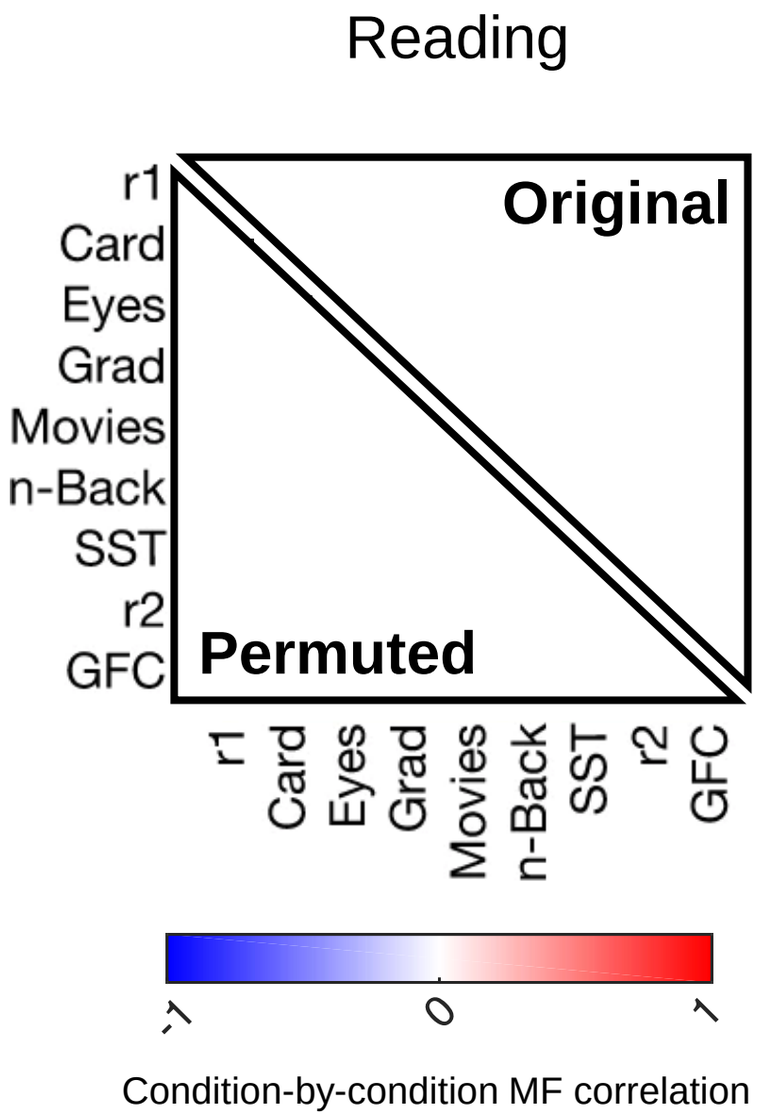
We can use brain data to predict phenotype (reading), but the model is making mistakes

Misclassification is stable and phenotype-specific

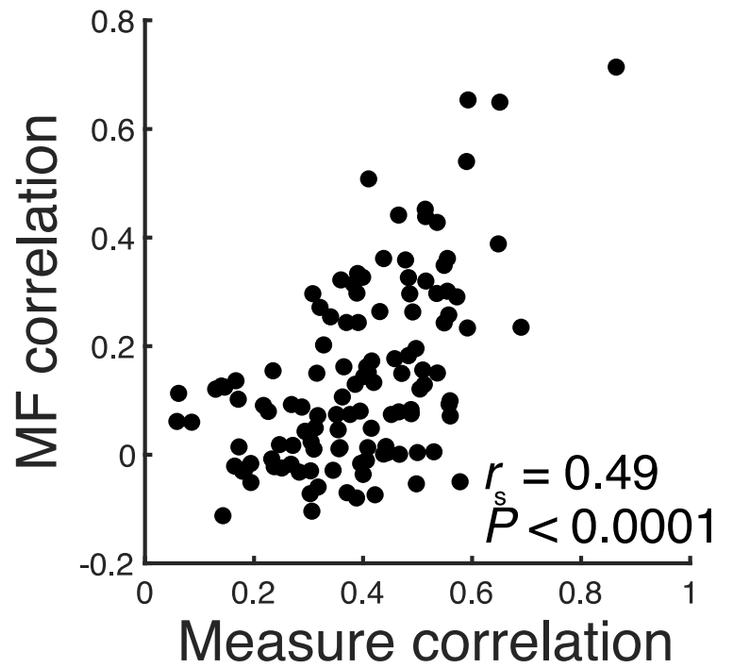
Across iterations

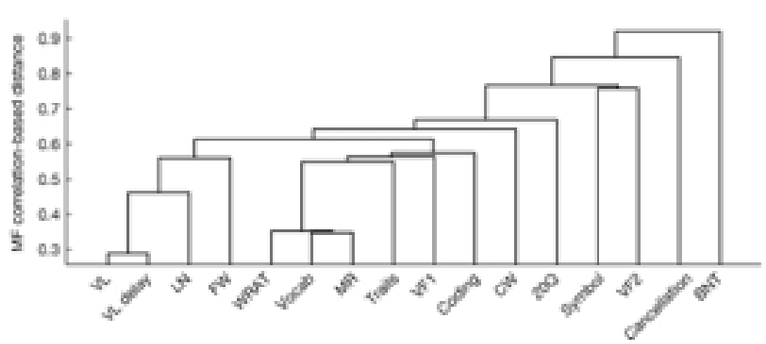


Across tasks

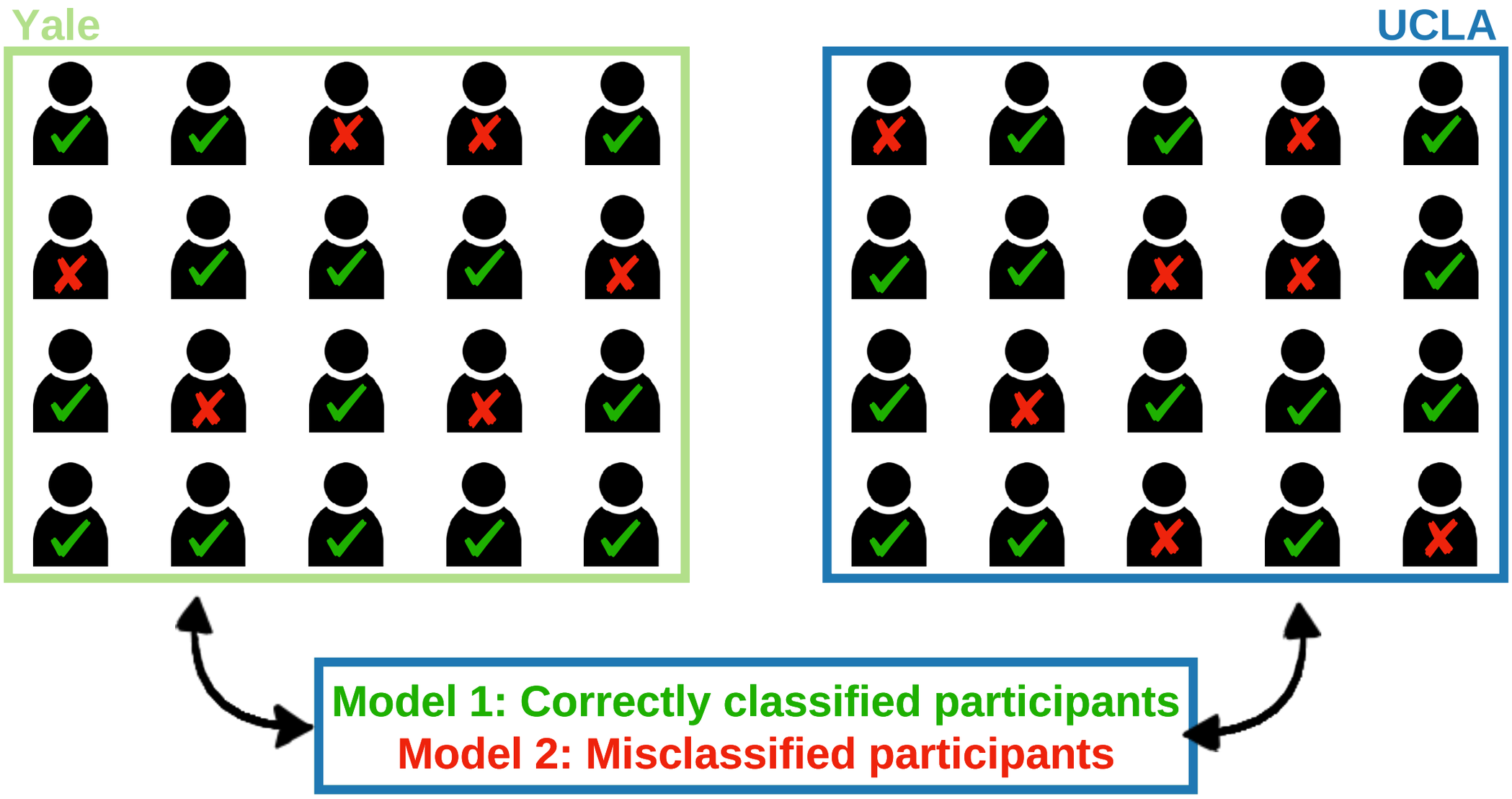


Across measures



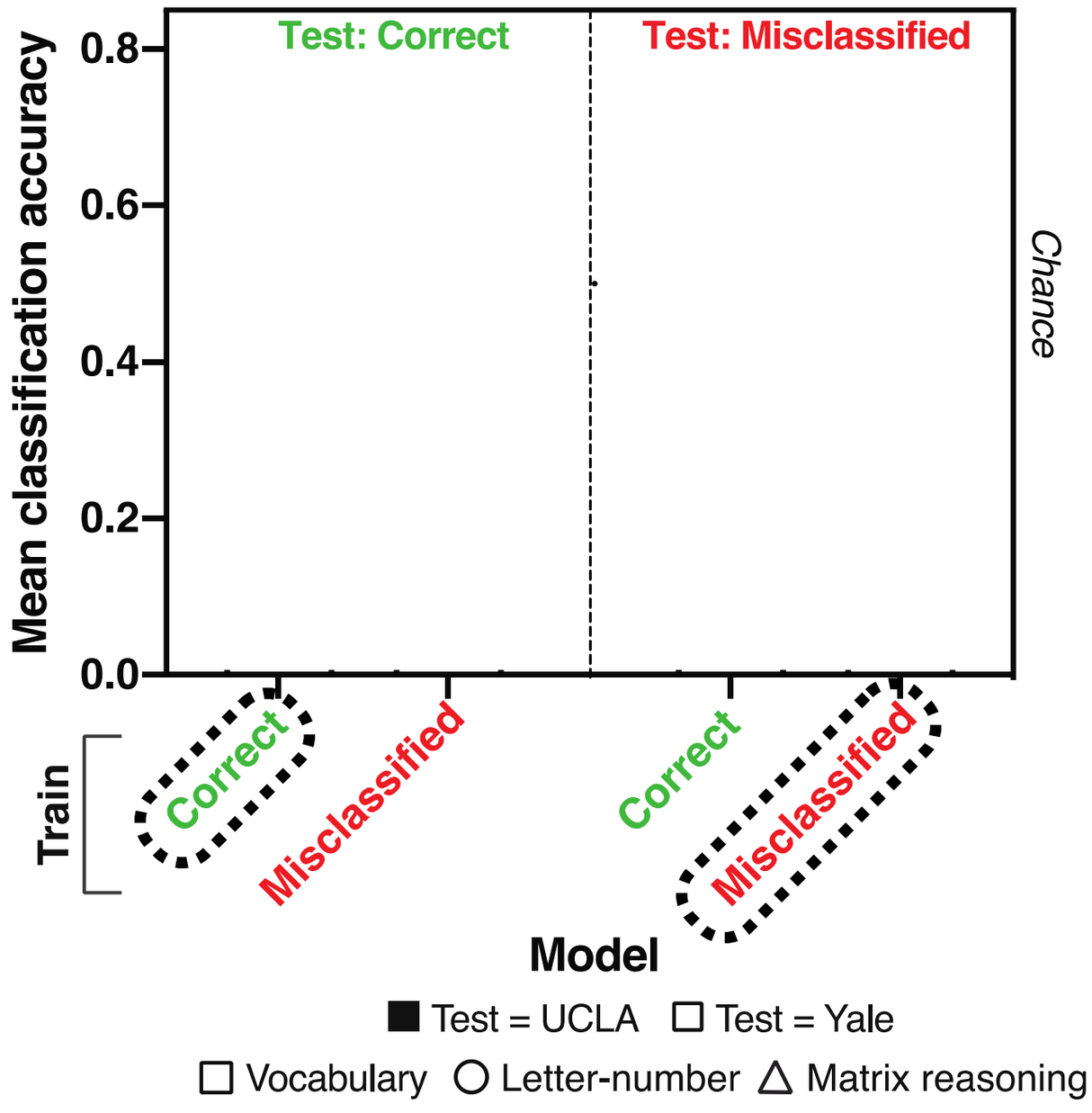


Validating misclassifications: Do models generalize across datasets?



How well does each model perform?

Models and “classification status” generalize across datasets



Brain-phenotype models fail for individuals who defy sample stereotypes

- Predictive modeling to relate brain to phenotype
- **Questioning the generalizability assumption**
- In whom do models fail, and why?

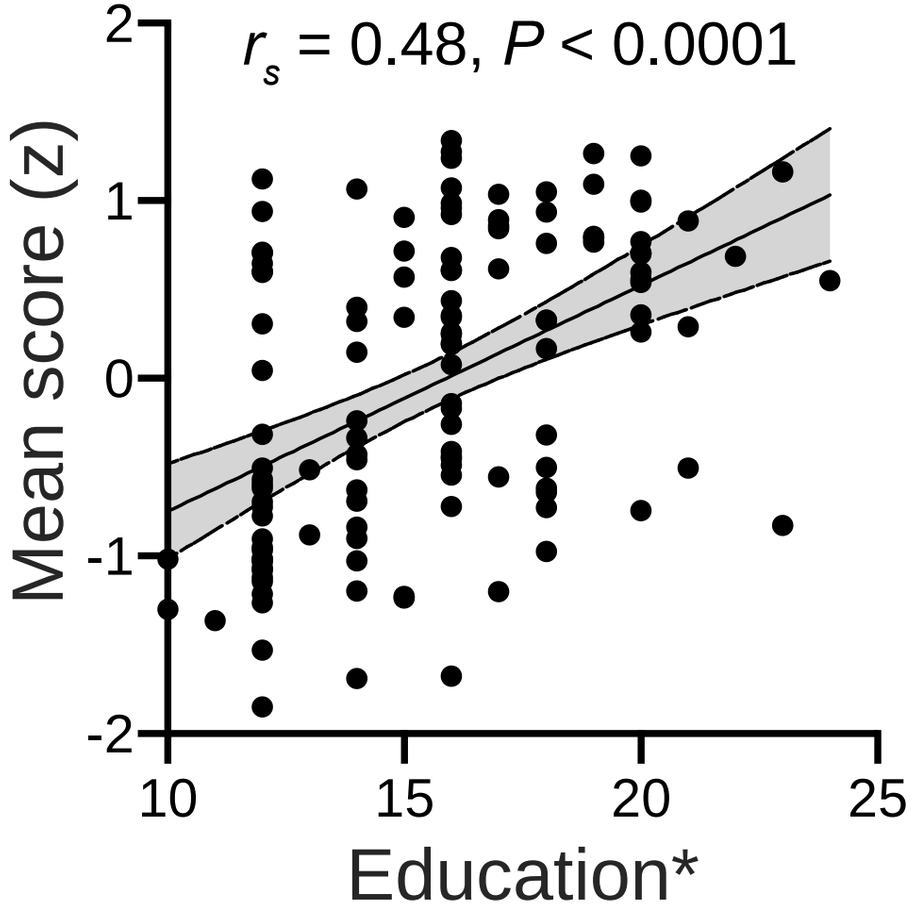
One model does not fit all: model failure is reliable, phenotype-specific, and generalizable across datasets.

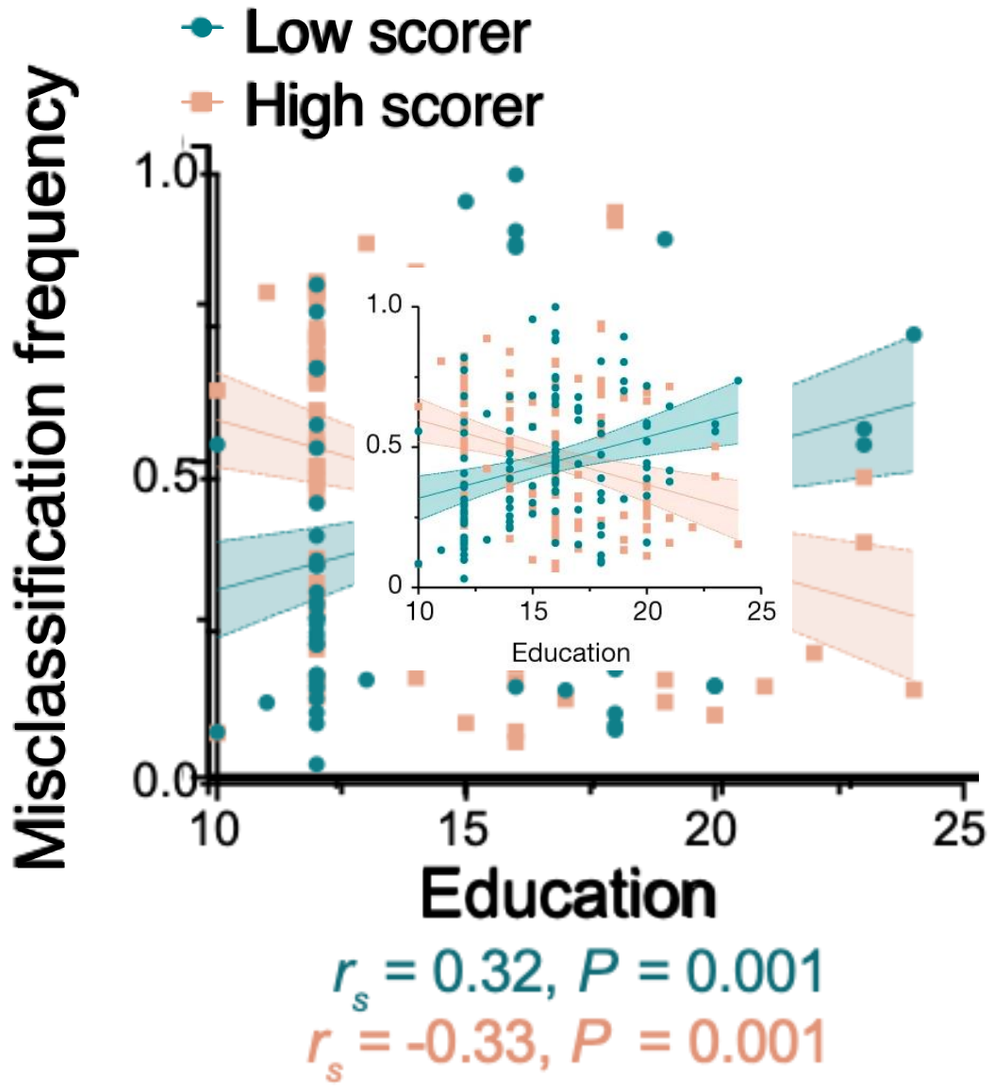
Brain-phenotype models fail for individuals who defy sample stereotypes

- Predictive modeling to relate brain to phenotype
- Questioning the generalizability assumption
- **In whom do models fail, and why?**

Who are the misclassified participants?

Those who defy stereotypical profiles



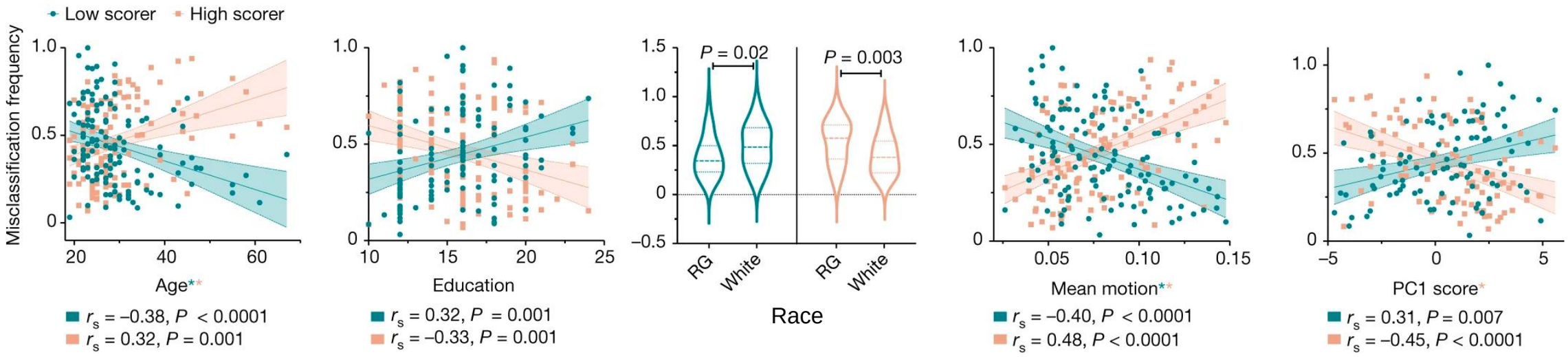


r_s = 0.32, P = 0.001

r_s = -0.33, P = 0.001

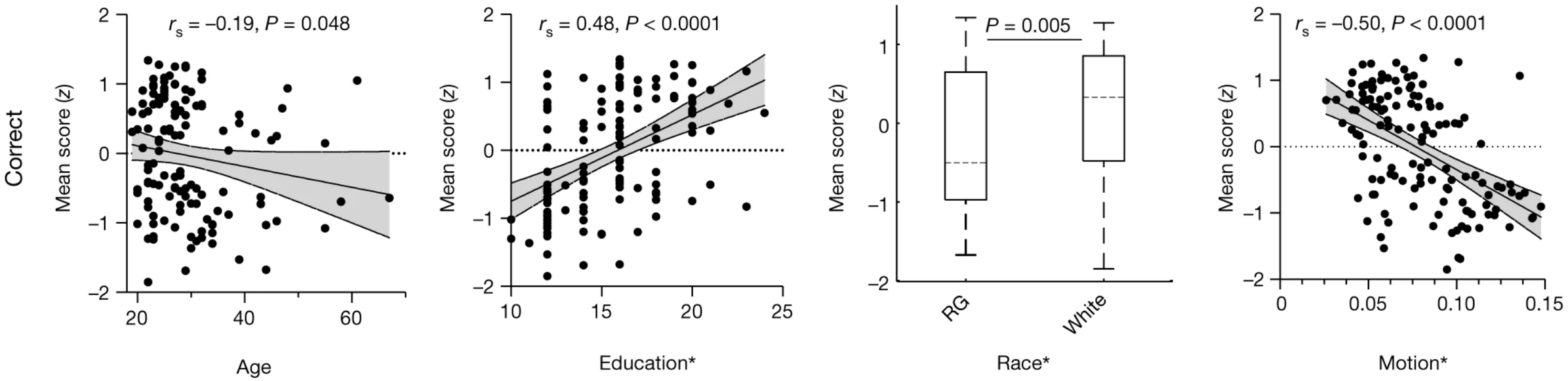
Who are the misclassified participants?

Those who defy stereotypical profiles



Who are the misclassified participants?

Those who defy stereotypical profiles



Brain-phenotype models fail for individuals who defy sample stereotypes

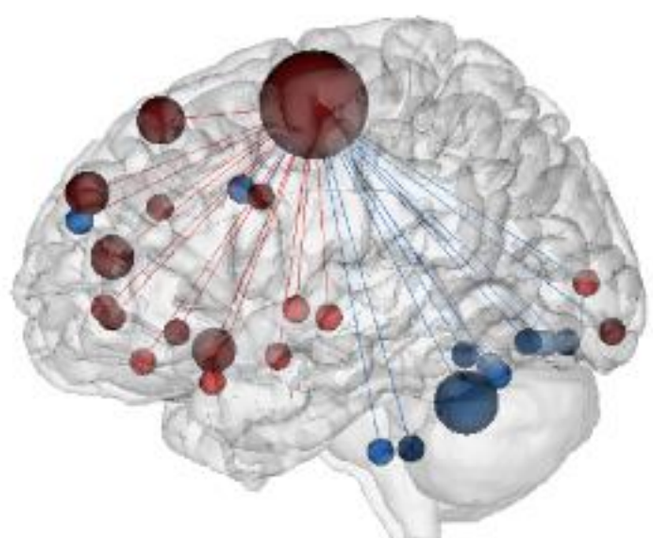
- Predictive modeling to relate brain to phenotype
- Questioning the generalizability assumption
- **In whom do models fail, and why?**

Brain-phenotype models predict stereotypical profiles, and fail for individuals who defy these profiles.

Looking forward: what can we learn from prediction?

	0.9	-0.2	0.3	0.7	-0.6
0.9		-0.7	0.8	-0.4	0.1
-0.2	-0.7		0.1	-0.1	0.6
0.3	0.8	0.1		0.8	-0.2
0.7	-0.4	-0.1	0.8		0.3
-0.6	0.1	0.6	-0.2	0.3	

+



Thank you!

Constable lab + collaborators

Todd Constable
Xilin Shen
Dustin Scheinost
Stephanie Noble
Daniel Barron
Marisa Spann
Alice Hahn
Fuyuze Tokoglu
Jagriti Arora
Eve Lake
Dave O'Connor

Siyuan Gao
Corey Horien
Xenios Papademetris
Emily Finn
Monica Rosenberg
Kangjoo Lee
Wendy Luo
Carmen Carrión
Mehraveh Salehi
Francesca Mandino
Daniel Barson

MD-PhD program
Barbara Kazmierczak, Reiko Fitzsimonds, Fred Gorelick, Cheryl DeFilippo, Sue Sansone, Christine Caminear, Alex Mauzerall

Thesis and qualifying committees
Daeyeol Lee, BJ Casey, Nick Turk-Browne, Mike Nitabach, Marvin Chun, John Murray

Clinical collaborators
Jerry Sanacora, Vinod Srihari, Scott Woods

INP
Charlie Greer, Ben Crome, Carol Russo

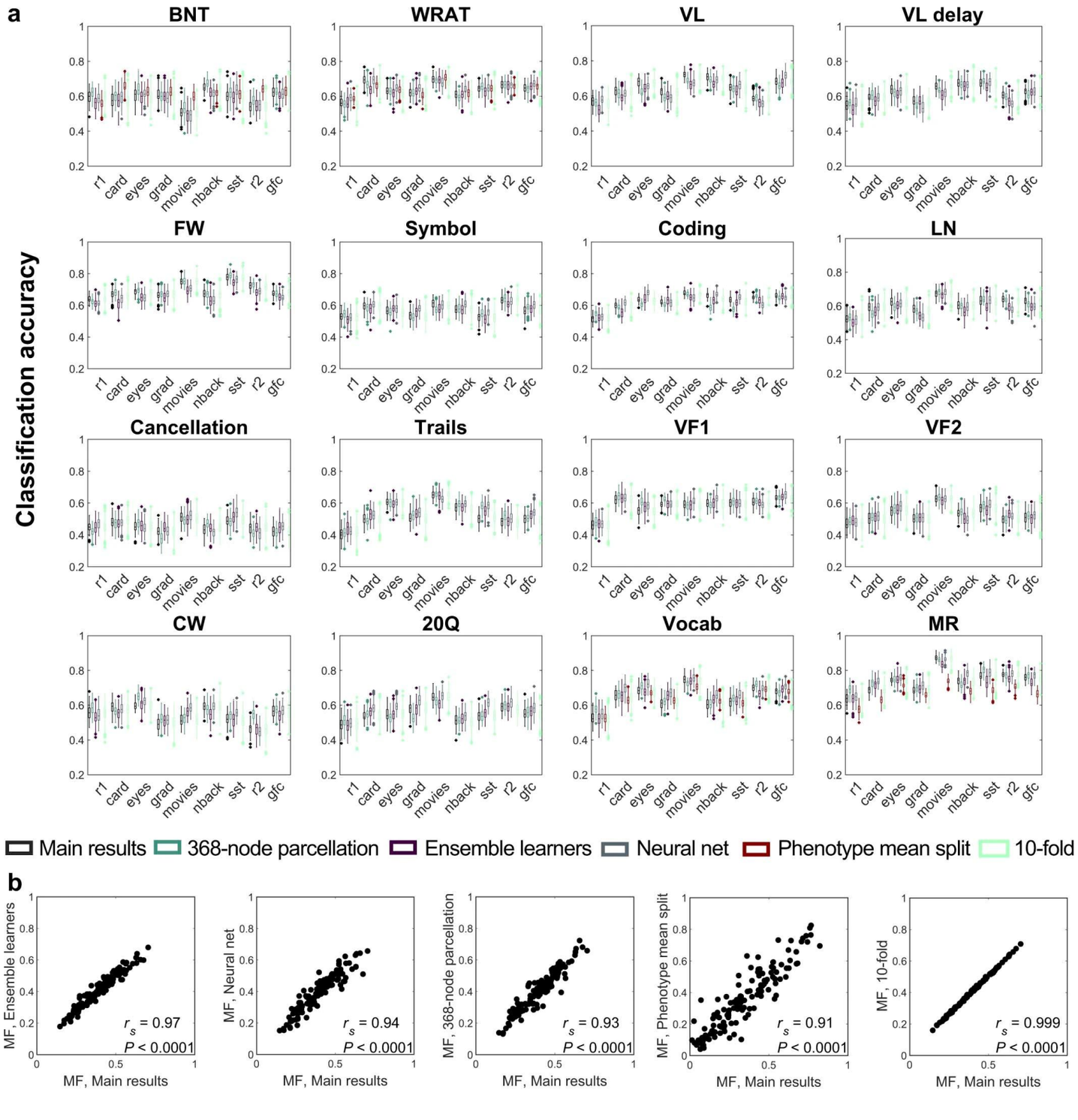
MRRC community

Study participants

Funding sources
NIH MSTP T32GM007205
CTSA TL1 TR001864
NIH R01 MH121095

Additional slides

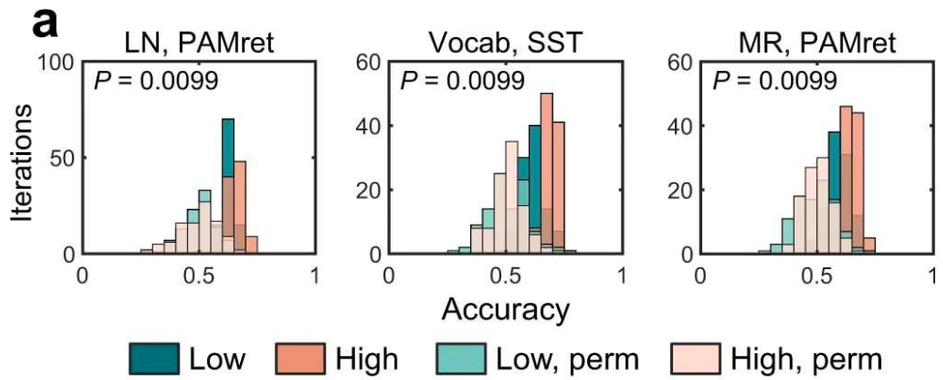
Model performance and misclassification frequency are robust to analysis approach

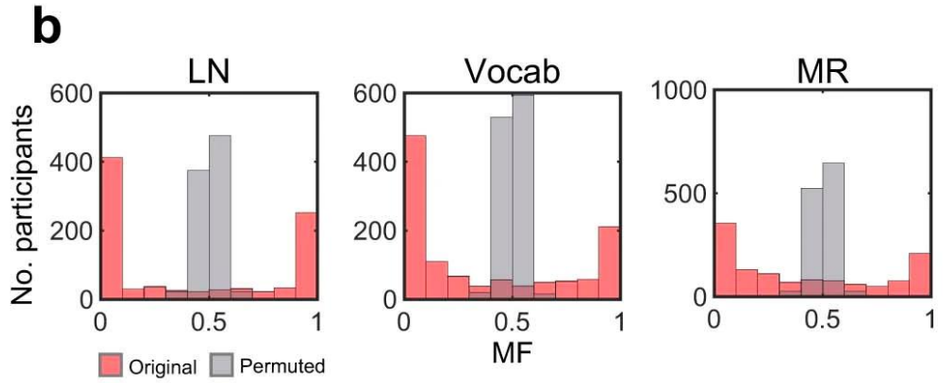


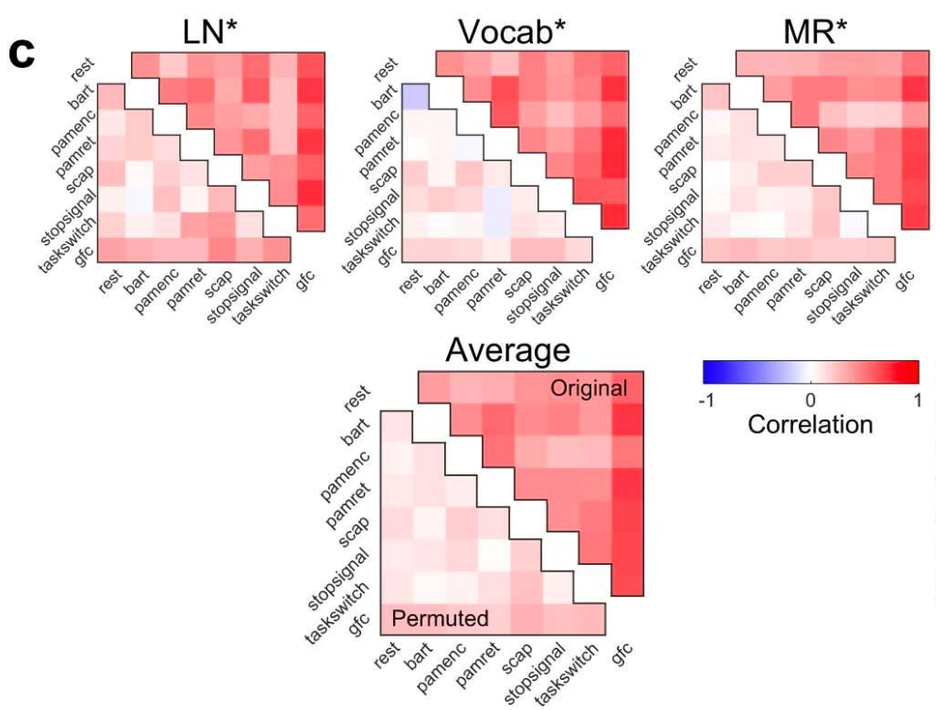
Model performance and misclassification frequency are robust to analysis approach

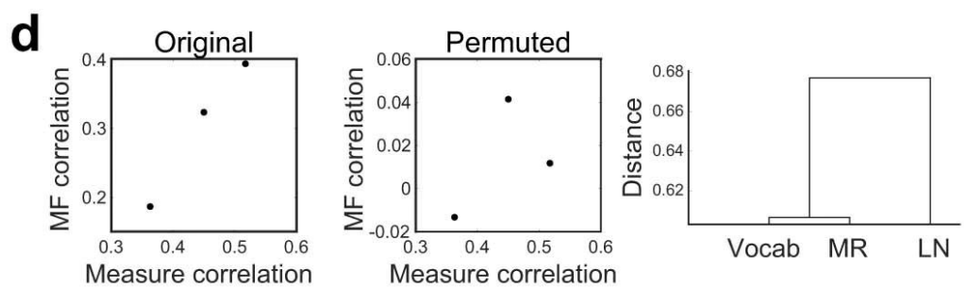
	Relationship to mean residualized prediction deviation
Sex	$P > 0.1$
Age	$r_s = -0.38$, $P_{FDR} < 0.0001$
Race	$\text{median}_{\text{RG-W}} = -0.78$, $P_{FDR} = 0.0005$
Symptom severity	$P > 0.1$
Education	$r_s = 0.29$, $P_{FDR} = 0.003$
Stress	$P > 0.1$
PSQI	$P > 0.1$
Positive affect	$P > 0.1$
Negative affect	$P > 0.1$
Rx	$P > 0.1$
Dx	$P > 0.1$
PC1 score	$r_s = 0.30$, $P_{FDR} = 0.007$
PC2 score	$r_s = -0.20$, $P_{FDR} = 0.11$
PC3 score	$r_s = -0.17$, $P = 0.09$
Mean motion	$r_s = -0.39$, $P_{FDR} < 0.0001$

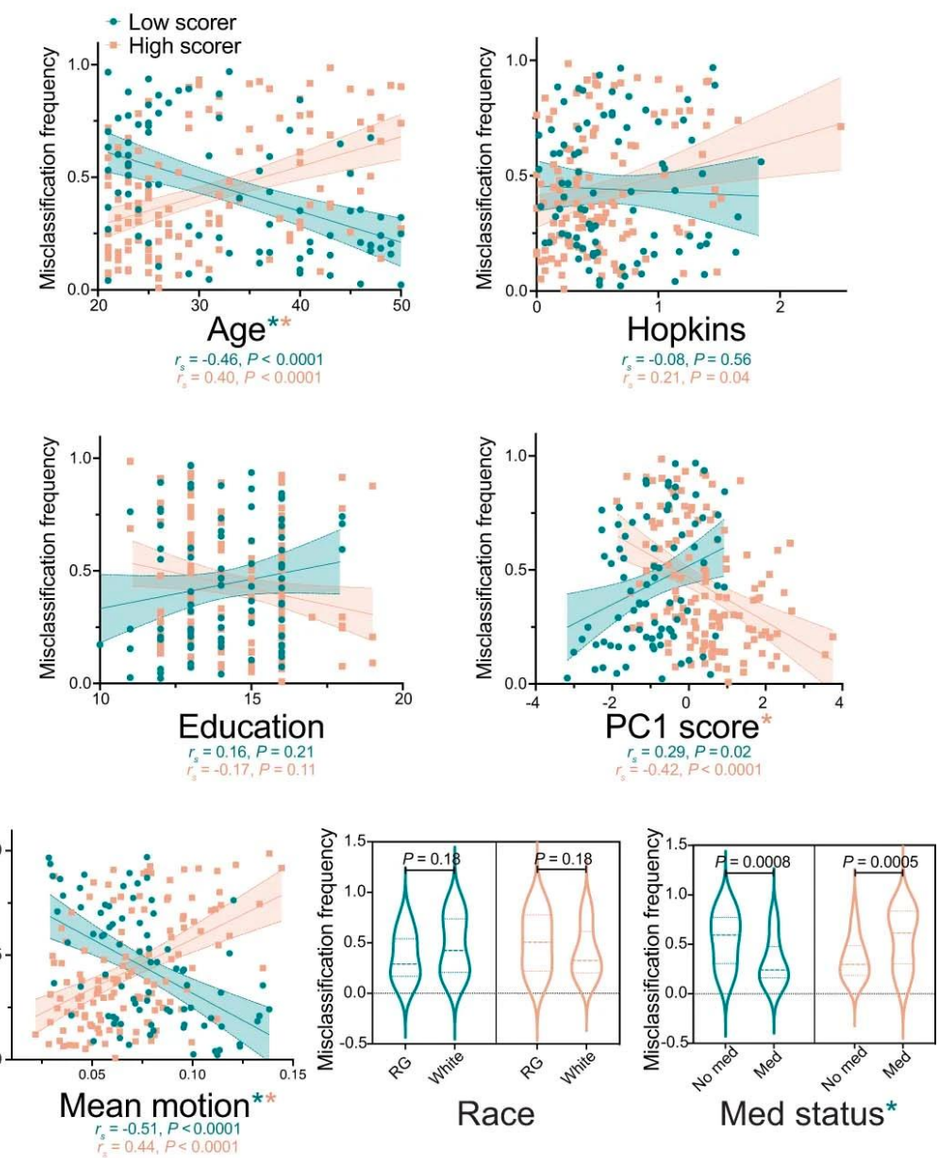
UCLA replication











HCP replication