# Data Explanation

### PhenotypicData\_AllABIDE.csv

This file includes the phenotypic data of all subjects in the ABIDE I dataset.

### ABIDE\_PHENOTYPE\_LEGEND.pdf

This document explains the different phenotypic categories described in PhenotypicData\_AllABIDE.csv.

# Data Acquisition

### MATLAB\_prep.ipynb

This jupyter notebook file fetches the appropriate data from the [Autism Brain Imaging Data Exchange (ABIDE)](http://fcon_1000.projects.nitrc.org/indi/abide/abide_I.html). It organizes the data from subjects with ASD and without in separate sections for easier error-checking and exploration. Example connectivity matrices are plotted. Connectivity matrices (M\_info.mat) and nuisance variable data (M\_confounds.mat) are exported as to MATLAB for analysis.

### matchedIDs\_female.csv

This file includes the subject IDs for all age-matched subjects that should be used analysis. This data is required for MATLAB\_prep.ipynb

### matchAge.csv

This file is created after running MATLAB\_prep.ipynb and it includes the ID, age, diagnosis, site, FIQ, and sex of all subjects exported to MATLAB for analysis. This is meant for error-checking only, not required for further analysis.

# Model Creation

### PredictiveModel.m

This script creates a predictive model for ASD using the connectivity matrices calculated in MATLAB\_prep.ipynb. The code is based off the work of Yahata et al. (2010).

### L1\_SCCA.m

This function performs L1-regularized Sparse Canonical Correlation Analysis on the input connectivity matrices and nuisance variable data.

### iterativeStratification.m

This function creates stratified groups of subjects based on input criteria.

### predict\_log.m

This function determines the probability of an instance of ASD from 0 to 1 using the sigmoid function.

### SB2\_Manual.pdf

This document explains the functions included in the predictive model.

### Other functions: biclsfy\_slrvar.m, calc\_label.m, calc\_percor.m, finputcheck.m, label2num.m, normalize\_feature.m, num2label.m, slr\_error\_table.m, slr\_learning\_var2.m

See SB2\_manual.pdf for explanation.

# Results

### FinalMode.mat

This file includes the weights/coefficients of all the features determined when training the model under the variable ‘feature\_weights’. The indices of the 15602 selected features out of the possible 19900 are included in the variable ‘selected\_features\_indices’. These can be used to predict diagnosis on new data using the predict\_log.m function.

## Bibliography

Yahata N, Morimoto J, Hashimoto R, Lisi G, Shibata K, Kawakubo Y, Kuwabara H, Kuroda M, Yamada T, Megumi F, Imamizu H, Náñez JE Sr, Takahashi H, Okamoto Y, Kasai K, Kato N, Sasaki Y, Watanabe T, Kawato M. A small number of abnormal brain connections predicts adult autism spectrum disorder. Nat Commun. 2016 Apr 14;7:11254. doi: 10.1038/ncomms11254. PMID: 27075704; PMCID: PMC4834637.