Explanation of variables and Matlab code made available accompanying the manuscript “Identification of a Resting Bold Connectome Associated with Cognitive Reserve” by Habeck, Varangis, and Stern NIMG-19-3083

**Matlab Data Archives**

**CUMC\_data.mat -** This is the data sample from Columbia University that was used in the manuscript. It functions as the derivation sample for the obtained IQ-connectome. It includes the connectivity, structural, demographic and cognitive data for 424 individuals. We list all variables with a brief explanation below with their dimensions and explanations below:

MNI 264x3 MNI coordinates for all 264 Regions of interest

MNI\_label 264x30 Network labels for all 264 Regions

age 424x1 Chronological age of participants in integer years

edu 424x1 Education in years

id 424x1 ID numbers (internal use only)

meanFD 424x1 mean framewise displacement for resting fMRI

nart 424x1 IQ measure

pairs 34716x2 edge assignment that explains the rows in zVec

paths 424x119 file paths for connectivity (internal use)

scrubFDpaths 424x106 scrubbed FD time series files (internal use)

scrubFrac 424x1 fraction of scrubbed data

scrubMaskPaths 424x89 scrubbing-mask paths (internal use)

sex 424x1 biological sex

thx 424x68 cortical thickness in 68 FreeSurfer ROIs

thxLabels 68x27 region labels for thickness ROIs

zG 424x1 total Cognition (=G)

zMem 424x1 Memory

zReason 424x1 Fluid Reasoning

zSpeed 424x1 Perceptual Speed

zVec 34716x424 FC-array: rows = edges, columns = participants

zVocab 424x1 Vocabulary

zincl\_index 1x26106 subset of edges included in all analyses

**NKI\_data.mat -** This is the data sample from Nathan-Kline Institute that was used in the manuscript. It functions as the replication sample for the obtained IQ-connectome. It includes the connectivity, structural, demographic and cognitive data for 286 individuals. We list all variables with a brief explanation below with their dimensions and explanations below:

ID 286x9 ID string (internal use only)

age 286x1 chronological age in years

edu 286x1 education in years

pairs 34716x2 edge explanation in terms of ROIs (=rows in zVec)

sex 286x1 biological sex

thx 286x68 cortical thickness in 68 FreeSurfer ROIs

wais 286x1 IQ measure

zG 286x1 total cognition (=G)

zMem 286x1 Memory

zReason 286x1 Fluid Reasoning

zSpeed 286x1 Perceptual Speed

zVec 34716x286 FC array: rows = edges, columns = participants

zVocab 286x1 Vocabulary

**CUMC\_results.mat -** This is the results file containing the derived functional connectome and 5,000 bootstrap connectomes. We list all variables with a brief explanation below with their dimensions and explanations below

MNI 264x3 MNI coordinates for all 264 Regions of interest

MNI\_label 264x30 Network labels for all 264 Regions

bpattern 26106x5000 5,000 bootstrap connectomes, masked only to the edge subset included in the analyses

distVec 34716x1 Euclidean length (mm) of all edges

fracPos 34716x1 fraction of bootstrap iterations > 0 for all edges

pairs 34716x2 edge explanation in terms of ROIs

point\_image 34716x1 point-estimate connectome, edges outside the included subset are set to NaN

zincl\_index 1x26106 subset of edges included in all analyses

**Matlab routines and sample analyses**

A few Matlab routines and functions are included. We will not list them all but only give the most important workflows, to recreate the results obtained for the paper. We annotated the code at the source to the best of our ability and hope it’s clear.

**Derivation of connectome with N-fold cross-validation and bootstrap estimation**

Type

**‘load CUMC\_data; loocv\_PCA ‘**

into Matlab console, the console returns an output to the routine that looks like:

Number of sources in the data, i.e. lambda > 1 : 167

Iteration and best-fitting PC-set: 1 1-4

Iteration and best-fitting PC-set: 2 1-106

Iteration and best-fitting PC-set: 3 1-101

Iteration and best-fitting PC-set: 4 1-101

Iteration and best-fitting PC-set: 5 1-101

Iteration and best-fitting PC-set: 6 1-101

Iteration and best-fitting PC-set: 7 1-11

…

Iteration and best-fitting PC-set: 424 1-32

Best-fitting set is PC 1-32

Perform bootstrap estimation too? (1==YES)

Now type ‘1’ for commencing the bootstrap routine, or any other key to stop If you type ‘1’ you will get the below TTY output next, and it will continue until all 5,000 iterations have been done

Perform bootstrap estimation too? (1==YES) **1**

Run 5,000 iterations of bootstrap test and report regular updates to TTY

Estimation takes 1-2 hours on reasonable lap top

Bootstrap iteration: 1

Bootstrap iteration: 2

Bootstrap iteration: 3

Bootstrap iteration: 4

Bootstrap iteration: 5

Estimation is continuing, reporting every 100th iteration ...

Bootstrap iteration: 100

Bootstrap iteration: 200

Bootstrap iteration: 300

Bootstrap iteration: 400

Bootstrap iteration: 500

…

**Creating table 3 and 4**

Tables 3 and 4 can be created, i.e. the brain-behavioral regression can be run and the results are displayed in the Matlab console. For table 3, the Columbia data file needs to be loaded, for table 4 it’s the NKI data file.

Type

**‘load Columbia\_data’** (table 3)

Or

**‘load NKI\_data ‘** (table 4).

Then follow with

**‘table3\_4’**

The results in table 3 and 4 are now displayed sequentially on the screen.

**Creating table 3 and 4**

Type **‘figure2’** into the Matlab console to create the two panels that appear in Figure 2.

**Getting a listing of the edges displayed in table 5**

Type **‘reportEdges’** into the Matlab console and enter the threshold for the positive and negative loadings. For the positive loadings it was 0.94 (implying that at least 94% of bootstrap iterations had a positive value), for the negative loadings it was 0.06 (implying that at most 6% of the bootstrap loadings had a positive value).