Heritability Analysis of voxel-wise traits in the UK BioBank

The following is a step-by-step tutorial of how to perform a voxel-wise heritability analyses in the UK BioBank dataset using SOLAR-Eclipse Imaging Genetics software found on our NITRC page: https://www.nitrc.org/projects/se_linux/. Before you begin the analysis, it is important to have the data labeled with the same IDs as used in the pedigree or genotype files to avoid confusion and error. SOLAR assumes that all subjects have been spatially normalized to a common space, so a voxel-wise trait is an estimate of a phenotypic variance across all subjects.

Note: In the following instructions, '$' denotes a Bash prompt, while '>' denotes the SOLAR software prompt. Commands preceded by '$' are for in the Bash terminal and should not be entered into SOLAR.

Step1: Download Multiband diffusion brain images from UKBB.
- $ ./ukbfetch -bukb20250.bulk -s1 -m5 -ak34077.key
  - Example of bulk file:

```
1   3987022 20250_2_0
2   2903723 20250_2_0
3   3117676 20250_2_0
4   2297630 20250_2_0
5   4814101 20250 2 0
```

Step2: Unzip all imaging data into the 'unzip' directory and copy FA images into the 'skeletonised_FA' directory.
- $ ./unzip.sh
  - Coding in the unzip.sh

```
1   #!/bin/bash
2   mkdir unzip
3   for f in `ls -1 *.zip`;
4       do
5           unzip -n $f -d `basename $f .zip`
6           mv `basename $f .zip` unzip
7   done
```

- $ ./extract_skeletonised_FA.sh
  - Coding in the extract_skeletonised_FA.sh

```
1   #!/bin/bash
2   path=`pwd`
3   mkdir skeletonised_FA
4   for i in `ls -d unzip`;
5   do
6       cd $i
7       for j in `ls -d */`;
8       do
9           id=${j:0:7}
10          echo "copy $id all_FA_skeletonised.nii.gz"
11          cp $path/$i/$j/dMRI/TBSS/stats/all_FA_skeletonised.nii.gz ../skeletonised_FA/"$id"_all_FA_skeletonised.nii.gz
12      done
13      cd ..
14  done
```

Step 3: Create a phenotype file that links the IDs, covariates, and names of the nifti files containing the voxel-wise traits

- This file can later be used for creating voxel-wise phenotype files, so it is important to include all covariates that will be used in the analysis; including additional covariates at later stages can be very complicated and tedious.
- An easy way to create a phenotype file:
  - cd to the 'skeletonised_FA' directory containing all nifti files
  - Use ls to write out the file names into a new text file
  - $ ls *.gz > nifti_phenfile.csv
  - Open the new .csv file in Excel and create columns for IDs, covariates, and trait names, which will be the column that contains the nifti file names. (Covariates can be extracted separately from the UK BioBank bulk files)

A1		fx	ID	
	A	B	C	D
1	ID	Sex	Age	fa
2	2297630	M	67	2297630_all_FA_skeletonised.nii.gz
3	2903723	F	72	2903723_all_FA_skeletonised.nii.gz
4	3117676	M	72	3117676_all_FA_skeletonised.nii.gz
5	3987022	F	70	3987022_all_FA_skeletonised.nii.gz
6	4814101	M	60	4814101_all_FA_skeletonised.nii.gz

Step 4: Start SOLAR-Eclipse

- $ ./data/mprc_data1/home/bdono1/solar-eclipse-8.5.1-experimental-dynamic-3/bin/solar

```
[gaosi@mprc04:/data/brutus_data17/Katie$ /data/mprc_data1/home/bdono1/solar-eclipse-8.5.1-experimental-dynamic-3/bin/solar

SOLAR Eclipse version 8.5.1 (beta), last updated on April 22, 2021
Developed at Maryland Psychiatric Research Center,
University of Maryland School of Medicine, Baltimore.
Visit our documentation and tutorial website www.solar-eclipse-genetics.org
Our download page https://www.nitrc.org/projects/se_linux
Our github page https://github.com/brian09/solar-eclipse
For questions email: pkochunov@gmail.com
Enter help for help, exit to exit, doc to browse documentation.
The software development is supported by NIH grant RO1EB015611
from The National Institute for Biomedical Imaging and Bioengineering.
Enter cite to see how to cite this software.

solar>
```

Step 5: Run the command 'nifti_to_csv' to convert nifti files to voxel-wise phenotype files

- This is a simple linear format in which voxel-wise values are stored under VOXEL_X_Y_Z traits
- Use the nifti_to_csv command:
  - > nifti_to_csv mask.nii.gz fa nifti_phenfile.csv csv_phenfile 200000
    - Where:
      - **mask.nii.gz** is the mask volume name
      - **fa** is the trait column name above the list of files (Figure 1)
      - **nifti_phenfile.csv** is the phenotype file that contains IDs, covariates, and the column of files names
      - **csv_phenfile** is the output file base name of the new phenotype files
      - **200000** indicates the max number of voxels per file. This number can be changed depending on the desired number of output files

Step 6: Regress the covariates from the phenotype data using the 'sporadic_normalize' command

- This is not necessary but encouraged as it reduces the time needed for performing multi-trait heritability analyses
- Load the phenotype file:
    - > load phen csv_phenfile.csv
- Then list the covariates:
    - > covar Age Sex
- Use the sporadic_normalize command
    - > sporadic_normalize -header csv_phenfile.header -out norm_phenfile.csv
        - Where:
            - **csv_phenfile.header** is the previously made header
            - **norm_phenfile.csv** is the name of the output file

Step 7: Run gpu_fphi or fphi

- Create a pedigree file
    - Calculate the SNP frequencies using 'plink_freq' in SOLAR:
      > plink_freq --plink UKBB_plink --o UKBB_freq
    - Use 'gpu_pedifromsnps' to create an empirical pedigree:
      > gpu_pedifromsnps--i UKBB_plink --o UKBB_ped.csv --thread_size 1024 --batch_size 5000 --snp_stride 10 --gpus 0 --freq UKBB_freq --normalize
      **Note: The '--i', '--o', '--freq', and '--normalize' options are required
- Load the pedigree file pedigree file create in the previous steps
    - > load ped pedigree.csv -t 0
        - Where **t** is the correlation coefficient
        - This step does not need to be repeated in future analyses run in the same directory
- Then load the normalized phenotype file
    - > load phen norm_phenfile.csv
- Create eigenvalue and eigenvector files
    - > trait VOXEL_27_99_79
    - > create_evd_data --o UKBB_evd
- Use the fphi command to analyze the heritability of all traits on CPUs:
    - > fphi -list csv_phenfile.header -evd_data UKBB_evd
        - where:
            - **csv_phenfile.header** is the header made through the nifti_to_cvs command
            - **UKBB_evd** is the basename of the eigenvalue and eigenvector files
- Use the gpu_fphi command to analyze the heritability of all of the traits in the phenotype file on GPUs with either:
    - > gpu_fphi -list csv_phenfile.header -evd_data UKBB_evd-out UKBB_h2.csv -gpus 0
        - This will save the voxel-wise results in a .csv file

- where:
  - **csv_phenfile.header** is the header made through the nifti_to_cvs command
  - **UKBB_evd** is the basename of the eigenvalue and eigenvector files
  - **0** is the ID of CUDA GPU