**Metabolomic studies of white matter hyperintensities and microstructural properties of the brain**

**Analysis plan, NeuroCHARGE Consortium**

Jean Shin (jean.shin@sickkids.ca)

Eeva Sliz (eeva.sliz@sickkids.ca)

Catriona Syme (catriona.syme@sickkids.ca)

Steve Tilley (stilley@hollandbloorview.ca)

Tomas Paus (tpaus@hollandbloorview.ca)

Zdenka Pausova (zdenka.pausova@sickkids.ca)

University of Toronto

Version: 2019-OCT-01-v1

**Aim**

*To identify circulating metabolites associated with white matter hyperintensities (WMH) and microstructural properties of the brain*

*Note:* This project is to be conducted in 2 phases: (*Phase 1*) to perform association tests between circulating metabolites and WMH, and (*Phase 2*) to perform association tests between circulating metabolites and brain microstructural properties

**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

**Circulating metabolites** (serum or plasma from the following platforms)

* Nightingale
* Metabolon (Global Metabolomics and/or Complex Lipids platform)
* Biocrates
* The Broad (named/identified metabolites only)
* Custom platform (named/identified metabolites only)

*Note*: All metabolites will be analyzed regardless of missing rates– the metabolites with high missing rates will be addressed in the meta-analyses.

**Brain outcomes**

*Phase 1*: 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

*Phase 2*: Microstructural properties of the brain:

* Diffusion tensor imaging, mean diffusivity of grey matter and white matter
* Diffusion tensor imaging, fractional anisotropy of grey matter and white matter
* T1-weighted normalized signal intensity of grey matter and white matter

*Note***:** FSL scripts can be requested if any of microstructural properties need to be derived.

**Covariates**

* Age (linear)
* Sex + interaction with age
* ‘Years’: years between brain MRI and blood drawing for metabolomic analysis
* ‘Fasting duration’: hours between the last meal and blood draw for metabolomic analysis (if available)
* Intracranial volume (ICV) or brain size (only for WMH analysis)
* Cohort specific covariates: e.g., MR-scanner, batch, etc.
* Additional covariates:

1. statin treatment (yes/no)
2. BMI
3. current smoking (yes/no)
4. estimated glomerular filtration rate, eGFR (mL/min per 1.73 m2)

<https://www.niddk.nih.gov/health-information/communication-programs/nkdep/laboratory-evaluation/glomerular-filtration-rate-calculators>

**Association analyses**

* Linear regression models will be fitted to the standardized (i.e., z-scored) brain outcomes and metabolites (within each platform separately).
* If multiple major ethnicities included (Caucasian, African American, Hispanics), run analyses stratified by ethnicity.
* Models (for WMH only, include ICV or Brain size as a covariate):

1. Whole sample

**M1:** Brain outcome ~ metabolite + Age + Sex + Age\*Sex + Years + Fasting duration (+ ICV or Brain size) + cohort specific covariates

**M2**: Brain outcome ~ metabolite + Age + Sex + Age\*Sex + Years + Fasting duration (+ ICV or Brain size) + cohort specific covariates + **all additional covariates (#1-#4)**

Fit models M1 and M2 within one group:

* all participants

1. Sex stratified

**M3**: Brain outcome ~ metabolite + Age + Years + Fasting duration (+ ICV or Brain size) + cohort specific covariates

**M4**: Brain outcome ~ metabolite + Age + Years + Fasting duration (+ ICV or Brain size) + cohort specific covariates + **all additional covariates (#1-#4)**

Fit models M3 and M4 within two groups:

* all females
* all males

1. Sensitivity analysis: statin stratified (yes/no)

**M5**: Brain outcome ~ metabolite + Age + Sex + Age\*Sex + Years + Fasting duration (+ ICV or Brain size) + cohort specific covariates + **additional covariates (#2-#4)**

Fit models M1 and M5 within two groups:

* all participants on statin
* all participants not on statin

*Note*: These statin-stratified analyses will exclude the individuals on non-statin lipid-lowering medication.

* R scripts for running association analyses will be provided: Instructions how to run the scripts are provided in Appendix 1.

**Study information tables**:

Please complete the following 3 tables in the excel file (Analysis\_information\_tables.xlsx) located here: https://www.dropbox.com/sh/9tyboajdyx786ek/AADSLrTfiaD3dRKbtvsg-Y5Ja?dl=0

Email the file to **Jean Shin (jshinb@gmail.com)**

1. General information of the study

Example



1. Participant characteristics table (one column per platform)

Example



1. Metabolite information table: names/units of metabolites included in the analyses as shown in the following example (one table per platform):

Example



If you have any questions about the analysis plan or the analysis itself, please contact the analysis group: Jean Shin (jean.shin@sickkids.ca), Eeva Sliz (eeva.sliz@sickkids.ca) and Catriona Syme (catriona.syme@sickkids.ca).

Analysis Deadline: ~~November 15, 2019~~ (New Deadline: March 15, 2020)

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 Phase 1 (WMH) analysis**

The R scripts and this analysis plan files can be downloaded from the following link: https://www.dropbox.com/sh/9tyboajdyx786ek/AADSLrTfiaD3dRKbtvsg-Y5Ja?dl=0

These modifiable R scripts are for *unrelated* participants. Please contact Jean Shin (jshinb@gmail.com) in the case of family data.

**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 ‘Example\_cohort\_specific\_inputs.txt’, for example).

**Step C. Prepare tab-delimited files (\*.tsv) for the following 3 sets of variables** (Note: all missing values must be coded as NA)**:**

1. Brain variables:

* Outcome: 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.
* Intracranial volume or brain size
* MRI-related cohort-specific variables (e.g., MR scanner site)

Example



1. Metabolites: \*\*one file per platform\*\*

* Metabolites: Names in this .tsv must match the metabolite names in Step C
* Platform-related cohort-specific variables (e.g., Batch).

Example



1. Covariates [Please include all individuals who have brain or metabolite data]

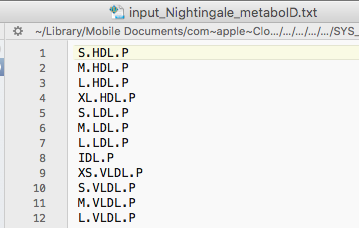
* age
* sex
* years between MRI and and blood drawing for metabolomic analysis
* fasting duration: hours between the last meal and blood draw for metabolomic analysis (if available)
* statin use (binary)
* other (i.e., non-statin) lipid lowering medication use (binary)
* BMI
* current smoking status (binary)
* eGFR (kidney function, estimated GFR (mL/min per 1.73 m2)
* other cohort-specific variables (not included in #2 and #3)

Example



**Step D. Prepare a list of the metabolite variables for each platform and save it as a text file [\*\*with no header\*\*]: one file per platform.**

Example

****

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

**Step F. Set the working directory:**

Edit line 17 in 'NeuroCHARGE\_CirculatingMetabolome\_Brain\_WMH\_association\_analysis\_2019-10-01.r'

**Step G. Run (i.e., source) the edited file**: 'NeuroCHARGE\_CirculatingMetabolome\_Brain\_WMH\_association\_analysis\_2019-10-01.r'.

**Step H. Compress the created output directory ('cohort\_name\_ancestry') and upload the compressed file to NeuroCHARGE Google drive link:**

[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=)