DOI 10.18699/SBB-2022-267

New genes predisposing to coronary artery disease detected by genome-wide gene-based association analysis

Zorkoltseva I.[1]*, Belonogova N.[1], Shadrina A.[1], Kirichenko A.[1], Tsepilov Y.[1, 2], Axenovich T.[1, 2]

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

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

[*] *zor@bionet.nsc.ru*

Key words: coronary artery disease, genome-wide gene-based association analysis, summary statistics

Motivation and Aim: Coronary artery disease (CAD) is one of the most common causes of death or disability worldwide, despite recent advances in its treatment and prevention. The etiology of CAD is complex and involves a genetic predisposition as well as various environmental factors. To date genome-wide association analyses (GWAS) have identified more than 300 single nucleotide polymorphisms at 163 genetic loci associated with CAD. However, there is no full understanding about the causal genes for CAD and the mechanisms of their action. We aimed to perform a detailed analysis of genes whose polymorphism may influence the risk of CAD.

Materials and Methods: Using the UK Biobank-based GWAS summary statistics (https://www.leelabsg.org/resources), we performed a three-stage gene-based association analyses of two CAD phenotypes, Ischaemic Heart Disease and Coronary atherosclerosis. We used four sets of genetic variants within a gene differing in their protein coding properties. We applied a combination of three methods, SKAT-O, PCA, and ACAT-V, implemented in sumFREGAT package (https://cran.r-project.org/web/packages/sumFREGAT/index.html) and used a 'polygene pruning' procedure to eliminate the influence of strong GWAS signals outside the gene.

Results: We found 123 genes significantly associated with CAD. Using the extended sample, we validated 63 of these genes. They contribute to CAD with their within-gene variants. Among 63 validated CAD-associated genes: 1) 26 genes did not contain any SNPs reaching the genome-wide significance threshold of $5.0 \cdot \times 10^{-8}$ in GWAS; 2) The proportion of genes being closest to the top GWAS signals was 0.52; 3) 17 genes were identified using protein-coding SNPs sets. Ten of them included 25 potentially pathogenic variants, 4) 44 genes were identified using a set of non-coding within-gene variants. For 10 out of them we found evidence of pleiotropy with gene expression. Many of identified genes are well known. Some known CAD genes such as *FURIN* and *SORT1* did not show the gene-based association because their variants had low GWAS signals or gene-based association signals were inflated by the strong GWAS signal outside the gene. For several known CAD genes, we demonstrated that their effects can be explained not only or not at all by their own variants but by the variants within the neighboring genes controlling their expression. Three genes, *CDK19, NCALD* and *ARHGEF12* were not previously associated with CAD. *Conclusion*: We identified 63 genes that contribute to CAD with their within-gene variants, prioritized genes in known loci and identified three genes in the new CAD loci. The role of these genes should be clarified in further studies.

Acknowledgements: This work was supported by the Federal Agency of Scientific Organizations via the Institute of Cytology and Genetics, SB RAS (project 0259-2021-0009/AAAA-A17-117092070032-4) and the Russian Foundation for Basic Research (project 20-04-00464).