

Research Meeting

UK Biobank, ImageNomer, AngleBasis, and fMRI Identifiability Results
in Memorization

Anton Orlichenko

July 2023



Outline

- 1 UK Biobank
 - Summary
 - Prediction and Feature Identification
- 2 ImageNomer: FC and Omics Visualization Software Detects Race Confound
- 3 AngleBasis: A Generative Model and Decomposition for Functional Connectivity
- 4 fMRI Identifiability Results in Memorization
 - Longitudinal or Contemporaneous Repeat Scans
 - Dynamic Connectivity



UK Biobank



Data Inventory Screenshot

- Dr. Ray Su helps acquire and access the data for Dr. Deng's group

Date Extracted: 2021-03-11T20:29:27
Data columns: 10963

Column	UDI	Count	Type	Description
0	eid	502486	Sequence	Encoded anonymised participant ID
1	19-0.0	324763	Categorical (single)	Heel ultrasound method Uses data-coding 100260
2	31-0.0	502485	Categorical (single)	Sex Uses data-coding 9
3	34-0.0	502485	Integer	Year of birth
4	36-0.0	497854	Text	Blood pressure device ID
5	36-1.0	20334		
6	36-2.0	43259		
7	36-3.0	2957		
8	37-0.0	497854	Text	Blood pressure manual sphygmomanometer device ID
9	37-1.0	20334		
10	37-2.0	43259		
11	37-3.0	2957		
12	38-0.0	497024	Text	Hand grip dynamometer device ID
13	38-1.0	20333		
14	38-2.0	47504		
15	38-3.0	3201		
16	39-0.0	497024	Text	Height measure device ID
17	39-1.0	20333		
18	39-2.0	47504		
19	39-3.0	3201		
20	40-0.0	497024	Text	Manual scales device ID
21	40-1.0	20333		
22	40-2.0	47504		
23	40-3.0	3201		
24	41-0.0	497024	Text	Seating box device ID
25	41-1.0	20333		
26	41-2.0	47504		
27	41-3.0	3201		
28	43-0.0	497024	Text	Impedance device ID
29	43-1.0	20333		
30	43-2.0	47504		
31	43-3.0	3201		
32	44-0.0	497024	Text	Tape measure device ID
33	44-1.0	20333		

TULANE®
UNIVERSITY

Summary



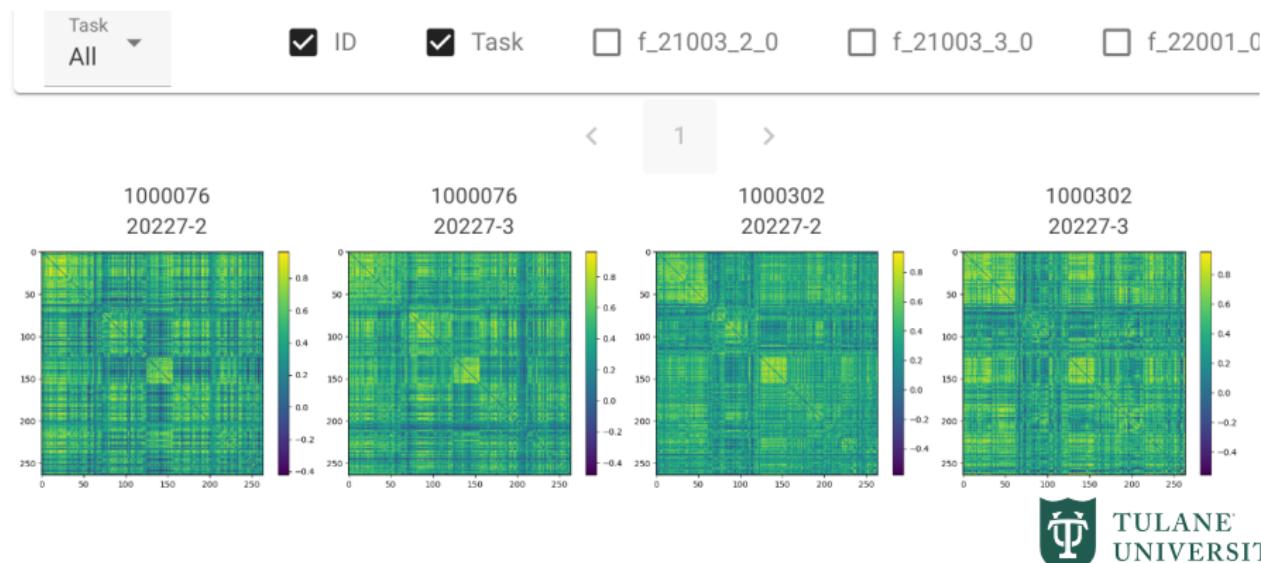
UKB Data

- The following fMRI-derived data are provided by UKB for 40,000+ subjects
 - ▶ 25 Component ICA FC, PC, and Amplitudes
 - ▶ 100 Component ICA FC, PC, and Amplitudes
 - ▶ “25” components are really 21
 - ▶ “100” components are really 55
- Additionally, we have pre-processed the data and now have the following items
 - ▶ 4D fMRI resting state volumes in MNI space
 - ▶ Power264 timeseries, FC and PC for resting state scans
 - ▶ Stored on the Wang lab NAS
 - ▶ Can take about 3 months to recreate on Cypress if lost
- There are also non-resting state scans which have not yet been pre-processed

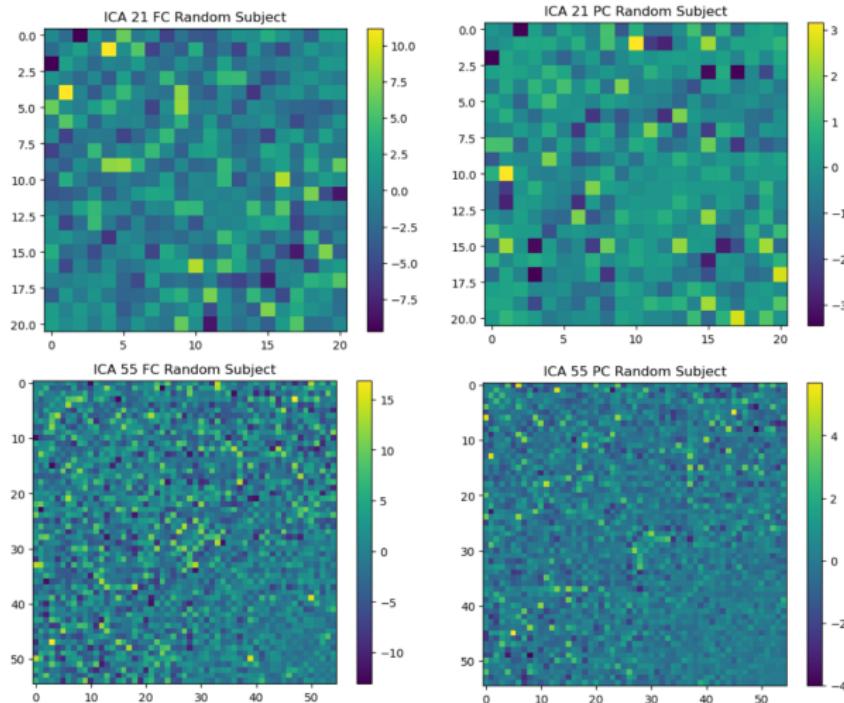


Longitudinal Data

- 2,722 subjects have longitudinal data (2 scans each)
 - ▶ 154 subjects have 2nd scan but not first (true missing, QC, etc.)
- fMRI scans taken about 2 years apart



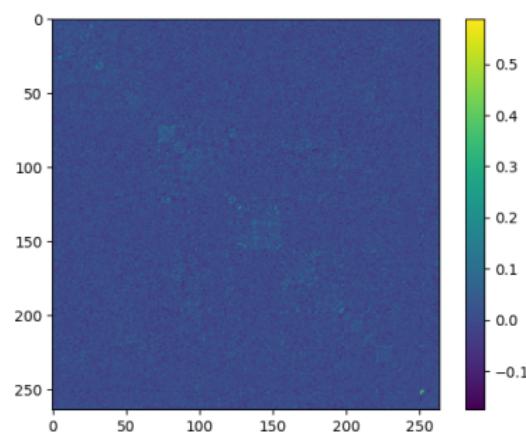
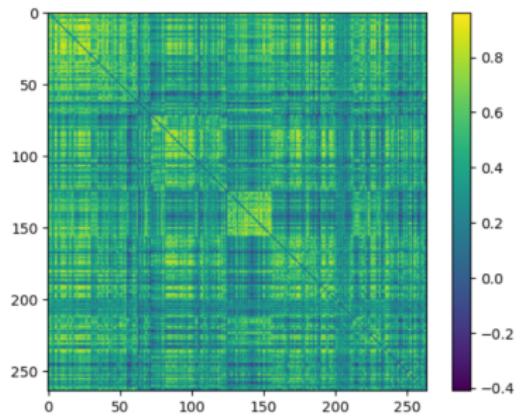
ICA-based FC and PC



Note: Data is upper or lower triangle, unknown indexing



Power264 FC and PC



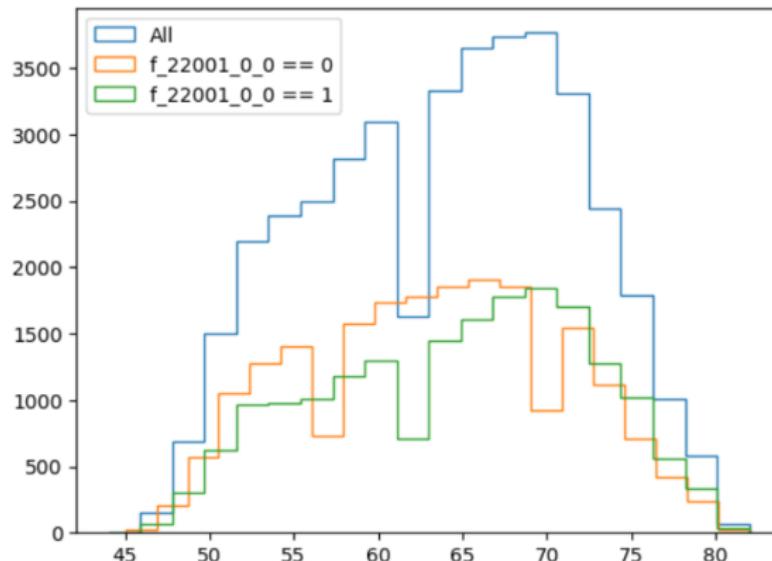
ImageNomer

- ImageNomer can load 40k subjects
- Can perform correlation with FC on 20k
- 32GB of RAM insufficient to perform correlation on full 40k



Demographics

- Mostly Caucasian race 50-80 years old
- Minor amount of other races
- Here is age distribution of males and females

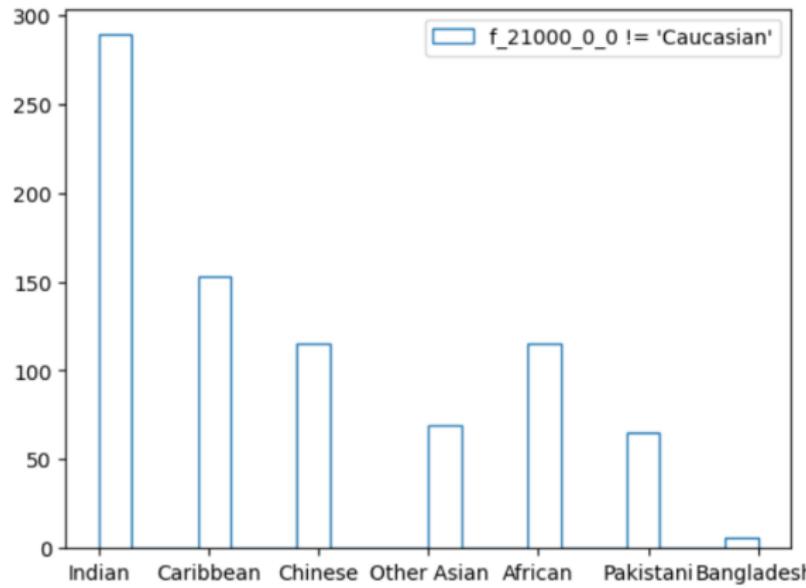


Phenotypes

Display phenotypes associated with the cohort

Race

- Out of 40,768 subjects with scans
- 39,811 (98%) Caucasian
- 957 other ethnicities



TULANE
UNIVERSITY

Interesting Additional Data

- Dr. Deng's group has access to over 10,000 phenotype and demographic fields
- fMRI for 40,000+ subjects, other fields for up to 500,000 subjects
 - ▶ Image-derived phenotypes (not just fMRI)
 - ★ Several thousand fields
 - ▶ Possibly SNPs or other genomic data
 - ▶ Metabolomics
 - ▶ Wearable data
 - ▶ etc.



Prediction and Feature Identification



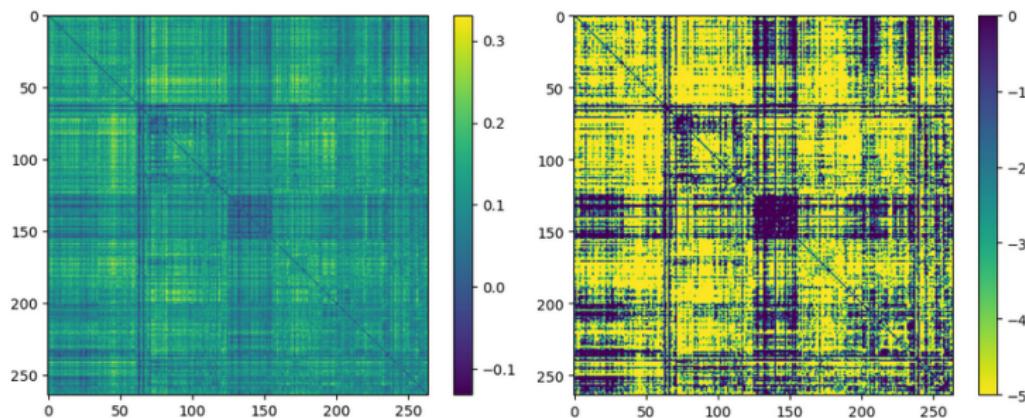
Sex Correlation

- Using our longitudinal subjects
- We see FC correlation with sex (M)
- Note uncorrelated visual network

Correlation

Find correlation between demographic features and FC for group

/analysis/corr/fc?cohort=UKB&query=All&field=f_22001_0_0&task=20227-3&fctype=fc&remap



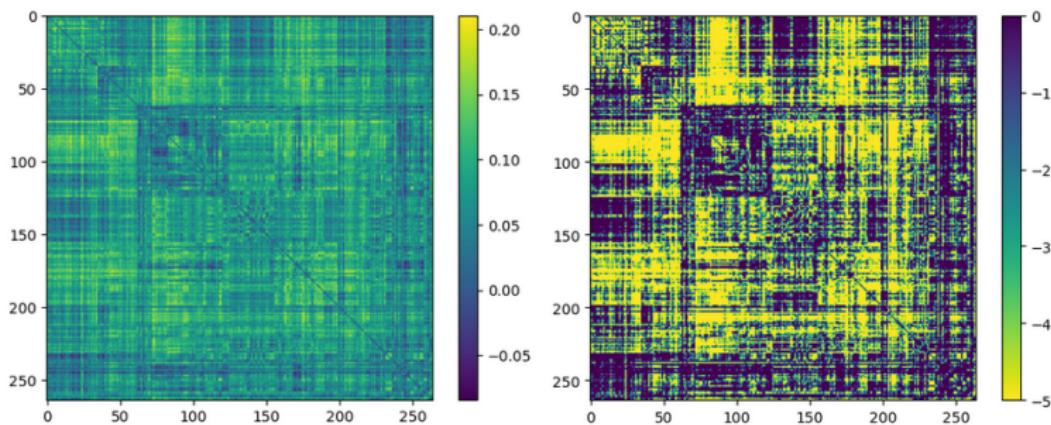
PNC Sex Comparison

- For comparison, here is the PNC dataset sex correlation
- Connections involving DMN and Frontoparietal are correlated with sex, as in the UKB

Correlation

Find correlation between demographic features and FC for group

/analysis/corr/fc?cohort=PNC&query=All&field=Sex&fctype=fc&remap&cat=M



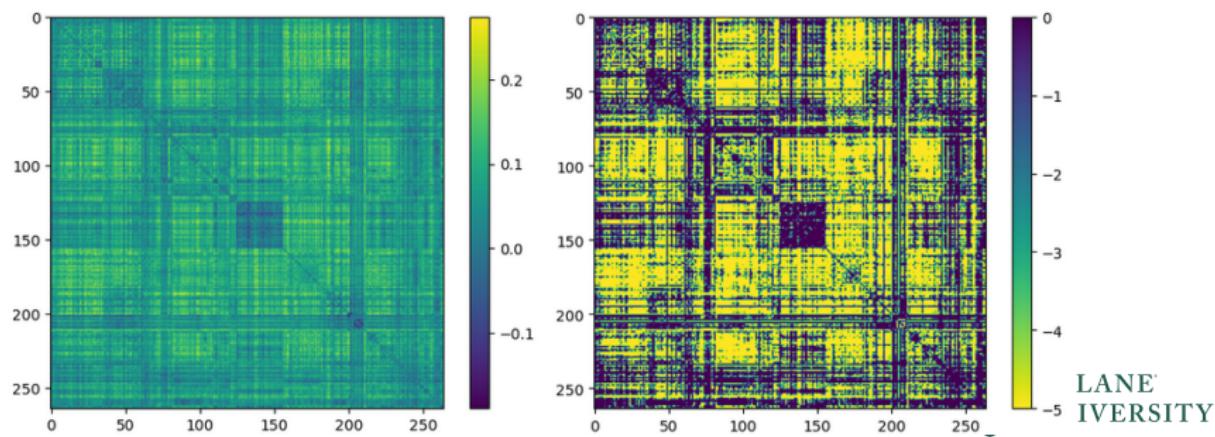
Age Correlation

- Still using longitudinal subjects
- Whereas age prediction is easiest task in PNC dataset of children and young adults, this is not the case in UKB older adults
- Visual network still uncorrelated

Correlation

Find correlation between demographic features and FC for group

/analysis/corr/fc?cohort=UKB&query>All&field=f_21003_3_0&ftype=fc&remap



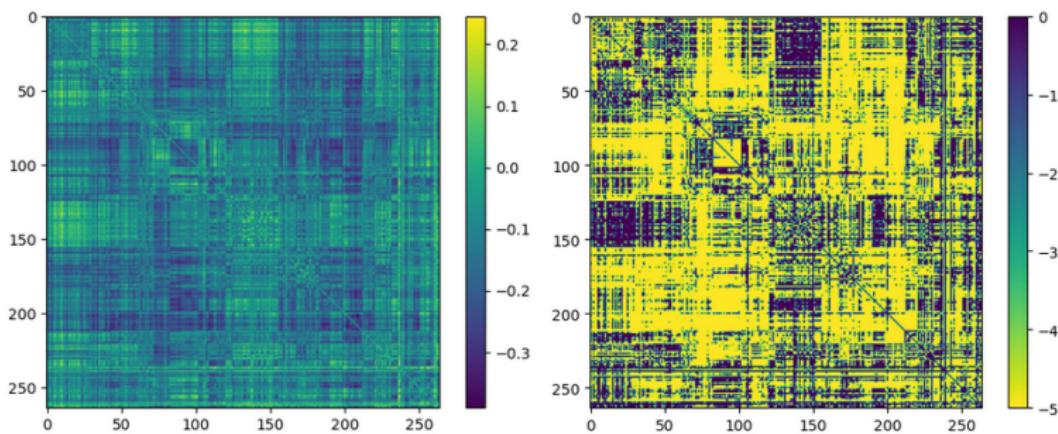
PNC Age Comparison

- For comparison, here is the PNC dataset age correlation
- In PNC, a "healthier" age is more mature, while in UKB, it is the reverse

Correlation

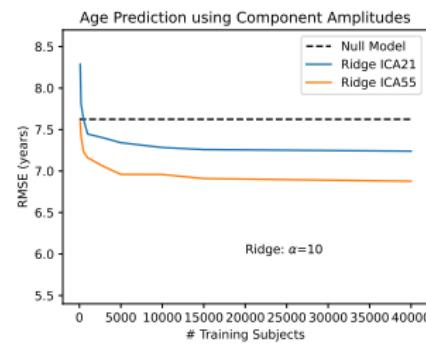
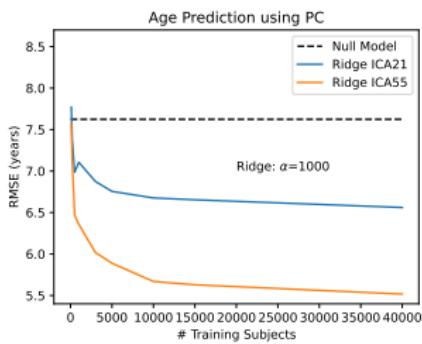
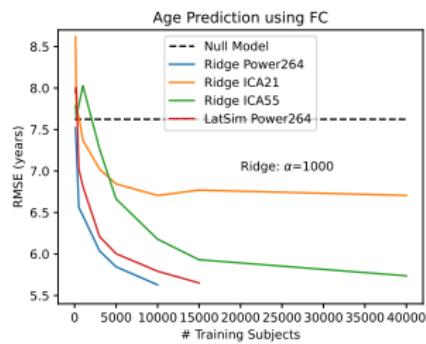
Find correlation between demographic features and FC for group

/analysis/corr/fc?cohort=PNC&query=All&field=age_at_cnb&fctype=fc&remap



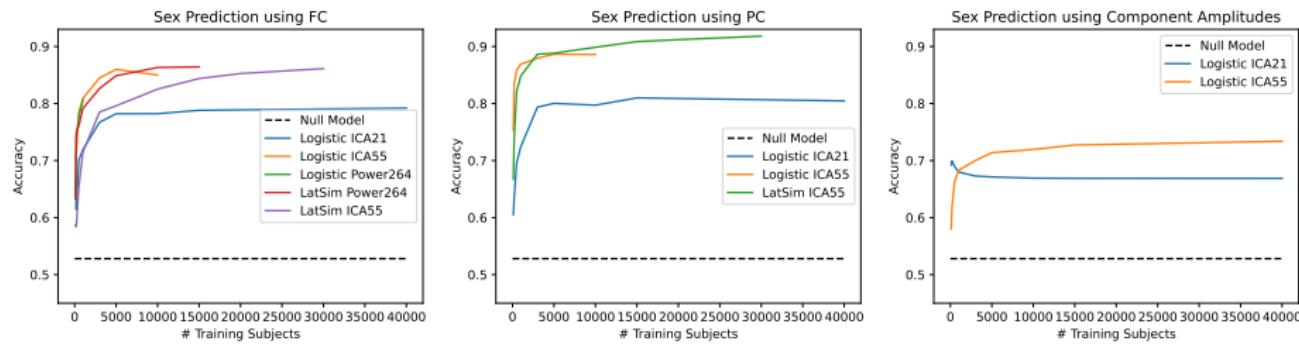
Age Prediction Validation

- Age prediction reaches an asymptote at large number of subjects - but at different rates
 - ▶ Null model 7.6 years RMSE
 - ▶ 5.6 years RMSE (Power264 FC, ICA55 FC)
 - ▶ 5.5 years RMSE (ICA55 PC)
 - ▶ 6.9 years RMSE (ICA55 Component Amplitudes)



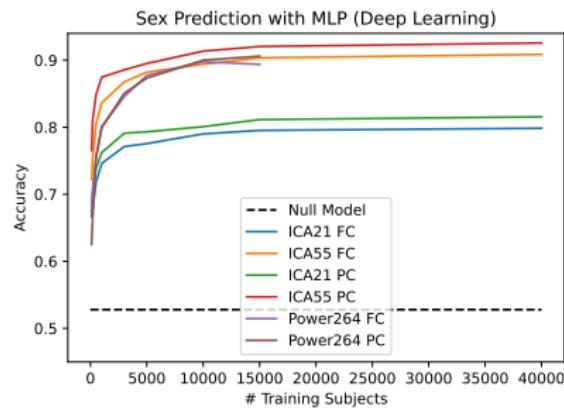
Sex Prediction Validation

- Sex prediction also reaches an asymptote around 5,000 subjects
- PC+LatSim is especially good
 - ▶ 86% accuracy (Power264 FC, ICA55 FC)
 - ▶ 89-91% accuracy (ICA55 PC)
 - ▶ 73% accuracy (ICA55 Component Amplitudes)



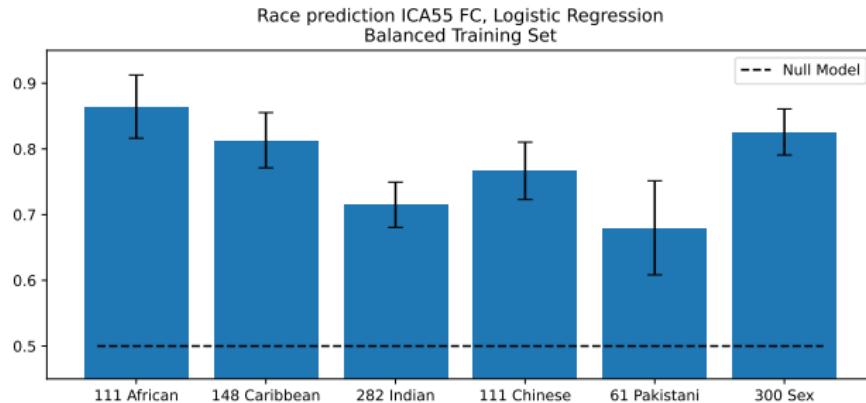
Sex Prediction (Non-linear Model)

- 2 layer MLP with 100 hidden neurons and ReLU activation function
- Slightly improved prediction, same trend
- Asymptote at 92.5% accuracy



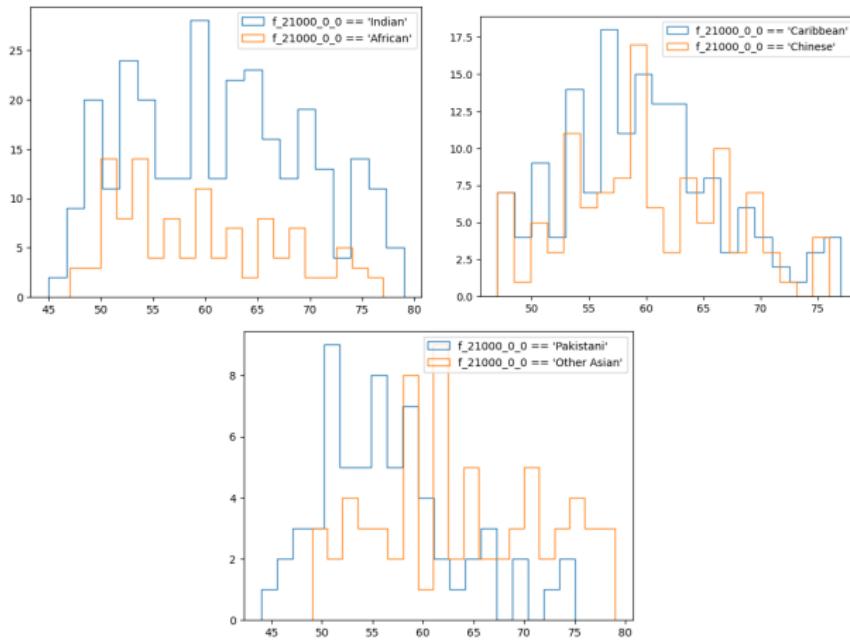
Race Prediction

- Can perform race prediction, although data more limited
- Similar but slightly better results for ICA55 PC
- Sex prediction included for comparison



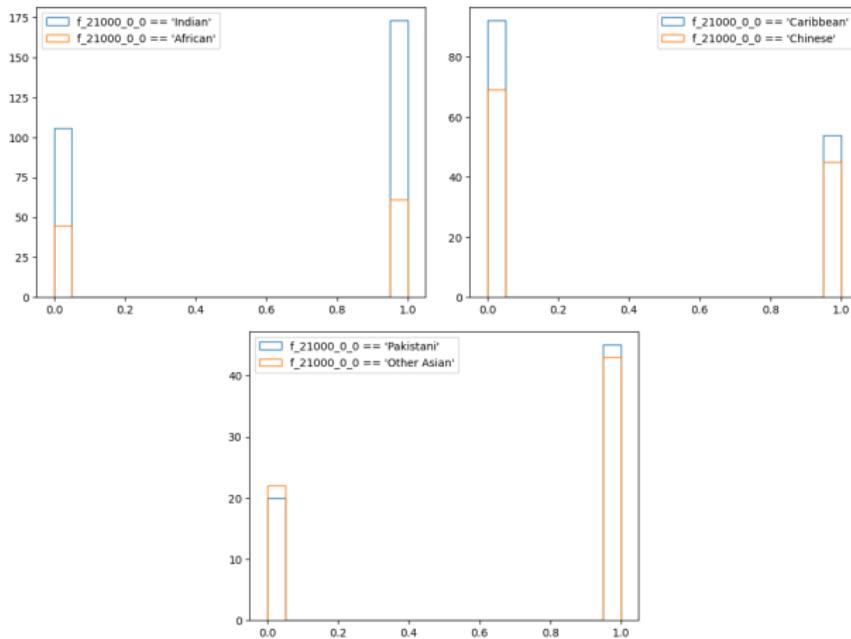
Examining Confounders for Race: Age

- All races with more than 50 members have approximately the same age distribution as the entire cohort



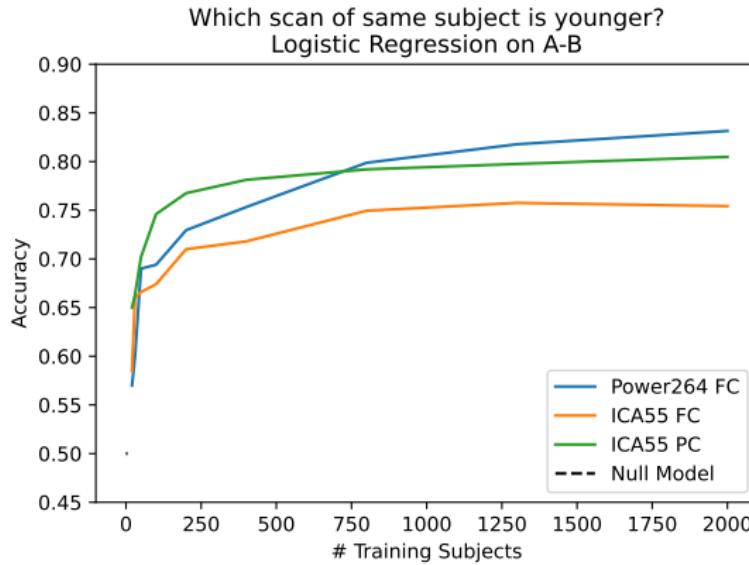
Examining Confounders for Race: Sex

- All races have skewed sex distributions compared to entire cohort
- Except for Pakistani, it isn't enough to fully explain race prediction (if sex was used as a replacement for race)



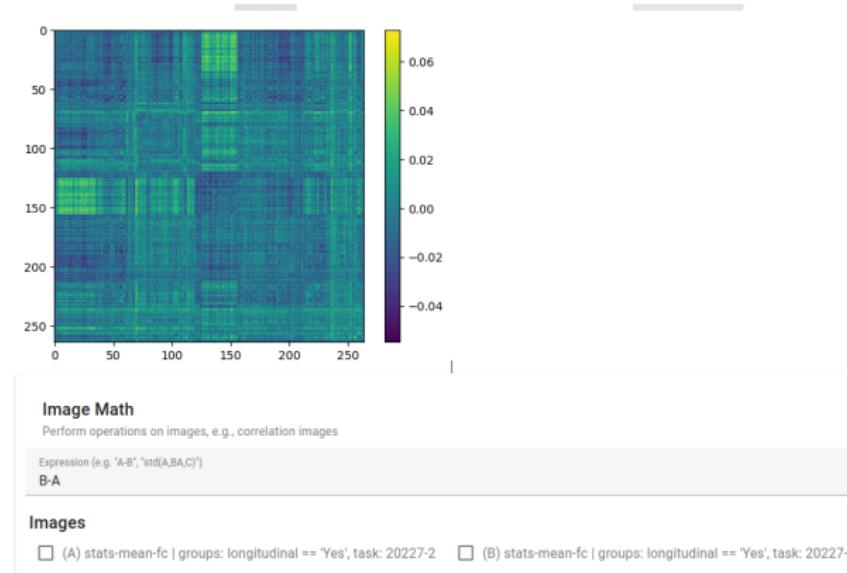
Longitudinal Scan Ordering

- Of the two longitudinal scans, which scan is taken earlier?
- Power264 FC asymptotically better at 82.5% accuracy



Visual-Somatomotor Connectivity Enhancement

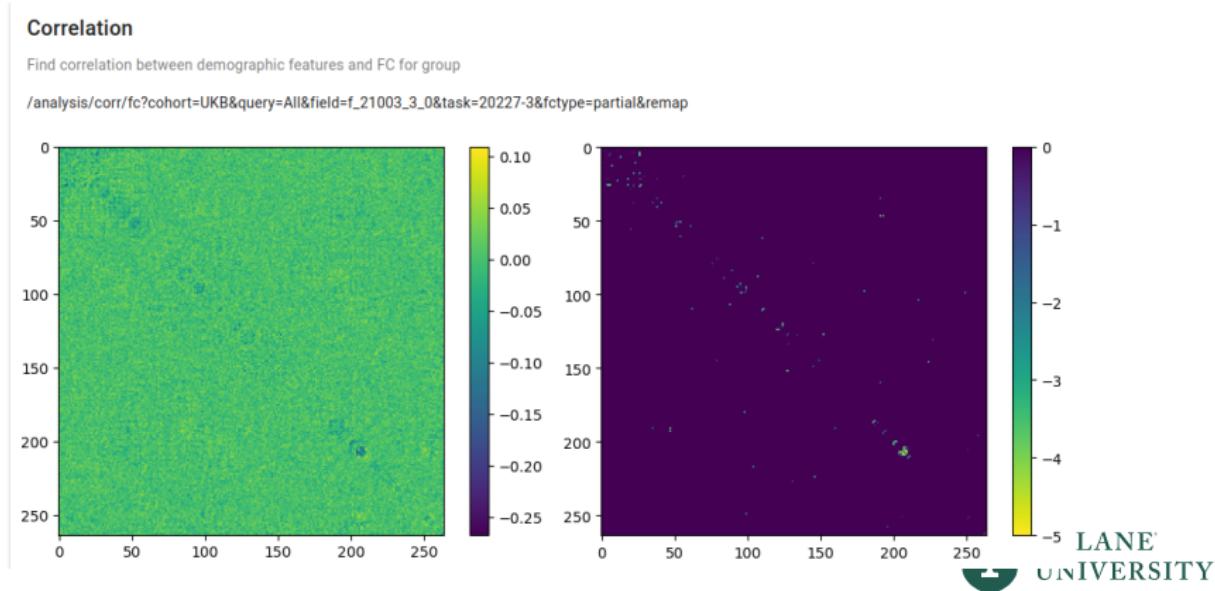
- Interesting enhanced connectivity between visual network and somatotmotor hand/mouth with aging
- Smaller increase between visual and DMN (Cingulo-Opercular, Auditory)



TULANE
UNIVERSITY

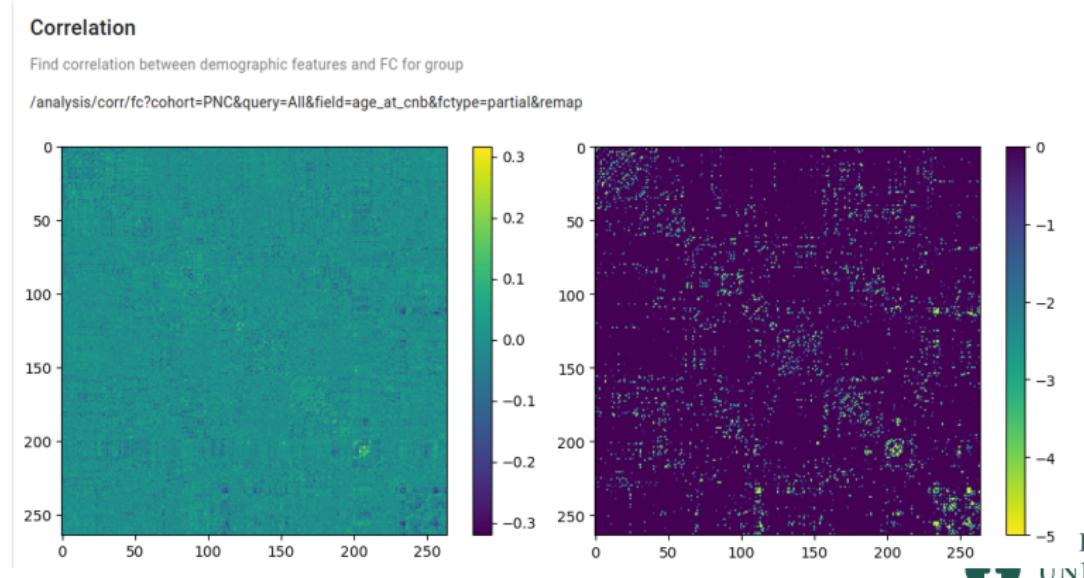
Power264 Partial Correlation

- “Partial correlation” is very good for provided ICA, but not as good for Power264
- Different from PNC dataset (see next slide)



Power264 PC PNC Comparison

- For example, here is the correlation between age-PC in the PNC dataset
- Approx. same number of subjects



ImageNomer: FC and Omics Visualization Software Detects Race Confound

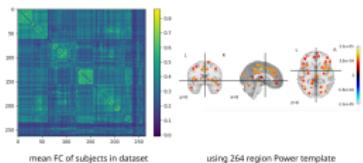


Graphical Abstract

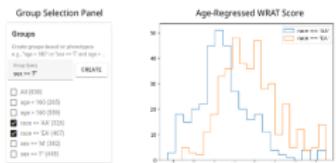
A. Our ImageNomer Software



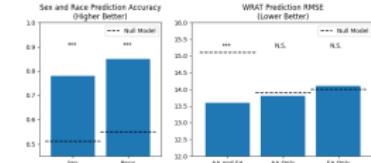
B. Using fMRI Functional Connectivity (FC)



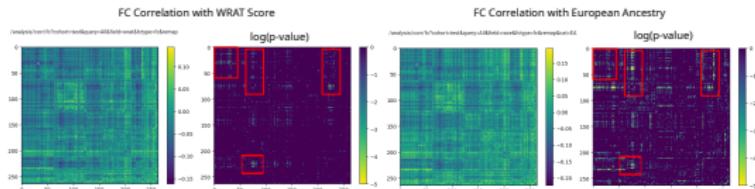
C. Finds Ethnic Bias in WRAT Score



D. FC Prediction of WRAT Actually Predicts Race



E. Makes the Case Against Unbiased Achievement-Related Features in FC



TULANE[®]
UNIVERSITY

Online Demo and Tutorial

- Online demo available
<https://aorliche.github.io/ImageNomer/live/>
 - Tutorial on OpenNeuro Fibromyalgia dataset ds004144

ImageNomer

latest

Search docs

Fibromyalgia Dataset Tutorial

FC View

Phenotypes View

Advanced Groups

Summary Images

Image Math

Phenotype Correlations

Phenotype-FC Correlations

Model Weights

Partial Correlation

Decomposition

SNPs

Further Analysis

Report Bugs

FC View

Select "All" in the "Groups" area and click on the "FC" (functional connectivity) tab.

Additionally, check "ID" and "Task" under "Display Options".

You should see the following:



The screenshot shows the ImageNomer software interface. At the top, there's a navigation bar with tabs like 'Home', 'About', 'Help', and 'Logout'. Below the navigation bar is a search bar labeled 'Search docs'. The main content area has a title 'FC View' and a sub-instruction: 'Select "All" in the "Groups" area and click on the "FC" (functional connectivity) tab.' Below this, another instruction says 'Additionally, check "ID" and "Task" under "Display Options".' At the bottom of the main content area, it says 'You should see the following:' followed by a large screenshot of the software's graphical user interface. The interface features a left sidebar with various menu items such as 'ImageNomer', 'Home', 'About', 'Help', 'Logout', 'Groups', 'Analyses', 'Reports', 'Visualizations', 'Tools', 'Data Sources', 'Modeling', 'Statistics', 'Machine Learning', 'Image Processing', 'Image Math', 'Phenotype Correlations', 'Phenotype-FC Correlations', 'Model Weights', 'Partial Correlation', 'Decomposition', 'SNPs', 'Further Analysis', and 'Report Bugs'. The main workspace displays a grid of 12 functional connectivity matrices arranged in a 3x4 layout. Each matrix is a heatmap showing correlations between different brain regions. The matrices are color-coded, with darker shades representing higher correlation values. The software's header also includes the text 'ImageNomer', 'Version 1.0.0', 'Build 1000', and 'Last updated: 2023-01-01'.



TULANE
UNIVERSITY

Docker Images

- We have a Docker image and instructions on how to use it
- Currently does not add more than live web demo
- To add own data you must install locally
- Looking for programmers to improve Docker, add upload page, and add CCA page

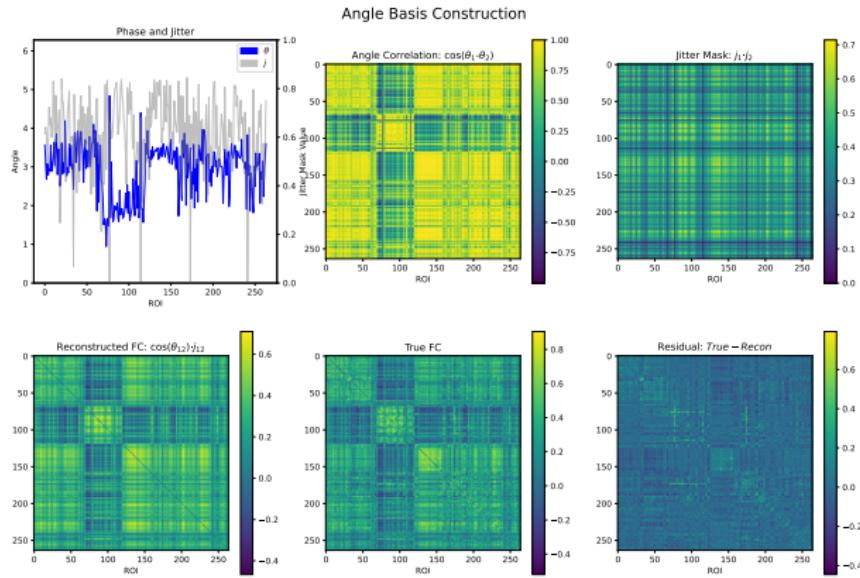


AngleBasis: A Generative Model and Decomposition for Functional Connectivity



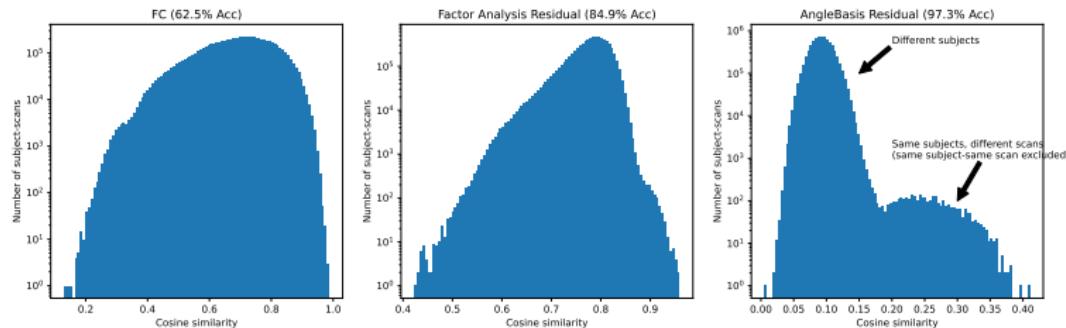
Graphical Overview

- Use combination of sinusoidal basis and low-rank jitter to approximate FC data (10x compression)



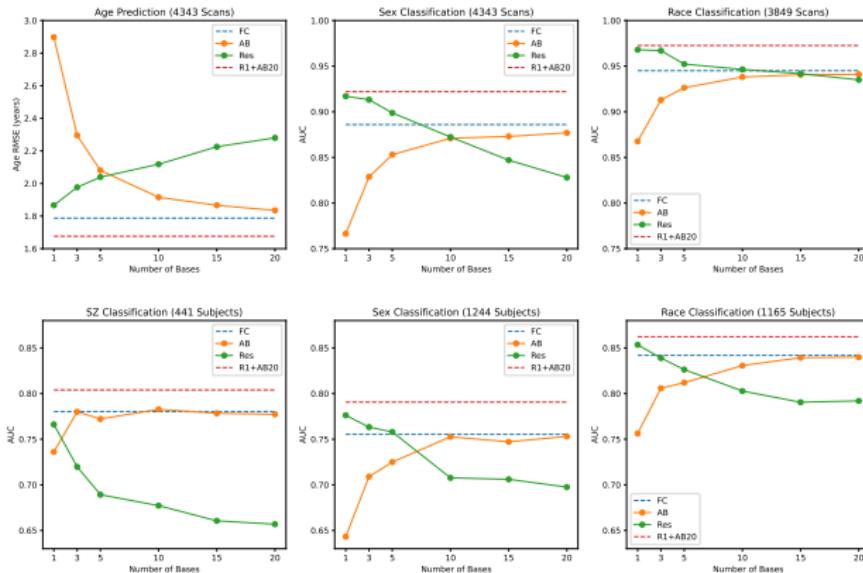
Identifiability

- Residual greatly increases identifiability (fingerprinting) compared to FC



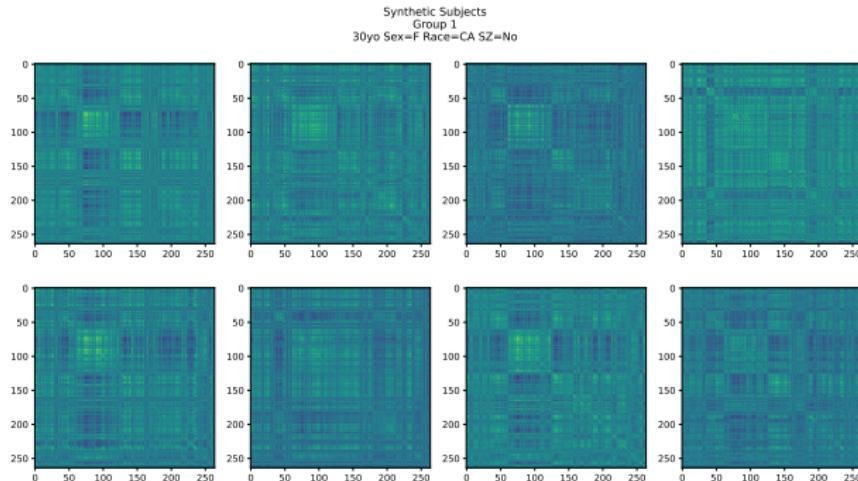
Reconstruction and Residual

- As recon accuracy goes up, residual accuracy goes down
- Ensemble gives superior results compared to FC



Synthetic Subjects

- We can use low rank approximation of FC and trained weak models to construct synthetic subjects



fMRI Identifiability Results in Memorization



Longitudinal or Contemporaneous Repeat Scans



Identifiability for Inflated Results

- We saw in the last section that FC can easily have over 60% same-subject different-scan identifiability
- With processing, this number can be increased to over 90%

Claim

fMRI identifiability can be used to artificially inflate model accuracy

- This can explain some of the very optimistic results found in the literature
- We demonstrate on UKB, Fibromyalgia, PNC, and BSNIP datasets

Identifiability also referred to as fingerprinting



Procedure

Two requirements:

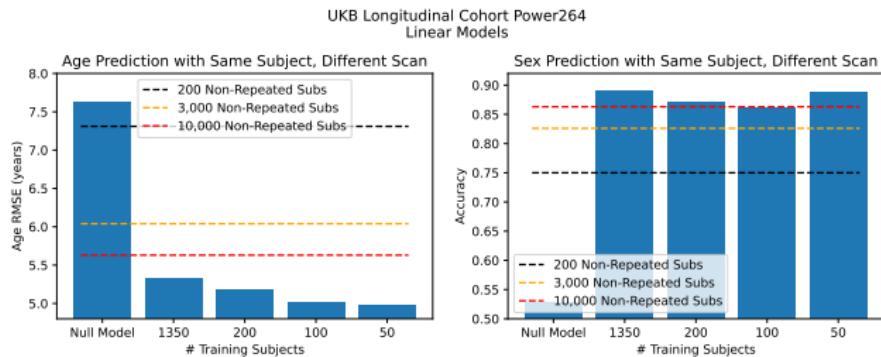
- ① Either longitudinal or contemporaneous scans of same subject
- ② These scans must treated as independent subjects (but with same response variables)

Note that we are not placing the same scan in the training and test sets.

- No individual scan can be in both training and test sets

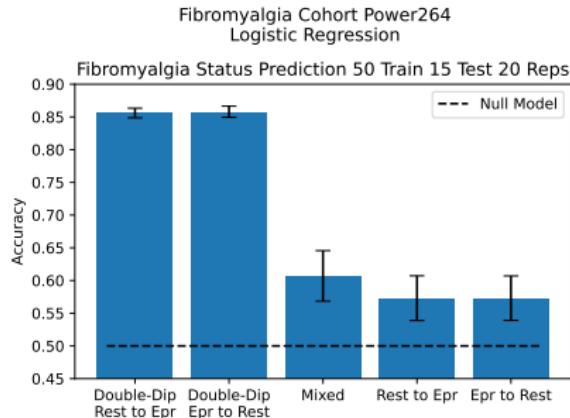
UKB Longitudinal Data

- Perform age and sex prediction on longitudinal UKB cohort (2,722 subjects)
- Identifiability enhancement exceeds anything possible with 10,000 training subjects
 - ▶ Using only 50 subjects having 2 scans...



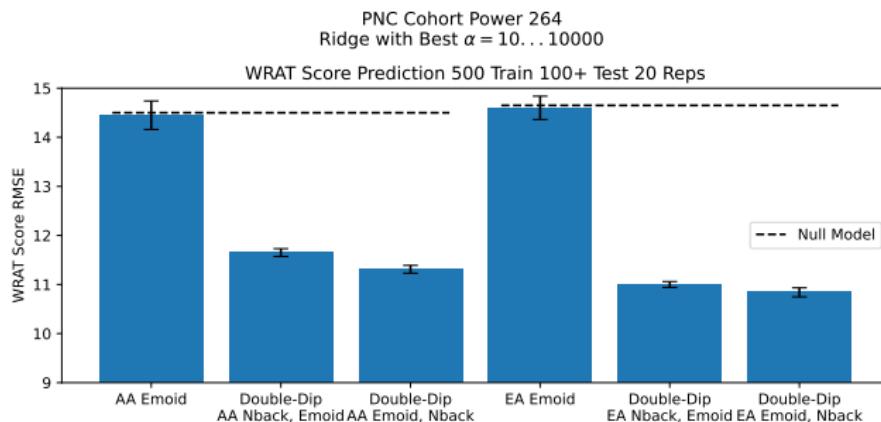
Fibromyalgia Dataset

- OpenNeuro dataset ds004144 with 66 female Fibromyalgia patients+controls
- 65 of whom have two scans: resting state and epr
 - ▶ Honest prediction is slightly but significantly better than chance
 - ▶ Using same subjects but different scans for train and test sets increases predictive ability to 85%



PNC Dataset WRAT Prediction

- PNC Dataset with 600+ scans of both nback and emoid tasks for both EA and AA ethnicities
- We have previously concluded that ethnic bias in WRAT score, combined with ability to predict race, leads to bias when predicting WRAT in mixed groups
 - ▶ We can still achieve even better prediction by repeating subjects but using different contemporaneous scans



Is this a problem?

Reported Features

The community's goal is to do science; features (ROIs, connectivity, etc.) that are reported along with misleading prediction results may confuse other scientists not intimately familiar with the field.

- This may or may not have been obvious to you
- However, given the very large numbers of papers trying to predict phenotype from fMRI metrics...
- ...and the very optimistic prediction accuracies or errors they report
- It is likely that someone, somewhere is taking advantage of this.



Dynamic Connectivity



Problem: Only One Scan Per Subject

- Is it still possible to inflate accuracy without having multiple scans per subject?
- Yes! You can try to use dynamic functional connectivity (or something similar).



Procedure

Example for windowed dynamic functional connectivity.

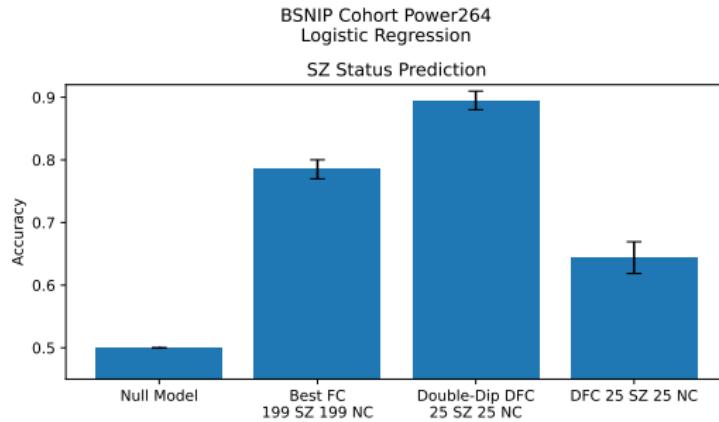
- For each scan, create several FC matrices using windows.
- Rest of procedure same as for static FC.
- Here you are actually putting same subject same scan (but different FC matrices) into training and test sets
- You can, e.g., perform k-means clustering and regress on the centroids to obfuscate¹



¹Rashid et al. 2016 10.1016/j.neuroimage.2016.04.051.

BSNIP SZ Prediction

- Predict schizophrenia status in BSNIP dataset
- We have 199 SZ scans and 243 NC (normal control) scans
- Use N=50 TR window size



Prior and Current Work

- Minor workflow anomalies can lead to superb results
- Bennett et al. 2009 found 16 voxels associated with emotion processing in a dead salmon
- Elliott et al. 2020² found a majority of fMRI studies had poor replicability
- More recently, two posters at OHBM 2023
 - ▶ Hamdan et al. 2023 “Cofound-Leakage: Confound Removal In Machine Learning Leads To Leakage”³
 - ▶ Rosenblatt et al. 2023 “Neuroimaging data trustworthiness: false enhancement of classifiers with minor data manipulations”⁴

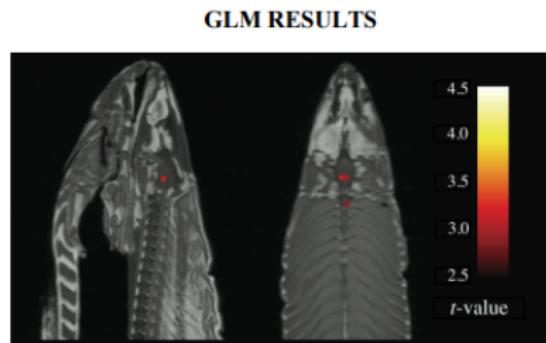
²<https://doi.org/10.1177/0956797620916786>

³<https://doi.org/10.48550/arXiv.2210.09232>

⁴<https://doi.org/10.1016/j.patter.2023.100756>

Dead Salmon

- The point of the Bennett et al. 2009 study was to highlight the need for multiple comparison correction
- Using Benjamini and Hochberg (1995) or family-wise error rate (FWER, Friston 1994) correction eliminated significant voxels



A t -contrast was used to test for regions with significant BOLD signal change during the photo condition compared to rest. The parameters for this comparison were $t(131) > 3.15$, $p(\text{uncorrected}) < 0.001$, 3 voxel extent threshold.



Classic Finance Example

- From “Learning from Data,” by Abu-Mostafa, Magdon-Ismail, and Lin (2012)
- A post-hoc algorithm found a 52.1% ability to predict exchange rate direction, in theory making a 22% profit over two years
- When used in live trading, the program loses money
- Explanation was data snooping
 - ▶ Original training data was normalized to zero mean and unit variance using both training and test set statistics
 - ▶ This minor inclusion of the test set was enough to poison results



Intentionality

- It may be that most researchers don't mean to deceive
- One may closely inspect something which reduces accuracy (in order to fix it)
- One may also not inspect procedures which increase accuracy
- A more complicated pipeline means more opportunities to make mistakes in your own favor



Questions

Thank you!
Any questions?

