Explanation of variables and Matlab code made available accompanying the manuscript

**PONE-D-20-24174 “Predictive utility of task-related functional connectivity vs. voxel activation” by Habeck et al.**

There are a lot of data files, code, and results used for this paper, and we provided everything used in the preparation of this paper. However, for brevity in these notes we describe only the main data and give the syntax for the figure computations /preparations.

However, to produce the figures, all Data and Code needs to be downloaded into the same Matlab working directory. The code also assumed that SPM 8 or 12 and all its components has been installed.

Any further information can be obtained from the first author directly at **ch629@columbia.edu**.

**Matlab Data Archives (.mat files)**

**BlCombo\_Nov2020.mat** - This is the main data set, containing all neural and behavioral data. All other .mat files are results that have been pre-run. We give variable names and dimensions with a brief explanation.

 age 240 x 1 age of participants

 allTaskID 2880 x 1 task index: (1-3 memory tasks, 4-6 fluid reasoning tasks, 7-9 speed tasks, 10-12 vocabulary tasks)

 behav 240x12 Z-scores task performance data, higher values denote better performance

 edu 240x1 Years of education

 fcData 34716x2880 Functional connectivity for 34,716 edges for all subjects and tasks

 file 1x1 NIFTI header information

 id 240x1 ID numbers, only internally relevant

 meanful\_set2 24596x1 probabilistic conjunction mask for voxel activation data

 ra\_label 1x4 Label for 4 cognitive domains: Memory(=MEM), Fluid reasoning (=FLUID), Perceptual Speed (=SPEED), and Vocabulary (=VOCAB)

 sex 240x1 sex/gender

 structData 240x136 regional cerebral volume and thickness data in 68 ROIs, concatenated within participant after z-scoring

 thx 240x68 regional cortical thickness in 68 ROIs

 vol 240x68 regional cortical volume in 68 ROIs

 voxData 24596x2880 voxel activation data in 24,596 voxels

**MEM\_resultsTrain192.mat, FLUID\_resultsTrain192.mat, SPEED\_resultsTrain192.mat, VOCAB\_resultsTrain192.mat, CPMC\_MEM.mat, CPMC\_FLUID.mat, CPMC\_SPEED.mat, CPMC\_VOCAB.mat** – These 8 files that contain all results for the 4 cognitive outcomes that were considered in the paper. These files serve as the inputs for creating the figures. We will not discuss all the variables saved in these files, but the main content is the Predictive-Residual-Sum-Of-Squares (=PRESS) statistic, and the patterns themselves.

**ConnVox\_Infra.mat –** This file contains some supporting information for visualizing connectomes, network labels, ROI listing etc. It is accessed by the code that produces Figure 5.

**Matlab code (.m files)**

**predictModule.m -** This is the center piece that contains all prediction computations in a split-sampling framework. The routine calls several functions. For repeating some of the computations, the code would have to be used as follows. Type the command:

RA=1; predictModule

This sets the cognitive domain to be MEM (RA=1), while RA=2/3/4 would set it to FLUID,SPEED, or VOCAB, respectively. Starting the command in an empty workspace will commence with the computation of the split-sample simulations (1,000 iterations) where several input modalities are compared in their predictive utility of multivariate models out of sample. The paper contains all details. This computation will take a little while (~several hours) and will plot intermediate results in a window with violin plots. There is no need to repeat the running of this program since the results files are given in this archive too. The code is mainly deposited for maximum transparency.

**Figure1.m -** This routine recreates Figure1 form the paper and plots histograms of the differences of PRESS values between the main analytic vehicle used in the paper, PCA regression, and Connectome Predictive Modeling (CPM). Just type

 Figure1

in a cleared workspace.

**Figure2.m -** This routine recreates Figure 2 from the paper and gives a cumulative scree plot of both voxel activation and functional connectivity up to PC 300, showing much stronger variance concentration in voxel activation compared to functional connectivity.

Just type Figure 2 in a cleared workspace.

**Figure3.m -** This routine recreates Figure 3 from the paper, summarizing the main results about predict out-of-sample success of PCA-regression to different modalities for different cognitive outcomes.

Just type Figure 3 in a cleared workspace.

**Figure4.m -** This routine computes voxel NIFTI images from the patterns that were generated and saved in the results files. Visualization of these figures requires an additional step and off-shelf packages are available to visualize these images. In the paper, we used the freely available suite MRICROn. Further, this routine requires SPM to be installed since it used the SPM function “spm\_write\_vol.m”.

Just type Figure 4 in a cleared workspace.

You will also notice a screen output of select loadings.

**Figure5.m -** This routine produces the four panels shown in Figure 5. It calls a host of other functions to first produce coarse-grained connectivity patterns, and transform those to square matrices, and plot them appropriately. These 4 panels were collated in PowerPoint to create the final version of Figure 5, but the content is the same here.

Just type Figure 5 in a cleared workspace.

You will also notice a screen output of select loadings.

These Matlab programs are only the top-level wrappers that you can execute from the command-line. However, many supporting functions and routines are used, which I list below:

PCA\_predict.m

lineplot\_ch.m

zVec2Mat.m

nancorrcoef.m

predictModule\_CPM.m

best\_fit2.m

networkComp\_gross.m

cpm\_predict.m

networkPlot2\_gross.m

eig\_habeck.m

find\_max2.m

patternZ.m

find\_max3.m

pca\_f.m

violin\_plot2.m