

FVGWAS-2.0 Manual

Hongtu Zhu @ UNC BIAS

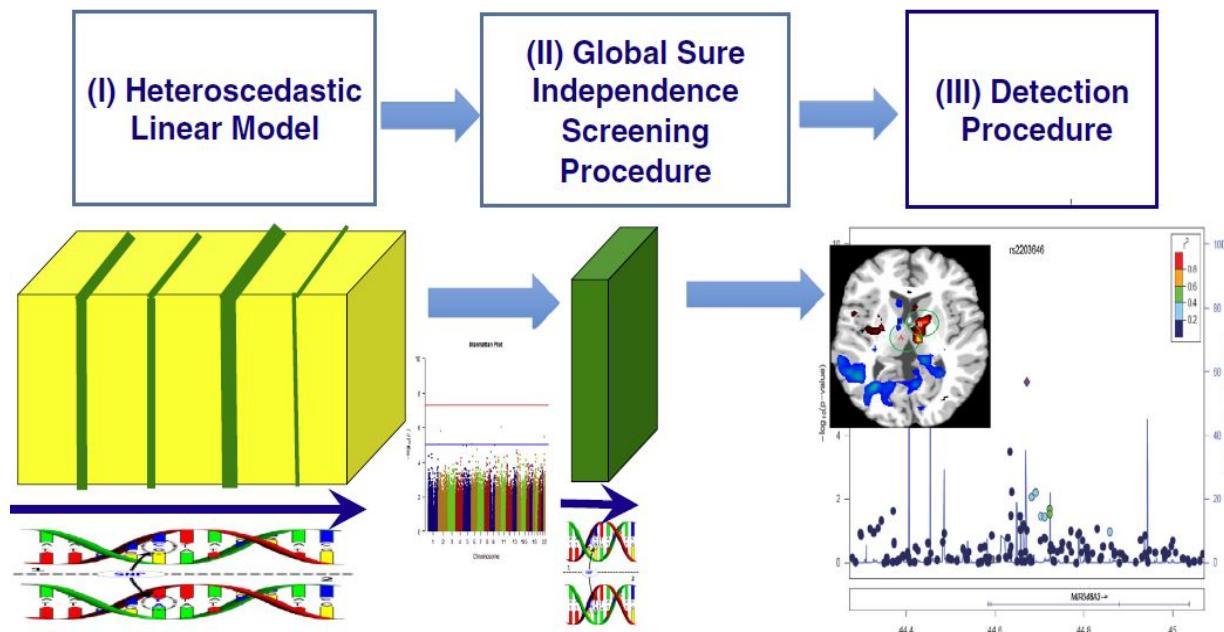
Chao Huang @ UNC BIAS

Nov 8, 2015

More and more large-scale imaging genetic studies are being widely conducted to collect a rich set of imaging, genetic, and clinical data to detect putative genes for complexly inherited neuropsychiatric and neurodegenerative disorders. Several major big-data challenges arise from testing genome-wide associations with signals at millions of locations in the brain from thousands of subjects. The aim of this software is to efficiently carry out whole-genome analyses of whole-brain data. FVGWAS consists of three components including a heteroscedastic linear model, a global sure independence screening (GSIS) procedure, and a detection procedure based on wild bootstrap methods. Our FVGWAS may be a valuable statistical toolbox for large-scale imaging genetic analysis as the field is rapidly advancing with ultra-high-resolution imaging and whole-genome sequencing.

1. Schematic overview of FVGWAS

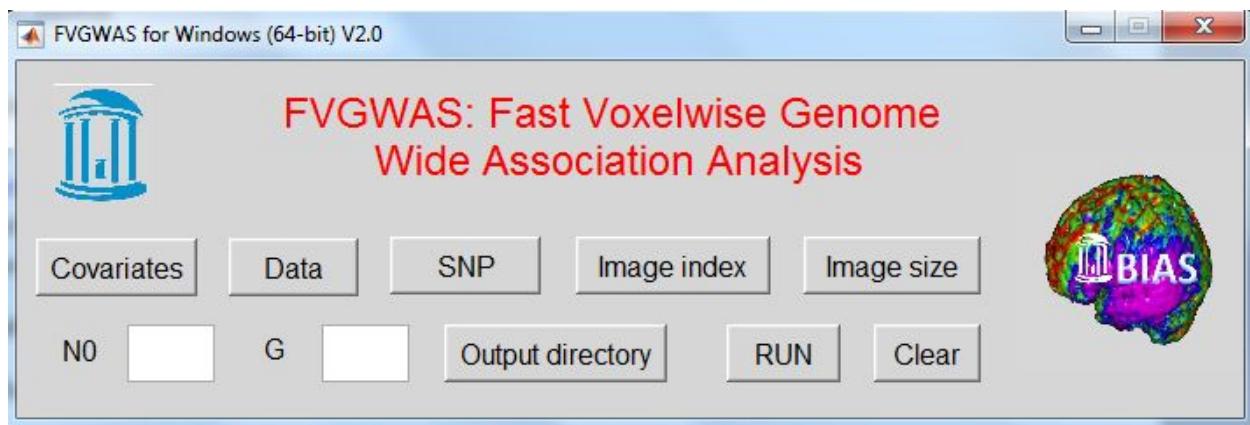
Fast Voxelwise Genome Wide Association analysiS (FVGWAS)



2. FVGWAS pipeline description

For FVGWAS, we provide two different version: interface and server-based pipeline. If you have a small data set, e.g., ROI data and SNP data from one chromosome, you can use the interface on your local desktop. If you have a large data set, e.g., whole brain imaging data and imputed SNP data, you can use the server-based pipeline which can also help you run all the tasks via parallel computing.

2.1. Interface-based FVGWAS



Covariates

$n \times d$ design matrix. n is the sample size and d is the number of covariates

Data

Image data ($n \times V$ matrix). V is the number of voxels (whole brain analysis) or the number of ROIs (volumetric analysis)

SNP

Genetic data ($n \times C$ matrix). C is the number of SNPs

Image index

$V \times 1$ vector. The index of voxels within the brain mask (Applicable only for whole brain analysis)

Image size

The size of image data, e.g., 1*3 vector means 3-D image data (Applicable only for whole brain analysis)

NO

The number of top SNPs

G

The number of bootstrap samples

Output directory

Set up the directory which is used to store all the results

RUN

Start running all the tasks

Clear

Clear all the input information on the interface

2.2. Server-based FVGWAS Pipeline

If you have a very big data set and prefer to run FVGWAS on the workstations or servers. There are five MATLAB functions and one R code which can help you accomplish all the tasks in FVGWAS.

A. FVGWAS_GSIS_server.m

This function is to run global sure independent screening (GSIS) procedure. After running this function, you will get the p values for all the loci with MAF no less than 0.05.

B. FVGWAS_Manhattanplot_server.m

This function is to summarize all the p values and SNP information into one file which will be used as an input file in plotting Manhattan plot and QQ plot. After running this function, you will get a file including the p values, SNP names, chromosome IDs, and BP values for all loci.

C. FVGWAS_Bootstrap_1_server.m & FVGWAS_Bootstrap_2_server.m

These two functions are to run the bootstrapping in the detection procedure. After running this function, you will get the significant locus-voxel pairs and significant clusters for top SNPs found in GSIS step. In this step, because the bootstrapping require a lot of memory, we adopt to run part of the code via parallel computing. If you have multiple cores on the server, you can finish this step very efficiently.

D. FVGWAS_Detection_server.m

This function is to run the bootstrapping in the detection procedure. After running this function, you will get the significant locus-voxel pairs and significant clusters for top SNPs found in GSIS step. In this step, because the bootstrapping require a lot of memory, we adopt to run part of the code via parallel computing. If you have multiple cores on the server, you can finish this step very efficiently.

2.3. GWAS Results Display

This function ([FVGWAS_ManPlot.R](#)) is to plot Manhattan plot and QQ plot with R package ([qqman](#)). The file generated in function B is treated as the input file in this step.

3. Examples

We consider both ROI volumes and RAVENS maps to illustrate the wide applicability of FVGWAS. We carried out two different FVGWAS analyses: one is to use the volumes of 93 ROIs as multivariate phenotypic vectors and the other is to use RAVENS maps as whole-brain phenotypic vectors. In both analyses, we included an intercept, gender, age, whole brain volume, and the top 5 principal component scores in SNPs as the covariates. Then, we tested the additive effect of each of 501,584 SNPs on either 93 ROI volumes or RAVENS maps. For the volumetric analysis, we run all the tasks via the interface pipeline, while for the Ravens map data, we conduct the analysis via server-based pipeline.

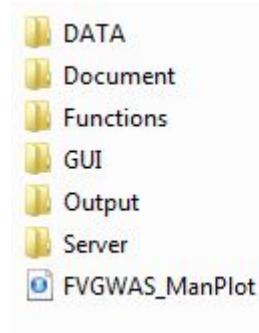
3.1. Volumetric Analysis

1. Download this package and unzip the file;
2. Start Matlab and set the following directory as the working folder

<C:\Users\Administrator\Desktop\FVGWAS-2.0>

Where `C:\Users\Administrator\Desktop` is the directory where you put the unzipped folder `FVGWAS-2.0`

3. In the folder `FVGWAS-2.0`, you can find six different folders and one R script.

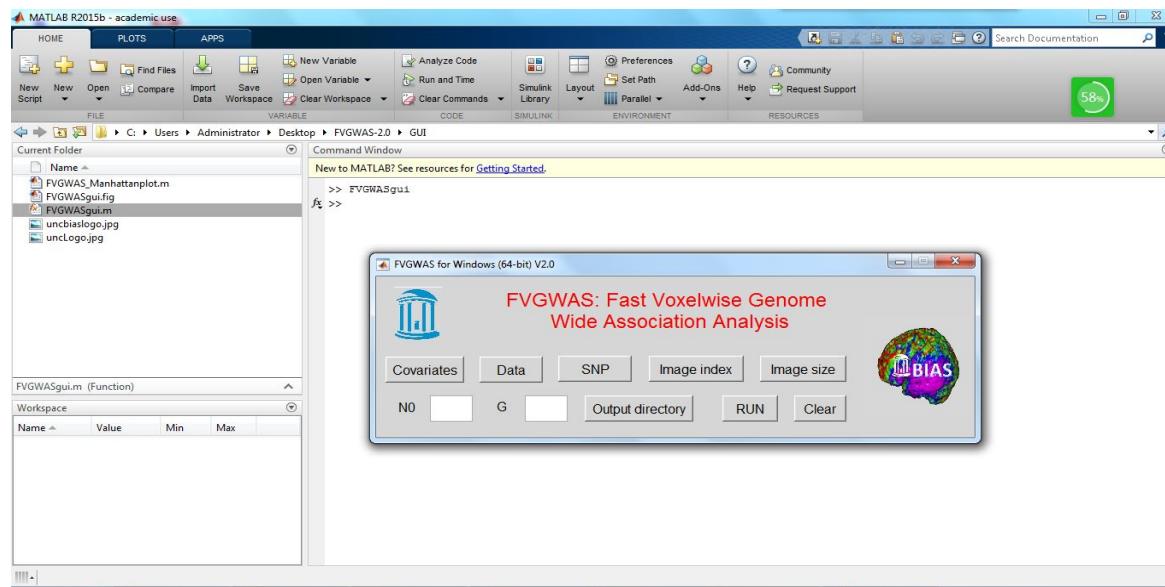


In the folder `DATA`, you can find the volumetric data set in the sub-folder `volumetric`. There are four files in that folder. `Covariate_Data.mat` includes all the demographic information; `Image_Data.mat` includes all the ROI thickness information; `SNP_Data.txt` includes the genetic data for all subjects; and `SNP_info.map` contains the SNP information including SNP names, chromosome IDs, and BP values for all loci.

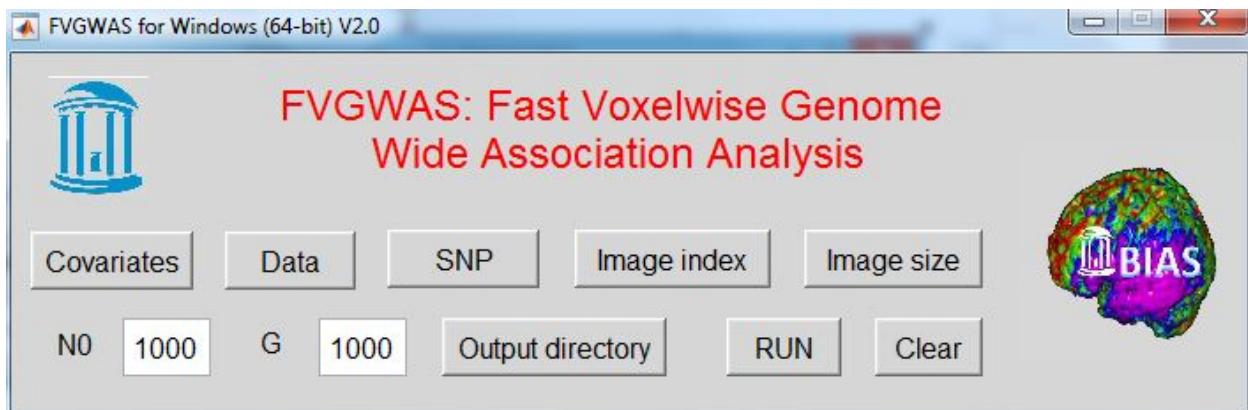
All the codes are stored in the folder `Functions`.

The interface can be found in the folder `GUI`.

4. Set `GUI` as the current folder and type `FVGWASgui` in command window to launch FVGWAS in MATLAB.



5. Load the data into FVFWAS



Click the button **Covariates**: load “[Covariate_Data.mat](#)”

Click the button **Data**: load “[Image_Data.mat](#)”

Click the button **SNP**: load “[SNP_Data.txt](#)”

Set **N₀**: 1000

Set **G**: 1000

Click the button **Output directory**: e.g. [./FVGWAS-2.0/Output/volumetric](#)

Click the button **RUN**

6. Output files

“[GSISresults.mat](#)”

results of GSIS step, variable “pp” contains the -log₁₀{p-values} of all SNP data, and “SNP_index” contains all the SNP ID which are removed in preprocessing.

“[VoxelclusterandSNP.mat](#)”

variable “rawpvalue” is a C*N0 matrix including the raw p-values of top N0 SNPs.

variable “pv” is a C*N0 matrix, including the corrected p-values.

7. Results display

(1) Run “[FVGWAS_Manhattanplot.m](#)” in MatLab to generate the GWAS results. Output file: “[SNP_GSIS.txt](#)”. One thing should be noted that, in [FVGWAS_Manhattanplot.m](#), there are two parameters should be modified according to your own settings.

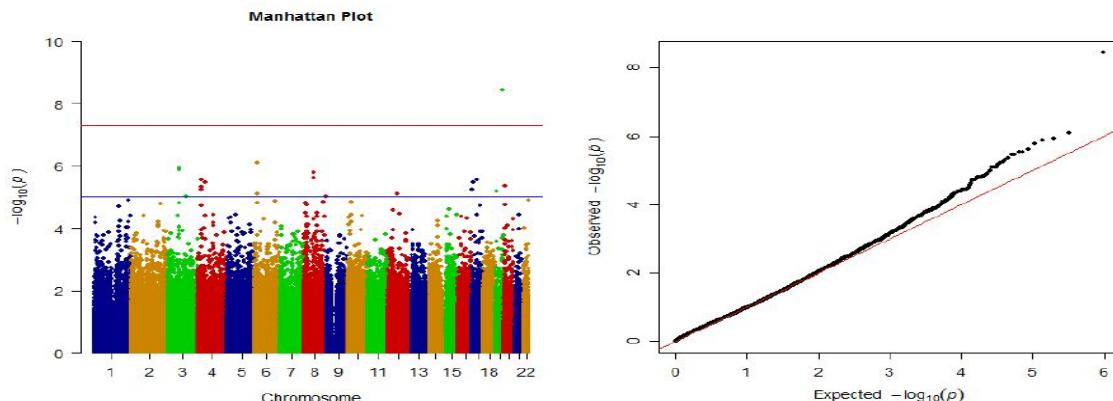
`mode='volumetric'; volumetric` is the name of the folder where the file [SNP_GSIS.txt](#) is saved.

`Flag=0;` where 0 for ROI-based imaging data; 1 for voxel-based imaging data

(2) Run “[FVGWAS_ManPlot.R](#)” in R to plot the Manhattan plot and QQ plot. One thing should be noted that, in [FVGWAS_ManPlot.R](#), there are two parameters should be modified according to your own settings.

`statName="volumetric"; volumetric` is the name of the folder where the file [SNP_GSIS.txt](#) is saved.

`HomeDir="C:/Users/Administrator/Desktop/FVGWAS-2.0";` which is the directory the folder FVGWAS-2.0 is located.



3.2. Volumetric Analysis

In the folder [DATA](#), you can find the RavenMap data set in the sub-folder [RavenMap](#). There are six files in that folder. [Covariate_Data.mat](#) includes all the demographic information; [Image_Data.mat](#) includes all the ROI thickness information; [ImageIndex_Data.mat](#) is a $V \times 1$ vector, which is the index of voxels within the brain mask; [ImageSize_Data.mat](#) is the size of image data, e.g., 1×3 vector means 3-D image data; [SNP_Data.txt](#) includes the genetic data for all subjects; and [SNP_info.map](#) contains the SNP information including SNP names, chromosome IDs, and BP values for all loci.

All the codes are stored in the folder **Functions**.

The server-based main functions can be found in the folder **Server**.

1. Set **Server** as the current folder and run **FVGWAS_GSIS_server.m**;

2. After Step 1 is finished, run **FVGWAS_Bootstrap_1_server.m**

and **FVGWAS_Bootstrap_2_server.m** simultaneously;

3. After Step 2 is finished, run **FVGWAS_Detection_server.m**;

8. Results display

(1) Run “**FVGWAS_Manhattanplot_server.m**” in MatLab to generate the GWAS results. Output file: “**SNP_GSIS.txt**”. One thing should be noted that, in **FVGWAS_Manhattanplot_server.m**, there are two parameters should be modified according to your own settings.

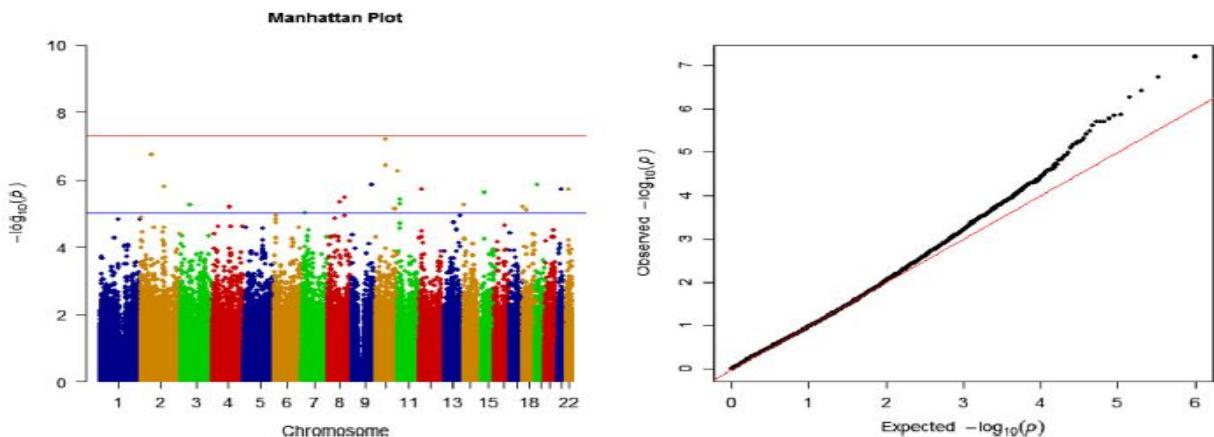
`mode='RavenMap'; RavenMap` is the name of the folder where the file **SNP_GSIS.txt** is saved.

`Flag=1;` where 0 for ROI-based imaging data; 1 for voxel-based imaging data

(2) Run “**FVGWAS_ManPlot.R**” in R to plot the Manhattan plot and QQ plot. One thing should be noted that, in **FVGWAS_ManPlot.R**, there are two parameters should be modified according to your own settings.

`statName="RavenMap"; RavenMap` is the name of the folder where the file **SNP_GSIS.txt** is saved.

`HomeDir="C:/Users/Administrator/Desktop/FVGWAS-2.0";` which is the directory the folder FVGWAS-2.0 is located.



4. PS

1. If you want to run volumetric analysis via server-based pipeline. You need to change two parameters in the following four main functions in folder [Server](#)

[FVGWAS_GSIS_server.m](#)

[FVGWAS_Bootstrap_1_server.m](#)

[FVGWAS_Bootstrap_2_server.m](#)

[FVGWAS_Detection_server.m](#)

mode='RavenMap'; should be changed to mode='volumetric';

Flag=1; should be changed to Flag=0;

2. When you run the server-based pipeline, please remain the name of each input file as the one shown in the folder [DATA](#). Otherwise, you also need to modify them in the four functions shown above.

3. If you have a new data set. You can put your data folder into the folder [DATA](#), modify the value of ‘mode’ in the four functions above, e.g., if the name of the new data folder is “[Newdata](#)”, you need to change the value of “mode” to mode='Newdata'. Also, if the new data set is voxel-type imaging data, you need to set Flag=1, otherwise, Flag=0.