Brain Wide Associations (BWAS) to model the link between brain features and behavior.

Oscar Miranda Domínguez, PhD, MSc
Assistant Professor | Department of Pediatrics | Medical School | University of Minnesota
Masonic Institute for the Developing Brain
Minnesota Supercomputer Institute

MASONIC INSTITUTE FOR THE DEVELOPING BRAIN
UNIVERSITY OF MINNESOTA
Driven to Discover®

BWAS/PNRS to model the brain-behavior

- Motivation
- Description of the method
- Using BWAS we can disambiguate between focal or globally distributed effects
- Detailed description of the figures and tables we use to validate every step of the method
- Other potential applications and future directions

Acknowledgments

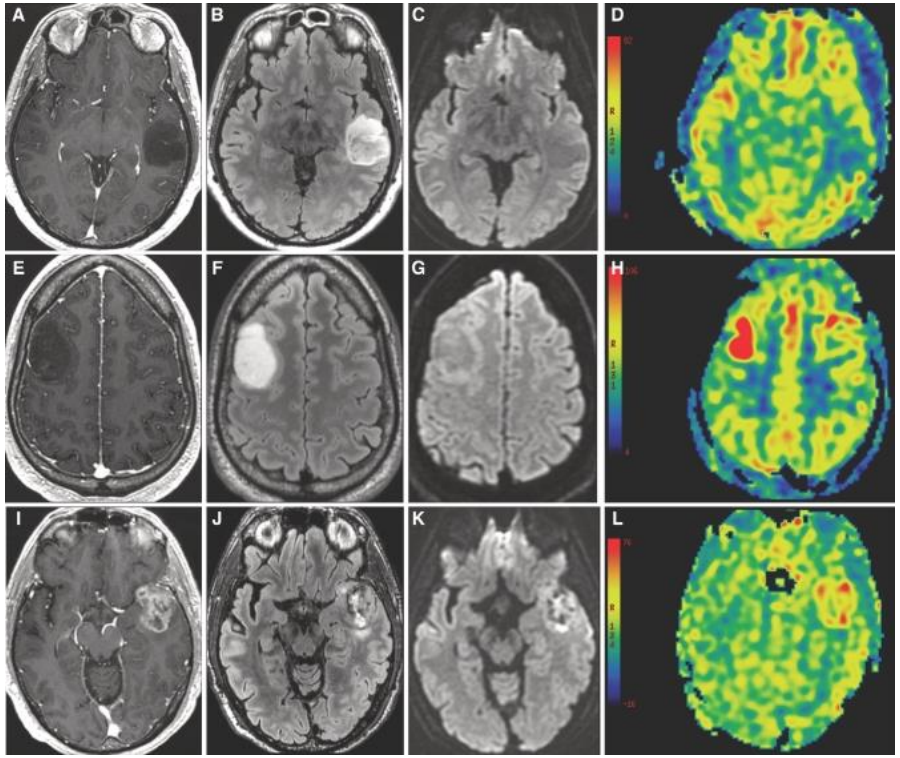
Development team
Damien Fair
Nora Byington
Gracie Grimsrud
Eric Feczko

Patient beta-testers
Amandine Marie Van Rinsveld
Dakota Kliamovich
Elina Thomas
Otiti Mayo

DCAN Lab
Alice Graham
Madeleine Allen
Amanda Rueter
Samuel Carpenter
Greg Conan
Trevor Day
Ekom Eyoh
Robert Hermosillo
Audrey Houghton
Kristina Hufnagle
Sanju Koirala
Timothy Hendrickson
Thomas Madison
AJ Mitchell
Lucille Moore
Cristian Morales Carrasco
Julia Moser
Samantha Papadakis
Anders Perrone
Anjanibhargavi Ragothaman
Anita Randolph
Paul Reiners
Kathy Snider
Amal Aaden

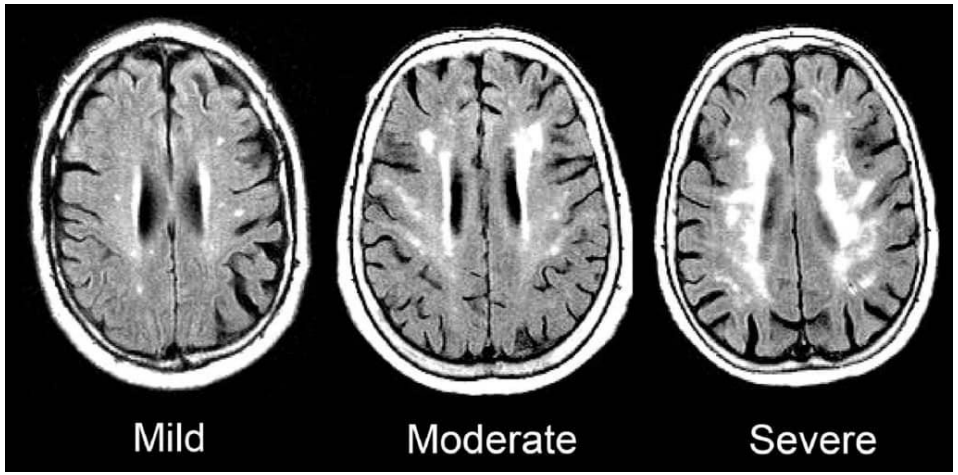
Neuroimaging is a very important tool in clinical practice

Diffuse astrocytic tumors



Presurgical MRI of 3 patients included axial T1 postcontrast **A, E, I**, axial FLAIR **B, F, J**, axial DWI **C, G, K**, and axial ASL perfusion **D, H, L** sequences. Villanueva-Meyer, et al 2017. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5581219/

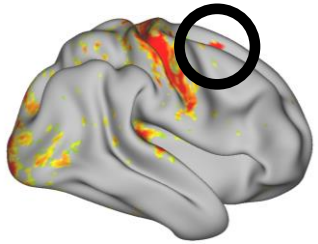
Signs of cerebral small vessel disease



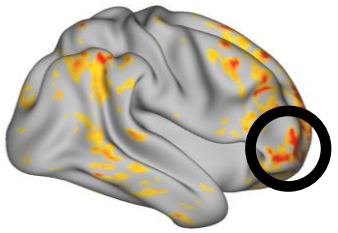
Signs of cerebral small vessel disease. From Inzitari et al, BMJ. 2009 Jul 6;339:b2477. doi: 10.1136/bmj.b2477
https://betterhealthwhileaging.net/cerebral-small-vessel-disease/

Associations between brain function and behavior, however, are dominated by small effects.

Given their small effects, it is hard to identify a “bright spot” in the brain indicative of a given behavior

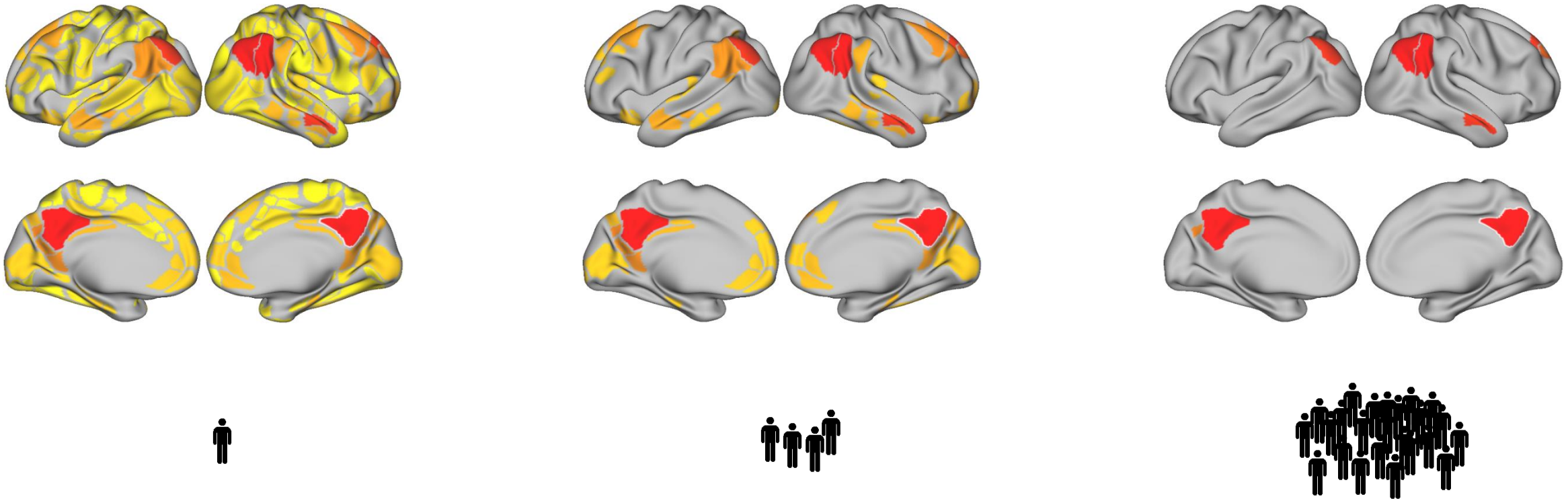


I see, rumination and internalizing behavior



Aha, an extrovert

One strategy is to pool data together from several participants to increase the signal to noise ratio and find those associations at the group level.



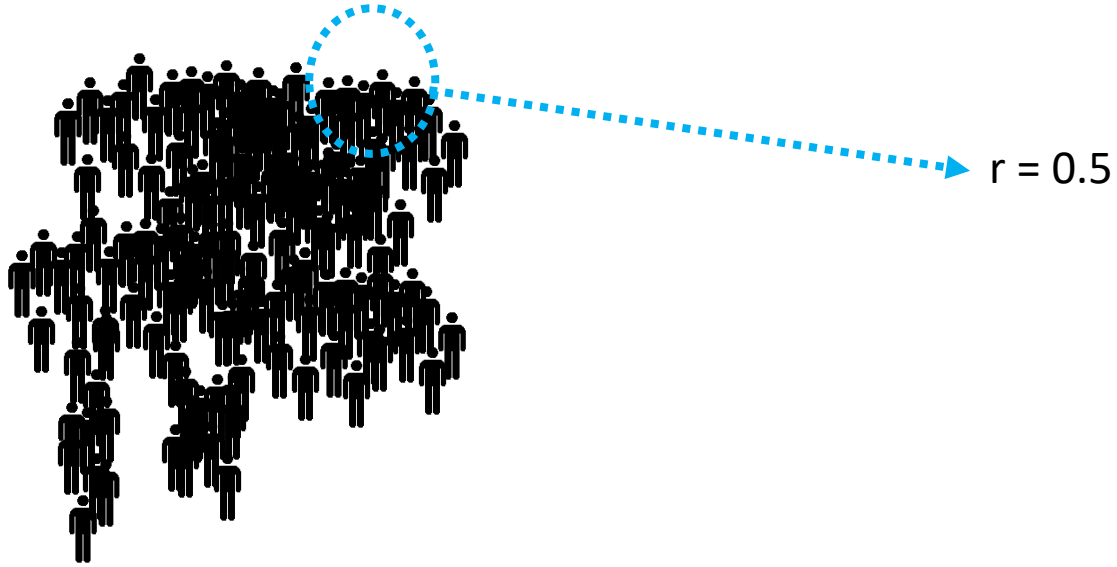
Following this approach, several groups have reported associations between atypical brain connectivity and different mental and neurological disorders.

Most of the studies, however, are underpowered to model small effects

Towards Reproducible Brain-Wide Association Studies
Scott Marek, Brenden Tervo-Clemmens, Finnegan J. Calabro, David F. Montez, Benjamin P. Kay, Alexander S. Hatoum, Meghan Rose Donohue, William Foran, Ryland L. Miller, Eric Feczko, Oscar Miranda-Dominguez, Alice M. Graham, Eric A. Earl, Anders J. Perrone, Michaela Cordova, Olivia Doyle, Lucille A. Moore, Greg Conan, Johnny Uriarte, Kathy Snider, Angela Tam, Jianzhong Chen, Dillan J. Newbold, Annie Zheng, Nicole A. Seider, Andrew N. Van, Timothy O. Laumann, Wesley K. Thompson, Deanna J. Greene, Steven E. Petersen, Thomas E. Nichols, B.T. Thomas Yeo, Deanna M. Barch, Hugh Garavan, Beatriz Luna, Damien A. Fair, Nico U.F. Dosenbach

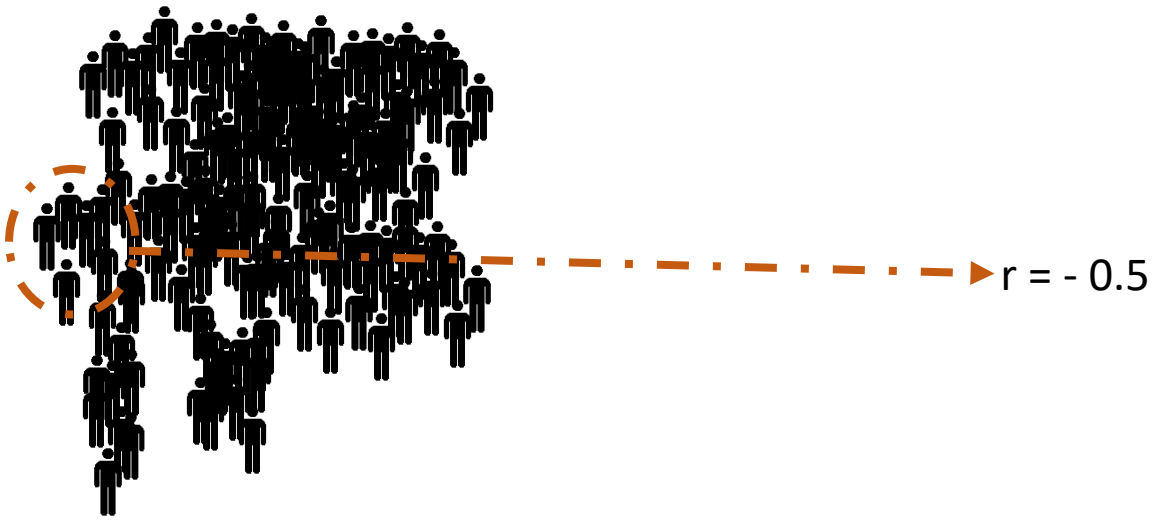
For example, a small sample of 25 participants can lead to strong positive correlations

Predicting cognitive ability
using functional connectivity



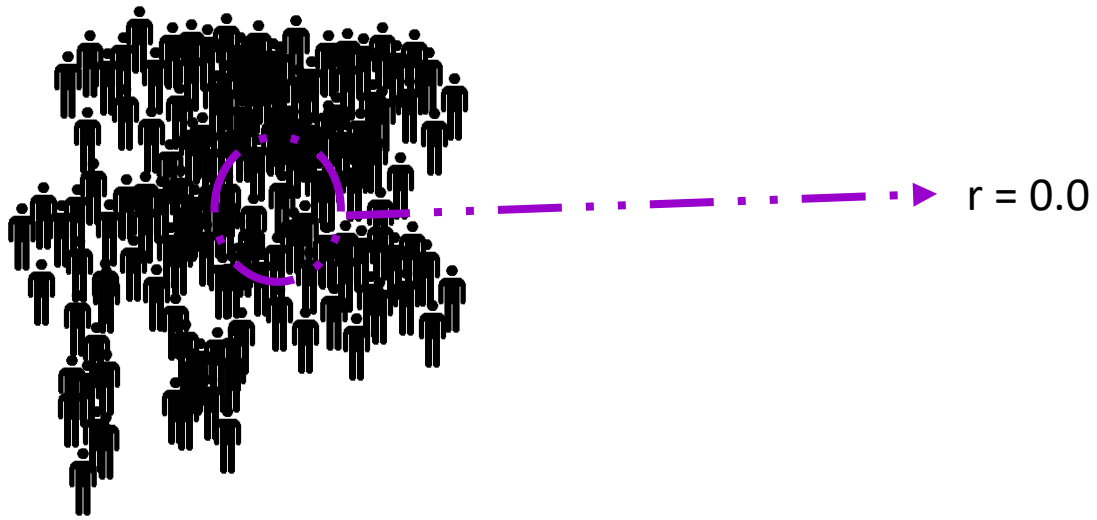
Marek and Tervo-Clemmens, *et. al.*
Towards Reproducible Brain-Wide Association Studies

Another sample of the same size can lead to a strong negative correlation



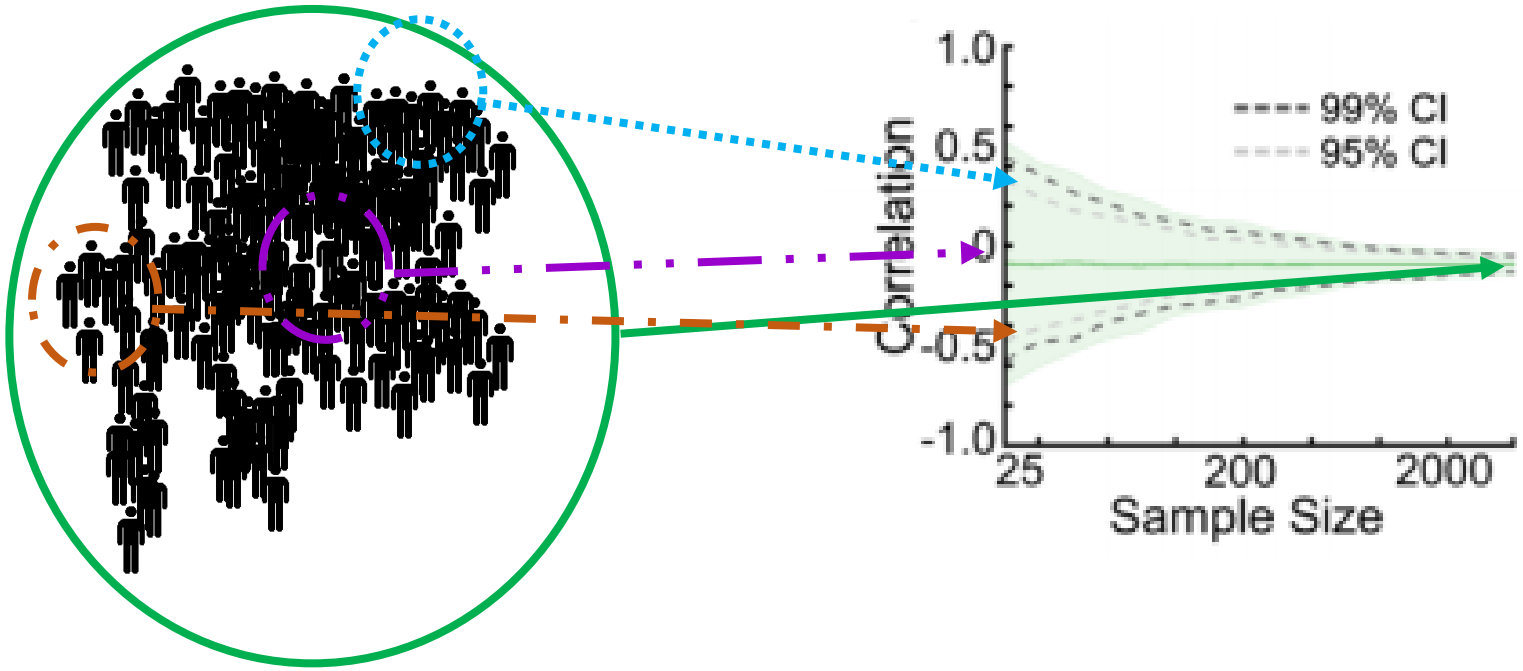
Marek and Tervo-Clemmens, *et. al.*
Towards Reproducible Brain-Wide Association Studies

Or even to a null correlation



Marek and Tervo-Clemmens, *et. al.*
Towards Reproducible Brain-Wide Association Studies

You need a sample size in the order of thousands to characterize the real strength of the association



Marek and Tervo-Clemmens, *et. al.*
Towards Reproducible Brain-Wide Association Studies

Underpowered studies and bias towards overreporting positive findings (among other factors) have led to a crisis in reproducibility

Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates

Anders Eklund[a,b,c,1], Thomas E. Nichols[d,e], and Hans Knutsson[a,c]

[a]Division of Medical Informatics, Department of Biomedical Engineering, Linköping University, S-581 85 Linköping, Sweden; [b]Division of Statistics and Machine Learning, Department of Computer and Information Science, Linköping University, S-581 83 Linköping, Sweden; [c]Center for Medical Image Science and Visualization, Linköping University, S-581 83 Linköping, Sweden; [d]Department of Statistics, University of Warwick, Coventry CV4 7AL, United Kingdom; and [e]WMG, University of Warwick, Coventry CV4 7AL, United Kingdom

Edited by Emery N. Brown, Massachusetts General Hospital, Boston, MA, and approved May 17, 2016 (received for review February 12, 2016)

Nature. Author manuscript; available in PMC 2014 Jun 16.
Published in final edited form as:
Nature. 2014 Jan 30; 505(7485): 612–613.

PMCID: PMC4058759
NIHMSID: NIHMS574872
PMID: 24482835

NIH plans to enhance reproducibility

Francis S. Collins, director and Lawrence A. Tabak, principal deputy director

Nat Rev Neurosci. Author manuscript; available in PMC 2019 Dec 13.
Published in final edited form as:
Nat Rev Neurosci. 2017 Feb; 18(2): 115–126.
Published online 2017 Jan 5. doi: 10.1038/nrn.2016.167

PMCID: PMC6910649
NIHMSID: NIHMS1060257
PMID: 28053326

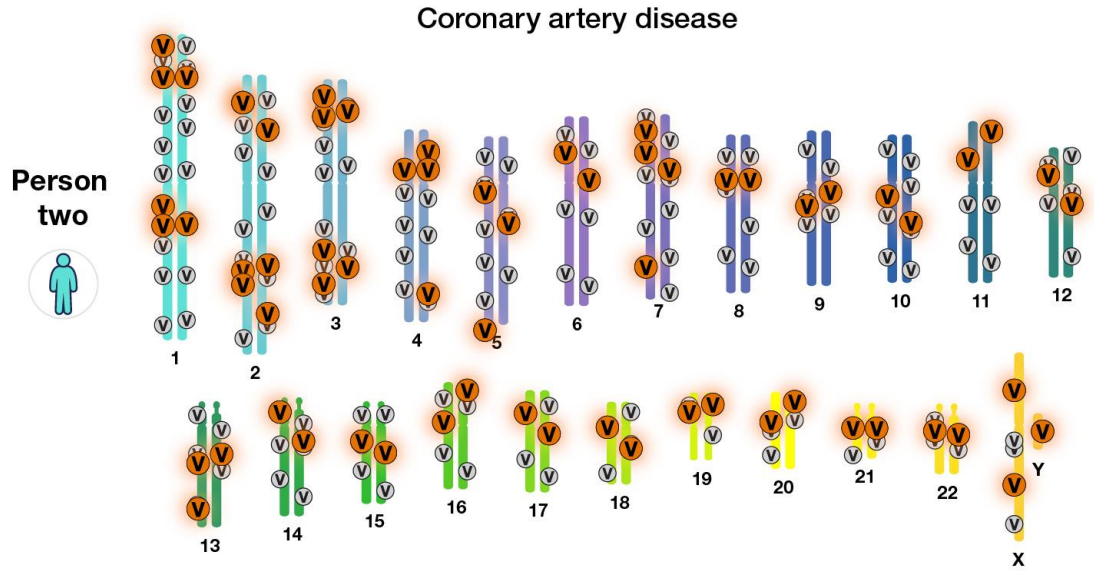
Scanning the horizon: towards transparent and reproducible neuroimaging research

Russell A. Poldrack,[1] Chris I. Baker,[2] Joke Durnez,[1,3] Krzysztof J. Gorgolewski,[1] Paul M. Matthews,[4] Marcus R. Munafò,[5,6] Thomas E. Nichols,[7] Jean-Baptiste Poline,[8] Edward Vul,[9] and Tal Yarkoni[10]

Weak associations between variables are very common in science.

In genetics, for example, several illnesses are associated with changes in either one or many genes (and frequently coupled with environmental factors)

In genetics, a “**polygenic risk score**” is a method to determine the risk of developing a disease, based on the total number of changes related to the disease



Each red “v” represents variants in an individual’s genome that is associated with coronary artery disease. Each smaller gray “v” is a variant that is also present in the person’s genome but is not implicated in disease.

https://www.genome.gov/Health/Genomics-and-Medicine/Polygenic-risk-scores

In that field, a large reference sample (~10^2 - 10^{10}), whose sample size depends on the frequency of the genetic variant to be identified, is used to characterize the association between gene expression and disease

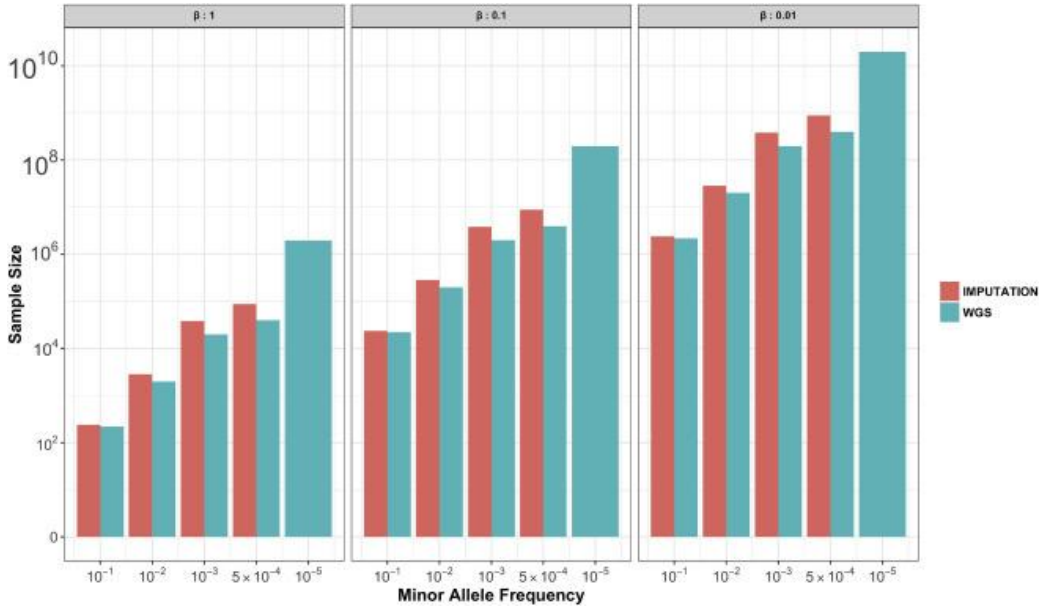
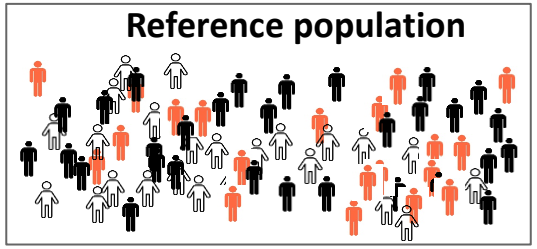


Figure 1. Minimum Sample Sizes for Detecting Trait-SNP Associations from Imputed and WGS Data

Visscher, Peter M., Naomi R. Wray, Qian Zhang, Pamela Sklar, Mark I. McCarthy, Matthew A. Brown, and Jian Yang. 2017. “10 Years of GWAS Discovery: Biology, Function, and Translation.” *American Journal of Human Genetics* 101 (1): 5–22.

They model each individual effect in the large reference sample to obtain a Genome Wide Association

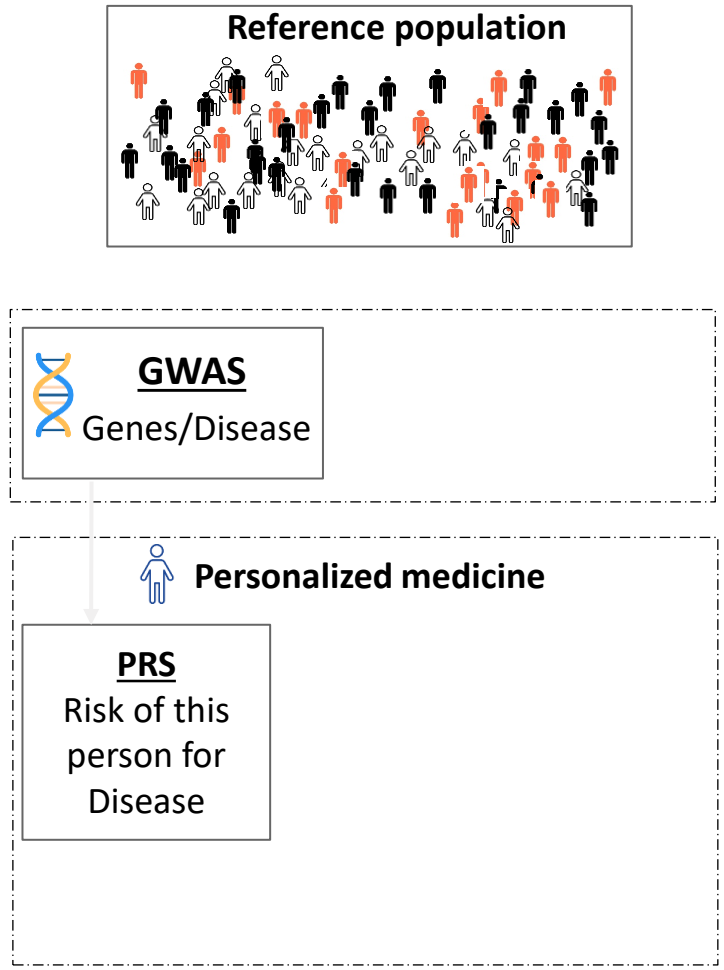
Those are known as Genome Wide Association Studies, or GWAS



Resulting models can be used to predict the risk of a given disease in individuals

Based on the GWAS, a person's gene expression is used to calculate a polygenic risk score for a disease

PRS: Polygenic Risk Scores



Zhao among others adapted that methodology to neuroimaging

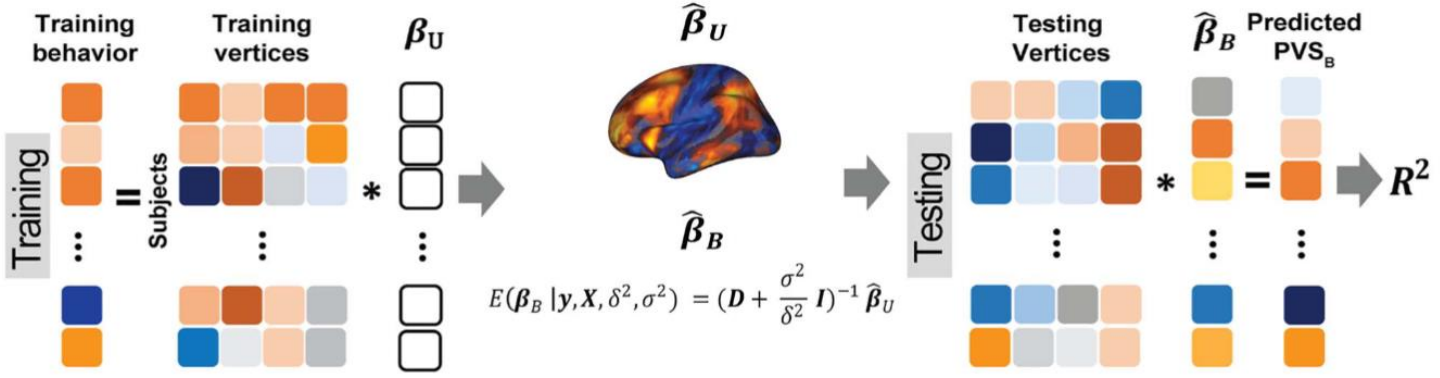


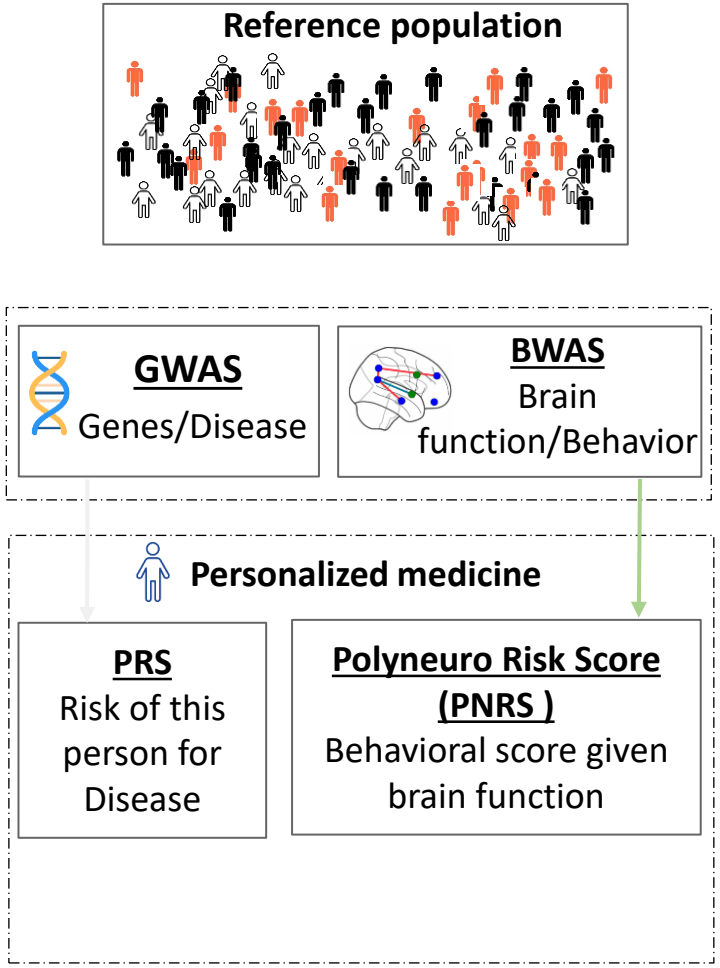
Figure 1. Overview of the PVS_B and the PVS_U algorithms. Ten-fold cross-validation was performed to obtain a PVS_B for each individual. For each fold, mass univariate summary statistics, $\hat{\beta}_U$, were obtained from the training set which contained 90% of the complete sample. Posterior mean effect sizes at each vertex, $\hat{\beta}_B$, were approximated by multiplying the mass univariate beta estimates, $\hat{\beta}_U$, by the inverse of the correlation structure of the brain, D, and a shrinkage factor that accounts for the number of vertices, V, and the total signal of the brain-behavior association. The PVS_B was subsequently calculated for the test set participants by multiplying their imaging phenotype with the $\hat{\beta}_B$. Simulations were conducted at three levels of total explainable signal, six levels of study sample size, and four levels of proportion of non-null vertices, yielding 60 instantiations of simulation conditions with 100 iterations per condition.

Zhao, Weiqi, Clare E. Palmer, Wesley K. Thompson, Bader Chaarani, Hugh P. Garavan, B. J. Casey, Terry L. Jernigan, Anders M. Dale, and Chun Chieh Fan. 2020. "**Individual Differences in Cognitive Performance Are Better Predicted by Global Rather Than Localized BOLD Activity Patterns Across the Cortex**." Cerebral Cortex 31 (3): 1478–88.

We are using the same approach

Brain-Wide Association Studies (BWAS)| PolyNeuro Risk Scores (PNRS)

We assume that behavioral traits emerge from the cumulative effect of functional motifs distributed globally across the brain

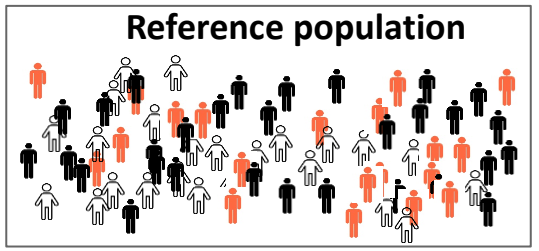


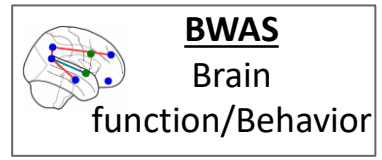
This is a two-step process:

1st, a Brain-Wide Association is estimated using a large reference sample

Massive Univariate models are fit to relate functional connectivity to behavior

Models are combined to estimate the beta-weights of the BWAS

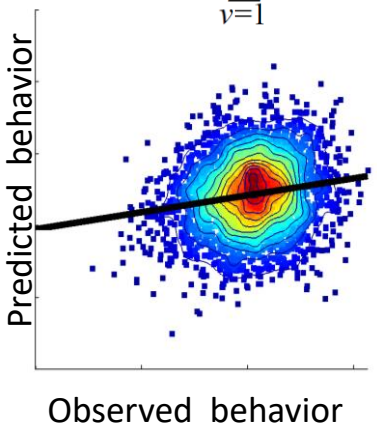




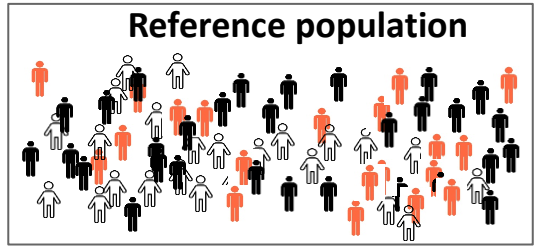
$$\hat{y}_v = x_v\beta_v + Covariate_1\beta_{v,1} + Covariate_2\beta_{v,2} + \cdots,$$

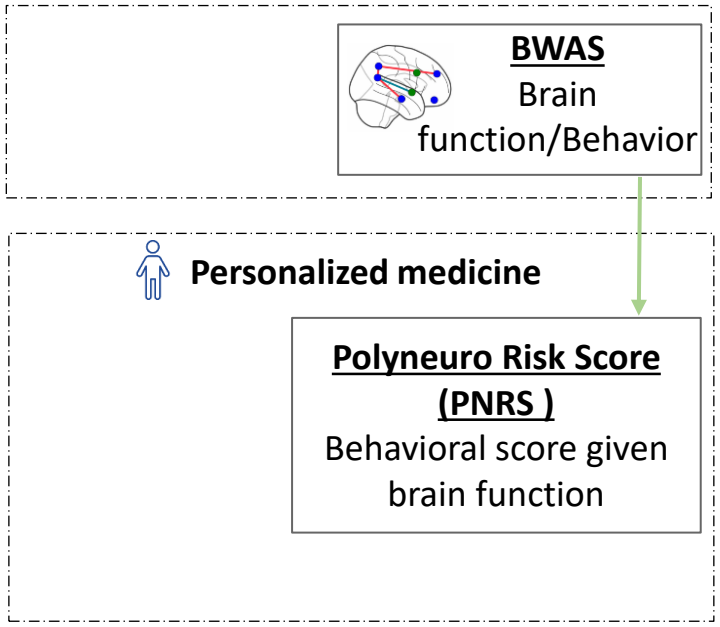
2nd, resulting models are selectively combined to estimate “behavioral” risk scores in individuals (PolyNeuro Risk Score, PNRS)

$$y_{score} = \sum_{v=1}^{N} x_v \beta_v$$



v are the preselected weights



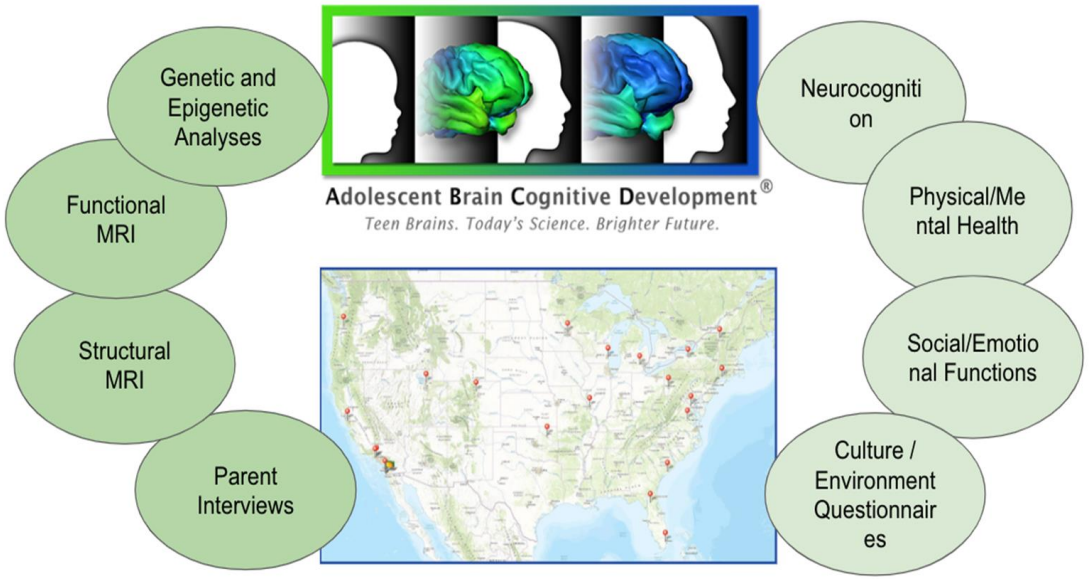


Example: Finding associations between functional connectivity and cognitive ability

Dataset: ABCD

- ABCD (N= 11,877)
  - Data split in 2 halves (5,786 each)

  ABCD Reproducible Matched Samples (ARMS [1])
- Covariates
  - site, gender, combined race, latin, highest parent education, interview age
- Motion censoring
  - Frame displacement <= 0.2
  - Time: 8 minutes
- Surviving participants
  - **ARMS-1: 3,383**
  - **ARMS-2: 3,286**



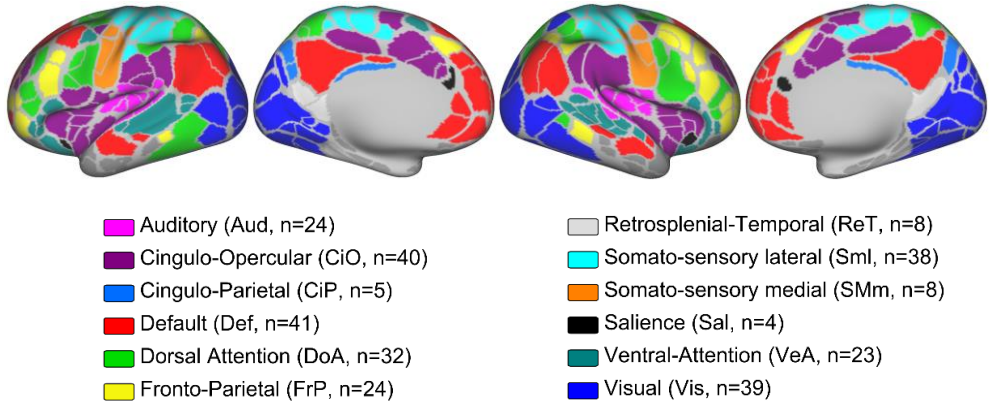
[1] Feczko, et al, 2021

Brain features: Resting-State Functional Connectivity

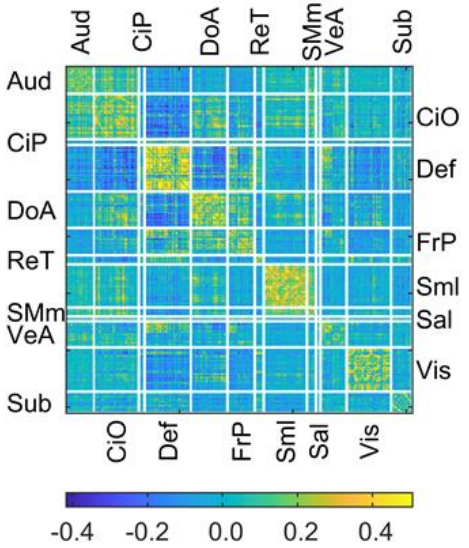
For each participant

- Parcellated functional connectivity data using Gordon's ROI set [2]
  - 352 ROIs
  - 333 cortical areas + 19 subcortical
  - 14 functional networks
- **Resulting number of connections (Brain features)**
  - **61,776 connections**
  - **Grouped in 105 functional network pairs**
    - Aud-Aud
    - Aud –DoA
    - …
    - Sub-Sub

Cortical projection of the Gordon ROI set in a *very inflated* brain



Parcellated connectivity matrix



[2] Gordon, et al, 2014

Behavior: Cognitive Ability

- PCA on neurocognitive assessments [3]
  - NIH toolbox (Picture Vocabulary; Flanker Test; List Sort Working Memory Task; Dimensional Change Card Sort Task; Pattern Comparison Processing Speed Task; Picture Sequence Memory Task; Oral Reading Test)
  - Rey Auditory Verbal Learning Task;
  - Little Man Task percent correct.
- Top 3 components explained most of the observed variance
  - **Cognitive ability: 21.1%**
    - Oral Reading, Picture Vocabulary, and List Sort Working Memory tasks
  - Executive function: 20.4%
  - Learning and memory: 18.05

[3] Thompson, et al, 2019

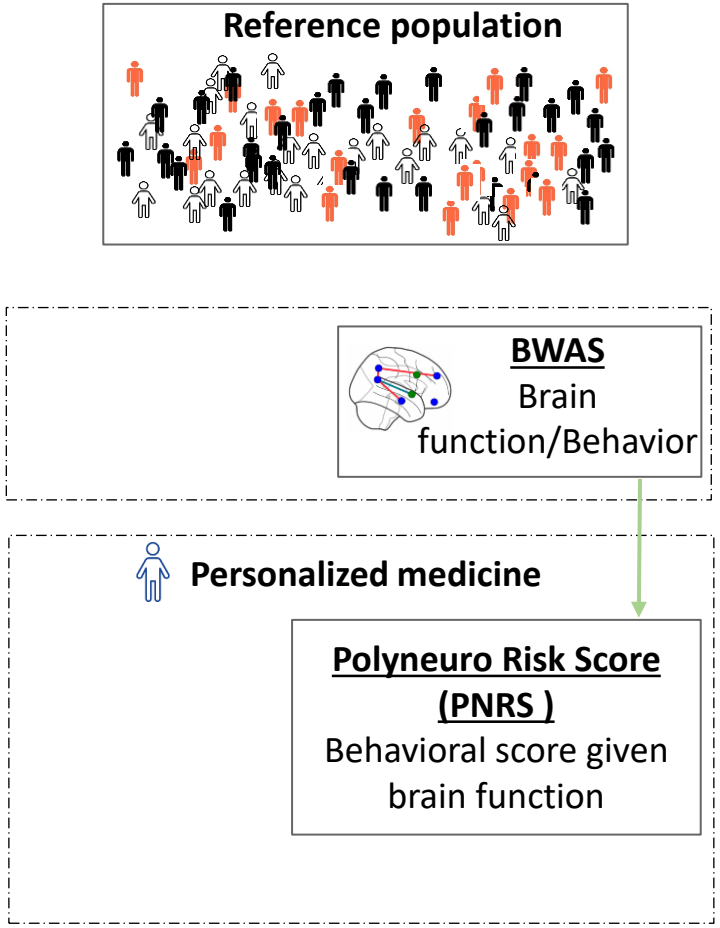
Table 3

Varimax Rotated Loadings for Three-Factor Model.

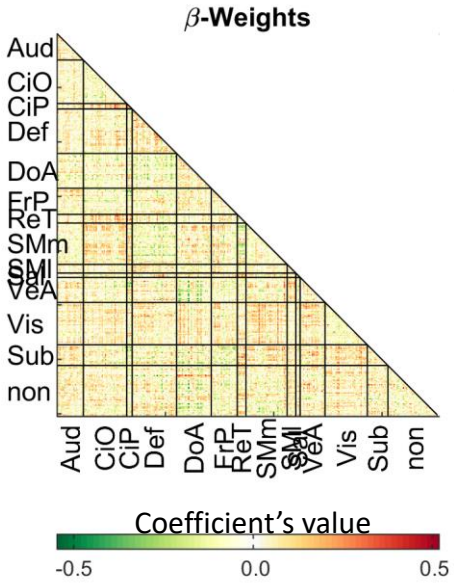
	PC1			PC2			PC3		
	.025	0.50	.975	.025	0.50	.975	.025	0.50	.975
Pic Vocab	0.706	**0.754**	0.799	0.029	0.065	0.102	0.133	0.19	0.252
Flanker	0.161	0.213	0.26	0.668	**0.712**	0.754	0.013	0.067	0.119
List	0.4	**0.471**	0.538	0.105	0.148	0.195	0.416	**0.493**	0.563
Card Sort	0.163	0.205	0.252	0.668	**0.71**	0.751	0.184	0.232	0.287
Pattern	−0.029	0.015	0.055	0.771	**0.813**	0.85	0.039	0.085	0.135
Picture	−0.023	0.012	0.049	0.102	0.135	0.171	0.816	**0.863**	0.904
Reading	0.782	**0.82**	0.86	0.084	0.12	0.16	0.067	0.122	0.173
RAVLT	0.253	0.306	0.364	0.085	0.125	0.163	0.663	**0.712**	0.76
LMT	0.424	**0.5**	0.57	0.246	0.299	0.36	0.002	0.068	0.144

Pic Vocab = Toolbox Picture Vocabulary; Flanker = Toolbox Flanker Test; List Sort = Toolbox List Sort Working Memory Task; Card Sort = Dimensional Change Card Sort Task; Pattern = Toolbox Pattern Comparison Processing Speed Task; Picture = Toolbox Picture Sequence Memory Task; Reading = Toolbox Oral Reading Test; RAVLT = Rey Auditory Verbal Learning Task, total correct; LMT = Little Man Task percent correct. For Toolbox measures, uncorrected scores were entered into the analysis. Loadings above 0.40 are highlighted; this is an arbitrary cutoff intended solely to assist with simple description of the factors, and does not enter into follow-up analyses in any fashion. Quantiles are from the posterior draws of the MCMC algorithm for each factor loading after varimax rotation and give the middle 95% of the distribution of the loadings (i.e., 95% posterior credible intervals).

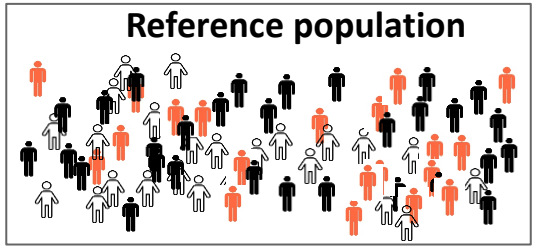
- In this example we first obtained a BWAS using data from the first half of the ABCD set
- We controlled for
  - site,
  - gender,
  - combined race,
  - latin
  - highest parent education
  - interview age

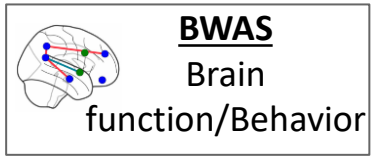


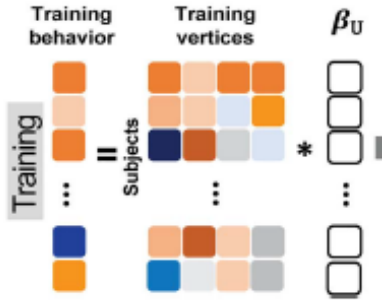
The first result of the BWAS corresponds to the β-weights per connection



Those weights (N=61,776) can be grouped per functional network pair

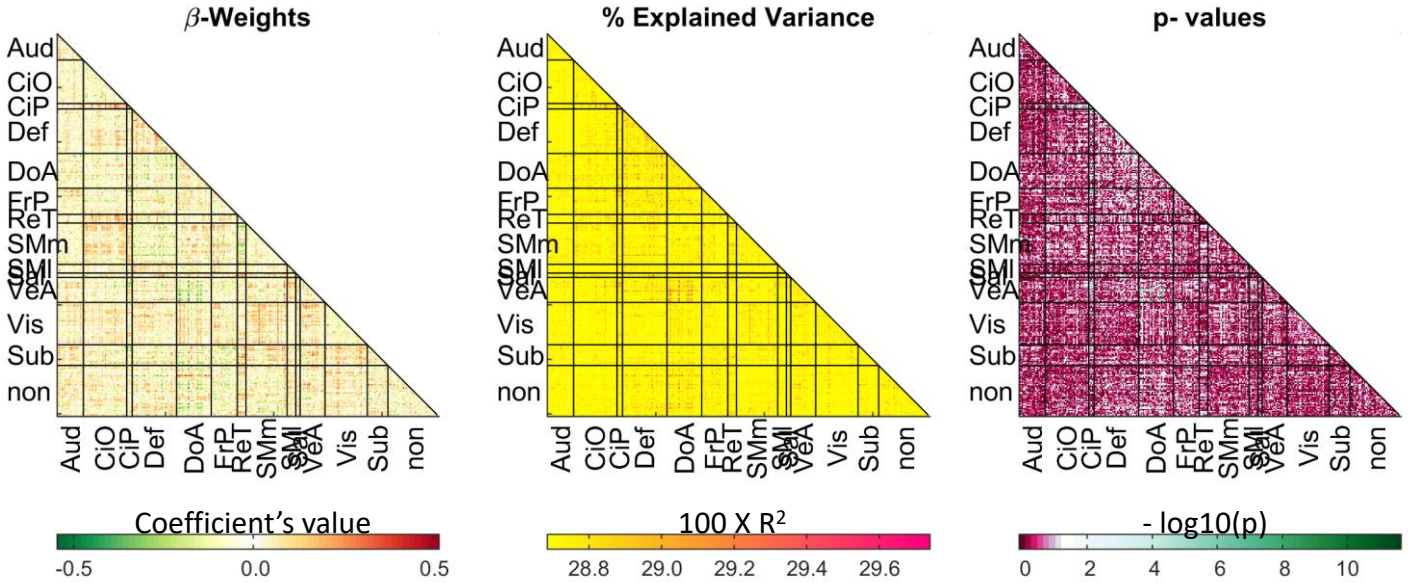






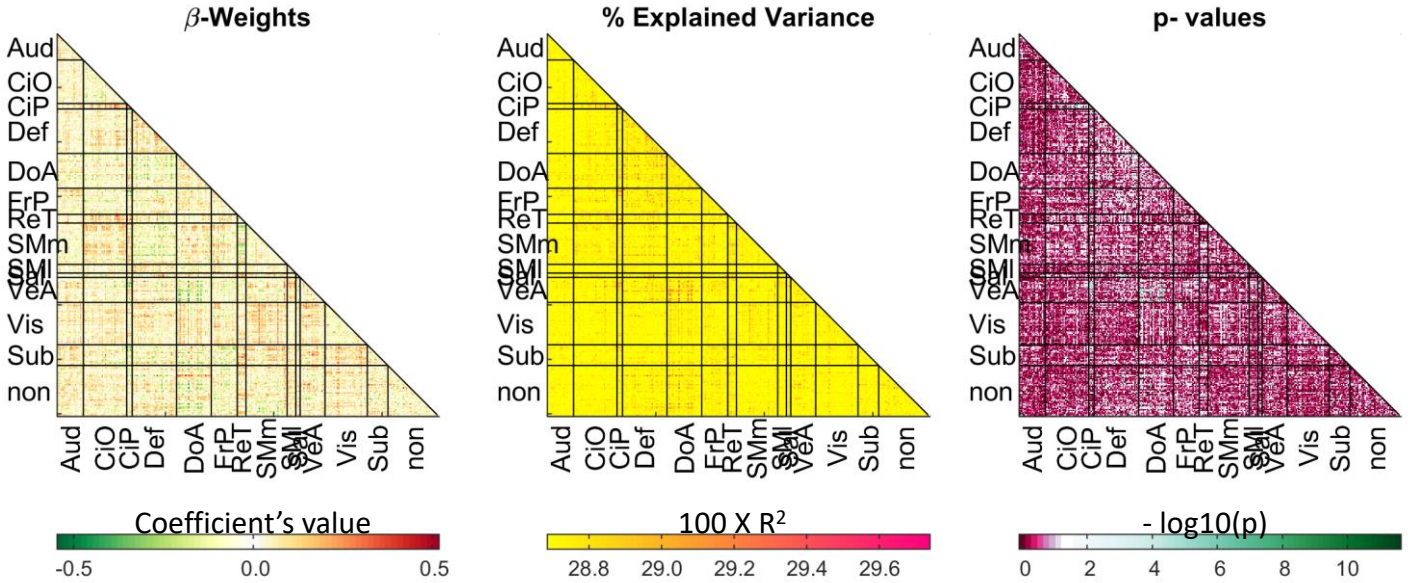
$$\hat{y}_v = x_v\beta_v + Covariate_1\beta_{v,1} + Covariate_2\beta_{v,2} + \cdots,$$

Each weight explains a given amount of variance in the training sample and has a corresponding *p-value*



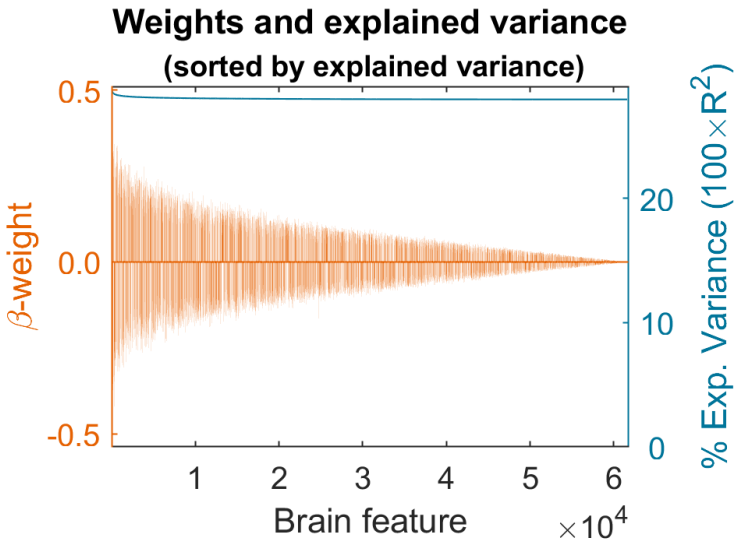
Gordon's ROI set

Visualization can be simplified by showing the weights and the explained variance in the same figure

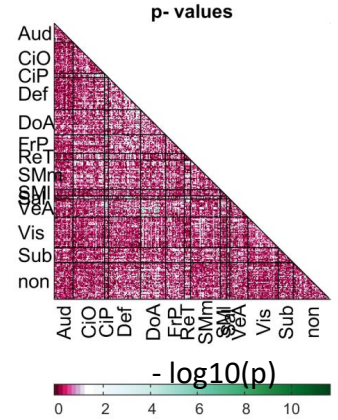


Gordon's ROI set

Results can be summarized as follows

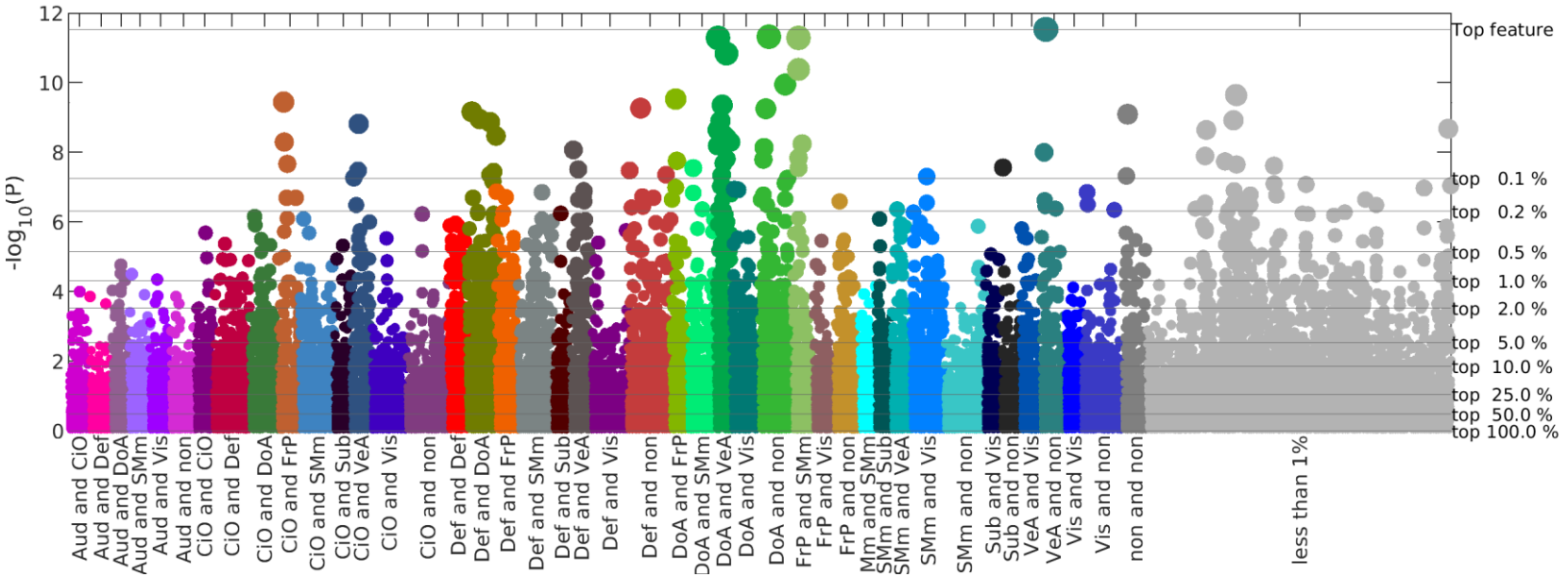


Strength of each connection can also be shown using Manhattan plots



Connections are color-coded by functional network pair

Negative *p-values* are shown in logarithmic scale

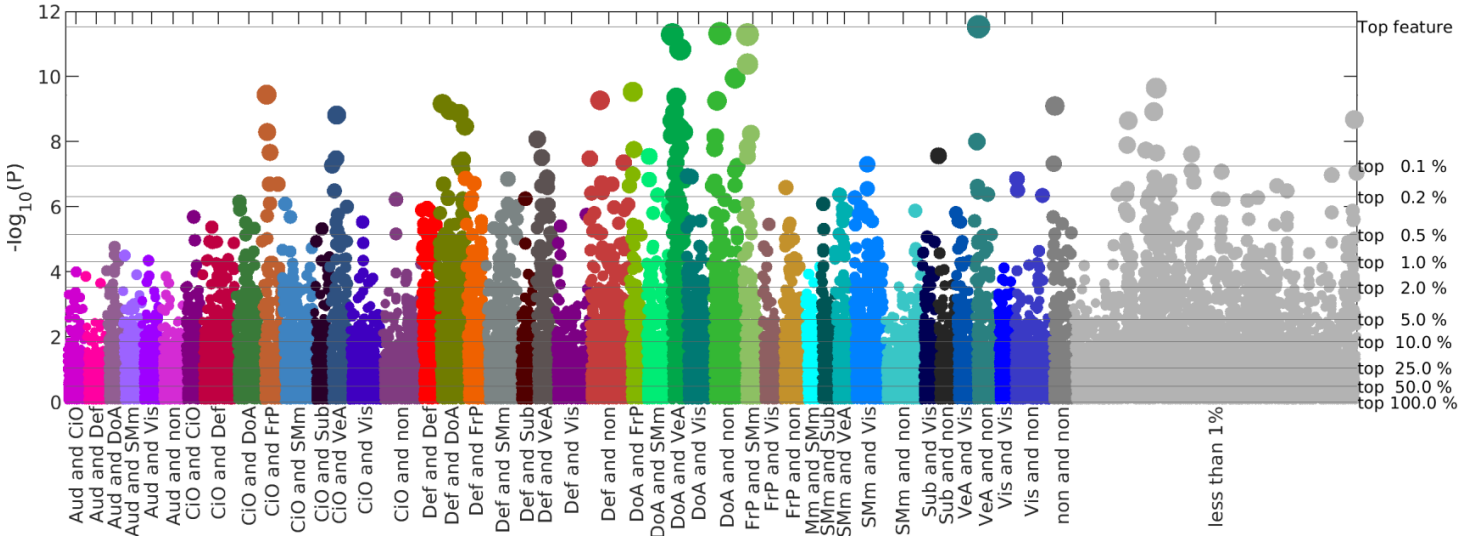


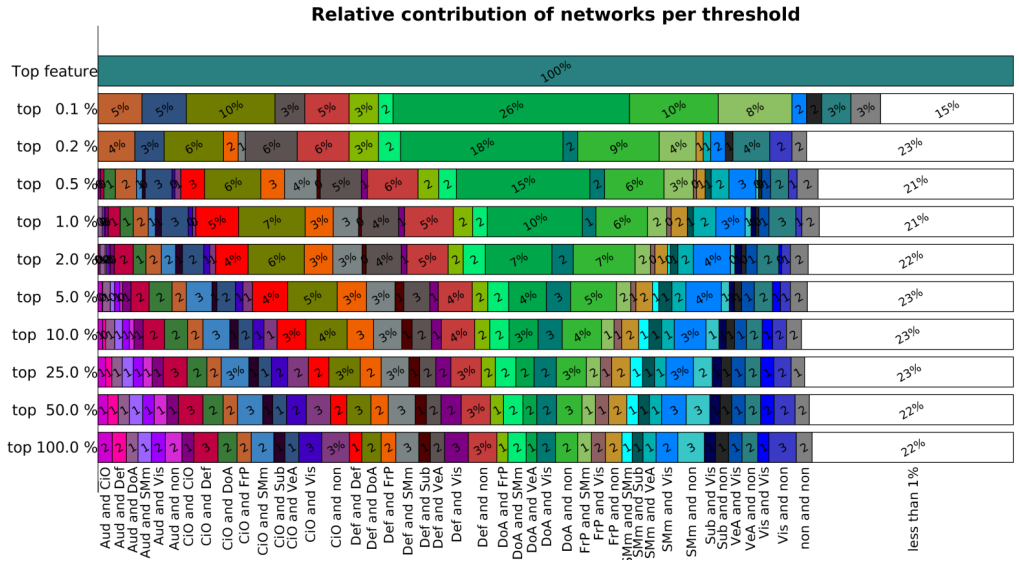
Horizontal lines indicate different thresholds based on p-values

Small networks are combined (additional figures show all the connections and only the small networks)

We also have “Relative Contribution” Figures

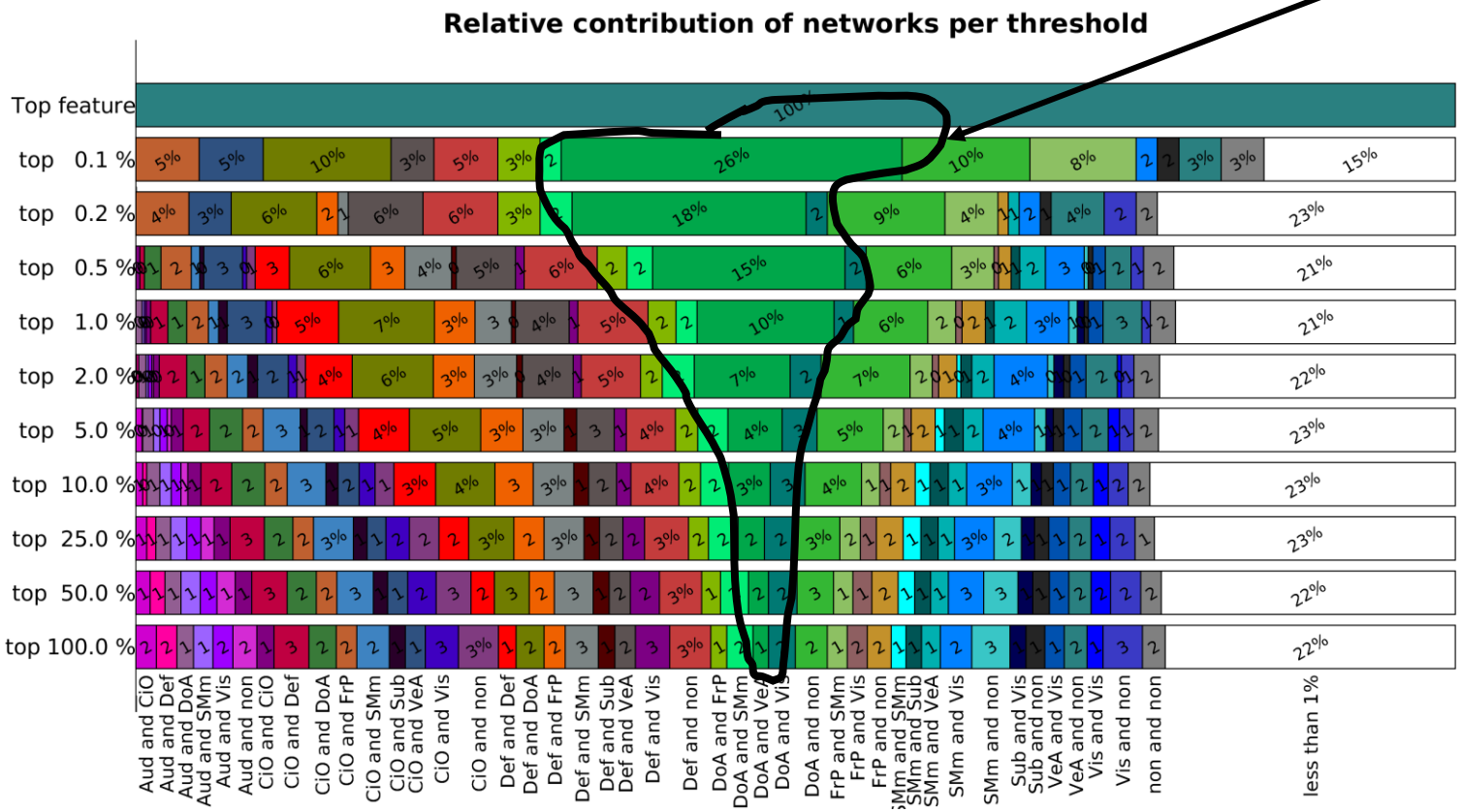
- There is one row per threshold
- Each block shows the relative contribution of each network per threshold
  - Calculated as 100 x (connections per network) / total count of connections at that threshold
- Blocks are color-coded by functional network pair



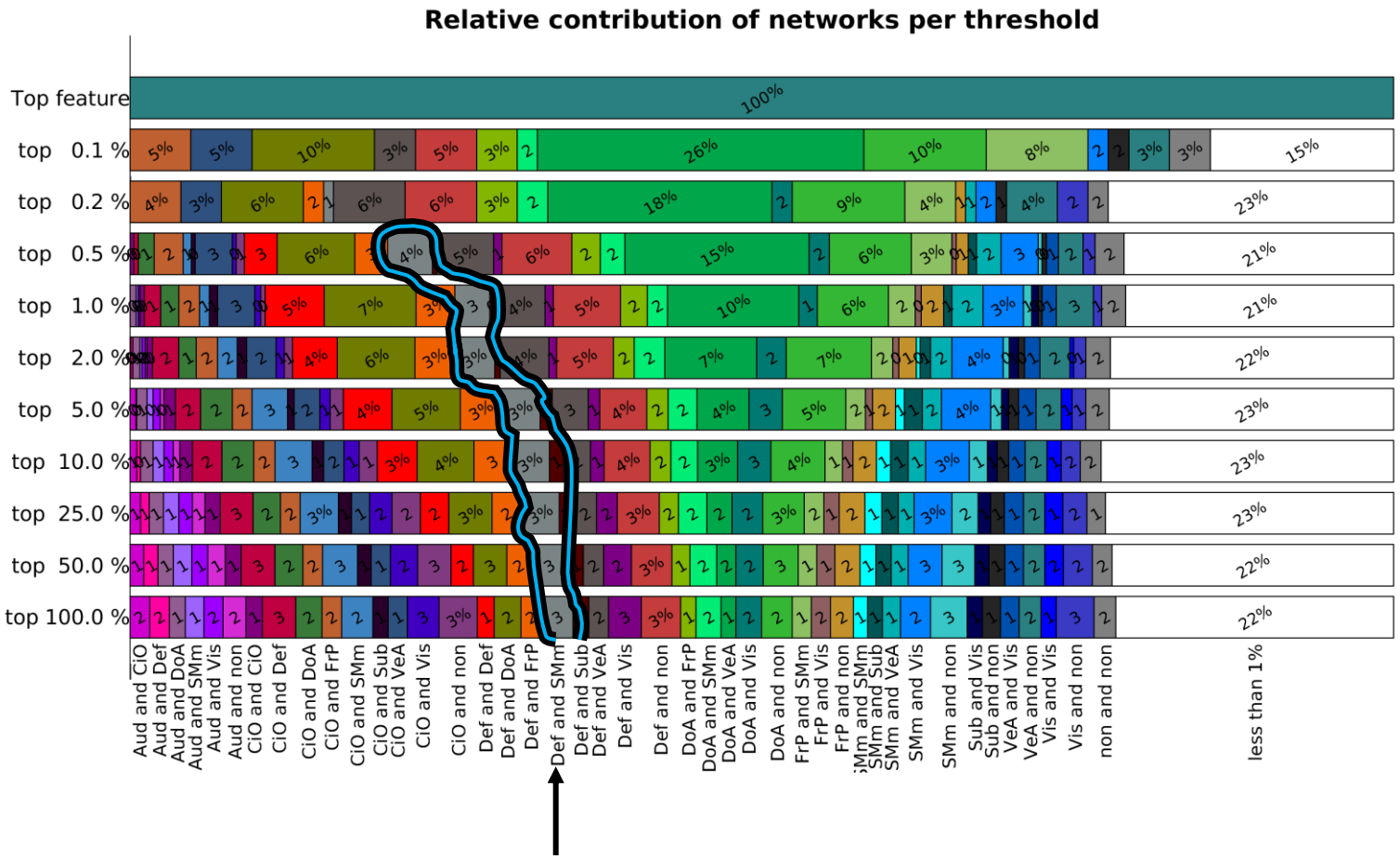


Relative Contribution Figures can help to visualize how stable the contribution of each network is across thresholds

DoA and Vea is diluted at higher thresholds

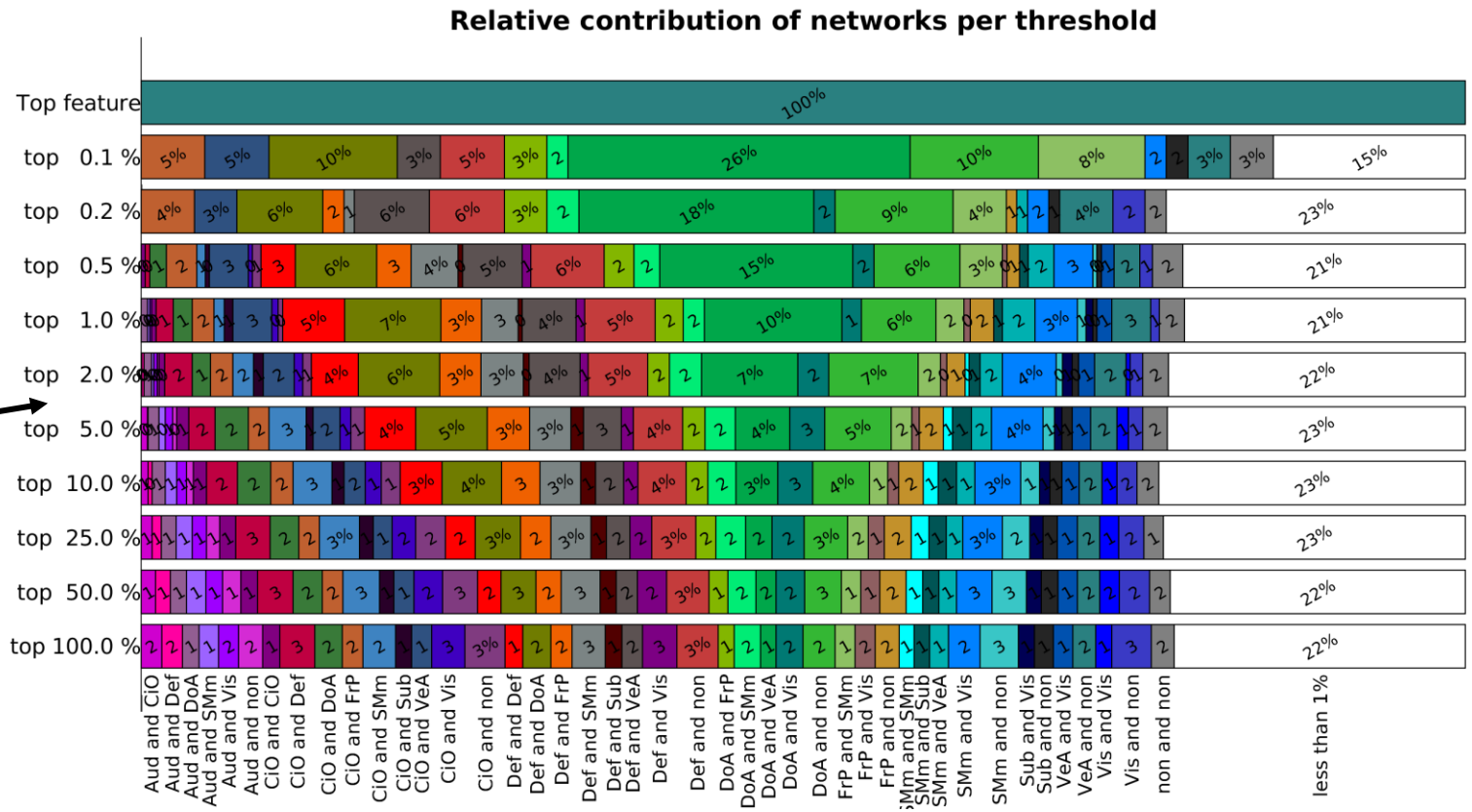


In contrast, the relative contribution of ***Def and SMm*** (among others) is stable across thresholds



We can also identify at which threshold all the networks are present

All the networks are present



Let’s apply this!

Motivation

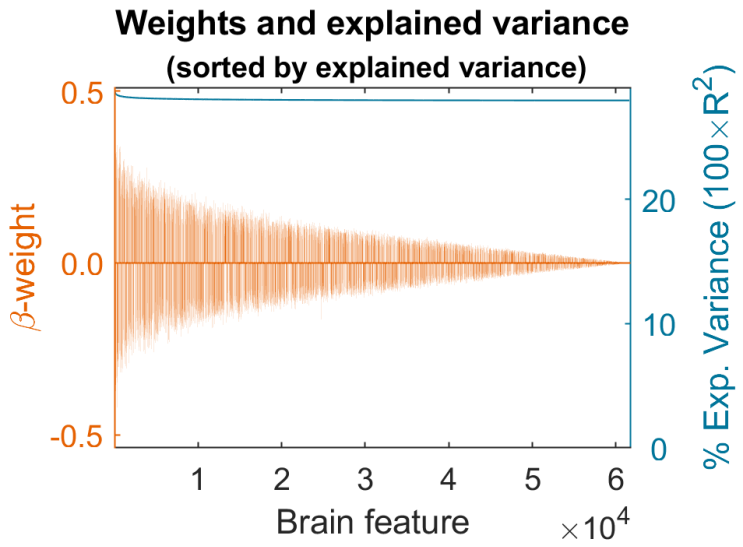
Description of the method

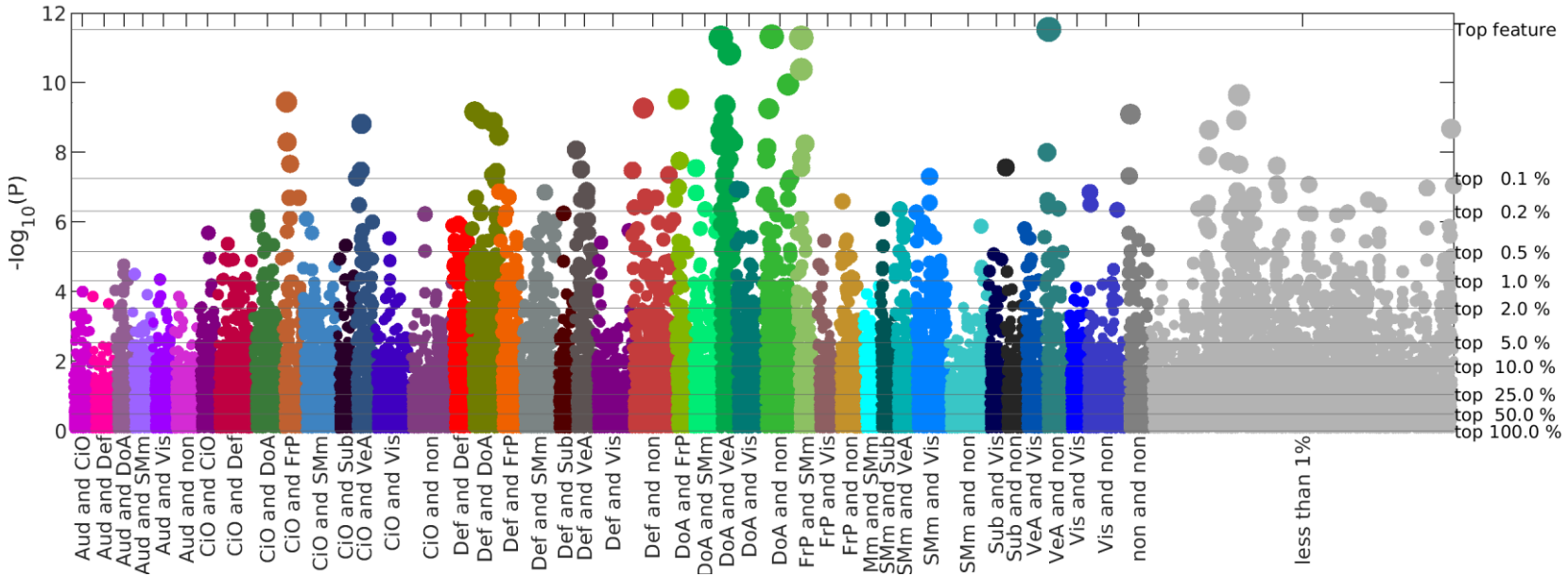
Using BWAS we can disambiguate between focal or globally distributed effects

Detailed description of the figures and tables we use to validate every step of the method

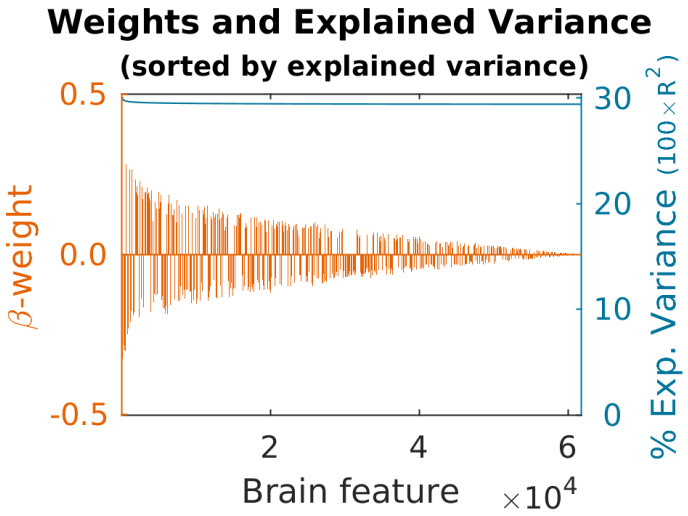
Other potential applications and future directions

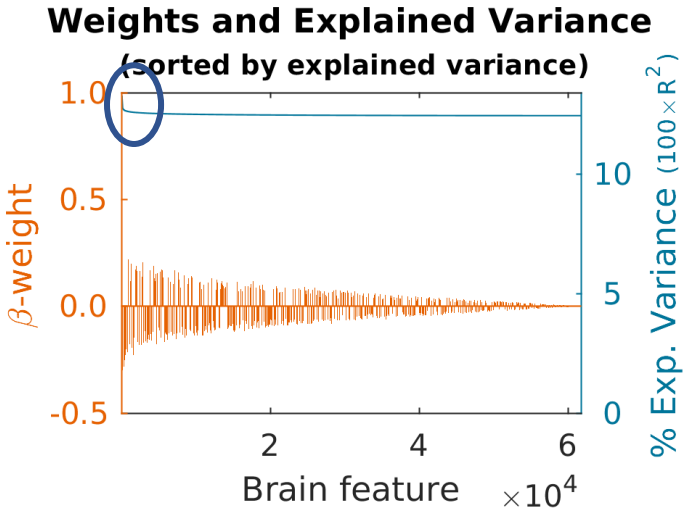
Manhattan-plots and Weights-and-Explained-variance figures can help with that (at least qualitatively)





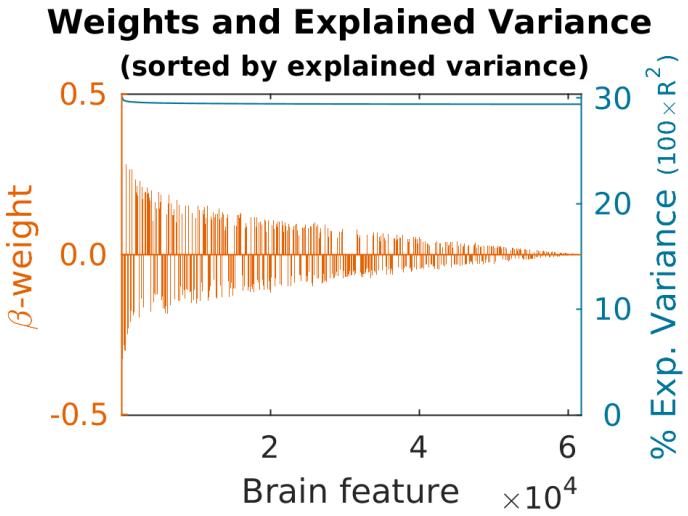
Compare the following results from two different BWAS

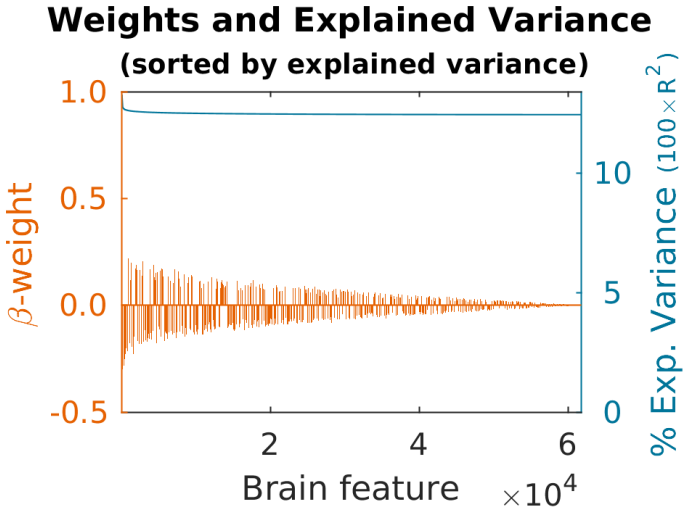


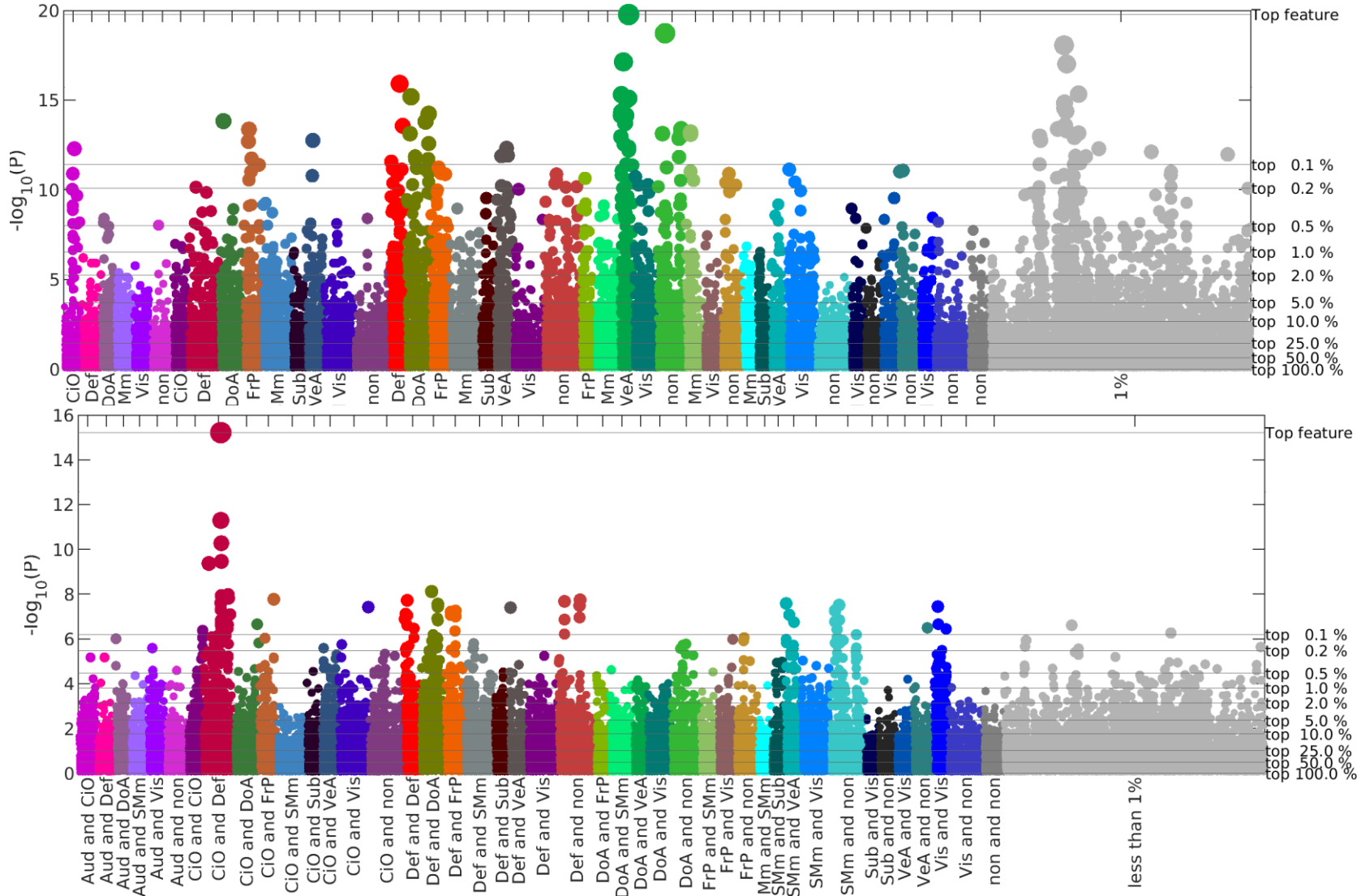


Bigger decay in predictive power, suggesting a more **focalized effect**

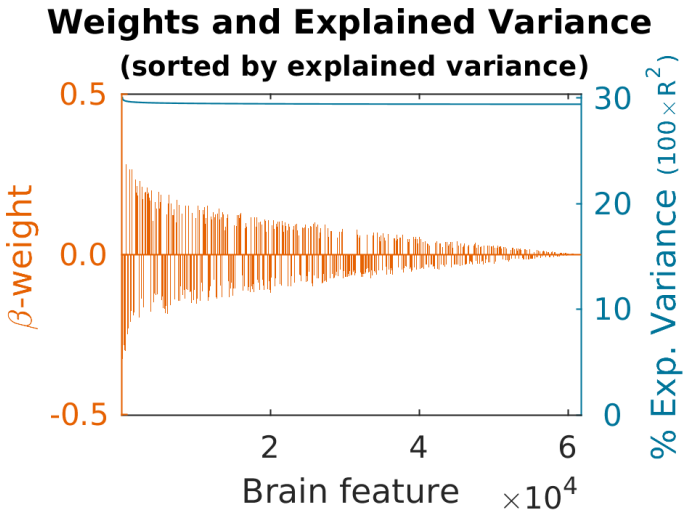
Height of the “towers” can also help to identify focalized versus global effects

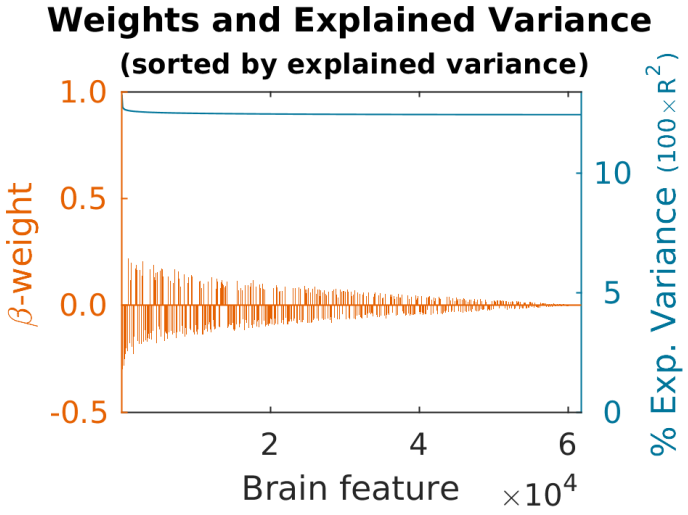


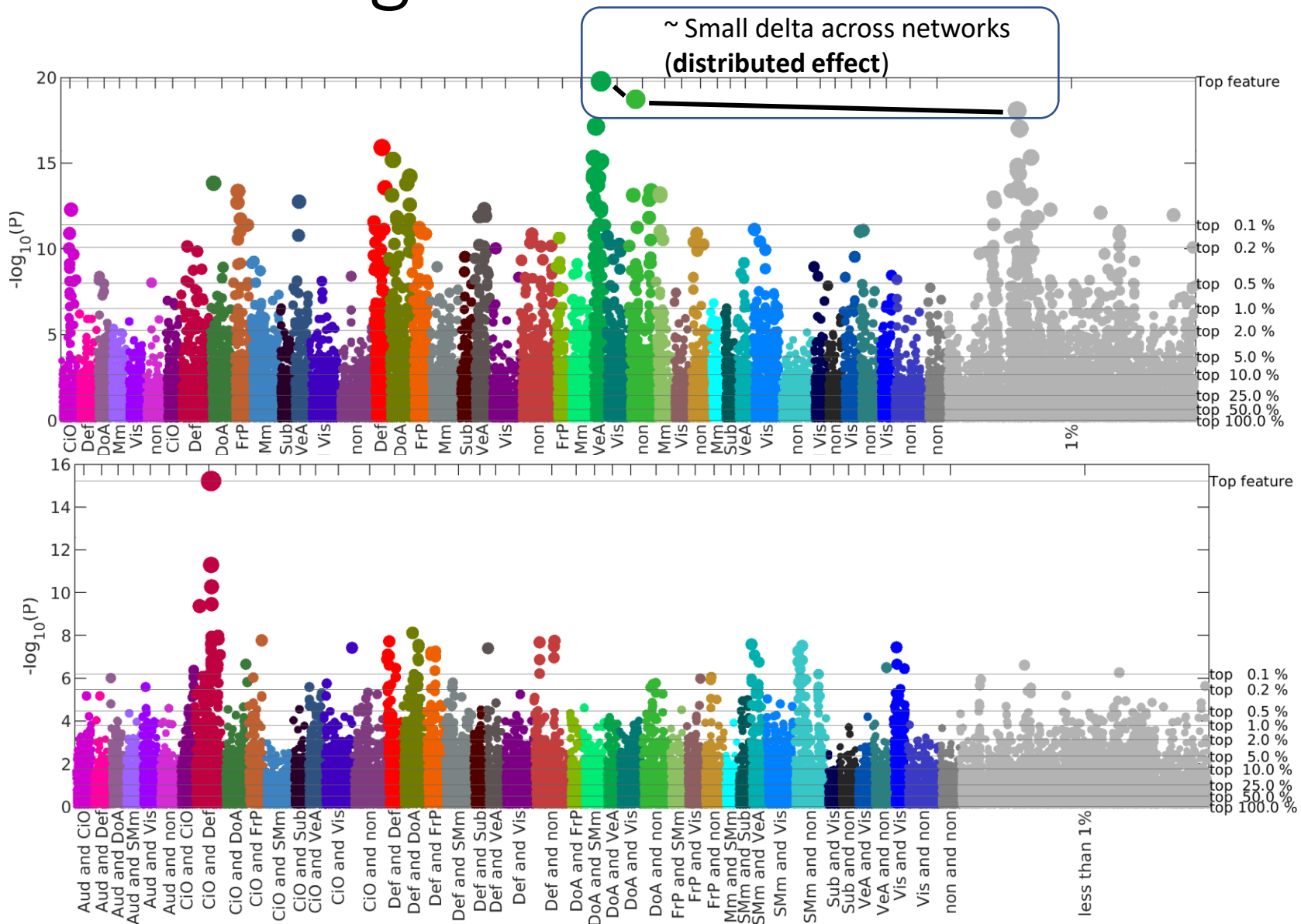




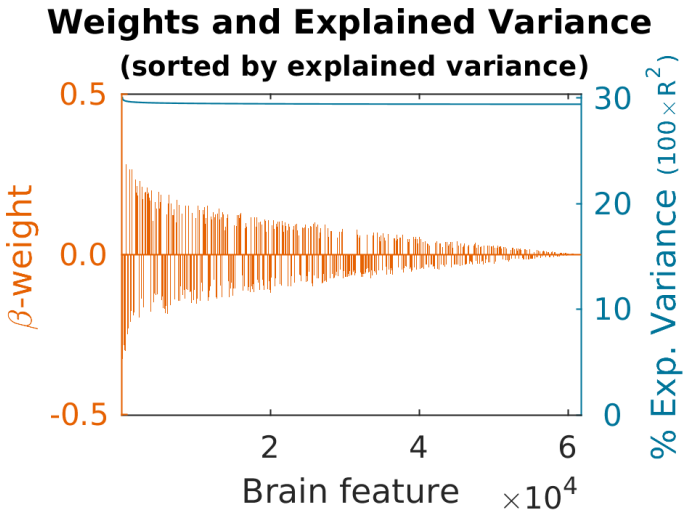
Height of the “towers” can also help to identify focalized versus global effects

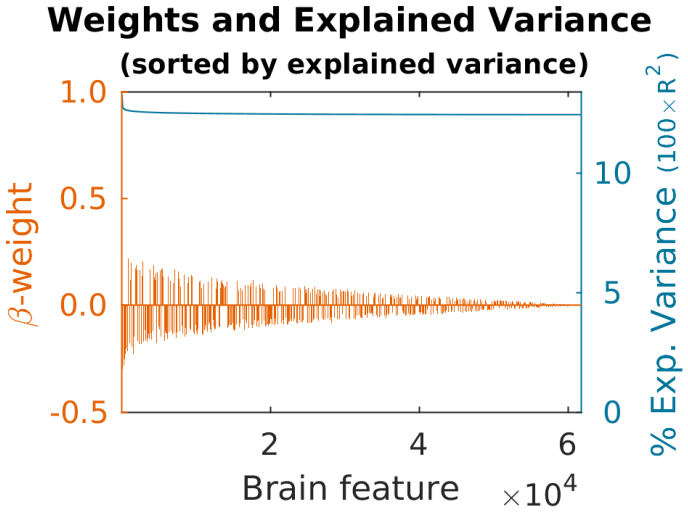


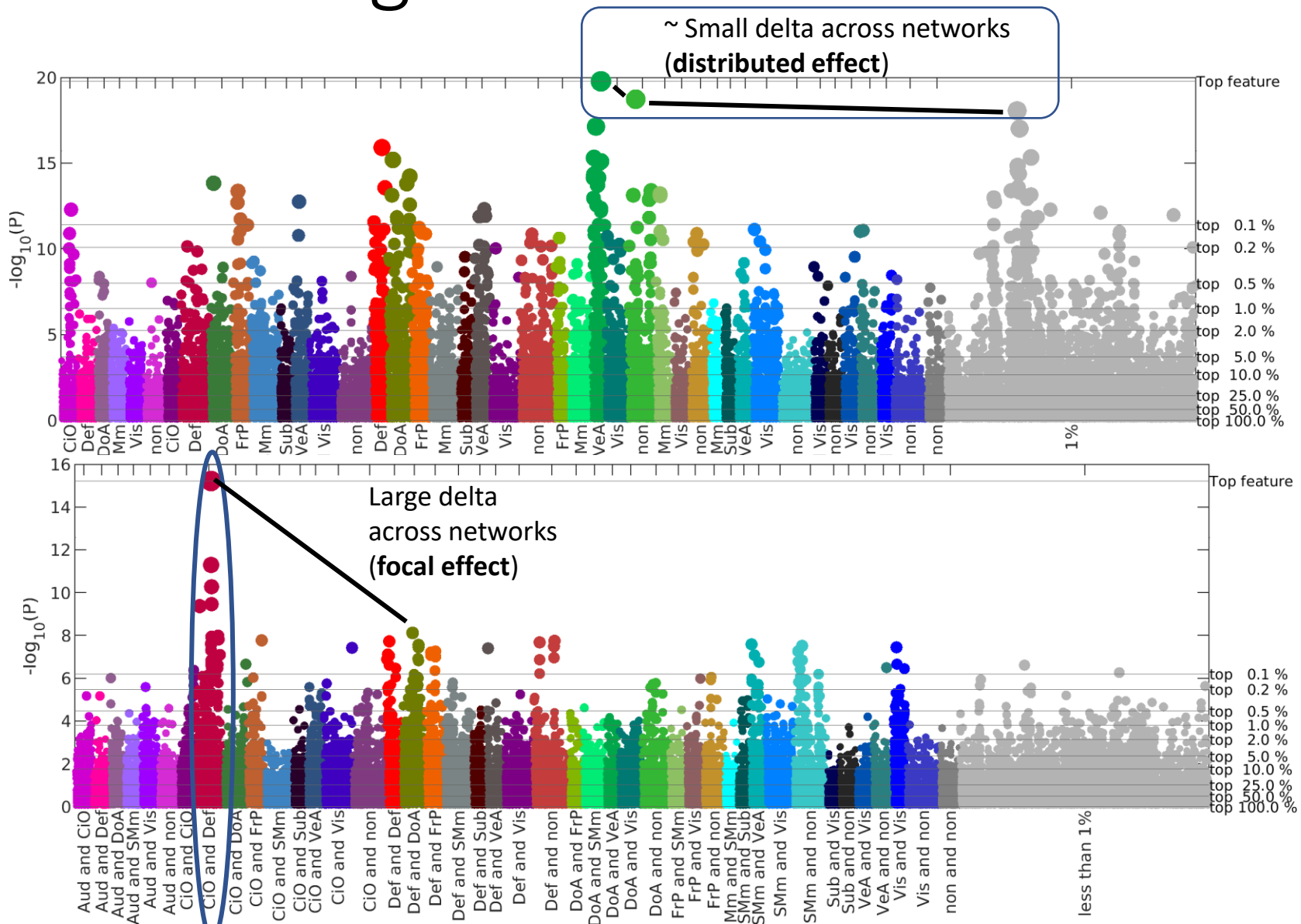




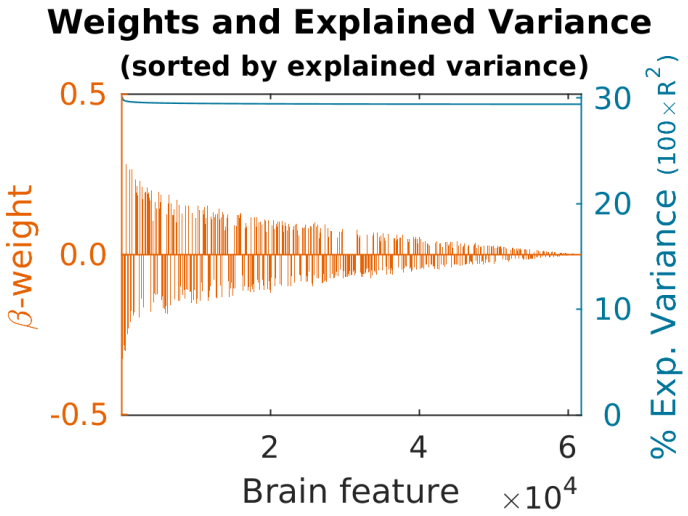
Height of the “towers” can also help to identify focalized versus global effects

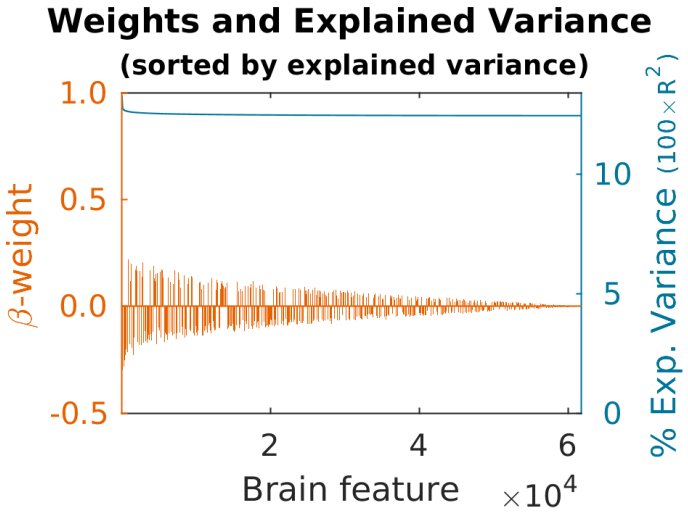


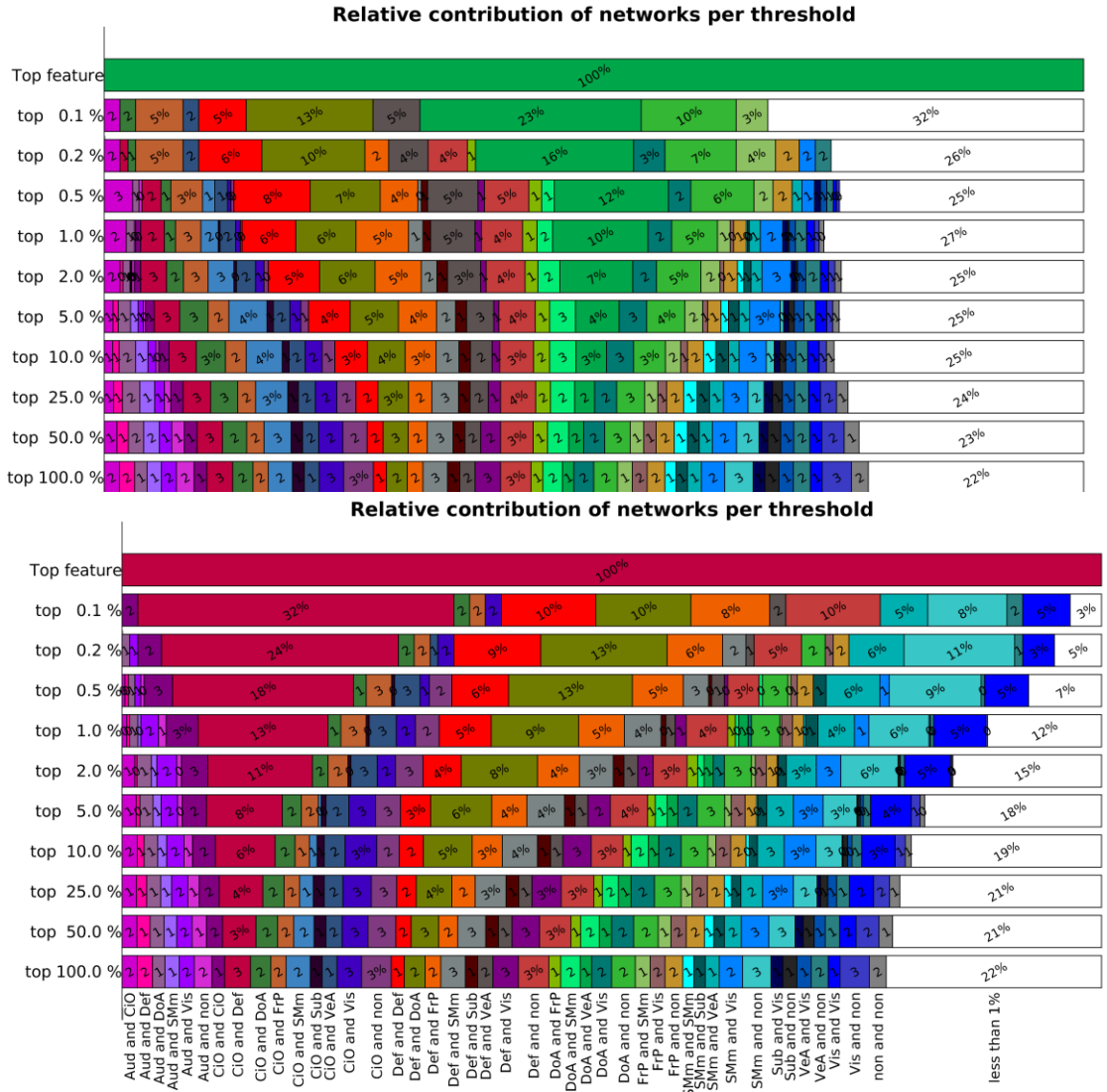




You can see how prominent each tower is at each threshold and how its relevance decays as more connections are added







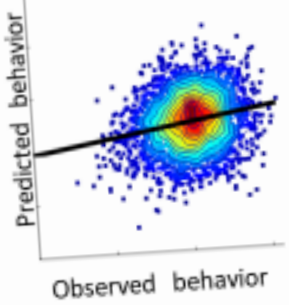
Now, let’s move to the second step of the approach, which is predicting scores in an independent sample

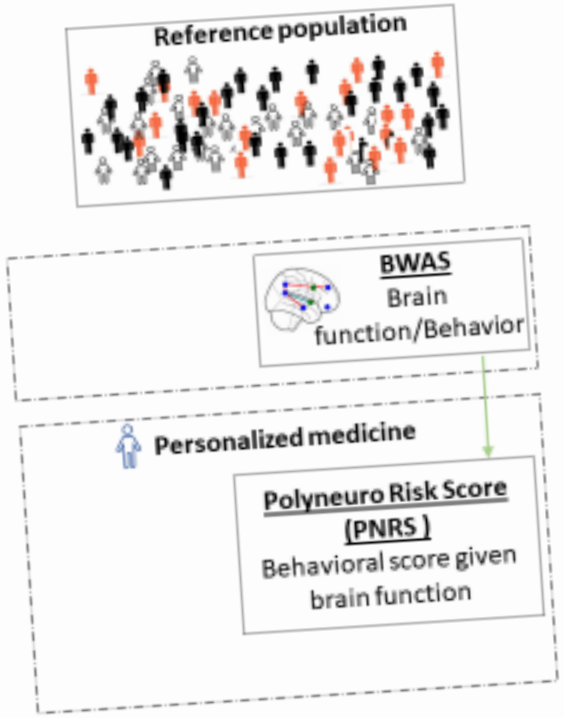
Do you remember this slide?

2nd, resulting models are selectively combined to estimate "behavioral" risk scores in individuals (PolyNeuro Risk Score, PNRS)

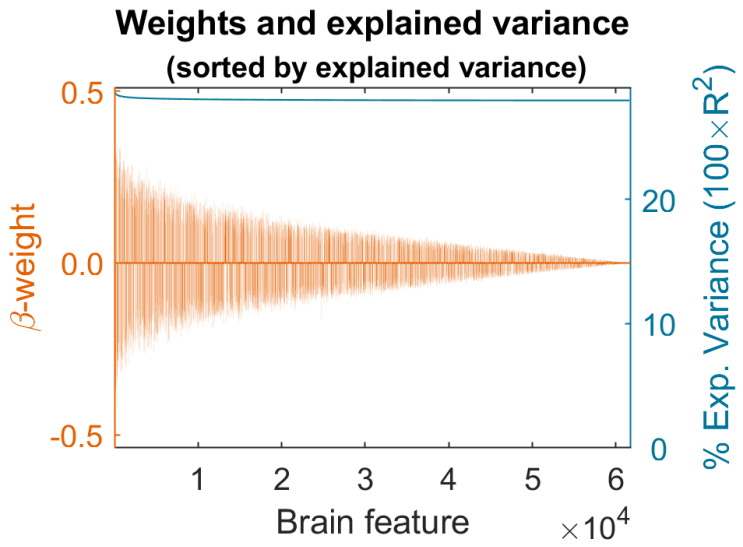
$$y_{score} = \sum_{v=1}^{N} x_v \beta_v$$

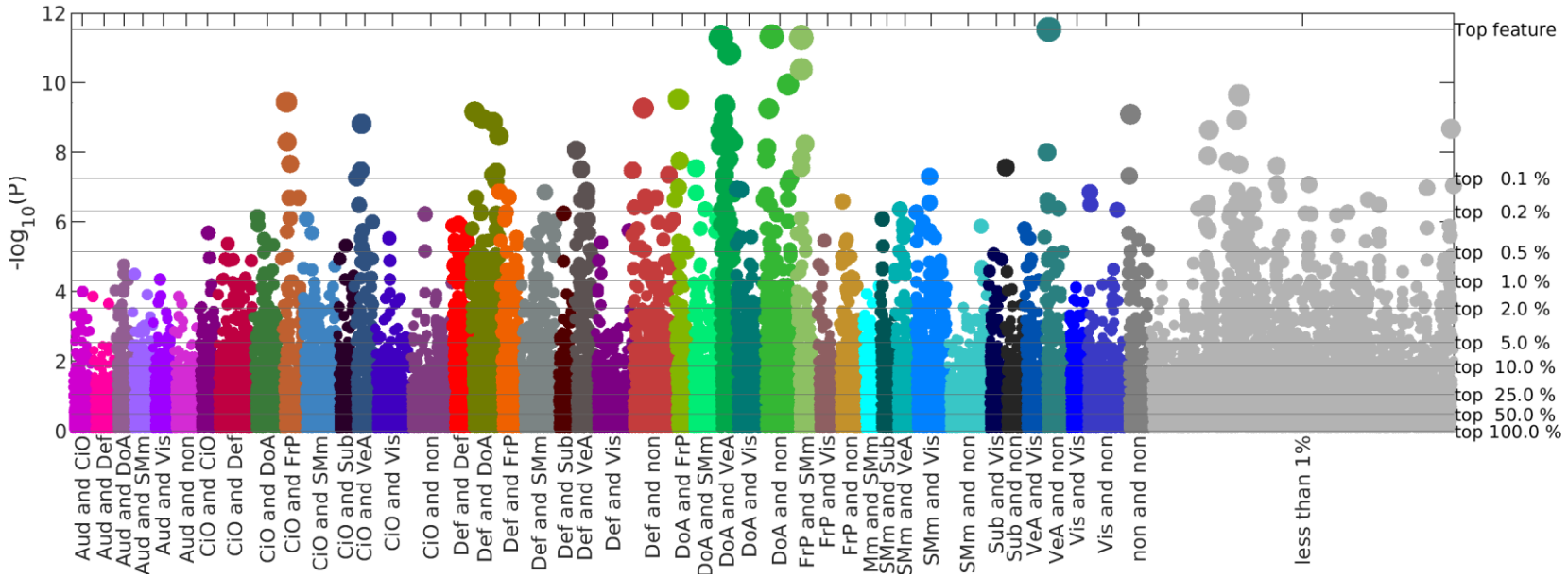
v are the preselected weights





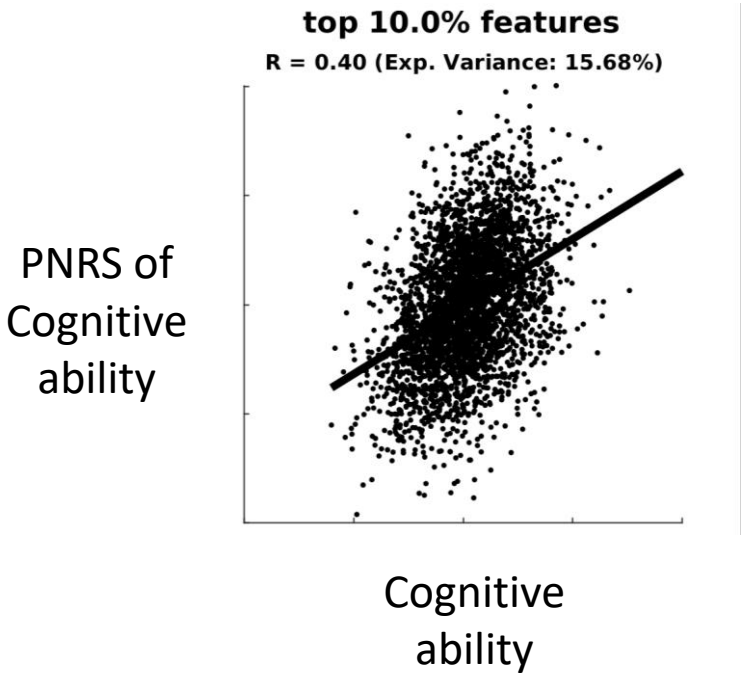
Beta-weights are selected by *top connections* (predictive power within the training sample) or *by networks*



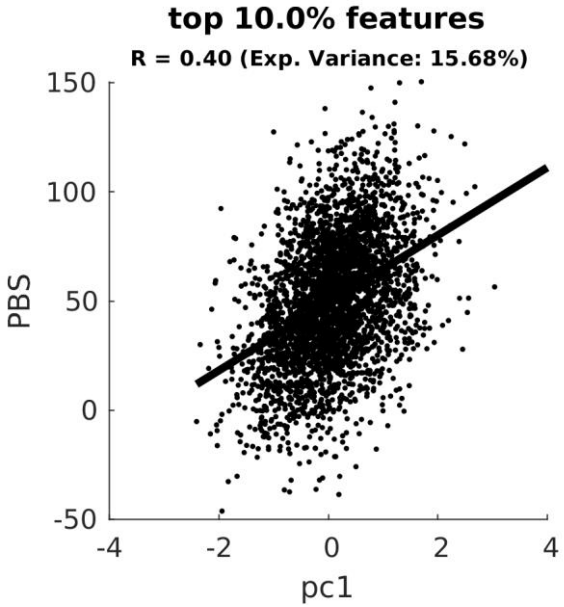


PNRS are calculated for each participant in the independent sample

- For each threshold (or network), a PNRS is calculated for each participant
- Pearson's correlation is calculated between the PNRS and the corresponding score of cognitive ability

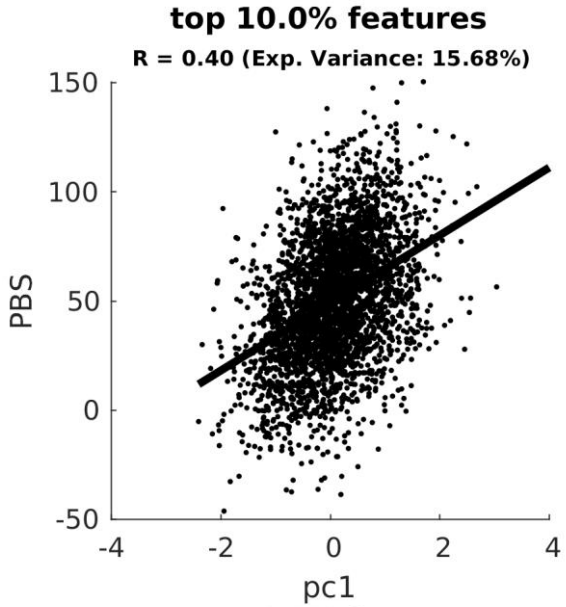


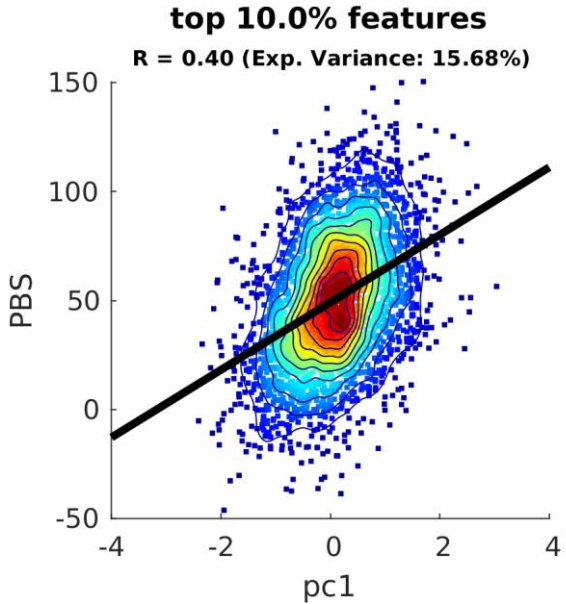
We generate scatter plots for each threshold



Just scores

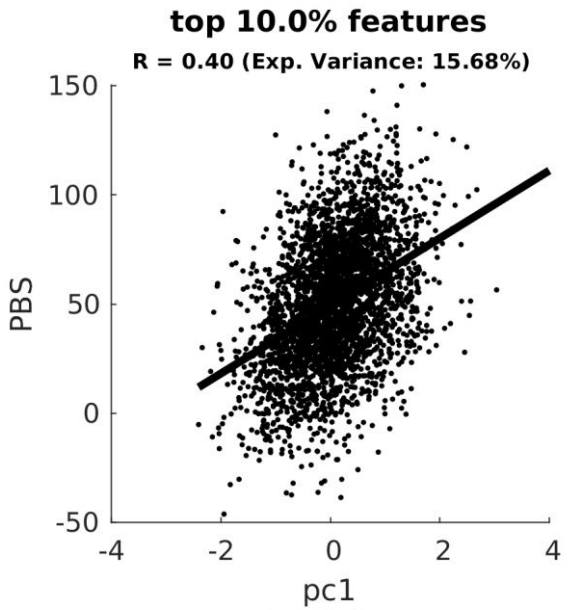
Showing a density map

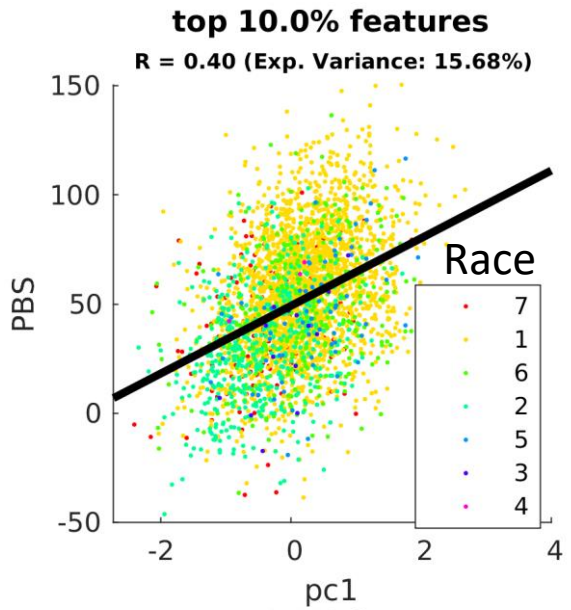


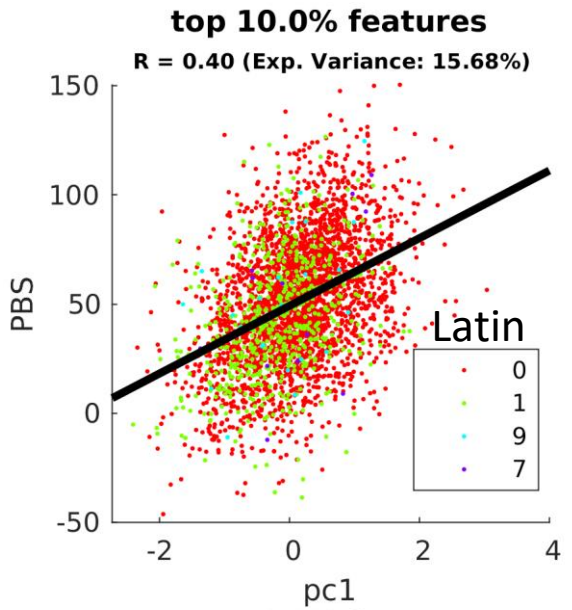


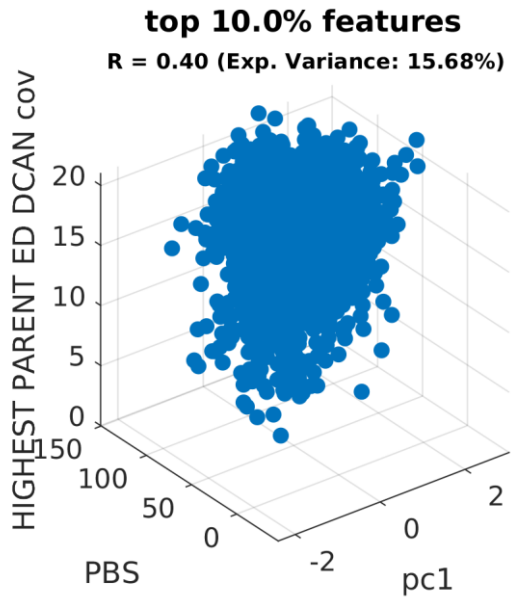
Showing a density map

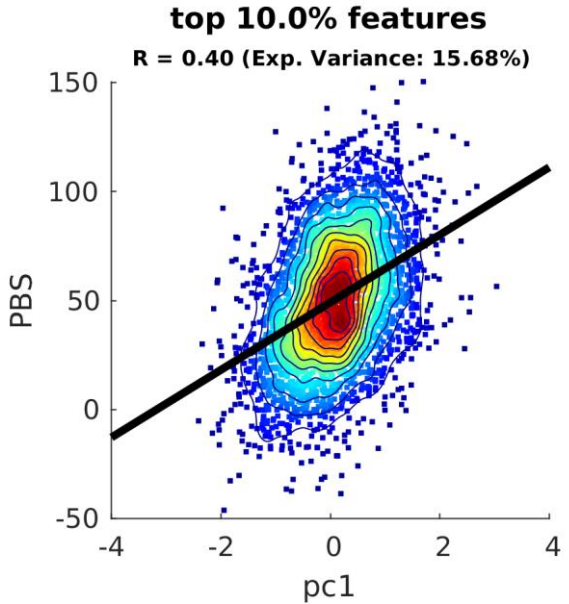
Showing covariates color-coding categorical values or using 3D scatter plots for continuous values

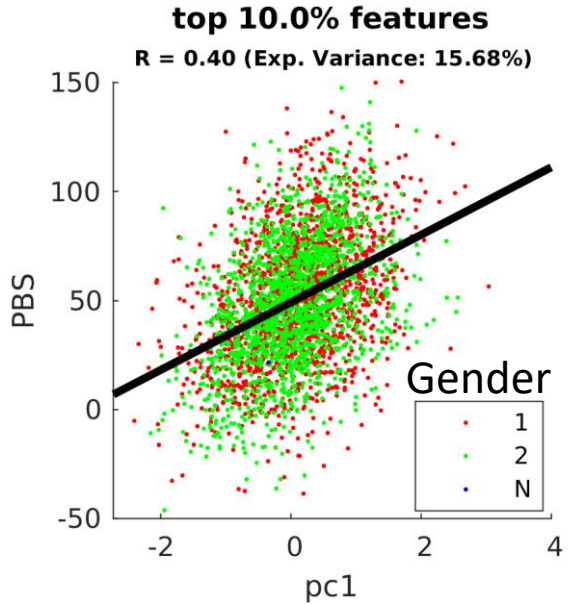


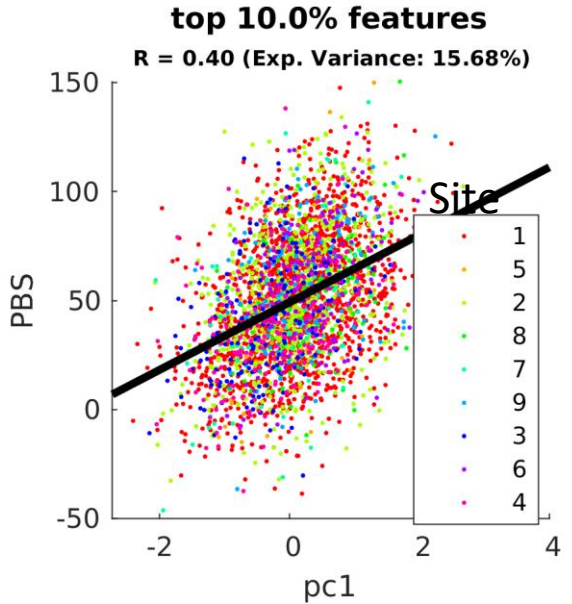


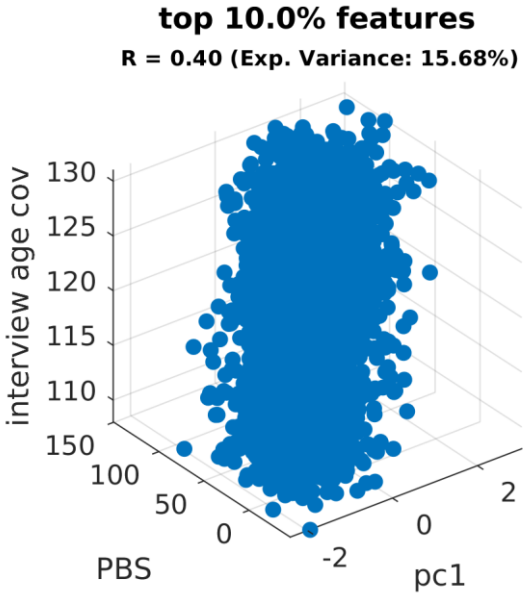




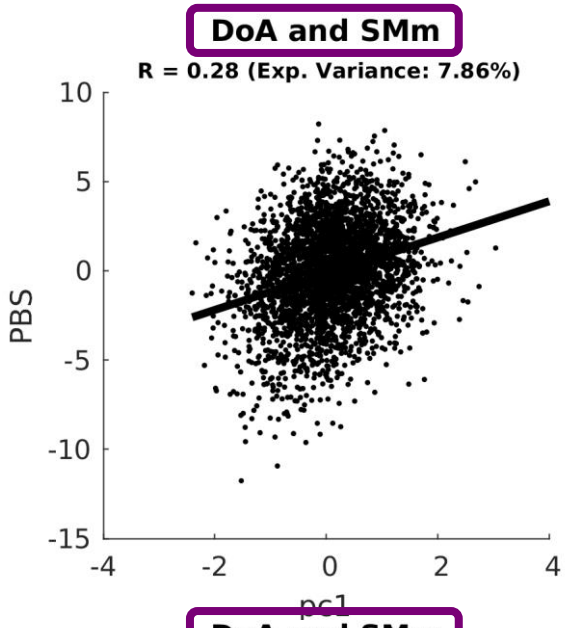


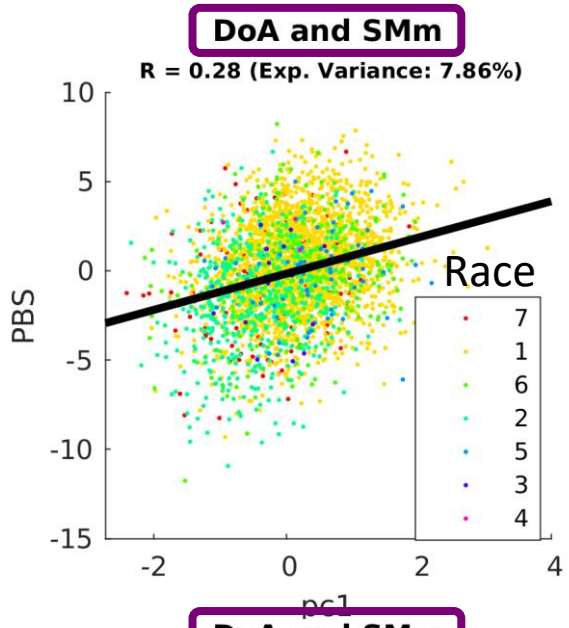


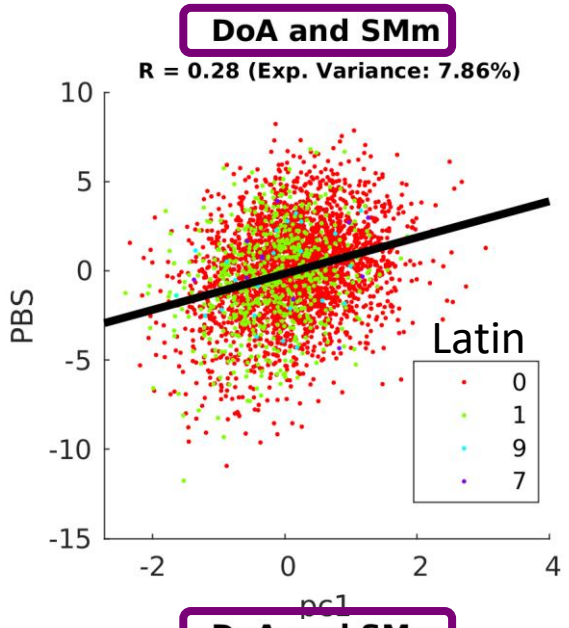


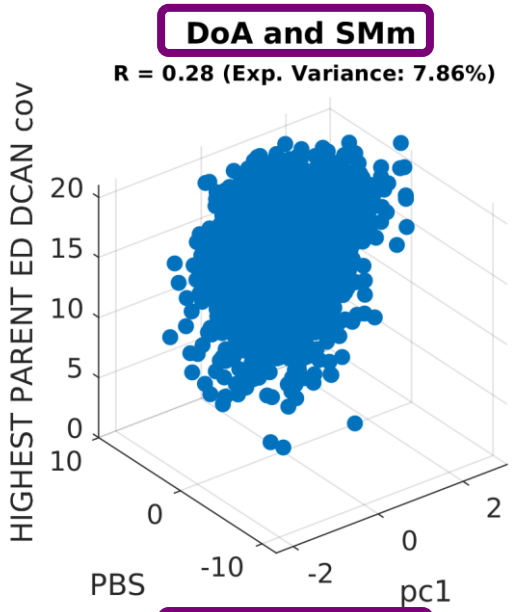


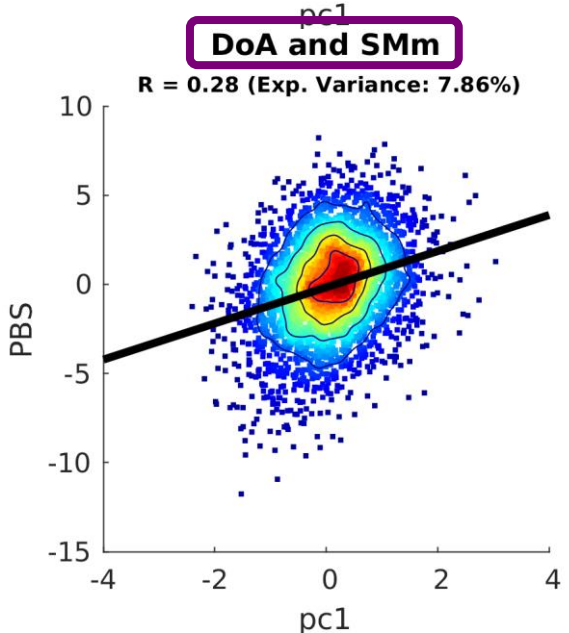
We also generate scatter plots for each functional network pair

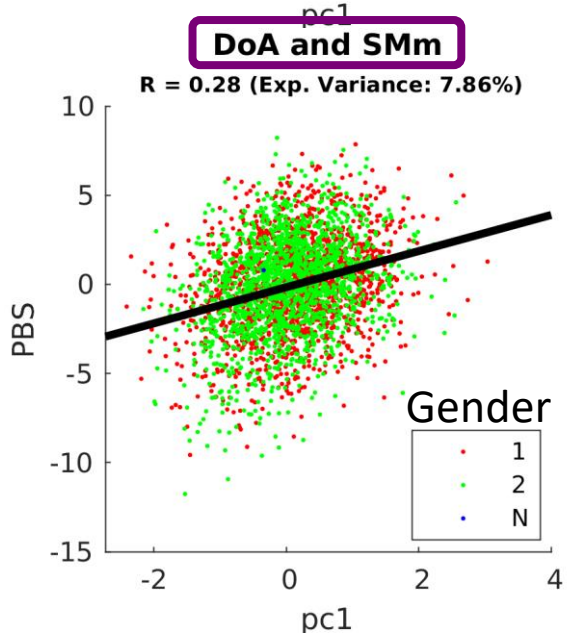


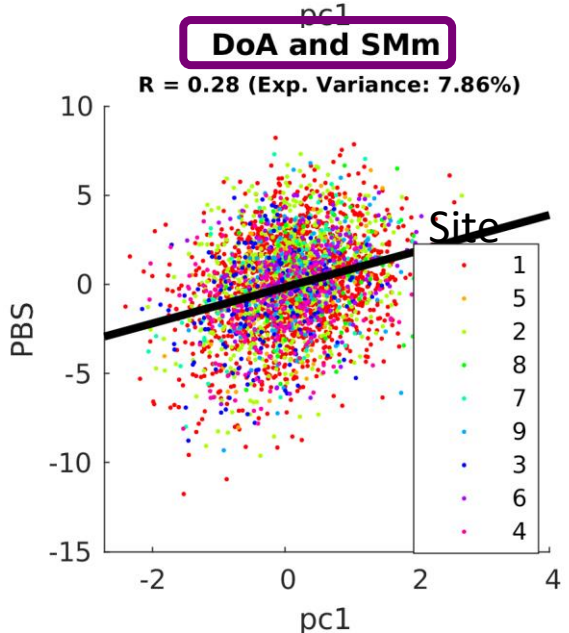


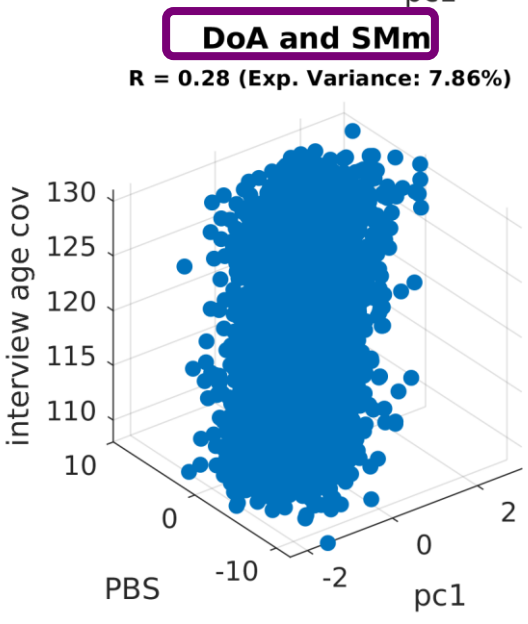






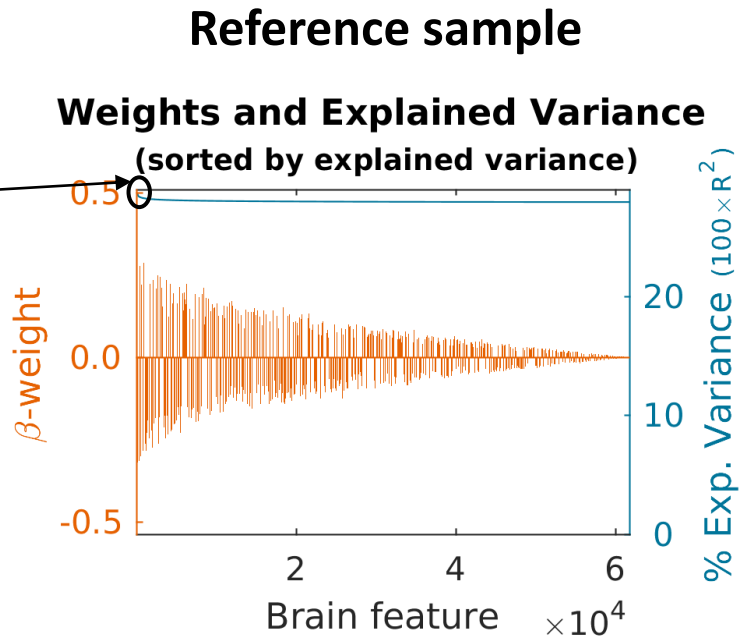






Results

First, we will use the best connection in the training sample to predict PNRS in the independent sample

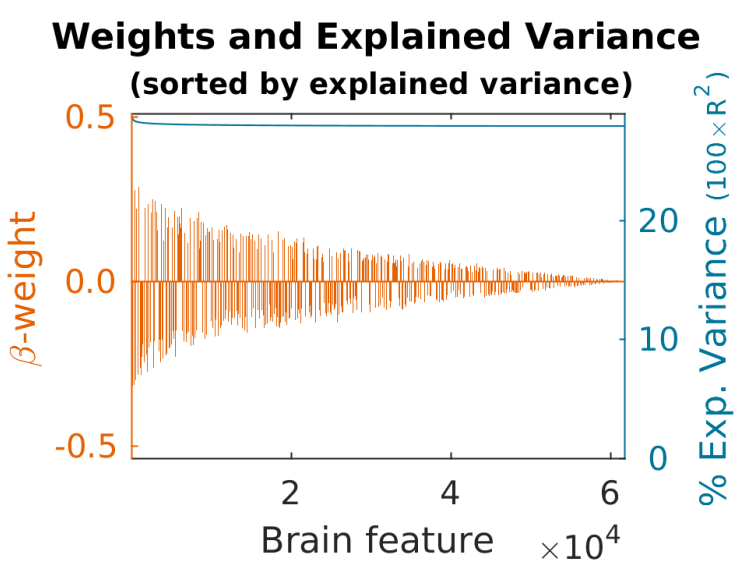


The best feature can only predict ~0.4% of the observed variance in the **independent sample**!

By top connections

Connections	Pearson's correlation	% Explained Variance (100 x $R^{2)}$
Top feature	0.062	**0.389**

Reference sample

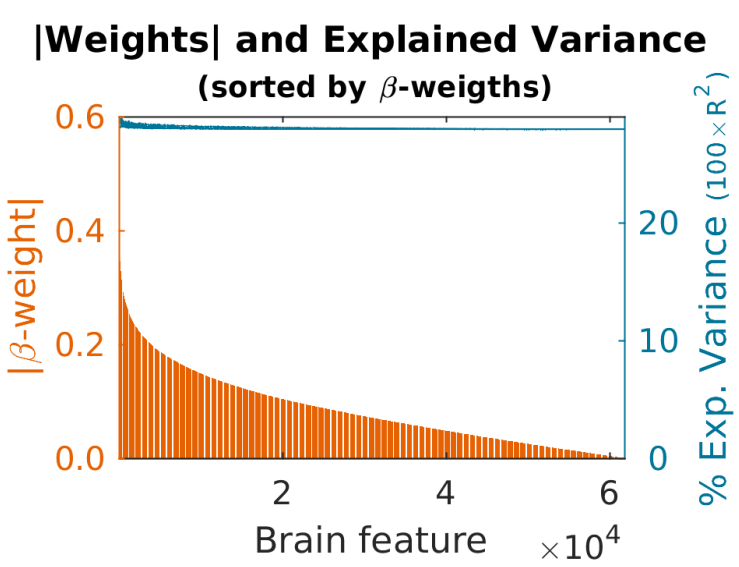


I will switch figures to facilitate the discussion

By top connections

Connections	Pearson's correlation	% Explained Variance (100 x $R^{2)}$
Top feature	0.062	**0.389**

Reference sample



Data sorted by the absolute value of the weights

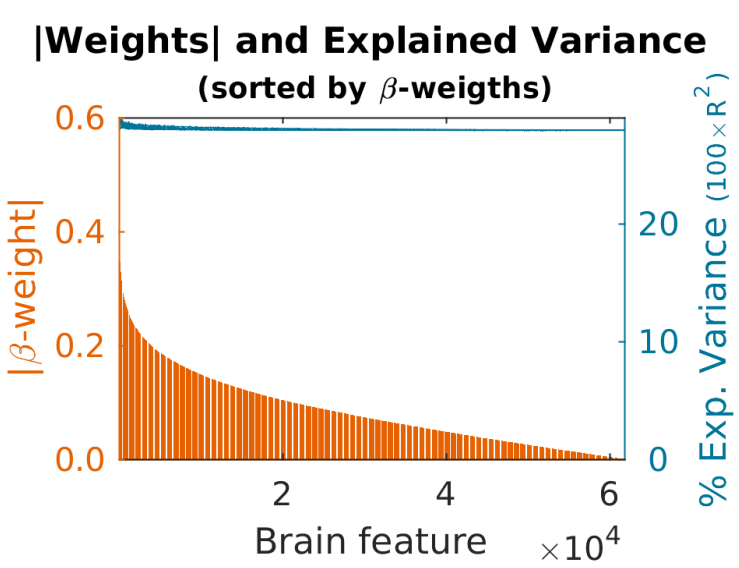
Now the question is, how much variance any other connection explains?

By top connections

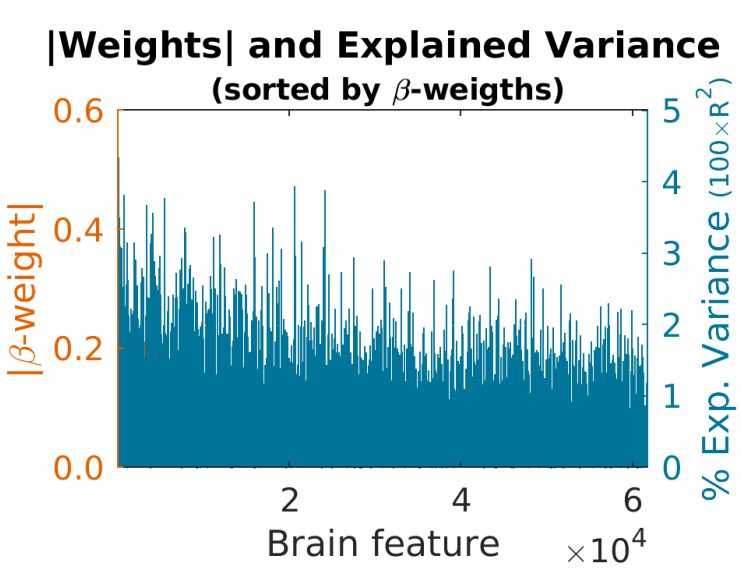
Connections	Pearson's correlation	% Explained Variance (100 x $R^{2)}$
Top feature	0.062	**0.389**

~0 to ~4%

Reference sample



Independent sample



What happen when we add more features?

By top connections

Connections	Pearson's correlation	% Explained Variance (100 x $R^{2)}$
Top feature	0.062	0.389
top 00.1% features	0.288	8.267
top 00.2% features	0.330	10.916
top 00.5% features	0.347	12.045
top 01.0% features	0.361	13.039
top 02.0% features	0.373	13.929
top 05.0% features	0.392	15.328
top 10.0% features	0.396	15.681
top 25.0% features	0.393	15.438
top 50.0% features	0.387	14.997
top 100.0% features	0.385	14.852

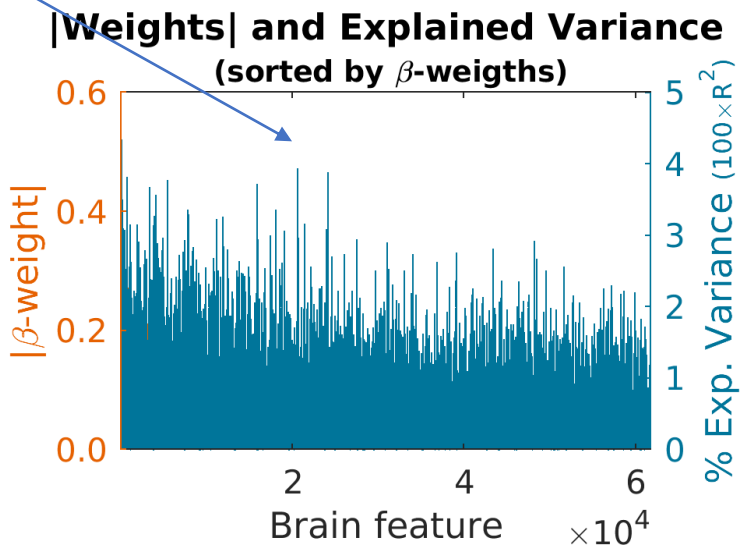
We can explain more variance!

Combined connections explain more variance than any individual connection

By top connections

Connections	Pearson's correlation	% Explained Variance (100 x $R^{2)}$
Top feature	0.062	0.389
top 00.1% features	0.288	8.267
top 00.2% features	0.330	10.916
top 00.5% features	0.347	12.045
top 01.0% features	0.361	13.039
top 02.0% features	0.373	13.929
top 05.0% features	0.392	15.328
top 10.0% features	0.396	15.681
top 25.0% features	0.393	15.438
top 50.0% features	0.387	14.997
top 100.0% features	0.385	14.852

(8.267 > 4%)



Peak at 10%

By top connections

Connections	Pearson's correlation	% Explained Variance (100 x $R^{2)}$
Top feature	0.062	0.389
top 00.1% features	0.288	8.267
top 00.2% features	0.330	10.916
top 00.5% features	0.347	12.045
top 01.0% features	0.361	13.039
top 02.0% features	0.373	13.929
top 05.0% features	0.392	15.328
top 10.0% features	**0.396**	**15.681**
top 25.0% features	0.393	15.438
top 50.0% features	0.387	14.997
top 100.0% features	0.385	14.852

Summary

- There is a decrease in predictive power across samples
- Adding connections improves the predictive power (~15.6% explained variance)

To test the specificity of the predictive power, we generate **null data** by selecting connections (or networks) randomly

By top connections

Connections	Pearson's correlation	% Explained Variance ($100 \times R^{2}$)
Top feature	0.062	0.389
top 00.1% features	0.288	8.267
top 00.2% features	0.330	10.916
top 00.5% features	0.347	12.045
top 01.0% features	0.361	13.039
top 02.0% features	0.373	13.929
top 05.0% features	0.392	15.328
top 10.0% features	**0.396**	**15.681**
top 25.0% features	0.393	15.438
top 50.0% features	0.387	14.997
top 100.0% features	0.385	14.852

Null data (N replicas)

for 1:N

Connection (or network) order is randomized

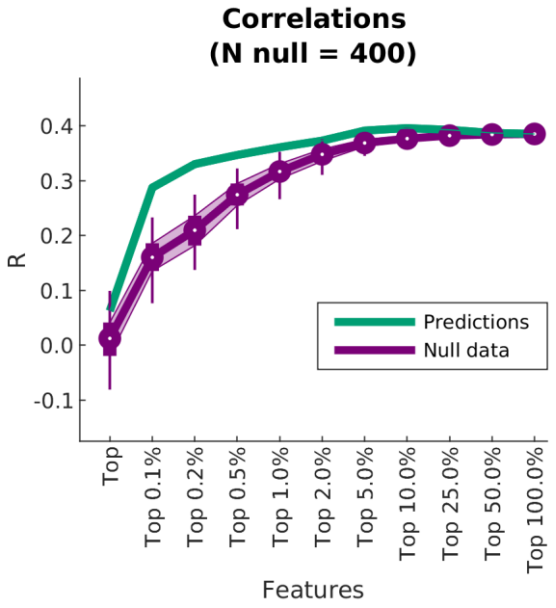
for 1: threshold (or network)

- *Predict scores*
- *Correlation between predicted and real scores*

Data visualization

By top connections

Connections	Pearson's correlation	% Explained Variance (100 x $R^{2)}$
Top feature	0.062	0.389
top 00.1% features	0.288	8.267
top 00.2% features	0.330	10.916
top 00.5% features	0.347	12.045
top 01.0% features	0.361	13.039
top 02.0% features	0.373	13.929
top 05.0% features	0.392	15.328
top 10.0% features	**0.396**	**15.681**
top 25.0% features	0.393	15.438
top 50.0% features	0.387	14.997
top 100.0% features	0.385	14.852



Predicted scores

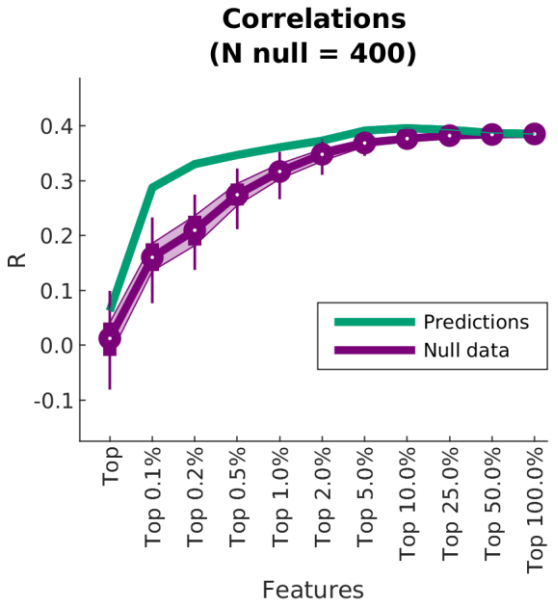
Boxplot showing distributions of null data

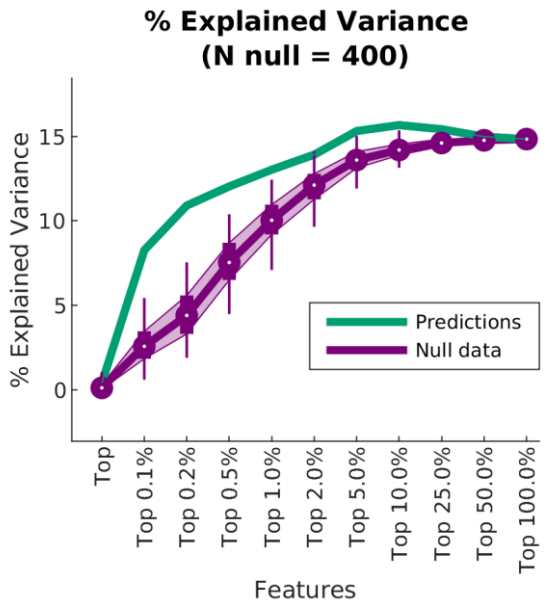
- *Circle: median*
- *Wide line, percentiles 25-75*
- *Thin line, percentiles 2.5-97.5*

We also generate a figure showing the % Explained Variance

By top connections

Connections	Pearson's correlation	% Explained Variance (100 x $R^{2)}$
Top feature	0.062	0.389
top 00.1% features	0.288	8.267
top 00.2% features	0.330	10.916
top 00.5% features	0.347	12.045
top 01.0% features	0.361	13.039
top 02.0% features	0.373	13.929
top 05.0% features	0.392	15.328
top 10.0% features	**0.396**	**15.681**
top 25.0% features	0.393	15.438
top 50.0% features	0.387	14.997
top 100.0% features	0.385	14.852



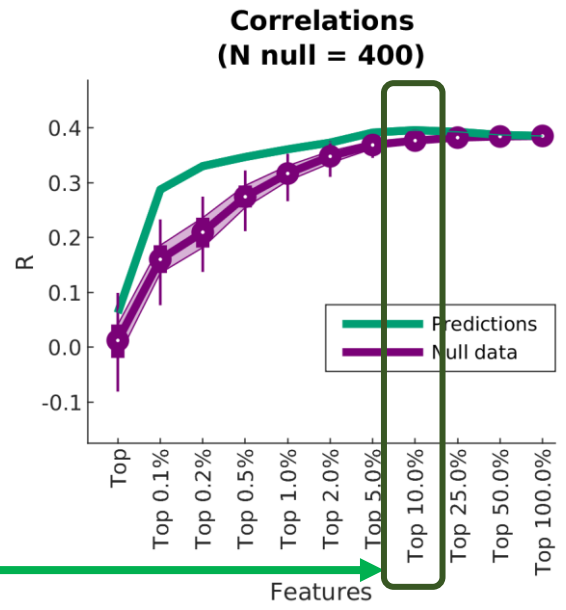


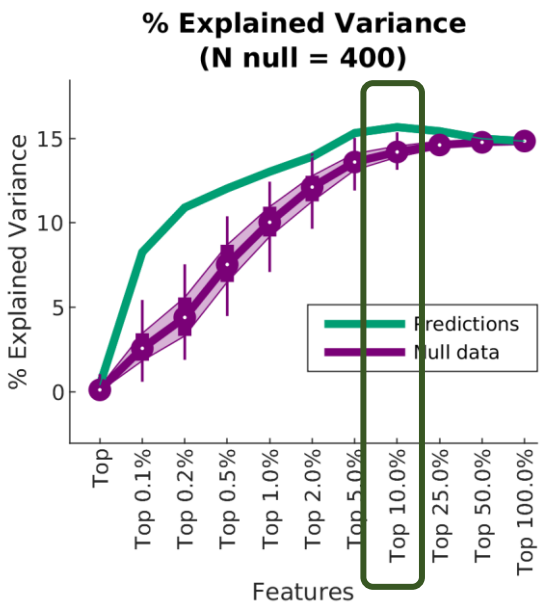
Using this figures, we can quickly visualize which threshold leads to the largest explained variance.
You can also see potential overlaps with any other set of random connections

Overlap might suggest global versus focal effect

By top connections

Connections	Pearson's correlation	% Explained Variance (100 x $R^{2)}$
Top feature	0.062	0.389
top 00.1% features	0.288	8.267
top 00.2% features	0.330	10.916
top 00.5% features	0.347	12.045
top 01.0% features	0.361	13.039
top 02.0% features	0.373	13.929
top 05.0% features	0.392	15.328
top 10.0% features	**0.396**	**15.681**
top 25.0% features	0.393	15.438
top 50.0% features	0.387	14.997
top 100.0% features	0.385	14.852

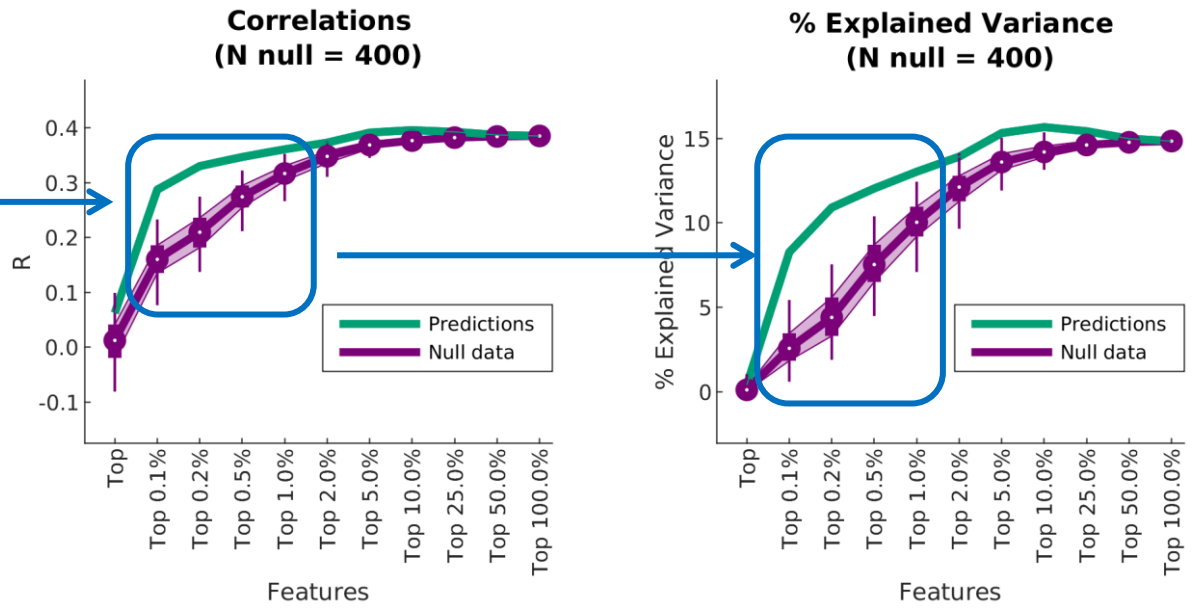




We can also identify the regions where no other set of connections lead to the same predictive power

By top connections

Connections	Pearson's correlation	% Explained Variance (100 x $R^{2)}$
Top feature	0.062	0.389
top 00.1% features	0.288	8.267
top 00.2% features	0.330	10.916
top 00.5% features	0.347	12.045
top 01.0% features	0.361	13.039
top 02.0% features	0.373	13.929
top 05.0% features	0.392	15.328
top 10.0% features	**0.396**	**15.681**
top 25.0% features	0.393	15.438
top 50.0% features	0.387	14.997
top 100.0% features	0.385	14.852

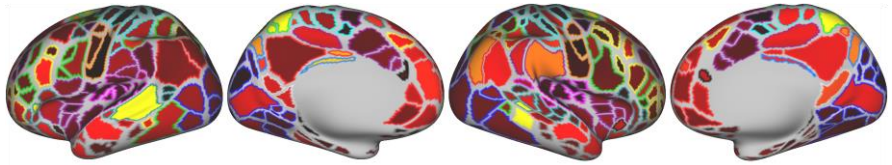


We also generate *cifti* files to make the corresponding brain figures

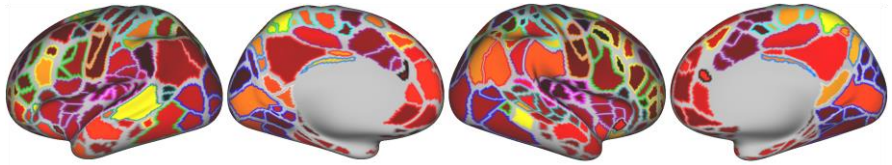
By top connections

Connections	Pearson's correlation	% Explained Variance (100 x $R^{2)}$
Top feature	0.062	0.389
top 00.1% features	0.288	8.267
top 00.2% features	0.330	10.916
top 00.5% features	0.347	12.045
top 01.0% features	0.361	13.039
top 02.0% features	0.373	13.929
top 05.0% features	0.392	15.328
top 10.0% features	**0.396**	**15.681**
top 25.0% features	0.393	15.438
top 50.0% features	0.387	14.997
top 100.0% features	0.385	14.852

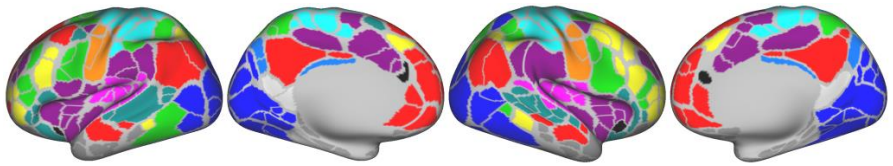
Top 10% brain features



All the brain features



Gordon's ROI set



- Auditory (Aud, n=24)
- Cingulo-Opercular (CiO, n=40)
- Cingulo-Parietal (CiP, n=5)
- Default (Def, n=41)
- Dorsal Attention (DoA, n=32)
- Fronto-Parietal (FrP, n=24)
- Retrosplenial-Temporal (ReT, n=8)
- Somato-sensory lateral (SmI, n=38)
- Somato-sensory medial (SMm, n=8)
- Salience (Sal, n=4)
- Ventral-Attention (VeA, n=23)
- Visual (Vis, n=39)

In addition, we also calculate the predictive power of each functional network pair

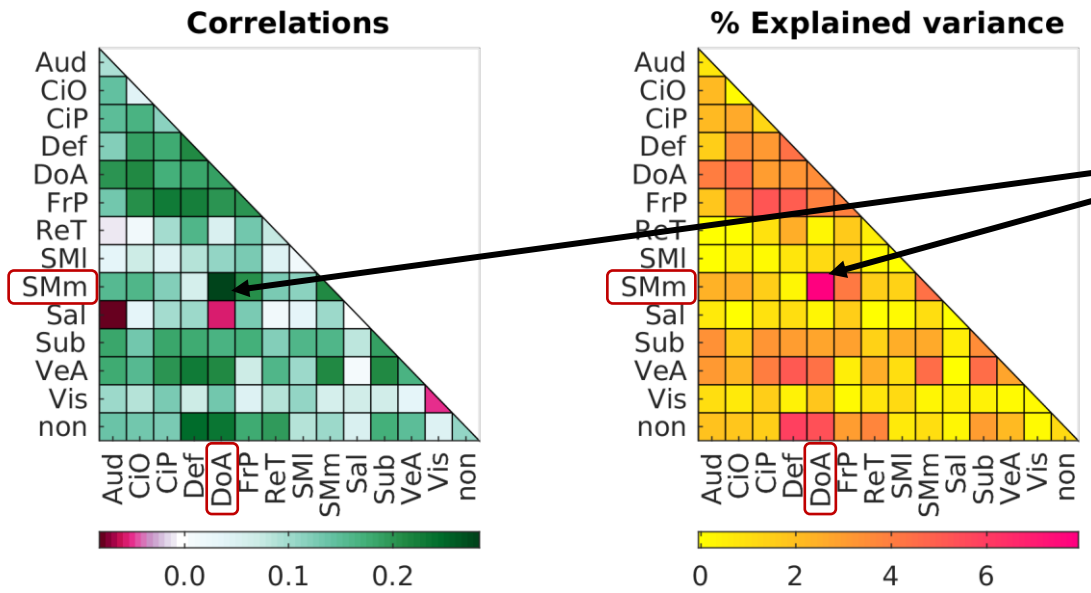
By top connections

Connections	Pearson's correlation	% Explained Variance (100 x $R^{2)}$
Top feature	0.062	0.389
top 00.1% features	0.288	8.267
top 00.2% features	0.330	10.916
top 00.5% features	0.347	12.045
top 01.0% features	0.361	13.039
top 02.0% features	0.373	13.929
top 05.0% features	0.392	15.328
top 10.0% features	**0.396**	**15.681**
top 25.0% features	0.393	15.438
top 50.0% features	0.387	14.997
top 100.0% features	0.385	14.852

By networks

Connections	Pearson's correlation	% Explained Variance (100 x $R^{2)}$
1) DoA and SMm	0.280	7.863
2) Def and non	0.242	5.837
3) DoA and non	0.232	5.404
4) CiP and FrP	0.228	5.217
5) Def and VeA	0.227	5.148
6) Def and FrP	0.223	4.990
7) CiO and DoA	0.213	4.517
8) Sub and VeA	0.212	4.475
9) SMm and VeA	0.211	4.469
10) SMm and SMm	0.210	4.425
11) DoA and VeA	0.209	4.369
…		
105) Aud and Sal	-0.079	0.624

And the corresponding visuals

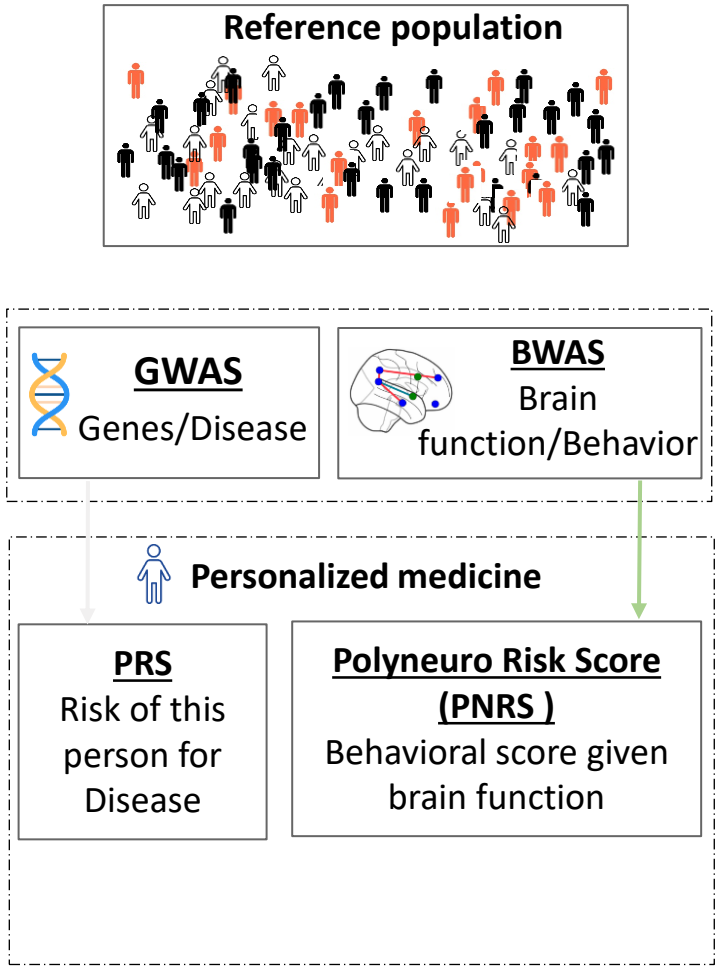


By networks

Connections	Pearson's correlation	% Explained Variance (100 x $R^{2)}$
1) DoA and SMm	0.280	7.863
2) Def and non	0.242	5.837
3) DoA and non	0.232	5.404
4) CiP and FrP	0.228	5.217
5) Def and VeA	0.227	5.148
6) Def and FrP	0.223	4.990
7) CiO and DoA	0.213	4.517
8) Sub and VeA	0.212	4.475
9) SMm and VeA	0.211	4.469
10) SMm and SMm	0.210	4.425
11) DoA and VeA	0.209	4.369
…		
105) Aud and Sal	-0.079	0.624

Summary

- BWAS/PNRS can leverage small effects across the brain
- Using this approach, we can identify if associations between brain function and behavior are focal or globally distributed



Other uses of BWAS

01

Study associations between PNRS of behavior and *disease severity*

02

Combine several PNRS to predict scores of *disease severity*

03

Combine several PNRS to study heterogeneity

Future directions

- Brain features here were connectivity matrices calculated via Pearson's correlations. We can use instead connectotyping [1,2])
- Add regularization to the estimation of the beta-weights, such as partial least squares regression [3,4]

[1] Miranda-Dominguez, O., Mills, B. D., Carpenter, S. D., Grant, K. A., Kroenke, C. D., Nigg, J. T., & Fair, D. A. (2014). Connectotyping: model based fingerprinting of the functional connectome. *PloS One*, *9*(11), e111048. https://doi.org/10.1371/journal.pone.0111048

[2] Miranda-Dominguez, O., Feczko, E., Grayson, D. S., Walum, H., Nigg, J. T., & Fair, D. A. (2018). Heritability of the human connectome: A connectotyping study. *Network Neuroscience (Cambridge, Mass.)*, *2*(2), 175–199. https://doi.org/10.1162/netn_a_00029

[3] Miranda-Domínguez, Ó., Ragothaman, A., Hermosillo, R., Feczko, E., Morris, R., Carlson-Kuhta, P., Nutt, J. G., Mancini, M., Fair, D., & Horak, F. B. (2020). Lateralized Connectivity between Globus Pallidus and Motor Cortex is Associated with Freezing of Gait in Parkinson's Disease. *Neuroscience*, *443*, 44–58. https://doi.org/10.1016/j.neuroscience.2020.06.036

[4] Silva-Batista, C., Ragothaman, A., Mancini, M., Carlson-Kuhta, P., Harker, G., Jung, S. H., Nutt, J. G., Fair, D. A., Horak, F. B., & Miranda-Domínguez, O. (2021). Cortical thickness as predictor of response to exercise in people with Parkinson's disease. *Human Brain Mapping*, *42*(1), 139–153. https://doi.org/10.1002/hbm.25211

Thank you!

Art at the MIDB

Brain Wide Associations (BWAS) to model the link between brain features and behavior.

Oscar Miranda Domínguez, PhD, MSc
Assistant Professor | Department of Pediatrics | Medical School | University of Minnesota
Masonic Institute for the Developing Brain
Minnesota Supercomputer Institute

MASONIC INSTITUTE FOR THE DEVELOPING BRAIN
UNIVERSITY OF MINNESOTA
Driven to Discover®