DOI 10.18699/SBB-2022-264

New framework for gene-based association analysis using GWAS summary statistics and functional annotations

Svishcheva G.[1, 2]*, Belonogova N.[1], Kirichenko A.[1], Zorkoltseva I.[1], Tsepilov Ya.[1, 3], Axenovich T.[1, 3]

[1] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[2] *Vavilov Institute of General Genetics, RAS, Moscow, Russia*

[3] *Novosibirsk State University, Novosibirsk, Russia*

* *gulsvi@bionet.nsc.ru*

Key words: gene-based association analysis, rare variants, summary statistics, genotypic correlations

Motivation and Aim: Gene-based association analysis using GWAS summary statistics is an effective tool for the mapping of common and rare genetic variants. However, none of the developed gene-based methods is a universal approach that would be effective for any trait with any genetic architecture. In this regard, the combination of different gene-based methods can be a universal and effective approach to the gene-based association analysis. Recently, several frameworks have been proposed to combine tests using various strategies for gene-based association analysis. They differ from each other in goals (one or more traits), input data (individual phenotypes and genotypes or the GWAS summary statistics) and ways to use functional annotations. However, these frameworks are not flexible, because they use a fixed set of methods, weights and combinations of functional annotations.

Methods and Algorithms: Here we propose a flexible framework for gene-based analysis using GWAS summary statistics. We implemented this framework in our sumFREGAT package [1], which now includes the aggregated Cauchy test, ACAT [2], and the probabilities of SNPs being causal [3] in Burden test, SKAT, SKAT-O, ACAT-V, PCA and FLM. For the new framework, the inputs are GWAS summary statistics, genotypic correlations, SNP-level weights defined by beta-distribution functions of MAFs and the probabilities of SNPs being causal, which are estimated by various functional annotations. First, the user arbitrarily defines a set of gene-based methods, a set of weights and a set of probabilities of SNPs being causal. Then all selected gene-based tests are run one-by-one, and the resulting test statistics are combined by ACAT.

Results: The new framework was successfully tested on neuroticism and coronary artery disease GWAS data from the UK Biobank and allowed us to identify several additional genes. The updated sumFREGAT is available at https://cran.r-project.org/web/packages/sumFREGAT/index.html.

Conclusion: The presented framework allows users to carry out a wide range of arbitrary defined gene-based association tests differing from each other by statistical method, weighting function, functional annotation, and then to combine these tests.

Acknowledgements: The study is supported by Russian Foundation for Basic Research (20-04-00464).

References

1. Svishcheva G.R., Belonogova N.M., Zorkoltseva I.V., Kirichenko A.V., Axenovich T.I. Gene-based association tests using GWAS summary statistics. *Bioinformatics.* 2019;35(19):3701-3708.
2. Liu Y., Chen S., Li Z., Morrison A.C., Boerwinkle E., Lin X. Acat: A fast and powerful p value combination method for rare-variant analysis in sequencing studies. *Am J Hum Genet.* 2019;104(3):410-421.
3. Li X., Li Z., Zhou H., Gaynor S.M., Liu Y., Chen H., Sun R., Dey R., Arnett D.K., Aslibekyan S. Dynamic incorporation of multiple *in silico* functional annotations empowers rare variant association analysis of large whole-genome sequencing studies at scale. *Nat Genet.* 2020;52(9):969-983.