Gladstone Outline Document

Project title: Genetic Insights into the cardiac contribution to small vessel disease and cognition

For any questions [email:Adama.saccoh.22@ucl.ac.uk](mailto:email:Adama.saccoh.22@ucl.ac.uk)

Background information

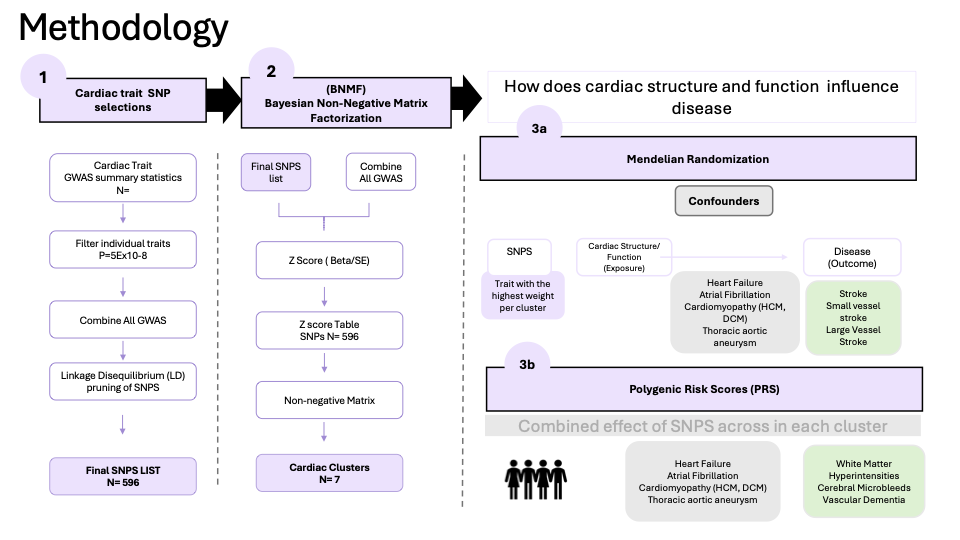
Bayesian non-negative matrix factorisation: A soft clustering method to group genetic variants and assign weights based on their contribution to the cluster. I have followed methods documented below .

Paper: https://www.nature.com/articles/s41591-024-02865-3

Github: <https://github.com/gwaspartitioning/bnmfclustering/blob/master/README.md>

GitHub repository of the files explained can also be found here: <https://github.com/AFSaccoh/Gladstone>

A visual workflow is detailed below



Cardiac traits and SNP selection

PRE-PROCESSING

Collating GWAS summary stats and preparing file for pruning

1. Use Step2 Regeinie summary statistics . Each cardiac trait has a separate summary statistics file (GWAS was carried out by Octavia)
2. Load these files and merge all GWAS summary statistics. You will need two versions

* Run pre\_pruned.py > This creates combined gwas summary stats file for each group of traits ["aortic\_area", "atrial\_volume","ventrical\_volume" ,"wall\_thickness"]. OUTPUT WILL BE four files {cardiac\_trait}\_prepruned.csv for all traits
* Run filtering.ipynb and this removes duplicates and creates a combined file ( all\_cardiac\_traits\_prepruned.csv) this can be the file used for pruning.
* Filter columns and creates Name=cardiac\_traits\_all\_final (Final TABLE USED FOR LD PRUNING)
* As gwas has LOG10P a new table is column pval is created and column. A new column is needed and duplicated SNP values displaying Chrom: pos:Allele1:Allele0 or Chrom:pos:Allele0:Allele1.
* Note: The combination that worked was Chrom: pos:Allele0:Allele1. Allele1 is the effect allele and Allele 0 is the alternate allele. This information will be useful when running MR analysis
* Using get\_full.ipynb for the full combination of summary statistics – this will be named RAW and not filtered and is needed for bnmf to obtain the z scores from the full table(( all\_cardiac\_traits\_raw.csv). This is the file that is used in the merge

1. LD pruning carried out (Hasan carried out the pruning)

Preparing z score table

1. Use extract SNP and chromosome to get list of all SNPS from each trait-> pruned\_df ( you are merging the SNPS from pruning with the gwas\_all\_file)
2. Merge with full gwas ( GWAS\_RAW)
3. Calculate z score
4. Filter only traits with correlation less than 0.8. A simple correlation of the z scores was done
5. Pivot table
6. Create non-negative matrix with function
7. Final file= all\_{date}\_bnmf matrix
8. Run the cut\_off script to get value and plot

BNMF

1. Use R file to run final\_ bnmf .R file
2. Final outputs are saved into a new folder folder
3. Use the L2EU.W.mat.7.txt as the file to create PRS scores (note would need to filter to only include scores that are above cut-off values)

Extra bnmf steps

Determine the cutoff value/appropriate weights to use and appropriate number of clusters

* The function for this can be found in the post\_bnmf script used by this paper
* For cut-off use the cut\_off function
* For elbow plot graph use the dist\_plot\_line

Mendelian Randomization

* Top weighted trait (significant SNPS only vs all SNPS) is run on the required outcomes
* Need to do the MR egger and follow up steps for results with heterogeneity

PRS SCORE based on bnmf clusters

PRS for CVD( Heart Failure, Atrial Fibrillation, HCM,DCM, TAAD, PVD)

* PRS in UKB in individuals without CMR
* PRS in MVP
* The result in UKB and MVP should be concordant

PRS for brain related traits

PRS in MVP

* Stroke – anything stroke related (subtype where phenotype definition possible: svs,is,nis (FILL this out)
* Alzheimers categories in MVP ( look for better phenotype definitions)
* Vascular Dementia

PRS in UKB

* WMH and brain MRI
* Stroke definitions and Alzheimers

PRS in SABRE cohort

* Stroke outcomes

This fact sheet contains information on classification of Alzheimer's disease which may be useful to get ICD codes for Vascular dementia codes

<https://allianceforcareathome.org/wp-content/uploads/Dementia_ICD10_Fact_Sheet.pdf>