Here is the link to dropbox folder:

<https://www.dropbox.com/sh/q6lfaft0xx50zq7/AADjW99S8kHWQaJzOxNfkxE0a?dl=0>

Please download all files and scripts to one folder. There is also atlas nifti file, which we need to split images by chunks of voxels (**brain\_vbm\_atlas.nii.gz**). **Please, check that all your images are the same size and in the same space as atlas.**

Everything below should be applied separately for every cohort.

1. Prepare working directory
   1. Make four new folders in your working directory:

./QC

./nparray

./np\_logs

./images

* 1. Move your VBM modulated and smoothed results images to separate folder ./images (they should look like {ID}\_GM\_to\_template\_GM\_mod\_s3.nii.gz).

1. Converting images and QC

**2.1**

* *python nii2np.py -i {full path to smoothed images folder} -o {folder to save nparray} -atlas {path to atlas (*brain\_vbm\_atlas.nii.gz*) } -logs {folder to save np logs} -code {chunk code}*

There are 188 chunks (-code parameter for nii2np.py script), you can run this script in a loop from 1 to 188 inclusive or submit to cluster in parallel, because this would take some time.

As a result in ***“nparray”***folder should be saved nparrays of voxel values for all these 188 chunks. ***“np\_logs”*** folder should contain files with images reading order per every chunk.

**2.2**

* *python QC\_vbm\_reg.py -mode region -i {nparray folder path} -o {QC folder path} -logs {np\_logs folder path} -code {chunk code} -q {quantile suggested threshold, set it to 10}*

Again there are 188 chunks, so you can run this script in a loop from 1 to 188 inclusive.

* Then run this script again but with different mode:

*python QC\_vbm\_reg.py -mode summary -i {nparray folder path} -o {QC folder path} -logs {np\_logs folder path} -q {quantile threshold I suggest set it to 10}*

It should print some info, plus save statistic into two tables in ***“QC”*** folder: ‘control.csv’ and ‘outlier\_mri\_id.csv'. The values in table represent the proportion of voxels in chunk with gray matter density below or higher defined quantile of all study images voxels distribution in chunk (for instance with threshold set to 10, proportion of voxels below then 10% and higher then 90% quantile).

Next step depends on how many images are with high proportion of outliers. If the proportion close to threshold, then it could be just statistical variation and it is not necessary to exclude images. So, it is better to check manually in viewer what’s wrong with them (at least several images). If the registration failed than it should be obvious. If the image looks fine, it could be that threshold for outliers was too strict and we should not exclude it.

* Make a list of images for exclusion from analysis (based on outlier\_mri\_id.csv data). Make a csv table with two columns (**id\_table\_example.csv** in dropbox folder):

id: ids identical to ids in genotype data

exclude: 1 if exclude, 0 if include to analysis

**The order and the length of id column should be identical the order of how images were read (check any file from ./np\_logs folder)!**

**2.3 make\_hase\_phen.py**  is a script from HASE. You can find it in HASE subdirectory. /tools/

* *python make\_hase\_phen.py -i {nparray folder} -id {csv data frame from 2.2 point 3 above}*

Script will make ***info\_dic.npy*** file for vGWAS analysis. This is the last step, if everything was fine, phenotype data in “nparray folder" should be ready for vGWAS analysis with HASE.