**GWAS of PC1 of white matter hyperintensities and thickness of the insular cortex**

**Analysis plan, NeuroCHARGE Consortium**

Ariana Tang (ariana.tang@sickkids.ca)

Jean Shin (jean.shin@sickkids.ca)

Tomas Paus (tomas.paus@umontreal.ca)

Zdenka Pausova (zdenka.pausova@sickkids.ca)

University of Toronto

Version: 2022-MAY-\_\_-v1

**Aim**

*To identify genetic loci associated with the shared variance between white matter hyperintensities (WMH) and thickness of the insular cortex*

*Note:* This project is to be conducted in 2 stages: (*Stage 1*) to calculate the first principal component (PC1) scores for white matter hyperintensities (WMH) and thickness of the insular cortex, and (*Stage 2*) to perform genome-wide association tests for PC1. For Step 1, R scripts will be distributed.

**Subject exclusion criteria**

* Dementia (≥mild severity) at time of MRI-scanning
* Stroke at time of MRI-scanning (use exclusion criteria if available for study, either based on clinical data or large artery strokes/lacunes in region of interest on MRI)
* Multiple sclerosis (if available)

Brain surgery (if available)

* Morphological abnormalities (e.g., cysts, brain tumors)
* Poor technical quality
* Non-White/non-Caucasian Participants

**Brain outcomes**

* White matter hyperintensities (WMH): T1, T2 or FLAIR, total load as a quantitative variable is preferable, but if not available, total load as a semiquantitative variable can be used.
* Cortical thickness (CT) averaged between the left and the right hemispheres in the 34 FreeSurfur-parcellated regions.
* First principal of WMH and thickness of the insular cortex.

**Covariates**

1. Age: linear and quadratic terms
2. Sex
3. Intracranial volume (ICV) or brain size
4. Current smoking (yes/no)
5. Hypertension (yes/no)
6. Type 2 diabetes (yes/no)
7. Cohort specific covariates: e.g., MR-scanner, etc.
8. Genetic principal components (cohort-specific numbers): Step 2 only

**Stage 1 (run with the provided scripts)**

* Descriptive statistics will be calculated for the brain phenotypes and covariates (#1-6)
* Brain outcomes will be first inverse-normal transformed (INT) and then adjusted for covariates:

**M1:** INT(WMH) ~ Age + Age2 + ICV or Brain size + hypertension + type 2 diabetes + current smoking + cohort specific covariates

**M2**: INT(insular CT) ~ Age + Age2 + hypertension + type 2 diabetes + current smoking + cohort specific covariate

Models M1 and M2 will be fitted within each sex.

* PC1 scores for WMH and insular CT will be calculated using the residuals obtained from fitting M1 and M2.
* R scripts for running principal components analyses will be provided: Instructions how to run the scripts are provided in Appendix 1.

**Stage 2: Genome-wide association study (in the sex-combined participants)**

* Use the PC1 scores returned after running Stage 1: the file is in ‘files\_for\_GWAS’ folder

**M3:** PC1 ~ SNP + Age + Age2 + sex + ICV or brain volume + hypertension + type 2 diabetes + current smoking status + cohort-specific covariates + genetic PCs

* Please format the GWAS result file with the following column names:

|  |  |
| --- | --- |
| **Column header/Variable** | **Description** |
| ID | The dbSNP or 1000 Genome Project ID number |
| CHROM | Chromosome number |
| POS | Base pair position of SNP on the chromosome, Build 37 (or hg19) |
| REF | Reference allele |
| ALT | Alternate allele |
| A1 | Allele for which the effect has been estimated |
| A1\_FREQ | Allele frequency of the effect allele |
| TEST | Not sure what this is, but in our GWAS the column says “ADD” |
| OBS\_CT | Number of individuals used in the association analysis |
| BETA | Effect size for A1, at least \_\_ decimal places (NA if not available) |
| SE | Standard error for A1, at least \_\_ decimal places (NA if not available) |
| T\_STAT | Is this the t-score for association analysis |
| P | P-value for A1, at least \_\_ decimal places (NA if not available |

**Study information tables**:

Please return the following 2 tables

Email the file to **Jean Shin (jean.shin@sickkids.ca)**

1. General information of the study

Example



1. Cohort characteristics table

|  |  |
| --- | --- |
| **Cohort** | **Cohort name** |
| **Basic characteristics** |  |
| Number of individuals analyzed (n) |  |
| Male, n (%) |  |
| Current smokers, n (%) |  |
| Hypertension, n(%) |  |
| Type 2 Diabetes Mellitus, n (%) |  |
| Age (years), mean (SD) |  |
| ICV or brain volume (mm3), mean (SD) |  |
| **White matter hyperintensities** |  |
| Description of method used for quantification |  |
| Reference describing the method, if applicable |  |
| **Cortical thickness** |  |
| Description of method used for quantification |  |
| Reference describing the method, if applicable |  |
| **GWAS analysis parameters** |  |
| Genotyping platform – SNP panel |  |
| Sample call rate threshold |  |
| SNP call rate threshold |  |
| MAF threshold (if applicable) |  |
| HWE P-value threshold |  |
| Assessment of population stratification |  |
| Other cohort-specific covariates used |  |
| Analysis software(s) used |  |
| **Other** |  |
| Other relevant details |  |

If you have any questions about the analysis plan or the analysis itself, please contact the analysis group: Jean Shin ([jean.shin@sickkids.ca](mailto:jean.shin@sickkids.ca)) or Ariana Tang (ariana.tang@sickkids.ca)

Analysis Deadline:

For uploading the results, provide us with your google ID to get the access to our CHARGE project google drive.

**Appendix 1: Instructions to run the R scripts for Stage 1 protocols**

The R scripts and this analysis plan files can be downloaded from the following link:

https://github.com/jshinb/neuroCHARGE\_scripts/tree/main/R

**Steps**

**Step A. QC all variables (i.e., brain outcomes, and covariates):**

For all variables, please remove technical and/or statistical outliers.

For brain data, please see exclude participants with

* Dementia (≥ mild severity) at time of MRI-scanning
* Stroke at time of MRI-scanning (use exclusion criteria if available for study, either based on clinical data or large artery strokes/lacunes in region of interest on MRI)
* Multiple sclerosis (if available)
* Brain surgery (if available)
* Morphological abnormalities (e.g., cysts, brain tumors)

**Step B. Specify cohort-specific naming of variables**

* Complete 'cohort\_specific\_inputs.txt', available in the dropbox, to specify the names you use for the variables required in Step C, and coding you use for binary variables (please see ‘cohort\_specific\_inputs\_example.txt’, for example).

**Step C. Prepare text files (e.g., tab- or comma-delimited) for the following 3 sets of variables** (Note: all missing values must be coded as NA)**:**

1. Brain outcomes:
2. Cortical thickness averaged between the left and the right hemispheres in 34 FreeSurfer-parcellated regions: Note – Please use the headers in (‘BrainOutcome\_Headers.txt’).
3. Total volume of white matter hyper- (or hypo-) intensities from T2 or T2\* (or T1): Total load as a quantitative variable is preferable, but if not available, total load as a semiquantitative variable can be used.

Example

Table

Description automatically generated

1. Covariates
2. age
3. sex
4. current smoking status (binary)
5. hypertension status (binary)
6. type 2 diabetes status (binary)
7. intracranial volume (ICV) or brain size

Example

Table

Description automatically generated

1. Other cohort-specific variables (not included in #2, if available)

Example

Table

Description automatically generated

**Step D. Put all the files in the working directory where the scripts are stored.**

**Step E. Run (i.e., source) the edited file after specifying the input file:**

Source after editting line 10 in ‘run\_protocols\_NeuroCHARGE.R'

**Step G. Run GWAS with the generated ‘PC1’ scores.**

**Step F. Upload the following two sets of output files to the provided NeuroCHARGE google link:**

1. Stage 1 file: Compress the created output directory and upload the compressed file with naming: 'Study\_Ancestry\_Stage1\_DDMMYYY.zip' (e.g., SYS\_EUR\_Stage1\_09052022.zip)
2. Stage 2 file: Gzipped GWAS result file with naming: ‘Study\_Ancestry\_PC1\_WMH\_insularTH\_DDMMYYYY.txt.gz’ (e.g., SYS\_EUR\_PC1\_WMH\_insularTH\_09052022.txt.gz)

[https://drive.google.com/open?id=1UiTSC9OV03lcO1zRvLxfdX5gnR7pKv95](https://urldefense.proofpoint.com/v2/url?u=https-3A__drive.google.com_open-3Fid-3D1UiTSC9OV03lcO1zRvLxfdX5gnR7pKv95&d=DwMGaQ&c=Sj806OTFwmuG2UO1EEDr-2uZRzm2EPz39TfVBG2Km-o&r=D3BEyk1TXLl4yW5gFWlCXks_wMQBa8FtJYfmXKaNtHo&m=piUVNiS3WmY5fsCaByNRenCCdjWZlYHgWnFwOH9FI2A&s=OJ7VFNbUDuJ8-R1HhyexNNUcKOo7lvKV7tNo4ZV_Wgc&e=" \o "https://urldefense.proofpoint.com/v2/url?u=https-3A__drive.google.com_open-3Fid-3D1UiTSC9OV03lcO1zRvLxfdX5gnR7pKv95&d=DwMGaQ&c=Sj806OTFwmuG2UO1EEDr-2uZRzm2EPz39TfVBG2Km-o&r=D3BEyk1TXLl4yW5gFWlCXks_wMQBa8FtJYfmXKaNtHo&m=piUVNiS3WmY5fsCaByNRenCCdjWZlYHgWnFw)