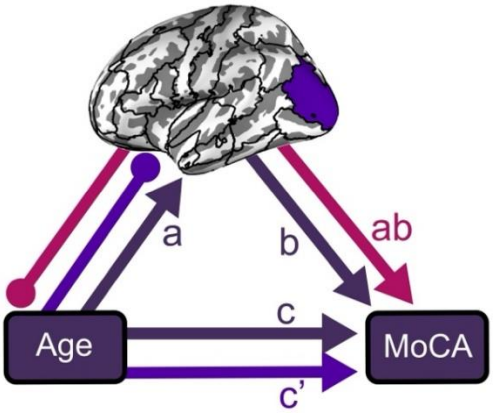
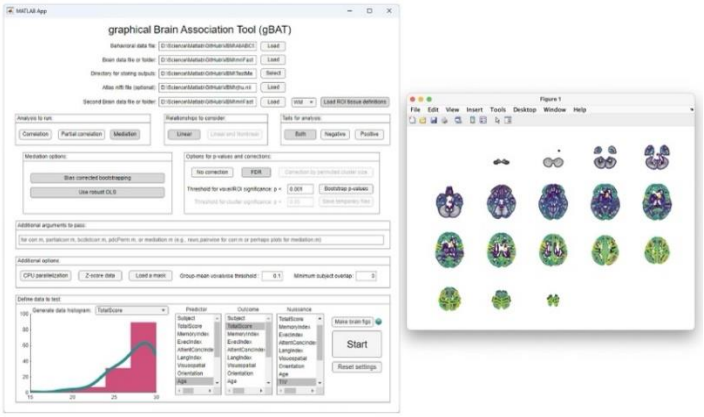
Graphical Brain Association Tool (gBAT)

User Manual





The Graphical Brain Association Tool (gBAT) is a MATLAB-based toolbox designed to map brain-behavior associations, with a particular focus on mediation analysis. Its purpose is to provide an easy-to-use graphical interface that simplifies exploring the trove of data generated in neuroimaging studies through powerful modeling techniques that can explain relationships between variables through their influence on brain structure or function. We tend to think of this analytic technique as generally underutilized in cognitive neuroscience, where brain data can help inform how different cognitive and biological processes are interrelated. By focusing on how these variables influence each other indirectly through neural pathways, gBAT offers researchers a valuable tool for uncovering complex interactions that might otherwise remain hidden. For example, consider the well-studied relationship between age and cognition. Mediation analysis can help uncover how age might influence cognition through alterations in specific neural pathways. Or consider the well-established relationship between stroke severity and rehabilitation. gBAT can help quantify how stroke severity leads to alterations in functional or structural connectivity within specific neural pathways, which in turn indirectly affect recovery outcomes.

While gBAT emphasizes mediation analysis, it is a highly versatile tool that can be applied to a wide range of brain-behavior analyses. To complement mediation results, the toolbox also supports independent analyses such as correlation, partial correlation, distance correlation, and partial distance correlation. These techniques can help determine whether mediation analysis is appropriate by testing for significant nonlinear relationships between variables. Additionally, these methods can be used on their own, allowing users

to explore associations more broadly, all within gBAT's user-friendly interface.

Core features

- Graphical interface for mediation and other brain-behavior analyses
- Automatically aligns brain and behavioral data across diverse naming conventions
- Automatically bins voxelwise data for atlas-based analysis
- Automatically plots volumetric brain images (with regional outlines for atlas-based analyses) to visualize results
- Provides pipeline for automatically projecting results onto fsaverage surface
- Only implementation of partial distance correlation in matlab, providing identical results to Python's dcor 0.6 using a different, recursive, formula: https://pypi.org/project/dcor/

1. Requirements

1.1. System requirements

gBAT runs in MATLAB (2019a and newer) in a Windows, MacOS or Linux environment. However, please be aware that some recent versions of Ubuntu have an incompatibility issue with GUIs in MATLAB: https://www.mathworks.com/matlabcentral/answers/1978579-why-do-some-matlab-features-fail-on-ubuntu-23-04-and-debian-12-with-exit-code-127-in-matlab-r2022b-a. RAM requirements will vary depending on the type of analysis performed, the dimensions of the images analyzed, and the number of individuals analyzed. We recommend at least 16 GB of RAM and caution that distance correlation requires the most RAM, followed by mediation analysis.

1.2. MATLAB toolboxes

gBAT relies on various MATLAB toolboxes depending on the type of analysis you plan to perform. We recommend installing the following toolboxes in order to enable all functionalities:

- Statistics and Machine Learning Toolbox for correlation and partial correlation (corr.m and partialcorr.m), some visualization elements (histfit.m), and bootstrapping *without* bias correction (bootstrp.m)
- Bioinformatics Toolbox for FDR correction (mafdr.m)
- Image Processing Toolbox for loading in and writing NIfTI files (niftiinfo.m, niftiread.m, niftiwrite.m), as well as creating montages of brain volumes (montage.m, imcontour.m), and finding clusters within volumes (bwconncomp.m)

1.3. Additional dependencies*

Certain analyses require external or redistributed toolboxes:

- Mediation Analysis: Mediation Toolbox from Canlab (Must be manually downloaded and added to matlab path. Available at: https://github.com/canlab/MediationToolbox).

- Path Diagrams and Partial Regression Plots: CanlabCore repository (Must be manually downloaded and added to matlab path. Available at: https://github.com/canlab/CanlabCore)

- Bcdistcorr.m (redistributed with gBAT, but can be found at: https://www.mathworks.com/matlabcentral/fileexchange/58445-bias-corrected-distance-correlation)

- Various colormap options for plotting brain data (redistributed with gBAT in ./gBAT/colormaps)

- Various atlas options for binning brain data (redistributed with gBAT in ./gBAT/atlases and usually combining elements that may be missing so please see the atlas section below).

- BrainSurfer if you want to project results onto the surface (Must be manually downloaded and added to matlab path. Available at: https://github.com/alexteghipco/brainSurfer)

*As gBAT incorporates open-source tools, we kindly request users to cite the original authors when using code from these external sources.

Note, the toolbox will remove options that are unavailable to you if you have not installed a certain dependency.

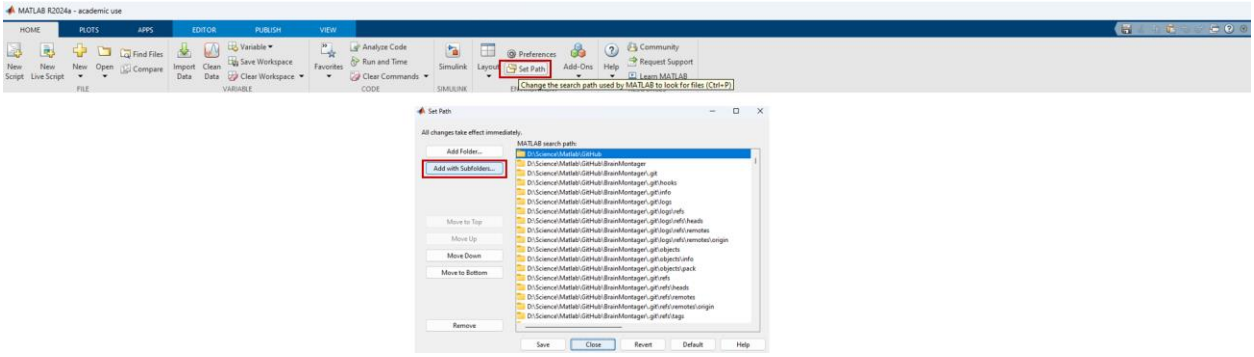
2. Installing the gBAT toolbox

The toolbox can be downloaded from:
https://github.com/alexteghipco/gBAT

After downloading the toolbox, make sure you add the folder (with subfolders) to your MATLAB path. An easy way to do this using the command window in matlab involves using addpath (replace with the file path to gBAT, wherever you placed it):

```
>> addpath(genpath('file/path/to/gBAT'))
```

Alternatively, click the HOME tab at the top of matlab, then find and click "Set Path", which sits next to the "Preferences" button, represented by a gear wheel. Then, click "Add with subfolders …" and navigate to the gBAT folder.

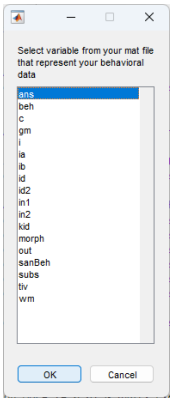


3. Organizing the data

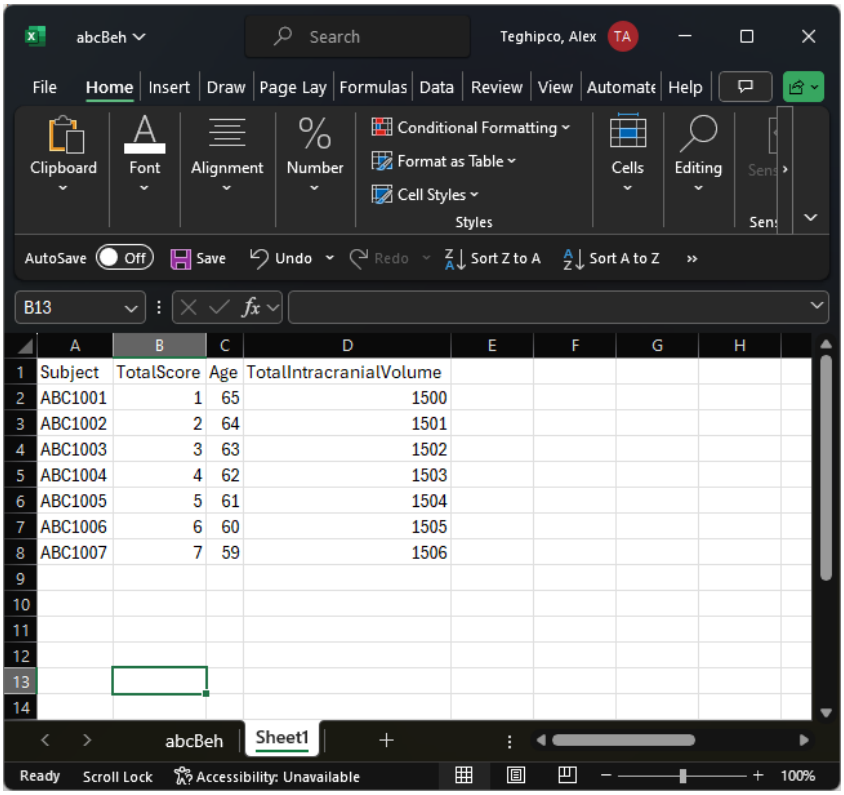
3.1. Behavioral data

You will most likely want to organize your behavioral data in a .csv, .xls or .xlsx format, with each row representing a participant and each column representing a behavioral variable. You can choose not to label your columns, but we recommend against this as it will result in assignment of potentially confusing default names (e.g., x1, x2, x3, etc).

Other file formats are supported as well (e.g., txt, tab, mat, etc). For example, you can load in a .mat file and gBAT will ask—as shown below—which of the potentially many variables contained in that file correspond to the behavioral data you would like to use (i.e., stored as either a matrix or a table variable).

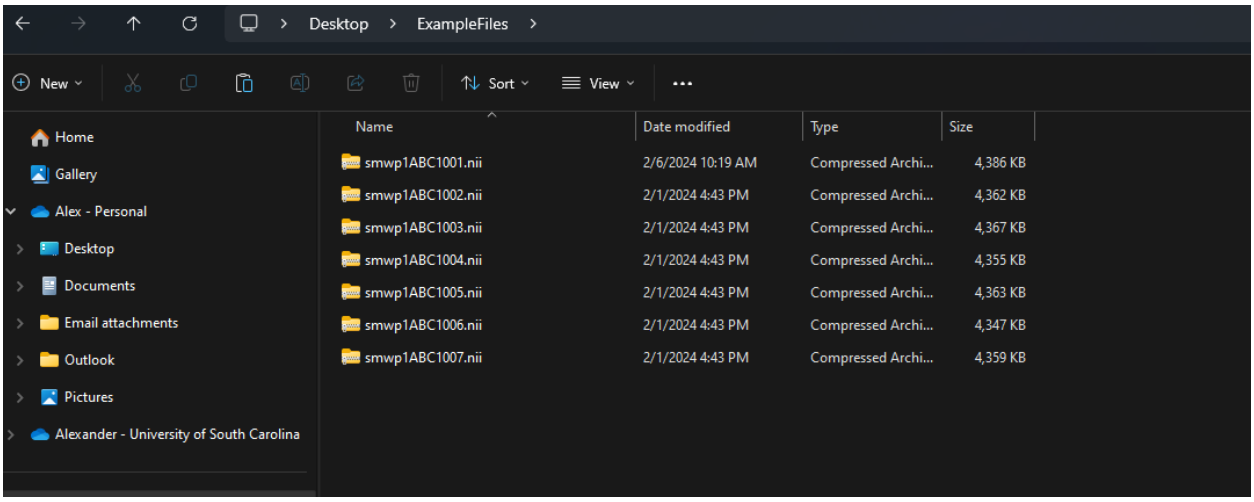


Here is an example of a csv file you might organize with three variables. Let's say we want to know how "total score" is related to "age" through regional brain volumes while controlling for intracranial volume:

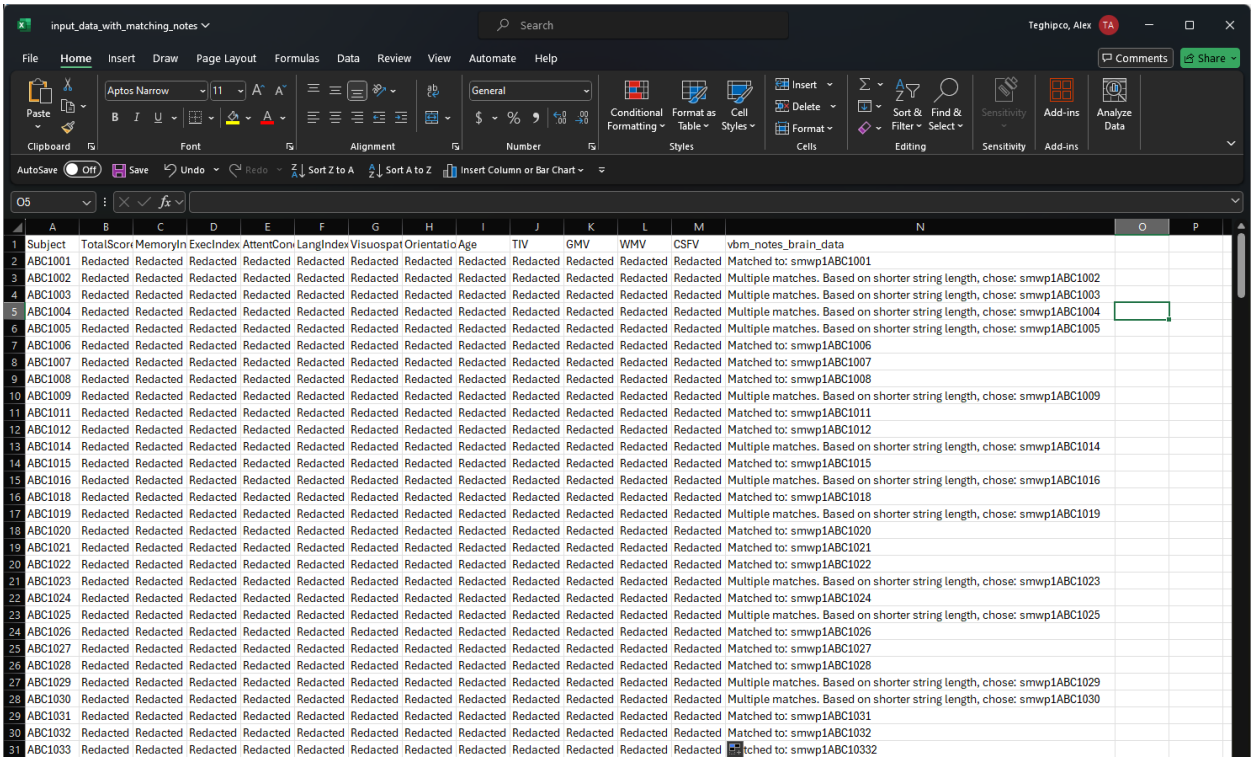


The subject ID in the behavioral data file is expected to be in the first column and must match the subject ID in the corresponding brain file you intend to analyze. While the filenames of the brain files do not need to be identical to the behavioral data entries, the subject IDs should be contained somewhere within the brain filenames, preferably separated by a character. For example, if your subject ID is ABC1003, an ideal corresponding brain file might be named something like: ?_ABC1003_?.nii.gz, ABC1003_?.nii.gz, ?_ABC1003.nii.gz, or ABC1003.nii.gz , where ? can be any other string that does not contain the exact subject ID, and _ can be any character that acts as a separator. If there are multiple matches, the shortest matching string is selected. This ensures that for example, ABC1, ABC11 and ABC111 are matched to the correct filenames ABC1.nii.gz, ABC11.nii.gz and ABC111.nii.gz, even though each of these filenames contains the string ‘ABC1’.

For the behavioral data shown above, here is what our corresponding file names might look like:



It is possible for matching to fail for some naming conventions we have not considered. Before the analysis proceeds, you will be asked to confirm that the matching was successful by checking a file written out by gBAT called input_data_with_matching_notes.csv. It will look something like the screenshot below, concatenating all of your behavioral data with a notes column indicating the chosen match (data redacted):



If there are discrepancies in the matching process, you will have to rename the problematic files before running the analysis (just rename, then restart gBAT and set up your analysis again).

Note, it is okay for some imaging or behavioral data to be missing. We will not analyze subjects without a matching brain-behavior component.

3.2. Brain data

Brain files should be volumetric and preprocessed according to the users' specifications. However, your brain data must be in some aligned space. In other words, gBAT will assume voxel *i,j,k* corresponds to the same voxel across all individuals. Brain files should be in NIFTI format but need not be compressed. Note, gBAT outputs compressed NIFTI files to save space. For more advanced users that are used to analyzing data within matlab, brain files may also be provided as a .mat file. In this case, gBAT will expect and ask which variable in the supplied .mat file stores the brain filenames (1 x *n* cell array of strings) and which variable stores their corresponding vectorized brain maps, contained in an *n* x *p* matrix where *n* is the number of brain filenames and *p* is the number of voxels in the maps. Note, you will also need to provide a 3D matrix that represents the shape of the maps before they were vectorized. The values within this matrix should not matter, we just need to know the shape of your data for writing out NIFTI files. Take care to make sure the vectorized maps are not masked out to remove, for example, non-brain components. You will be able to provide a mask later on during analysis setup.

As detailed above in the behavioral data section, the subject IDs in the brain file names must contain the subject IDs within the behavioral file that is supplied.

Brain data can be any kind of voxelwise neuroimaging data and need not be morphometry-based. Maps could be structural or functional connectivity values to a particular seed region, FA, MD, CBF, BOLD-contrast maps, anything you want provided it does not contain a 4^{th}+ dimension (e.g., time). It may even be lesion data, which combined with a regional analysis, would amount to looking at lesion load (i.e., by averaging binary lesion masks within regions we would be getting regional lesion load). Note, it probably will not make sense to analyze binary data in a voxelwise fashion since most types of analysis that you can set up in gBAT will expect continuous data.

3.3. Atlas and mask files

The emphasis in gBAT is on regional analyses, which require supplying an atlas containing regions of interest. However, it is possible to perform voxelwise analyses as well, in which case an atlas is not required.

For the purposes of regional analyses, an atlas is any NIFTI file that contains whole integers where each unique value corresponds to a different region. That means that your

atlas file can cover the entire brain, or it can contain just one or two regions from a whole brain atlas—say, 1 for inferior frontal gyrus and 2 for superior temporal gyrus.

In regional analyses, the average of your brain data is taken over all voxels within a region. This is done for each subject to create an *n* x *p* matrix of *n* samples with *p* regions.

It is critical that the atlas you choose to use aligns to the same space as the subject-level data you plan to analyze. It is also important to note that if your atlas is in the same space but a different resolution, you will need to resample it to match the shape of your subject-level data before passing it on to gBAT.

Given the proliferation of standardized atlases we provide several options for you to use in ./gBat/atlases. These do not need to be resampled if your subject-level data is in MNI152 2mm space and has the shape: [91, 109, 91]. Before using, please visualize the atlas you intend to use over an individual subject in your dataset to check that they comport. Our atlases come with our own classifications of each region according to tissue type (more on that later).

The same applies to any mask files you supply to gBAT—they should be in the same space as and have the same shape as the subject-level data you plan to analyze.

3.4. Atlas choices

The atlas options distributed with gBAT are detailed below along with the appropriate references for more information (please cite the appropriate paper if using the atlas). Many of the atlases do not contain subcortical regions so we have filled those in using atlases that provide this coverage. Note, these can be removed by deleting their corresponding rows from the text files before starting an analysis with multiple sets of brain maps, or if using a single set of brain maps, by using the ‘Make new atlas’ option in the postprocessing menu of gBAT (see make new atlas section below).

Each atlas comes with a corresponding text file that describes the integer associated with each region, the region’s name if available, and our characterization of which tissue type it belongs to. The name-region mappings may be helpful for analyzing outputs from gBAT, which will often but not always be based on the integer assigned to each region (more on that below).

Tissue classification is based on maximal overlap between each region and binary tissue maps segmented with FSL’s FAST for the MNI152_2mm brain (contained in ./tissue). For example, if a region has 80% overlap (in voxels) with the white matter segmentation and 20% with grey matter segmentation, it is classified as belonging to white matter. Within

these text files, 1 corresponds to grey matter, 2 corresponds to white matter, and 3 corresponds to CSF.

These text files are required if your analysis will need to combine two different types of data. Using a hidden (right-click) menu detailed below, it is possible to supply a text file such as this, which will also reveal an option for uploading a second set of brain files. ROIs tagged as 1s and 2s will be separately associated with the first set and second set of brain files you supply. Don't worry, you will be able to assign whether 1s or 2s should be associated with the first or second batch of brain files, so the class labels are flexible. However, regions classified as belonging into any other group (e.g., 3 for CSF) will not be analyzed. While gBAT assumes you may want different data for grey and white matter specifically, in practice, you can define 1s and 2s in your text file to mean anything you want in terms of classifying the regions in your atlas. The point is that different data will get pulled for the two groups. This can be useful. For example, white and grey matter volume may be contained in different maps, and voxels may overlap between the two maps. This option would streamline estimating white matter volume only for white matter regions and grey matter volume only for grey matter regions. You could also use this method to combine different modalities—say simultaneously analyzing tract-based statistics in white matter and functional connectivity in grey matter (e.g., for a particular seed).

The text files we provide can serve as a template for setting up your own custom atlases. Note, it is possible to run a truncated regional analysis using only select ROIs (Regions of Interest) contained within a whole brain atlas by removing rows or elements of the text files that contain ROI names you wish to exclude from analysis. This may be easier for some users to do than removing all of those regions directly from the atlas file.

Atlas Options

See atlas_scratch_work.m in ./atlases for more details on how we created the atlases. This code may help you create custom atlases. For all atlases, tissue classification was based on strongest overlap with white and grey matter segmentations (partial volume) of the MNI152_2mm brain as obtained with FSL's FAST (see aforementioned code and ./tissue for maps). However, you may want to check these text files and adjust them for your purposes.

- Harvard-oxford: macronatomical regions defined by overlap across subject-level segmentations: Makris N, Goldstein JM, Kennedy D, Hodge SM, Caviness VS, Faraone SV, Tsuang MT, Seidman LJ. Decreased volume of left and total anterior insular lobule in schizophrenia. Schizophr Res. 2006 Apr;83(2-3):155-71; Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale

AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage. 2006 Jul 1;31(3):968-80.

- We combine the subcortical and cortical Harvard-oxford atlases (see ho.nii.gz for combined and hocorticalOnly.nii.gz and hosubcorticalOnly.nii.gz for the components that go into the combined atlas). Note, cortical regions are given preference, so only voxels missing from the cortical atlas are filled in with white matter voxels. We use the 20% thresholded version of the cortical atlas (i.e., each probabilistic region map is thresholded at 20%; see: https://neurovault.org/collections/262/). We also split left and right cortical regions manually.
- This atlas does not cover the cerebellum.
- You may want to remove the left and right Cerebral Cortex ROIs that are from the subcortical version of this atlas. The addition of cortical regions results in these voxels mapping onto CSF in the outer periphery of the brain.
- A projected version of the atlas onto fsaverage surface is provided for you without the subcortical regions (see ?_LH.nii.gz and ?_RH.nii.gz files)
- See ./atlases/ HO
- **Choose this atlas if you want to focus on intuitively defined macroanatomical regions and do not care very much about white matter structures.**

- JHU: an atlas combining cortical regions with white matter tracts. The cortical regions have fairly close alignment with the Harvard-Oxford atlas. See: Faria, A. V., Joel, S. E., Zhang, Y., Oishi, K., Van Zjil, P. C. M., Miller, M. I., Pekar, J. J., & Mori, S. (2012). Atlas-based analysis of resting-state functional connectivity: Evaluation for reproducibility and multi-modal anatomy–function correlation studies. *NeuroImage*, *61*(3), 613–621. https://doi.org/10.1016/j.neuroimage.2012.03.078
  - This atlas does not fully cover the cerebellum.
  - This is a version of the atlas from niiStat but instead of just resampling, this is registered directly to the MNI152_2mm brain for better alignment.
  - A projected version of the atlas onto fsaverage surface is provided for you without the subcortical regions (see ?_LH.nii.gz and ?_RH.nii.gz files)
  - See ./atlases/ JHU
  - **Choose this atlas if you want to focus on intuitively defined macroanatomical regions and want to have more information about white matter structures.**

- AICHA: homotopic functional connectivity atlas based on resting-state data: Joliot, M., Jobard, G., Naveau, M., Delcroix, N., Petit, L., Zago, L., ... & Tzourio-Mazoyer, N. (2015). AICHA: An atlas of intrinsic connectivity of homotopic areas. *Journal of neuroscience methods*, *254*, 46-59.
    - This atlas excludes white matter, but we will have a version of this available with some subcortical regions in the future (from the JHU atlas).
    - See ./atlases/ AICHA
    - **Choose this atlas if you want to focus on more intuitively defined functional regions, and do not care about the dimensionality of the atlas (i.e., regions are defined by having similar connectivity patterns)**

- Multiresolution local-global parcellation (Schaefer atlas) that is based on resting state data but provides atlases of varying complexity (i.e., number of parcels): Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X. N., Holmes, A. J., ... & Yeo, B. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral cortex*, *28*(9), 3095-3114.
    - This atlas is provided at multiple resolutions. That is, usually the dimensionality of a functional atlas is determined by model-order selection of some data-driven algorithm. Selecting the right “dimensionality” or “complexity” or “resolution” is an art and usually relies on heuristics. This atlas is provided at multiple resolutions, making it ideally suited if you want to experiment with atlas dimensionality.
    - We use a version of this atlas from the Lesion Quantification Toolkit, which adds subcortical/cerebellar regions from the AAL atlas. As such, if using this atlas, please cite the additional works:
        - Griffis, J. C., Metcalf, N. V., Corbetta, M., & Shulman, G. L. (2021). Lesion Quantification Toolkit: A MATLAB software tool for estimating grey matter damage and white matter disconnections in patients with focal brain lesions. *NeuroImage: Clinical*, *30*, 102639.
        - Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*(1), 273-289.
    - Fsaverage projections for all resolutions are provided in ./atlases/ multires\fsaverage_projections
    - **Choose this atlas if you want to focus on more complex functional regions (i.e., regions defined by having similar connectivity patterns as most functional atlases, but also showing local areal transition**

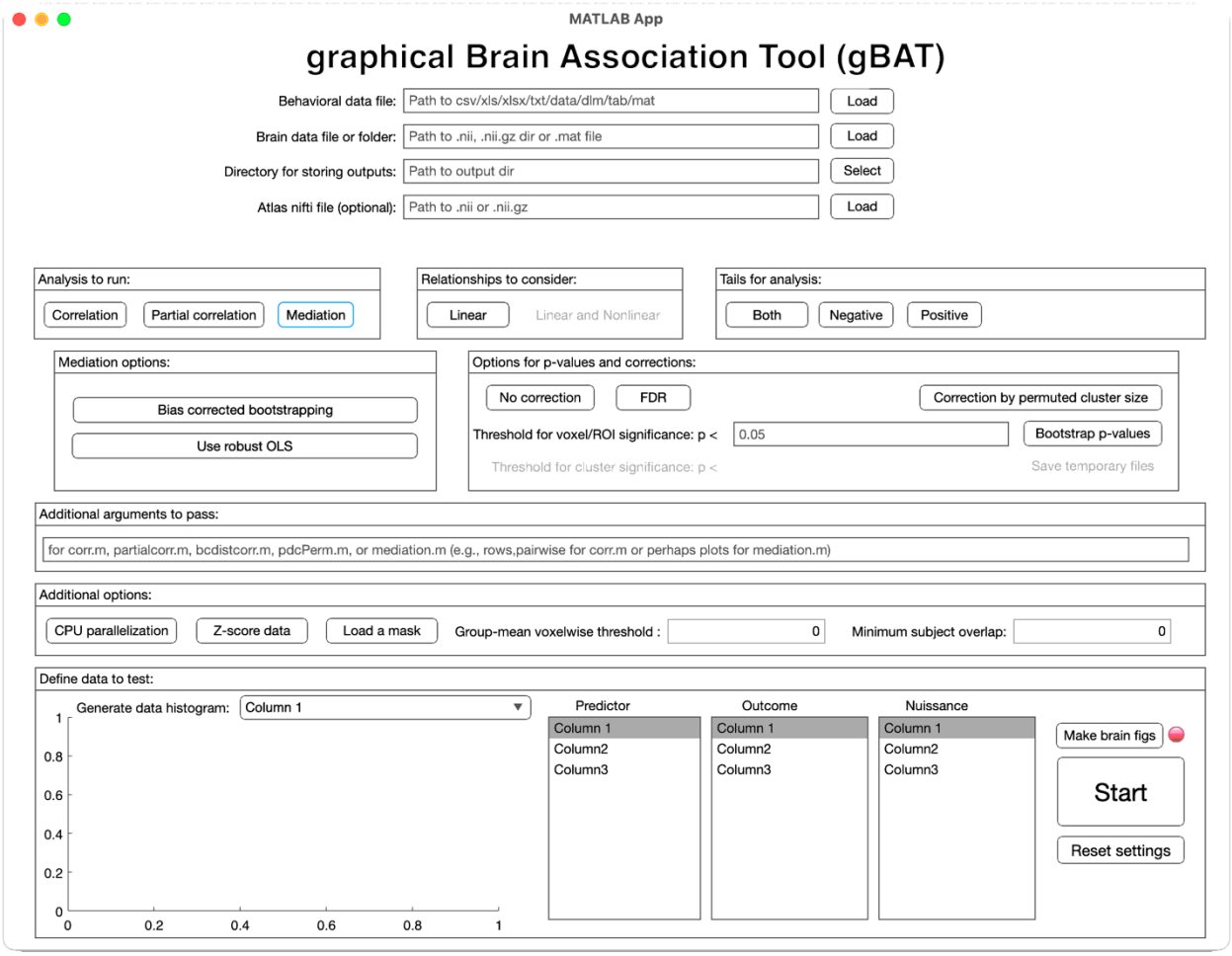
gradients a la human connectome project's atlas mapping efforts), or if you want to experiment with different atlas complexities (i.e., number of parcels).

- Julich-Brain histological atlas based on microstructural properties: Amunts, K., Mohlberg, H., Bludau, S., & Zilles, K. (2020). Julich-Brain: A 3D probabilistic atlas of the human brain's cytoarchitecture. *Science*, *369*(6506), 988-992.
    - Histological atlas with "gap maps" (unmapped areas)
    - Only in grey matter (cortical ribbon)
    - We have resampled the data to 2mm MNI152 and combined the left and right atlas files that are provided into one NIFTI
    - Projections are provided for convenience, but will require relabeling (update forthcoming)
    - **Choose this atlas if you want to focus on more granular regions histological regions.**

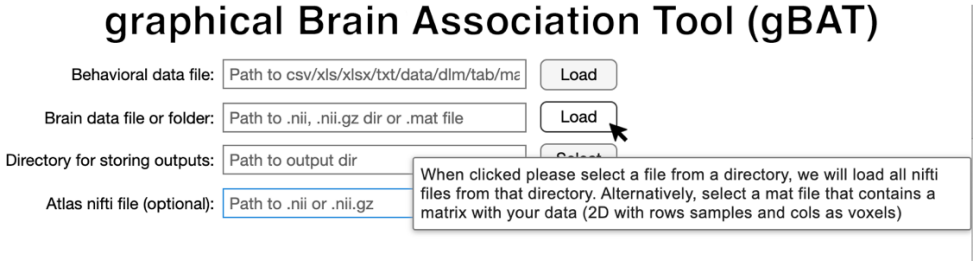
- XTRACT Human Connectome Project probabilistic tractography atlas: Warrington, S., Bryant, K. L., Khrapitchev, A. A., Sallet, J., Charquero-Ballester, M., Douaud, G., ... & Sotiropoulos, S. N. (2020). XTRACT-Standardised protocols for automated tractography in the human and macaque brain. *Neuroimage*, *217*, 116923.
    - This atlas contains white matter tracts defined over a very large number of individuals (1000 from the human connectome project) and ex vivo macaque brains, and validated in another large number of individuals (UK biobank)
    - See ./atlases/ XTRACT
    - **Choose this atlas if your primary interest is in white matter tracts.**
    - **You cannot project this atlas onto the surface.**

4. Using the toolbox

Once the toolbox is installed and added to the path, open MATLAB and navigate to the folder where you downloaded gBAT. Type "gBATGui" into the command window. This will open the GUI, which looks like this:

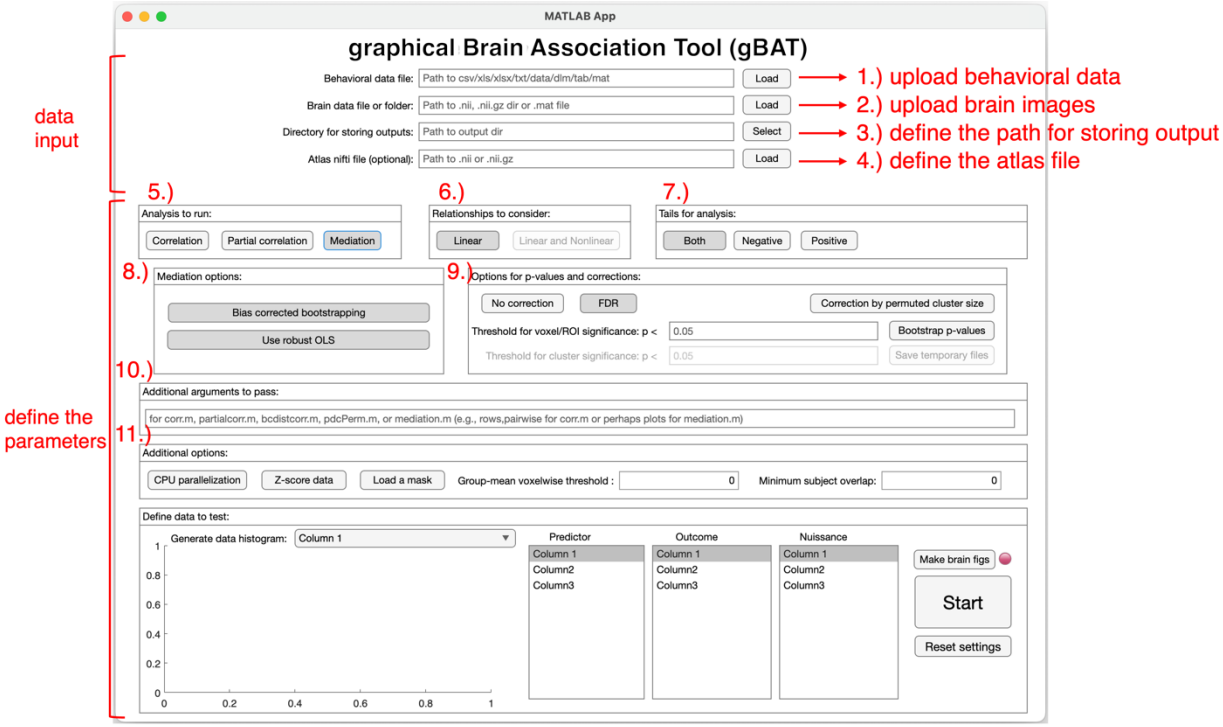


The gBAT GUI provides an intuitive interface for conducting analyses. To learn about each feature, simply hover your cursor over any button in the GUI. A tooltip will appear, providing detailed information about the function of that element, instructions for correct usage, and any specific requirements or options. See example below:



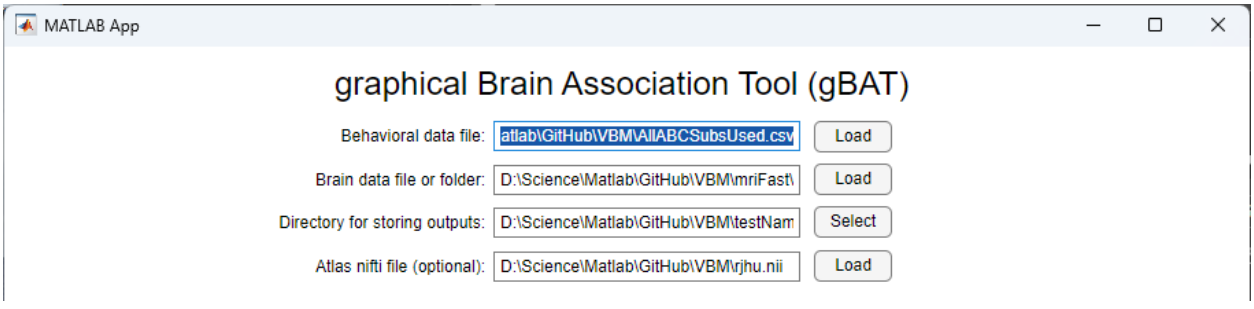
4.1. Graphical User Interface GUI

How to use the GUI – overview of the buttons:



The top panel of the GUI allows you to specify all the necessary data for analysis. In each file type that you can import or define, you have two options for supplying inputs. You may use the load/select buttons, or you can paste in a direct filepath or a filepath with a wildcard that catch multiple files with a similar naming convention.

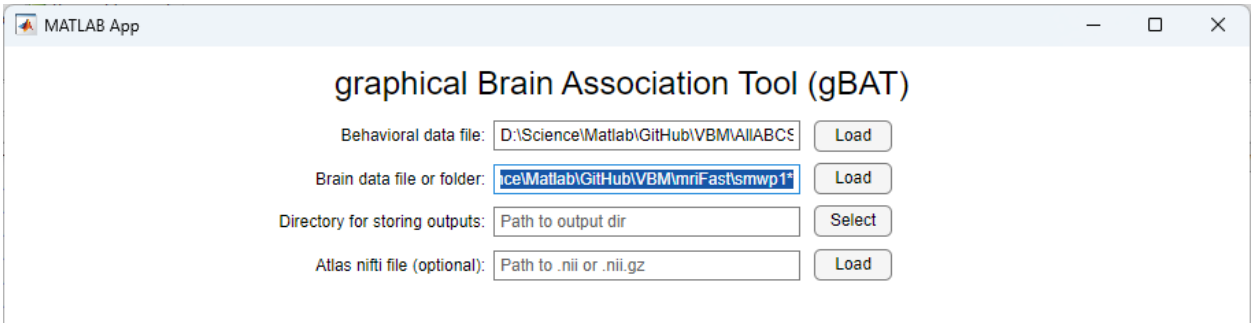
1.) Behavioral data file: Accepted formats include .csv, .xls, .xlsx, .mat, or any other file type compatible with *readtable.m*. Using the load button will prompt you to navigate and select the appropriate file. Alternatively, you can supply a full file path directly into the text box. For example: D:\Science\Matlab\GitHub\VBM\AllABCSubsUsed.csv as shown below:



Supplying a .mat file will prompt for additional information (i.e., which variable stored in the .mat file contains the behavioral data).

2.) Brain data file or folder: Import brain images (NIFTI files). If using the load button, you will be asked to select a single file from a directory containing all the files you would

like to load. That means your directory should contain only the NIFTI files you expect to analyze. Alternatively, you can use a wildcard if the directory contains different sets of NIFTI files that are sensibly separated. For example, I have a directory with two sets of files on each individual. One set that starts with the string smwp1 and one that starts with smwp2. If I only want to analyze the files that start with smwp1, I can paste this regular expression into the text box like so: D:\Science\Matlab\GitHub\VBM\mriFast\smwp1*

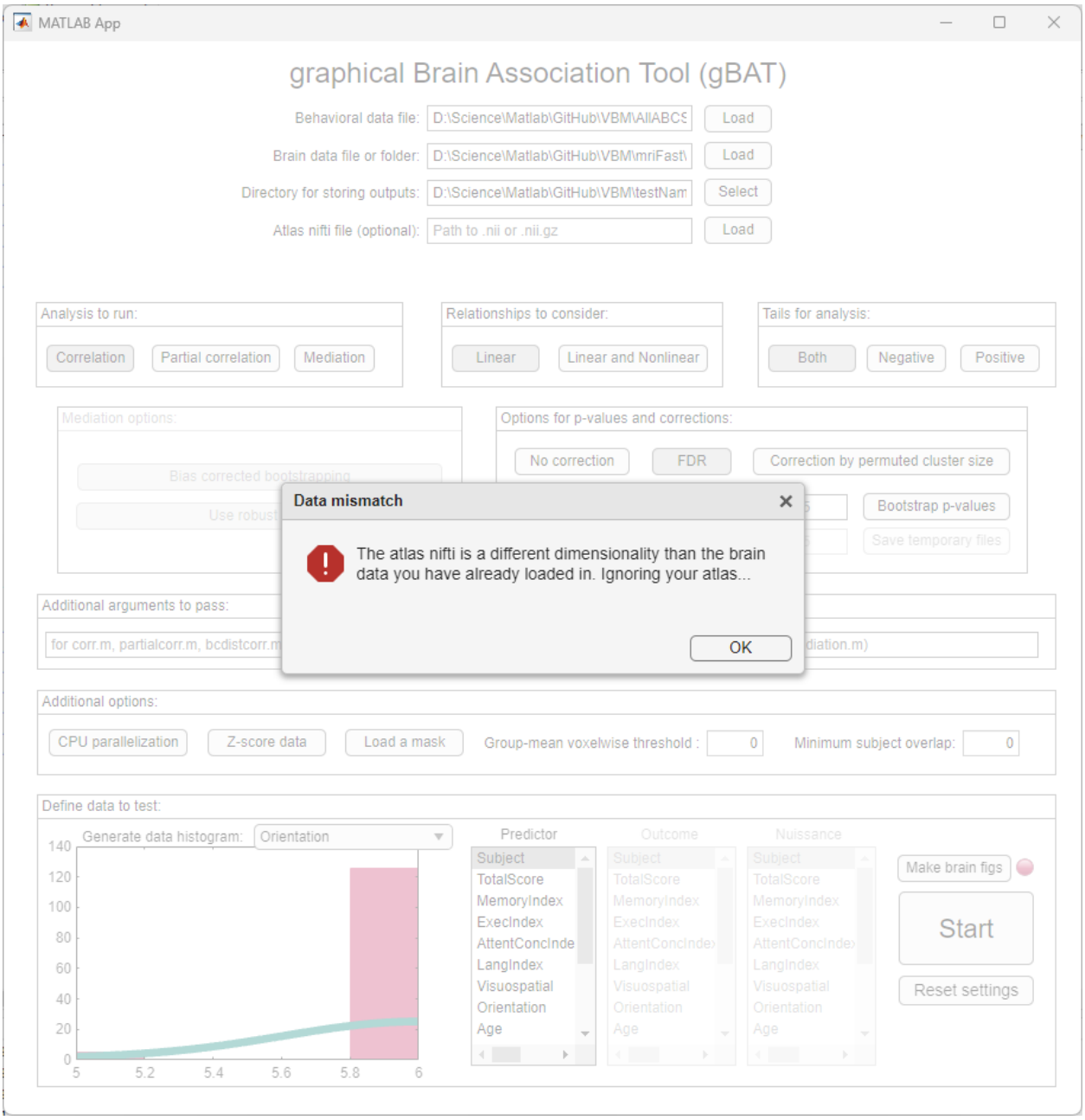


You do not need to worry about the file extension.

Alternatively, you can select a .mat file containing a 2D matrix with samples as rows and voxels as columns. These are vectorized brain maps that must not be masked, and you will be asked to supply the variable names for different data that will be required for this route of loading data. For example, we will also need the shape of the original data as reflected in any 3D matrix stored in the .mat file (see how to organize brain data above).

3.) Directory for storing outputs: Define the directory where you want your results to be saved. You can either select a directory using the button, or type out the path to a directory you would like to use (if it does not yet exist, it will be created). The GUI will warn you if you will be overwriting an existing directory.

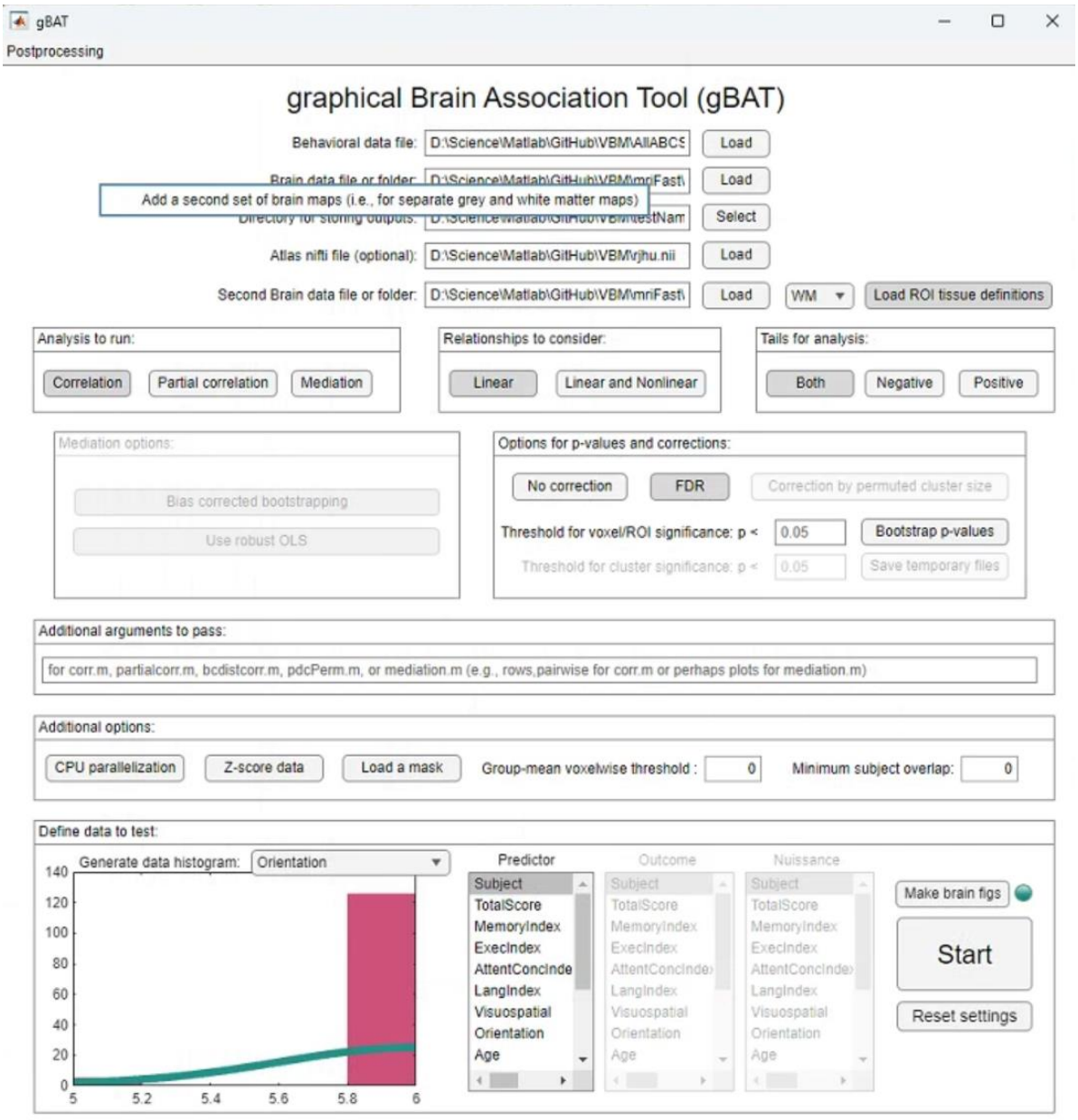
4.) Atlas nifti file (optional): For ROI-based analyses, users can select a brain atlas file (see atlas organization section above). Again, this can be selected directly using the load button or a filepath can be pasted in. If the atlas does not match the shape of the brain data, you will be given an error and the atlas file will not be loaded as shown below:



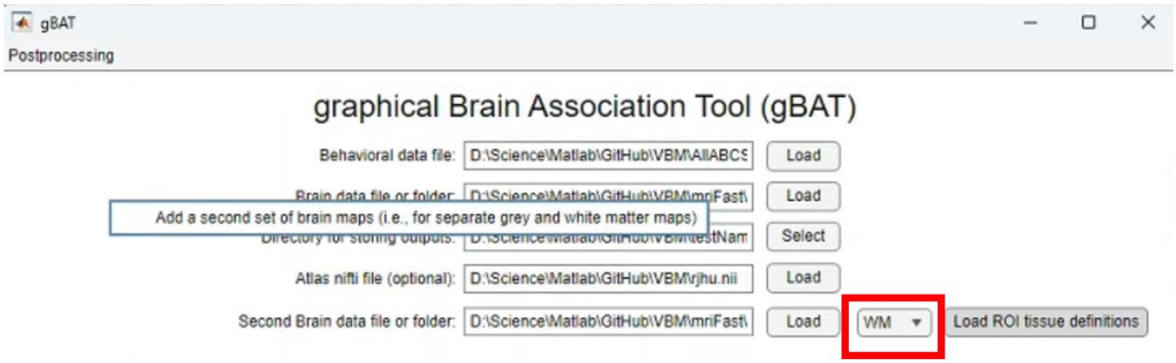
A hidden contextual menu for loading multimodal data during regional analysis

Right clicking on any of the fields associated with brain data will allow you to import a second set of brain images. The purpose of this is to use different data for ROIs classified to two different groups (see atlas organization section above). Ultimately, the most straightforward way we can imagine users relying on this is to pull separate data for grey matter and white matter when generating regional data for analysis. Regional classifications are supplied in a text file accompanying the atlas, and indicating which class each region belongs to (see examples in: ./atlases). When a regional analysis is performed, voxels within different sets of brain maps can be averaged depending on the

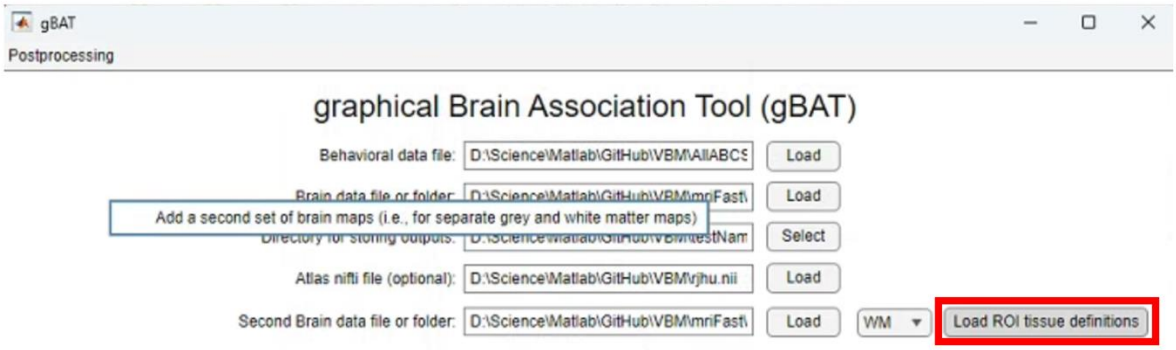
region class. We assume the class ‘1’ corresponds to Grey Matter ,which maps onto the primary set of brain files imported, and the class ‘2’ to White Matter, corresponding to a secondary set of brain files. However, this can be flipped by toggling the menu to the right of the load button for the second brain data folder. The button further right of that is used to load the text file that classifies regions in the atlas (again, see atlas organization section for additional information).



As shown below, toggling the radio button in red will switch which class label (i.e., ‘1’ or ‘2’) is associated with the second set of brain files. Here, ‘1’ in the atlas text files is interpreted as ‘GM’ or grey matter and ‘2’ in the atlas text files as ‘WM’ or white matter.



The button to the right of this toggle can be clicked to add and clicked again to remove the atlas text file that specifies label.



Once the data is loaded, parameters can be defined in the bottom panel of the GUI.

5.) Analysis to run: Choose between correlation (correlating brain data with a predictor from the behavioral data), partial correlation (correlation including one or more covariates), or mediation (relating the predictor to the outcome variable using regional or voxelwise brain data as a mediator).

6.) Next, define whether only linear or both linear and nonlinear relationships should be considered. If detecting linear relationships, by default, the Pearson correlation is used. However, you can change this by supplying a different argument into the “additional arguments to pass” section detailed below. For detecting linear and nonlinear relationships, a distance correlation will be used. This requires computing Euclidean distances between all samples, so you may run out of RAM if your dataset is large enough.

7.) Specify, whether the analysis should be one-tailed or two-tailed. If you chose mediation analysis, you can specify if bias corrected bootstrapping should be performed to test the mediation models’ significance. Hover over this button for more information as there are some conditions under which we cannot do one-tailed testing.

8.) Mediation options: choose between bias corrected bootstrapping (using mediation.m) and robust ordinary least square (OLS) regression. Bias corrected bootstrapping is a standard for mediation analysis. Robust regression can help address outliers in the data, making the results more robust. Without bootstrapping, the sobel test is used to estimate p-values. Note, however, that you can untick the “Bias corrected bootstrapping”, which is only available for mediation analysis, and enable “Bootstrap p-values” in the p-values and corrections section (section 9) to get bootstrapped mediation p-values without bias correction. This may be preferable, as some recent work suggests bias correction may inflate false positives: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.810258/full

9.) Define options for p-values and corrections. We recommend choosing FDR. Set the threshold for voxel/ROI significance (default: $p < 0.05$), and specify whether p-values should be bootstrapped. Choosing “Bootstrap p-values” will initiate the bootstrapping process and will take much longer to run, as well as require much more RAM. This bootstrapping option can be applied to any analysis and is not bias corrected.

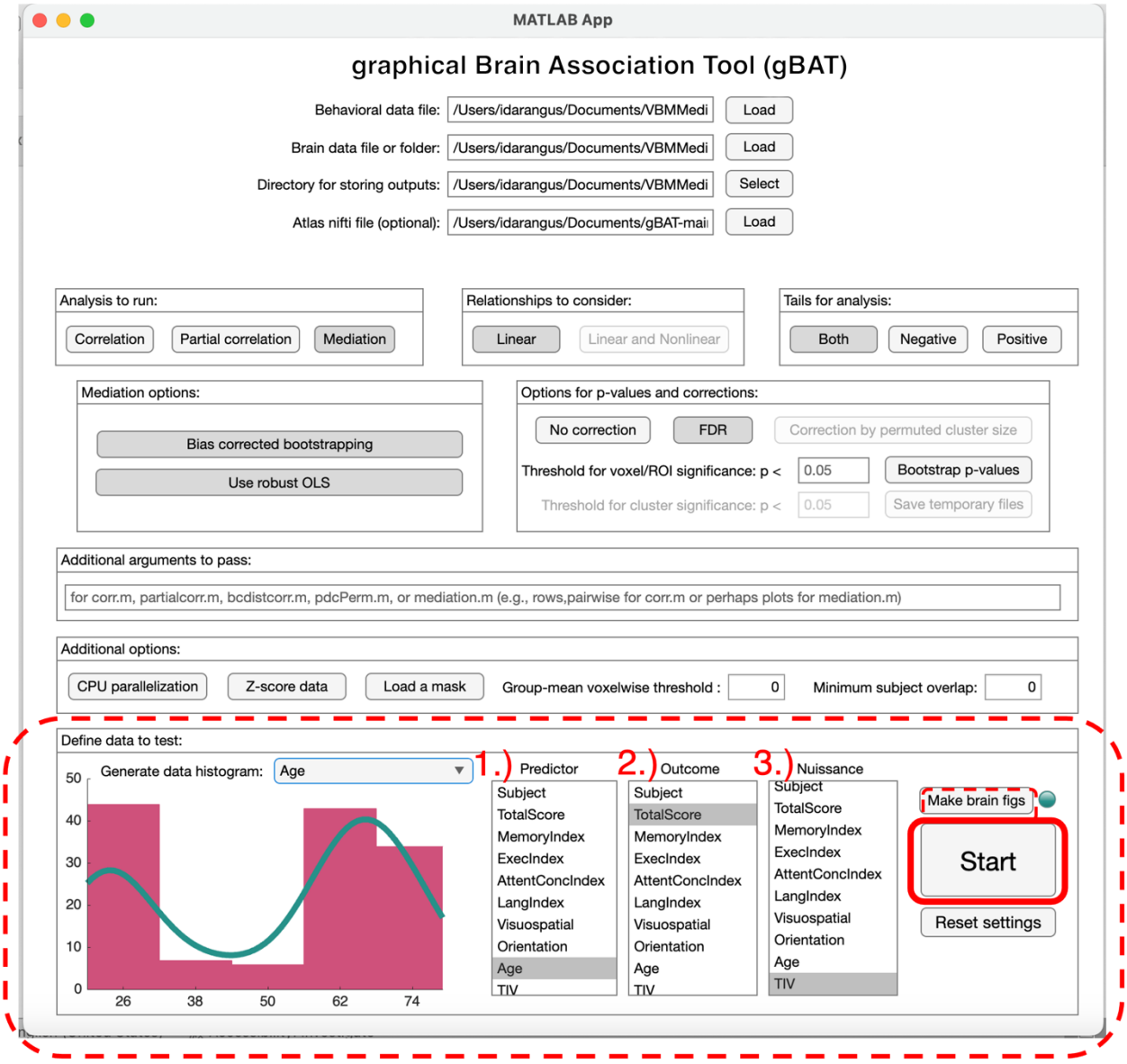
If you are performing a voxelwise analysis, you will have the additional option to perform cluster correction through permutation testing, where we will permute the data, rerun our analysis and check the maximum cluster size that is retrieved under the null distribution (using bwconncomp.m). Note, this will take a very long time and will require a lot more RAM.

You may pass additional arguments into **(10.)** e.g. plots for mediation.m (see tooltip). This essentially passes any string you supply onto whatever function is performing the analysis you have selected (see functions mentioned above). As another example, if using linear correlation, you can specify the type of correlation (e.g., by inputting: Type, Spearman).

See the underlying functions used for the analysis for full set of options that you may use (e.g., corr.m).

As a special note, you can pass in plots to mediation.m and gBAT will create a folder within your **working directory** called temporaryFolderForMediationFigs that will contain mediation path schematics and residual scatter plots.

Additional options **(11.)** include CPU parallelization (this leverages all CPU cores, but may actually slow processing for high-dimensional data), Z-scoring (normalizing the variables before analyzing them by z-scoring them with normalize.m) and loading a mask (restricts analysis to non-zero voxels in a specified nifti file). You can also define a Group-mean voxelwise threshold, which excludes voxels exceeding a specified average value across subjects; in ROI-based analysis, this affects which voxels are averaged within ROIs. Minimum subject overlap removes voxels lacking a specified minimum number of overlapping subjects, applied before ROI averaging in ROI-based analyses.

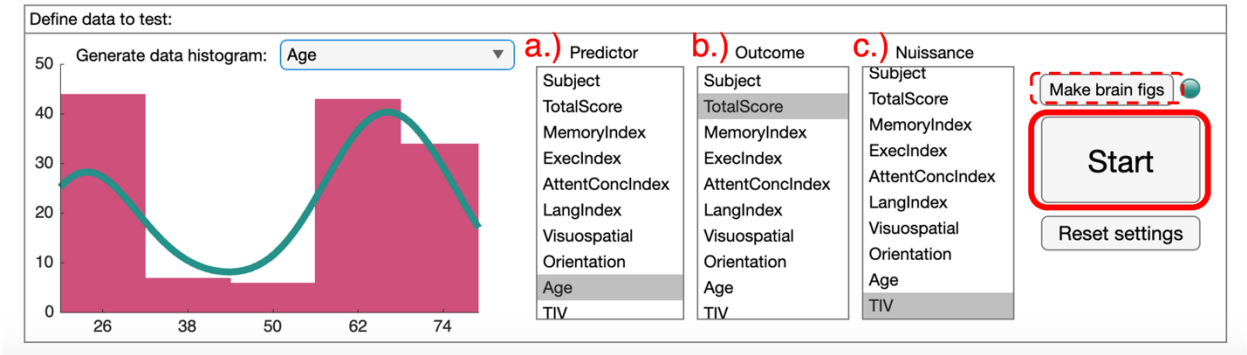


After the data has been loaded and the settings have been adjusted, the behavioral variables that will be used in the analysis can be specified in the final panel. The predictor variable **(a.)** and the outcome variable **(b.)** are specified here. Optionally, nuisance variable **(c.)** can be specified. Here, we control for total intracranial volume (TIV). The distribution of the data can be checked on the left bottom side, you can choose different variables from the drop-down menu.

Note, for many analyses you will be able to select variables in order to loop the analysis over them (e.g., selecting multiple outcomes above). Note, some of these listboxes will not be available to you depending on the analysis. For example, only predictors are available for other analyses, which will associate brain and predictor data. Nuisance

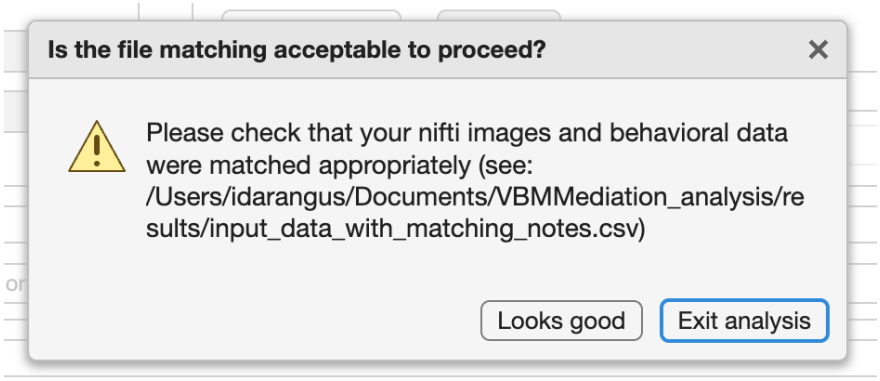
variable options will only be available when using partial correlation, partial distance correlation, or mediation analysis.

After all data has been loaded and the settings have been adjusted as needed, you can press “Start” to begin running the analyses. If you would like to make brain figures, you can click on the button above the “Start” button:



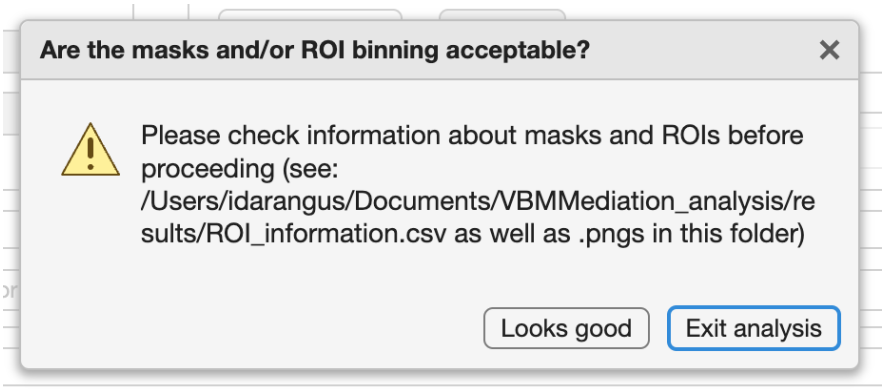
By the way, the teal circle above the start button indicates this is a regional analysis. If this circle was red, it would indicate that the analysis was setup as a voxelwise analysis.

After clicking the “Start” button, the following window will appear:



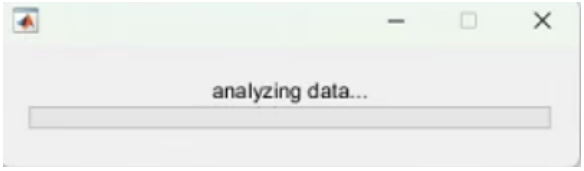
This means that the brain files and behavioral datasets were matched according to the subject ID and saved in the folder you specified previously (the path to the file is shown in this window). Navigate to the folder and open this file. Check whether the matching is accurate (i.e. whether the name/subject number of behavioral data was assigned to the correct brain file; see last column of the spreadsheet). If the matching is accurate, click on “Looks good”. If the matching was unsuccessful, click “Exit analysis”. In the latter case, double-check for ambiguous names and consider renaming the files to resolve any issues.

After clicking “Looks good”, this warning should appear:



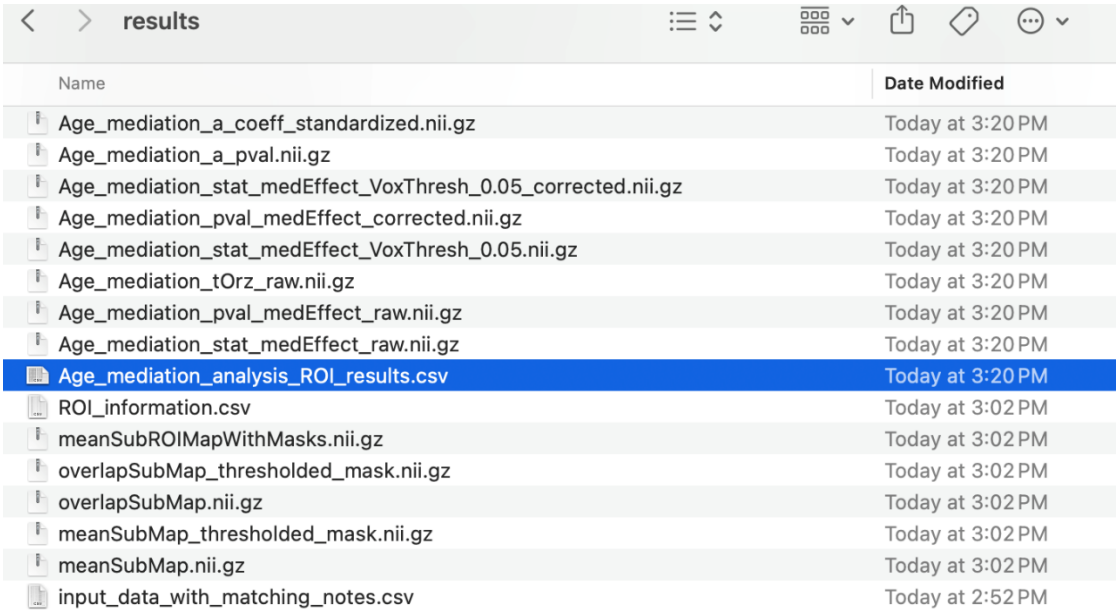
Head to the specified folder and review the file showing mean values across subjects for each ROI (this will only be available for regional analyses). Verify that the data appears correct. If satisfied with the results, click “Looks good”.

Once the analysis starts, this window will appear:



You can monitor the progress via the advancing bar. This can take a while depending on your analysis and hardware.

Upon completion, output files will be generated in your prespecified folder. A summary of results is available in the ‘*_results.csv’ file.



For each ROI corresponding to the atlas file's numbering, the analysis results are listed, including the effect of each path and associated p-values. Here is what this looks like for mediation analysis, where we get all of the path coefficients or parameters for each ROI:

ROI index	indirect path coeff (ab)	path a coeff	path b coeff	path c' coeff	path c coeff
1	-0.016846100817369	-0.000532659553691959	31.626393820604	-0.0399075944017199	-0.0555069219648755
2	-0.0171347935810345	-0.000549489502479917	31.183113605816	-0.0392804170794123	-0.0555069219648755
3	-0.018587824730949	-0.000313105562076143	59.3659997851737	-0.038552047826809	-0.0555069219648755
4	-0.0185486663287041	-0.000334541220658283	55.4450847408448	-0.0380854678264521	-0.0555069219648755
5	-0.0157772908996551	-0.000349352696279417	45.161497442791	-0.0410192443223636	-0.0555069219648755
6	-0.0157128054306932	-0.00031503779464195	49.8759377380459	-0.0412765191695527	-0.0555069219648755

As you can also see, other files will be available to you. These are very straightforward for correlation and distance correlation-based analyses, where you will see a 'stat' image showing raw coefficients (e.g., Age_person_stat.nii.gz if age was the main variable you were correlating), a corresponding image with raw p-values (?_pval.nii.gz), and a thresholded image based on your p-value threshold (?_stat_VoxThresh?). There may also be a thresholded image based on corrected p-values if FDR correction was selected. This would also mean that you would see a corrected p-value map that is untresholded (e.g., pval_?_corrected.nii.gz).

For mediation analysis, more images will be available. The main ab indirect effect path is shown in the ?_stat_medEffect_? images. However, you will also see ?_a_coeff_?, ?_b_coeff_?, ?_c_coeff_?, ?_c'_coeff_? Images showing the other paths in the mediation

model. All of these will be provided raw, but also thresholded. These will also have corresponding p-values and corrected p-values. Note, the tOrz map will either be t-statistics or z-statistics depending on whether you are bootstrapping (see mediation.m).

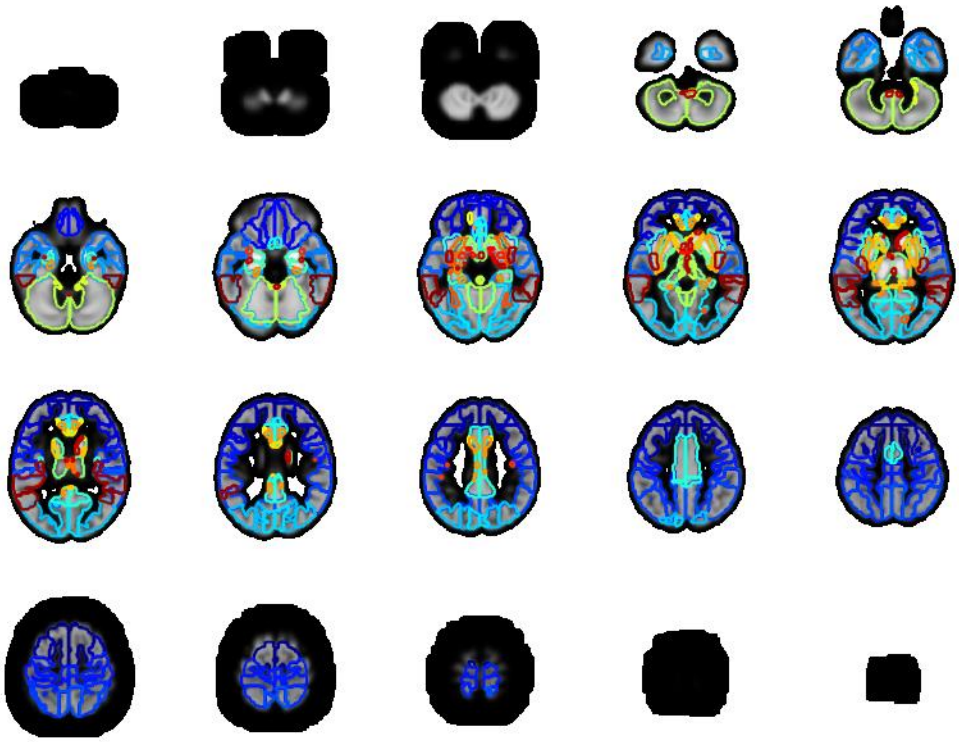
Finally, you will see some intermediary images showing masks. For example, the mean of all your images after being masked by any mask you supply, and/or any threshold you chose for the mean image is shown in meanSub?WithMasks.nii.gz. A version of this before applying any user-input mask is shown in meanSubMap.nii.gz. A map showing voxels that overlap among all input brain images is shown in overalpSubMap.nii.gz. Note, you will see two mean images of your inputs if you chose to supply two different sets of brain images for analysis (these will be appended WM and GM to correspond to the regional labels 1 and 2).

Generating images

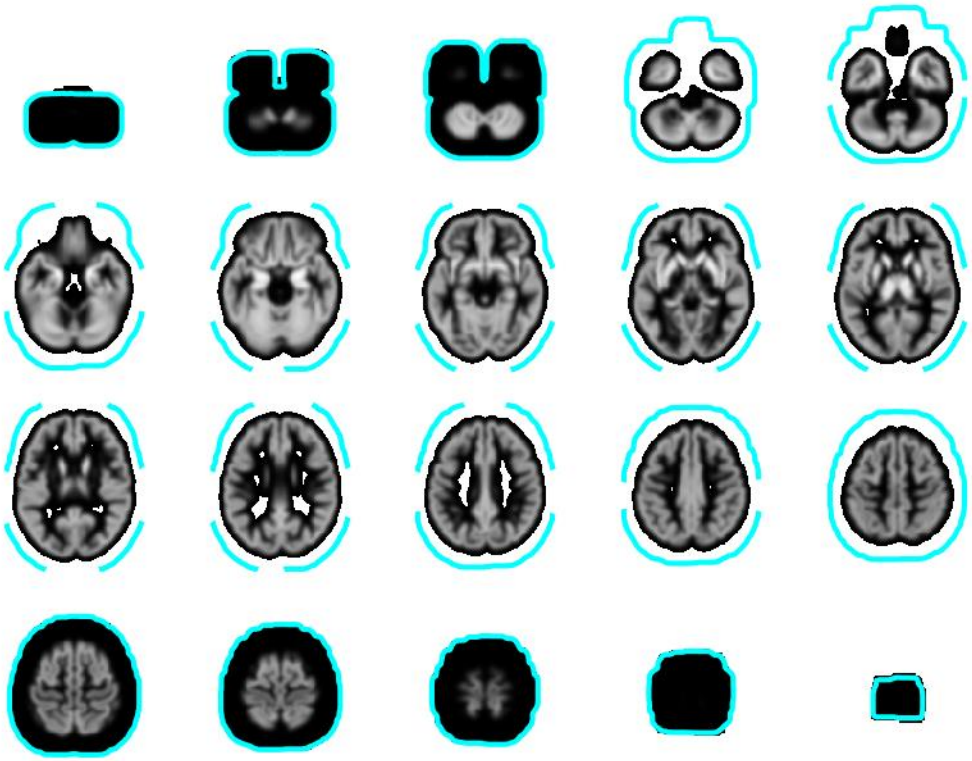
By clicking on “Make brain figs” in the bottom GUI panel, gBAT will automatically create visualizations of the results. This will take more time for the analyses to finish. The time it takes to generate these is so long because we plot each regional contour independently. This will take less time in the future, but it’s the way gBAT works for now.

Several images will be generated automatically. In all of them, an average of your analyzed maps (after applying any chosen threshold to the mean maps) will be used as the background or underlay. The overlay will involve various masks and regions.

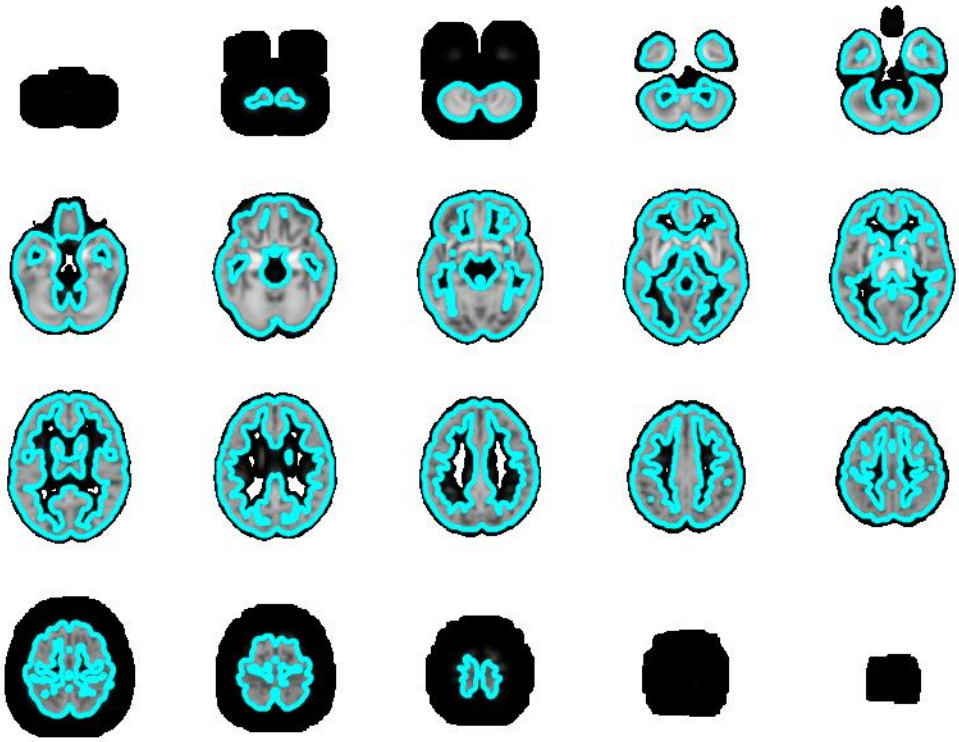
Images will include an overlay of a chosen atlas on the mean of all of the input images (meanSubROIMapWithMasks.png). This provides a quick visualization of your atlas with your data.



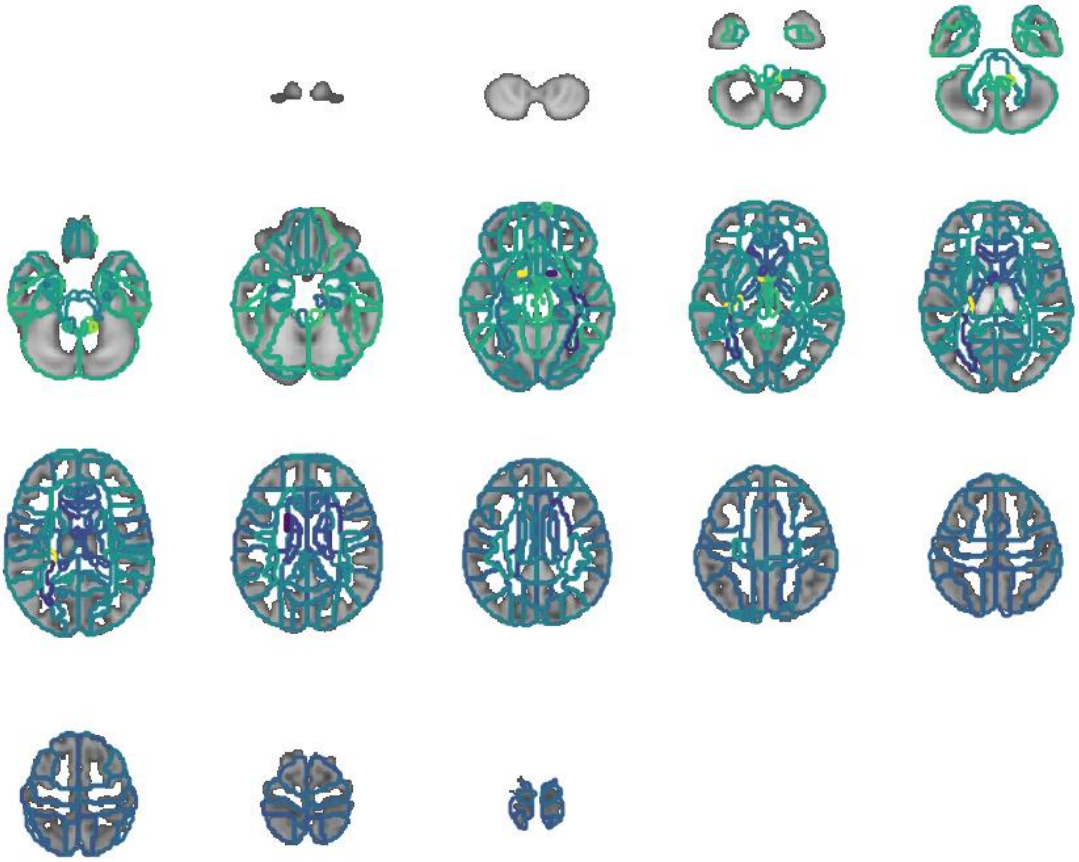
Images will also include the voxels that overlap across all input images with the mean image in the background (overlapSubMap.png). The overlap map is an outline. As you can see below, it is possible for overlap to be so extensive it is outside of the ‘view’ defined by thresholding the average map.



Another image you will see is the actual threshold you applied to the average image (meanSub.png). Here, you can see our threshold nicely removes areas of low overlap (ar3eas outside the teal outlines).



Finally, you will see thresholded statistics (if these options were enabled during analysis). Only for this case, we also save out the matlab figure so you can double click the corresponding fig file to see the montage and zoom in if needed. Here, the underlay is the thresholded average image (teal outline above) and the outlines correspond to atlas regions, with the color mapping onto the effect strength.



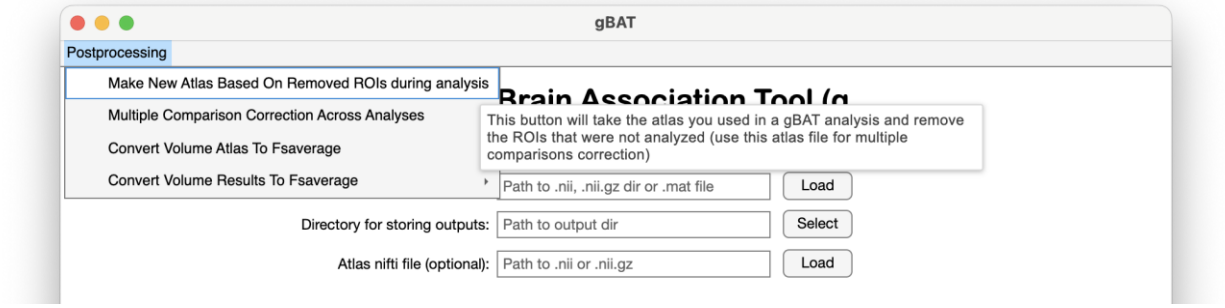
By default brighter colors will mean higher statistics but you can change this. If you would like to generate similar montages outside of gBAT or tinker with the settings for visualizing the results that have been output by gBAT, please view the live code tutorial (postprocessing_examples.mlx).

In case you would like to apply a new p-value threshold, you can use the files in your output folder to generate new maps. This can be done using ./postpro (again see postprocessing_examples.mlx). If you did not select FDR correction during analysis, you can still apply FDR or Bonferroni correction to maps using the postprocessing menu in gBAT (see multiple comparisons option below).

Visualizing atlases with regions excluded from analysis

If you perform a region-based analysis it is possible that you selected options that removed some ROIs from analysis (e.g., a mask, a threshold, etc). To visualize the resulting atlas with the regions removed, you can use the following postprocessing menu button in gBAT. Clicking this button will show you prompts for loading in the atlas file you

used, and some outputs automatically generated during a gBAT analysis to exclude ROIs. Note, all postprocessing menu options can be used without loading any data.



The *analysis_ROI_results.csv file that is required has a simple organization, with the first column corresponding to ROI indices used in the analysis.

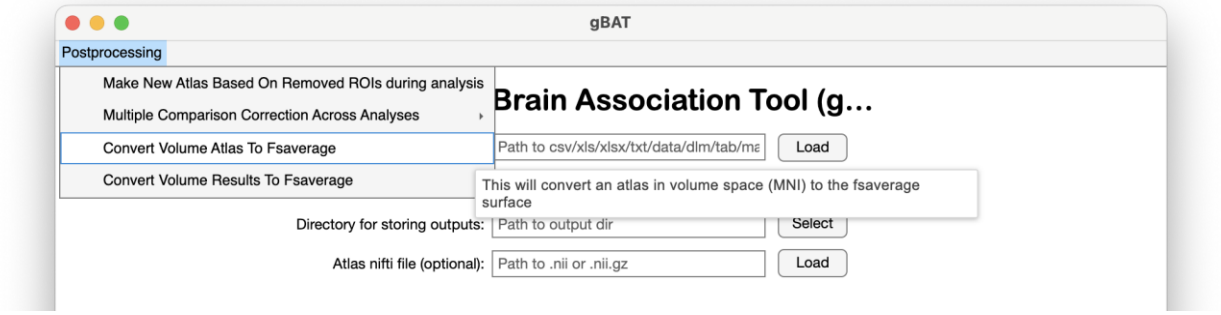
ROI index	indirect path coeff	path a coeff	path b coeff	path c' coeff	path c coeff	p
1	-0.00466431772138276	-0.000417424291426194	11.1740447721583	-0.0289875319290398	-0.0315150501161902	
2	-0.00604952025547251	-0.000432199719216198	13.9970480925888	-0.0282645758473245	-0.0315150501161902	
3	-0.00300120161216097	-0.000241459247516265	12.4294333020267	-0.0306867472520047	-0.0315150501161902	
4	-0.00315728782767298	-0.00025449373170426	12.4061516428309	-0.0306830454310978	-0.0315150501161902	
5	-0.00102367795129424	-0.000261818296683863	3.90987934861669	-0.0318557853910882	-0.0315150501161902	
6	-0.000581499942095803	-0.000237724328811692	2.4461103539656	-0.0324381994544656	-0.0315150501161902	
7	-0.00368378266396534	-0.00030070011513357	12.2506859112076	-0.0299622565906538	-0.0315150501161902	
8	-0.00410186604861354	-0.000301923480275568	13.5857802277244	-0.0293998031868515	-0.0315150501161902	
9	-0.00437367273088138	-0.000226257288878077	19.3305274387788	-0.0296851825216229	-0.0315150501161902	

That means that if you want to create *any* altered atlas with just a few regions removed, you can use this button with a csv file just like this, where you simply have a column of integers corresponding to the ROIs of the atlas you would like to keep.

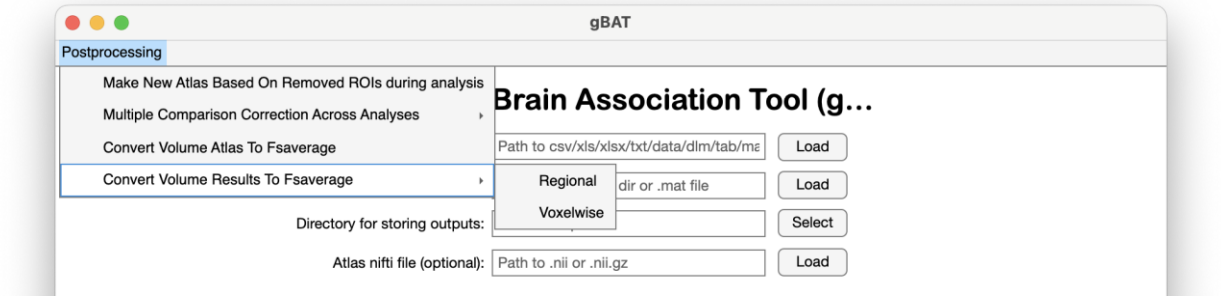
Surface projections

If a visualization in surface space is desired, atlasVol2Surf can be used to copy a volume space result for ROIs within some atlas onto an existing surface space version of that atlas, converting the ROI result into surface space without having to individually project ROIs. You may either use this as a standalone function, or you can use the menu options in gBAT.

If you don't have your atlas in surface space, you can use the postprocessing menu in gBAT to perform this projection as shown below. Just follow the prompts that appear over gBAT once you hit this button. This will require brainSurfer, so please make sure you have that downloaded (see documentation in brainSurfer for projection method). Note, all postprocessing menu options can be used without loading any data.

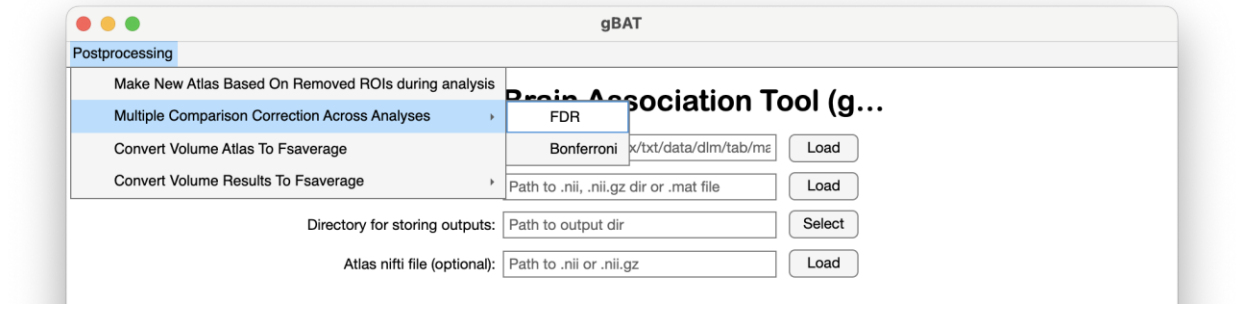


Once you have your atlas in surface space, you can use atlasVol2Surf through the GUI by clicking the 'convert volume results to fsaverage' button as shown below. If you performed a voxelwise analysis, we will simply project your results onto fsaverage instead of remapping ROI values from volume to surface parcels. Again, just follow the prompts that appear after you click this button. You will need brainSurfer for this option.



Multiple comparisons correction

By default, gBAT corrects for multiple comparisons within each output map separately. That means that if you performed multiple analyses (pressed the start button multiple times), or you are interested in more than just the indirect mediation effects during mediation analysis, **you are not correcting for all of the comparisons you have made**. This can be resolved through the file menu, where clicking the multiple comparison correction button will summon a separate GUI asking you to select all files that you would like to correct together. Here, you can specify any number of brain files for correcting *together*. For files output by gBAT, you will want to make sure that you supply uncorrected p-value maps. To simplify the process, you will want to copy all of the files you want to correct together into a separate directory that will be easier to import through gBAT. Again, all postprocessing menu options can be used without loading any data.



In case your analysis is regional, you will also be asked to provide the atlas that you previously used. If you want to perform voxelwise correction, simply hit the cancel button on this second request.

If you select the option to correct using FDR, we use the original Benjamini & Hochberg (1995) procedure. This avoids the complexities and potentially counterintuitive results that can arise from adaptive corrections in neuroimaging data (see: Reiss, P. T., Schwartzman, A., Lu, F., Huang, L., & Proal, E. (2012). Paradoxical results of adaptive false discovery rate procedures in neuroimaging studies. *NeuroImage*, *63*(4), 1833-1840.)

Some notes on distance correlation

For standard distance correlation, we use bcdistcorr.m, which implements Székely, G. J., & Rizzo, M. L. (2013). The distance correlation t-test of independence in high dimension. Journal of Multivariate Analysis, 117, 193-213. By default, we use the p-values output by this method as it is much more efficient and won't be as memory-intensive. However, we recommend using permutation analysis if possible. You can use our pdc repository, redistributed here, for this (see also https://github.com/alexteghipco/partialDistanceCorrelation). The partial distance correlation option automatically implements a permutation based p-value estimation. The method for partial distance correlation can be found here: Székely, G. J., & Rizzo, M. L. (2014). Partial distance correlation with methods for dissimilarities. The Annals of Statistics, 42(6), 2382-2412. Also see these slides: https://stat.wisc.edu/wp-content/uploads/sites/870/2020/03/SzekelyGabor.pdf. Note, pdcPerm uses a different method for computing partial correlation (recursive formula), but will produce the same output as the only other toolbox we are aware of implementing partial distance correlation—dcor in python (see: https://dcor.readthedocs.io/en/latest/index.html)