Heritability Analysis of a Single Trait in the UK BioBank

The following is a step-by-step tutorial of how to perform single trait heritability on the UK BioBank data using SOLAR-Eclipse Imaging Genetics software found on our NITRC page: https://www.nitrc.org/projects/se_linux/. Please already have all the UKBB bulk imaging and genetic data downloaded and unpacked onto your servers. Also please download PLINK software before starting the heritability analysis as SOLAR requires genetic data to be in the PLINK format: https://www.cog-genomics.org/plink/1.9/.

**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: Filter and combine all genetic data into a single set of PLINK bed files using the --merge-list option. Ensure that SNPs pass quality control using Minor Allele Frequency and Hardy-Weinberg Equilibrium thresholds.

$ plink --bfile chr1 --merge-list file_names_text --maf 0.01 --hwe 0.001 --make-bed --out UKBB_plink

- If there are merge errors, please consult the PLINK documentation. Usually there are missnps that need to be flipped: https://www.cog-genomics.org/plink/1.9/data#merge3

Step2: Create a pedigree file

1. Calculate the SNP frequencies using 'plink_freq' in SOLAR:
   > plink_freq --plink UKBB_plink --o UKBB_freq
2. 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

Step 3: Create a phenotype file from the UKBB bulk data and normalize the trait data using 'sporadic_normalize' in SOLAR. Covariates usually include age and sex, which should be columns in your phenotype files.

1. Phenotype files should in a csv format and include all covariate and traits you would like to include in your analyses. In this example we will use grey matter volumes. The easiest way to do this with the UKBB data is to include the age and sex category ids when you unpack the UKBB data:
   $ ukbconv ukb12345.enc_ukb csv --itrait_ids_list --ophenotype_file
   Where trait_ids_list is your list of UBK data ids you want to include in our phenotype file.

Rename the column headers. In this example we will use grey matter volume of the thalamus as our trait of interest.

2. Create a header file
    - First use:
      $ head -n 1 phenotype_file.csv > phens.header
    - Then remove the id and covariate text, and then convert commas into spaces such that the file only contains the traits of interest delimitated by spaces:

      ```
      thalamus_volume Caudate_volume Putamen_volume Palladium_volume
      ```

3. Normalize the data by first loading the phenotype file, then using sporadic_normalize:
   > load phen phenotype_file.csv
   > covar Age Sex
   > sporadic_normalize -header phens.header -out normalized_phenotypes.csv

Step 4: Load your previously create pedigree and set your kinship threshold:

> load pedi UKBB_ped.csv -t 0

**Note: This may take a couple of minutes, but this command only needs to be run once per workspace.

Step 5: Create your eigenvalue and eigenvector decomposition file. The UKBB imaging dataset is so large that creating a separate file to store the eigenvalues/vectors significantly decreases run time.

- First load your phenotype file:
  > load phen normalized_phenotypes.csv
- Then select your trait:
  > trait thalamus_volume
- Run 'create_evd_data':
  > create_evd_data --o thalamus_evd

Step 6: Run the heritability analysis on your single selected trait:

- Use fphi with the evd_data option for optimized run time:
  > fphi --evd_data thalamus_evd