Pleiotropic Effects of CSF Levels of Alzheimer’s Disease Proteins

Olga A. Vsevolozhskaya  
*University of Kentucky, vsevolozhskaya@uky.edu*

Ilai Keren  
*Washington Department of Fish & Wildlife*

David W. Fardo  
*University of Kentucky, david.fardo@uky.edu*

Dmitri V. Zaykin  
*National Institute of Environmental Health Sciences*

*Click here to let us know how access to this document benefits you.*

Follow this and additional works at: [https://uknowledge.uky.edu/biostatistics_present](https://uknowledge.uky.edu/biostatistics_present)

Part of the [Biostatistics Commons](https://uknowledge.uky.edu/biostatistics_present)

Repository Citation  
[https://uknowledge.uky.edu/biostatistics_present/4](https://uknowledge.uky.edu/biostatistics_present/4)

This Presentation is brought to you for free and open access by the Biostatistics at UKnowledge. It has been accepted for inclusion in Biostatistics Presentations by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.
Pleiotropic Effects of CSF Levels of Alzheimer’s Disease Proteins

Background & Motivation

Cerebrospinal fluid (CSF) analytes harbor potential as diagnostic biomarkers for Alzheimer’s Disease (AD). Quantitative measures of CSF proteins comprise a set of often highly correlated endophenotypes that have previously shown promise in genetic analyses (Cruchaga et al., 2013; Kauwe et al., 2014). Pleiotropic impact of genetic variations on this set may provide additional insights into AD pathology at its earliest stages. To determine which specific endophenotypes are pleiotropic, one can employ methods based on the reverse regression of genotype on phenotypes. Recently, we proposed a method based on functional linear models (Vsevolozhskaya et al., 2016) that utilizes reverse regression and simultaneously evaluates all variants within a genetic region for an association with multiple correlated phenotypes. Here we apply our novel methodology to explore pleiotropic effects of CSF analytes using Alzheimer’s Disease Neuroimaging Initiative (ADNI) data.

Materials

- CSF levels of β-amyloid (Aβ) and β-Tau proteins are potentially early diagnostic markers for probable AD. We investigated the influence of genetic variation on these markers for two candidate genes: APOE and CLU.
- Data were obtained from the ADNI database. A total of 309 (AD = 79, MCI = 148, NL = 82) individuals were included.
- 526 APOE single nucleotide polymorphism (SNP) and 569 CLU SNPs were available for the analysis.

### Results: Exploring the estimated $\hat{\beta}_{\text{Tau}}(t)$ and $\hat{\beta}_{\text{Abeta}}(t)$

#### APOE: no ε4 adjustment

- APOE genotype is the established risk factor for LOAD. The two previously identified APOE SNPs (rs429358, rs7412) that define the ε2/ε3/ε4 alleles for AD susceptibility are among the top SNPs that drive the association with the elevated Aβ levels in our model. Additionally, TOMM40/APOE rs7258492 that was recently identified to be associated with longer human longevity (Lin et al., 2016) is clearly associated with lower Aβ levels in our analysis. TOMM40/APOE rs2075650 is also associated with lower Aβ levels under the functional inference, confirming prior report by Kim et al., 2011.

- The ε4 adjustment did not have a significant impact on the association between T levels and AD, for Aβ the region over APOE SNPs rs429358, rs7412 is not longer significant. Nonetheless, the second interval still contains a statistically significant signal.

#### Methods

- Our method is gene- or region-based. What makes it stand apart is the fact that a genetic region serves as an outcome in the model, not a predictor (i.e., the so-called reverse regression).

- The genetic information does not enter the model as discretized allele counts (e.g., 0, 1, 2) but rather as a smooth function $G(t)$, with $t$ indexing a genetic variants’ position over a genetic region.

- Fig. 1 illustrates genotypic functions fitted for each subject, $G(t)$, $i = 1, \ldots, 309$, over the APOE region.

- Statistical Analysis

- To examine the effect of SNPs on the 2 CSF biomarkers, we fit the following functional linear model:

$$E[G(t)] = \beta_0(t) + \beta_1(t)X_{\text{Tau}} + \beta_2(t)X_{\text{Abeta}} + \beta_3(t)X_A$$

$$\beta_4(t)X_{\text{Sex}} + \beta_5(t)X_{\text{Age}} + \beta_6(t)X_{\text{Diag}} + \sum_{j=1}^{3} \beta_{6+j}(t)X_{\text{PC}_j}$$

- The model is looking for an association with a CSF biomarker, while adjusting for the level of the other protein, gender, age and diagnosis status. Also, to ensure the significance is not due to population stratification, we incorporated the top three principal components to further control for the population structure.

- Two analyses were performed: with APOE ε4 genotype (number of APOE ε4 alleles; 0, 1, 2) as a covariate and without.

- The analysis is Bayesian; the estimated $\hat{\beta}(t)$’s are obtained as posterior expectations; the confidence bands are obtained as 95% posterior credible intervals.

Cervantes et al., 2011, tested APOE rs79420216 and rs67214972 for an association with AD, but after multiple test correction their allelic P-value were found to be non-significant (0.747 and 0.053 respectively). Nonetheless, in our analysis, the credible interval for the genetic effect of both of these SNPs does not capture zero and they were found to be associated with the elevated T levels.