1. THIS IS THE BETA VERSION. Please check your analyses and results!! The statistical analysis scripts will create log files in the **stats\_uni\_logs and stats\_bil\_logs** directories. If you note an "Error" in a log file, please double check the analysis and send questions to us (Dmitry, Boris or Neda) at [enigma@ini.usc.edu](mailto:enigma@ini.usc.edu).
2. TheRscripts folder consists of several bash scripts. Only one of these need to be customized for your study and server. This is the script: **configure\_shape\_4stats\_{some postfix}.sh**

The following sections of the script above need to be modified. Except for the cohort name, all files names must have the full (absolute) path as part of the name:

Section 1:

This section will need to be edited for each new study.

**cohort=your\_cohort\_name:** This name must be unique if you are running this as part of a meta analysis, e.g. in ENIGMA, but there are no other specific requirements (of course, no spaces)

**fsdir=your\_freesurfer\_output\_directory**: This is the same path you must have set for the medial demons scripts, pointing to the folder with FreeSurfer output. It is assumed that the folder has subfolders named after the subject id’s in your covariates file, as is standard.

**covariates=your\_covariates\_file**: This is the .csv file, frequently named Covariates.csv, which contains all the demographic and clinical information needed for statistical analysis.

**outdir=your\_shape\_output\_directory**: Again, this must have already been set in the medial demons scripts – this is the directory where all shape output is written. As with the FreeSurfer folder, it is assumed that the folder contains subfolders named as the subject id’s.

**exclude\_file=your\_quality\_score\_file**: This is the full path to your QA\_Status.csv file, generated by the autoQA scripts, and modified by a human rater. It is assumed that the file will have ratings 1-3, 1 being unacceptable quality, 2 - questionable, and 3 - good. **IMPORTANT –** your QA\_Status file must have headers: SubjID,T1,R10,R11,R12,R13,R17,R18,R26,R49,R50,R51,R52,R53,R54,R58 Allowed variations are ROI\_{id} and ROI{id}. Else, your QC will not be used.

If you simply marked failed subjects using the alterative format, you will need to convert this to a rating format, e.g. by adapting ENIGMA\_shape\_qa\_rearrange.R

Section 2:

You should only have to set these values once for all studies on your server:

**demons\_dir=your\_medial\_demons\_directory**: this is the directory where you copied your medial demons scripts. Same as what was set in the medial demons wrapper script (for example, shape\_group\_run.sh, provided in the medial demons package)

**Rscript\_dir=your\_path\_to\_the\_R\_scripts\_in\_this\_package**: This folder is where the R scripts from the shape stats package (and by default, where the script you are currently editing) are.

**R\_binary=your\_path\_to\_the\_R\_executable**: This is the R executable file that opens the R environment. Remember, the version you choose must have several libraries listed in this link connected to it: <https://github.com/ENIGMA-git/ENIGMA/blob/master/WorkingGroups/EffectSize_and_GLM/README.md>

Section 3:

**DO NOT CHANGE THESE VALUES IF YOU ARE PART OF AN ENIGMA-SHAPE STUDY (E.G. MDD, SCZ GROUPS)**

These values define the set of linear models which are run vertex-wise on your data. If you are running your own study, you may wish to alter these to fit your needs. Please create your own google docs as per instructions in the github link above, and provide links/correct names for them here:

**model\_list=your\_model\_identifier**: This is the same thing as the RUN\_ID variable in the github link above. It serves as an indicator of the specific set of models (another google doc, see github README) you wish to run. Note that if you run this script, the scripts described in the readme above will have your desired value set automatically.

**config\_link=your\_master\_google\_doc\_with\_stats\_model\_lists**: This is the same thing as the CONFIG\_PATH variable in the github link above. If you are running your own stuff, you will need to make one of these with the same headers as the current link. Note that if you run this script, the scripts described in the readme above will have your desired value set automatically.

**3.** Once you have created a customized version of the **configure\_shape\_4stats** script, go ahead and it. You should get 6 bash scripts in your shape output directory called:

**i. configure\_data\_{your cohort name}.sh**

**ii. & iii. mass\_uv\_regr\_IGC\_shapes\_{your cohort name}\_[uni/bil].sh**

**iv. & v. concat\_mass\_uv\_regr\_{your cohort name}\_[uni/bil].sh**

**vi. collect\_shape\_stats\_{your cohort name}.sh**

You will need to run 3 of these from command line, and a 4th one if you are contributing group-level results to a meta-analysis. Below is the order in which you run these

**4.** Run **configure\_data\_{your cohort name}.sh**

This will do several things:

**i.** It will create a set of bilateral vertex-wise measures of shape to complement the unilateral (left/right) measures. In this case, the maps you get will correspond to

(a) vertex-wise mean of left and right shapes for ROI id 10-26 and

(b) absolute symmetry ( |left-right| ) for ROI id 49-58

This is useful b/c the measures don't change even if some sites have a left/right flipping issue. The original measures are still used as well.

**ii. & iii.** A pair of files called **Covariates\_[vols/bil].csv** - the original clinical + demographic data concatenated with (ii) original FS volume data and (iii) the bilateral (r/l mean) volume measures. This is needed for the statistical analysis of unilateral bilateral measures.

**iv.** **QA\_status\_bilat.csv** - the analogue of the original human-rated quality of each shape, but taken as the bilateral minimum (so if left hippo is rated 2 and right as 3, they are both "2" for each of the bilateral measures)

**5.** Next, you should run the actual statistical models. The models are defined by the google docs link, and you can read about specifics on how to define them in the github repository (again): <https://github.com/ENIGMA-git/ENIGMA/blob/master/WorkingGroups/EffectSize_and_GLM/README.md>

You can follow the current link to see what models we have selected to run. Remember, only models with 1's in the "active" column on the google doc model file will be actually run. If you wish to change what models are run in your own independent study, you will need to alter the variables in section 3 of the **configure\_shape\_4stats** script, and run it again to regenerate the R-wrappers. You will **NOT** need to run the **configure\_data** script twice, unless the data itself (shape or QA) have changed.

To run the statistical models, you will need to submit using qsub:

**qsub -t 1-14 .../{shape output directory}/mass\_uv\_regr\_IGC\_shapes\_{your cohort name}\_uni.sh**

**qsub -t 1-14 .../{shape output directory}/mass\_uv\_regr\_IGC\_shapes\_{your cohort name}\_bil.sh**

*Note that you can also run these serially, which will take considerably longer. You will need to alter the mass\_uv... files in lines 53-54, and set Nnodes=1.*

The scripts should generate logs, which you may need to check if something goes wrong. If all goes well, the results will be in these two folders:

**.../{shape output directory}/stats\_unilateral**

**.../{shape output directory}/stats\_bilateral**

The per-vertex statistics are printed to csv files in the folders above. We have additional scripts to compute FDR corrections, and make p- and beta- maps for all models which pass FDR per subcortical shape. These scripts are written in Matlab, and available upon request.

If, however, you are only participating in a meta-analysis, the only thing left to do is to run is the  **collect\_shape\_stats\_{your cohort name}.sh** script. This will generate two tar.gz files in your shape output folder.

**shape\_stats\_2send\_${model\_list}\_${cohort}\_uni.tar.gz**

**shape\_stats\_2send\_${model\_list}\_${cohort}\_bil.tar.gz**

These files contain all the data needed for meta-analysis. Please send them to the IGC shape team.

Thank you, we look forward to finalizing our ENIGMA shape projects!

**–** The Team: Dmitry Isaev, Boris Gutman, Neda Jahanshad, Artemis Zavaliangos-Petropulu, Chris Ching