Team Meeting

29 APRIL 2022 / 11:00 AM / ZOOM

Attendees

Zexi, Stefan, Xavier, Ate, Nancy, Jiadong, Shuai, John, Vivienne

Agenda

Summary of last few weeks

1. We examined carefully on the paper and realized the results of 4 classifiers trained on different input sets, MZ vs DZ, and Mz vs the others
2. We tried to read and process big files, and directly train a classifier on 450K predictors (in progress)
3. We found some codes of the existing classifier for testing, but need to confirm with the client more carefully about what the classifier means in the code, and how to use it.

Part 1:

1. Is the predictor provided in Software 1

   #### Code availability

   An R-script (EpiPredictorMZtwin.R) and accompanying R data object to apply the epigenetic predictor of MZ twinning is provided in Supplementary Sofware 1. The pipeline for DNA methylation-array analysis developed by the Biobank-based Integrative Omics Study (BIOS) consortium are available here: https://molepi.github.io/DNAmArray_workflow/ (https://doi.org/10.5281/zenodo.3355292). All other analysis code is available upon request from the corresponding author.

2. In the classifier, it has 883 columns, which was claimed in

In the training data, zygosity was regressed on all methylation sites ($N = 381{,}376$) that (1) were present both on the Illumina 450K and EPIC array (2) survived quality control in the training set (NTR-blood) and in the test data sets (NTR-buccal, Australia-blood). Second, we tested training on the subset of epigenome-wide significant CpGs from the meta-analysis (833 CpGs that were also present on the EPIC array). The alpha parameter of glmnet was set to 0.5 (elastic net regression) and the lambda value was selected by taking the minimum lambda using 10-fold cross-validation on the training data with the AUC method (R command: cv.glmnet($x =$ methylation, $y =$ zygosity, alpha $= 0.5$, nfolds $= 10$, family $=$ "binomial", type.measure $=$ "auc")).

3. If so, how do you use predict() in, "they combine different variables to get the beta value, use the beta value to predict the MZ/NZ?

4. The weights of the two models are provided, can we use them, and how to use them?

> File Name: Supplementary Data 13
> Description: Weights from the elastic net regression model trained on meta-analysis significant probes to predict zygosity

> File Name: Supplementary Data 14
> Description: Weights from the elastic net regression model trained on meta-analysis significant probes to distinguish MZ twins from DZ twins and non-twins

Part 2 (building own classifier):

5. Problem with merging, no computing power => HPC

Notes

- The structure of the data is discussed
- Based on the code that is presented in the paper, the classifier is discussed.

Action Items

1. Filter with 833 predictors line by line, load it in R, easier to manipulate,
2. Use around 200 number of coefficients,
3. Send the link of this paper to clients,

Next Meeting Agenda