

Guide to running GREGOR (with helper scripts)

This guide lays out the steps for running GREGOR using custom helper scripts to make the process much easier. First, disease/traits will be established from GWAS. It will be shown how to include additional lists of SNPs to run in conjunction with these. Next, GREGOR will be configured, and output scripts will be generated for seamlessly running GREGOR using the Slurm workload manager. Finally, the GREGOR outputs will be summarized for further analyses.

Step 1: Establish disease/traits SNP associations from GWAS catalog.

To establish a background, we first prepare disease/trait associations with dbSNP reference ids (rsIds).

1. Download the associations (All associations v1.0) from ebi:
<https://www.ebi.ac.uk/gwas/docs/file-downloads>
2. Run the helper script from the Singularity image:
3. `singularity run RunGREGOR.sif BuildGWAS [options] <associationfile> <outputdirectory>`

associationfile: This is the file downloaded from EBI. This should have .tsv as the file extension. If it is compressed, uncompress it.

outputdirectory: Specify an output directory to store the associations. These will be .txt files for each trait in the form (Trait_x.txt). The trait descriptions are available in *traitmapping.txt*.

Options The following options can be used to change what associations will be kept:

- m Exclude multiple SNP associations (rsid1; rsid2; etc.).
- x Exclude snp interactions (rsid1 x rsid2).

The output directory will contain a directory “traits” with multiple files of the form Trait_x.txt. Each file contains a list of rsIds. Additionally, the following two files will be provided:

stats.txt: Statistics on how many associations were read.

traitmapping.txt: A two-column tab delimited file specifying the filename (Trait_x) and the disease/trait association.

Step 2: Incorporating additional SNP associations

To include additional SNP associations, generate .txt files for each additional disease/trait. Each line of these files corresponds to either an hg19 chromosome position (chr:pos) or rsId. Include these files in the trait directory that will be used in the next step.

Note: setting up the trait directory can also be used to exclude/select traits from the analysis. Refer to traitmapping.txt to determine which traits/diseases are the most applicable.

Step 3: Configuring GREGOR

After setting the disease/trait associations, the next step is to configure the GREGOR run files. For this use the following command:

```
singularity run RunGREGOR.sif ConfigureGREGOR [options] <bedfilelist> <snpsdir> <refdir> <configdir> <outdir>
```

bedfilelist: This is a list of paths to the bed files that will be used for checking SNP-Disease associations. This is essentially a file where each line is an absolute path to a bed file that will be processed. Each bed file must be in tab delimited format with chromosome, start, end, and a final column with “.”.

Example:

```
chr1 100 200 .
chr1 3000 4000 .
```

snpsdir: The directory of SNP-Disease associations generated in the previous steps.

refdir The directory containing the compiled reference files obtained from <http://csg.sph.umich.edu/GREGOR/index.php/site/download> .

configdir The path to the directory that will store the configuration files to be generated.

outdir The output directory that will store the outputs when running each of the run configurations

Options The following options can be used to adjust the configuration files:

--r2	Adjust the r^2 threshold (Default: 0.7, must be at least what is specified by the reference)
--ldw	LD window size. (Default: 1000000)
--min	Minimum number of neighbors. (Default: 500)
--pop	The reference population to use from the refdir. (Default: EUR)

This will populate the directory specified by configdir with GREGOR run configurations that can be run independently. A list of all the configuration files is found with “configfiles.txt”. Additionally, two Slurm scripts are generated in the configuration directory: “RunGREGOR.sh” and “CleanGREGOR.sh”. These scripts will be used to run GREGOR with the sbatch command.

Step 4a: Running GREGOR with Slurm

To run GREGOR, proceed to the configuration directory and run the RunGREGOR.sh file with sbatch.

```
sbatch RunGREGOR.sh <pathtosingularity>
```

where **pathtosingularity** is the absolute path to the singularity image file. It will be helpful to create a new directory and run this command from there to have the slurm outputs organized accordingly.

Step 4b: Running GREGOR *via* looping

If the Slurm workload manager is not available, GREGOR can be run by looping over the configuration files (the list of these files is in “configfiles.txt”) using a shell script.

Step 5: Summarizing results

After all the runs have completed, a single table of the SNP-Disease associations for each of the bed files can be generated using the following command:

```
singularity run RunGREGOR.sif SummarizeGREGOR <outdir>
```

where **outdir** is the output directory specified in the GREGOR configuration step. This will generate a file: **Results.txt** in the output directory. This file contains five columns of information:

Phenotype:	The trait/disease
Bed File:	The BED file analyzed.
Observed	The number of SNPs observed
Expected	The expected number of SNPs to observe
P-Value	The significance P-Value for observing the observed number of SNPs

Note: Additional processing is required (i.e., multiple hypothesis testing).

Step 6a: Cleaning output directories with Slurm

Once the analyses are finished. The output directory will take some time to remove. It is recommended to remove these directories using the clean script:

```
sbatch CleanGREGOR.sh
```

IMPORTANT: Be very careful with this script! Please review it before running to ensure it does not remove unintended directories!

Step 6b: Cleaning output directories *via* looping

If the Slurm workload manager is not available, create a shell script that loops through all the output directories (the list of these directories is in “configouts.txt”). Alternatively, you can remove the entire directory. This process will take a long time due to the high volume of files generated by running GREGOR.

Appendix:

Building the singularity image file (.sif):

1. Download the latest version of GREGOR from: <http://csg.sph.umich.edu/GREGOR/>
2. Unzip the contents into a directory labeled GREGOR
3. In the same directory add the helper scripts (.sh and .py) available from:
<https://github.com/UcarLab/GREGORHelper>
4. Modify the .def file available from the github in step 3 and change the path in %files to the path of the GREGOR directory from the second step
5. Build the singularity image file using the following command:
`sudo singularity build RunGREGOR.sif RunGREGOR.def`