GENE-LEVEL PATHOGENICITY SCORES FOR ALZHEIMER'S DISEASE USING GENOMIC VARIANTS FROM RNA-SEQ DATA

Fatma Betül Bozkurt, Atılay İlgün, Dilara Uzuner, Tunahan Çakır

Department of Bioengineering, Gebze Technical University
Gebze Technical University, 41400, Kocaeli, Turkey
email: f.bozkurt2021@gtu.edu.tr

ABSTRACT

Alzheimer's disease (AD) is a complex neurodegenerative disorder affecting millions of people worldwide. Next-generation sequencing technologies such as whole-exome/genome sequencing have been widely used for detecting the variants in the genome to understand the disease etiology and unravel underlying molecular mechanisms. Alternatively, RNA-Seq data can also be used to detect variants. Since AD is a complex disease, several variants are involved in the disease pathogenesis. By using scoring algorithms, it is possible to estimate gene pathogenicity combining the pathogenicity scores of all variants in that gene into a single score. In this project, three large RNA-Seq datasets comprising 704 AD patients and 313 controls in total were used to determine gene pathogenicity: The Religious Orders Study/Memory and Aging Project (ROSMAP) [1], Mayo Clinic [2], and Gene Expression Omnibus (GSE125583) [3]. To identify variants, the raw data was analyzed, and variant call format (VCF) files were generated. The VCF files annotated by ANNOVAR [4] were used for calculating gene pathogenicity scores (per-gene, per-individual) by GenePy algorithm [5]. These scores reflect the combined pathogenic burden of non-synonymous, frameshift, and stop-loss/gain variants. One-tailed Mann-Whitney u-test was performed to detect genes with significantly greater GenePy scores in AD patients compared to controls. Gene Ontology (GO) enrichment analysis of genes with high pathogenicity identified GO terms known to be associated with AD such as neuron differentiation, neurogenesis, neuron projection morphogenesis, nervous system development, and several terms related to synapse, axon, and dendrite. Among the predicted pathogenic genes, several genes were detected to be reported as AD-related by previous studies, and as AD risk loci by genome-wide association studies (GWAS). This work provides an example for using RNA-Seq data to detect variants in individuals with AD and using gene pathogenicity scoring algorithms to predict pathogenic genes.

This study was financially supported through a grant by TUBITAK (Project Code: 120S824).

References

1. de Jager, P. L., Ma, Y., McCabe, C., Xu, J., Vardarajan, B. N., Felsky, D., Klein, H.-U., White, C. C., Peters, M. A., Lodgson, B., Nejad, P., Tang, A., Mangravite, L. M., Yu, L., Gaiteri, C., Mostafavi, S., Schneider, J. A., & Bennett, D. A. (2018). A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Scientific Data*, *5*(1), 180142. https://doi.org/10.1038/sdata.2018.142

2. Allen, M., Carrasquillo, M. M., Funk, C., Heavner, B. D., Zou, F., Younkin, C. S., Burgess, J. D., Chai, H.-S., Crook, J., Eddy, J. A., Li, H., Logsdon, B., Peters, M. A., Dang, K. K., Wang, X., Serie, D., Wang, C., Nguyen, T., Lincoln, S., … Ertekin-Taner, N. (2016). Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Scientific Data*, *3*(1), 160089. https://doi.org/10.1038/sdata.2016.89

3. Srinivasan, K., Friedman, B. A., Etxeberria, A., Huntley, M. A., van der Brug, M. P., Foreman, O., Paw, J. S., Modrusan, Z., Beach, T. G., Serrano, G. E., & Hansen, D. v. (2020). Alzheimer's Patient Microglia Exhibit Enhanced Aging and Unique Transcriptional Activation. *Cell Reports*, *31*(13), 107843. https://doi.org/10.1016/j.celrep.2020.107843

4. Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, *38*(16), e164–e164. https://doi.org/10.1093/nar/gkq603

5. Mossotto, E., Ashton, J. J., O'Gorman, L., Pengelly, R. J., Beattie, R. M., MacArthur, B. D., & Ennis, S. (2019). GenePy - a score for estimating gene pathogenicity in individuals using next-generation sequencing data. *BMC Bioinformatics*, *20*(1), 254. https://doi.org/10.1186/s12859-019-2877-3