DOI 10.18699/SBB-2022-217

Search for novel genetic variants associated with the variability of cognitive functions in the elderly according to whole exome sequencing data

Bocharova A.[1]*, Vagaitseva K.[1], Zarubin A.[1], Makeeva O.[2], Zhukova N.[2, 3], Zhukova I.[2, 3], Stepanov V.[1]

[1] *Research Institute of Medical Genetics, Tomsk National Research Medical Center, Tomsk, Russia*

[2] *Nebbiolo Center for Clinical Trials, Tomsk, Russia*

[3] *Siberian State Medical University, Tomsk, Russia*

** anna.bocharova@medgenetics.ru*

Key words: cognitive function, whole exome sequencing, genome, cognitive decline

Motivation and Aim: Cognitive decline with the age, both in the normal aging and in the pathological manifestations in form of dementia, is an important public health and social challenge. Next generation sequencing (NGS) is increasingly being used to identify known and novel gene mutations. In this study we used deep (> 50x) whole-genome sequencing (WGS) to search for novel potential markers of cognitive functions in the coding and regulatory parts of the genome in 86 elderly individuals without established diagnoses of neurodegenerative and mental diseases.

Methods and Algorithms: The process for enrichment of genomic DNA sample libraries was carried out using SeqCap EZ Library Kit (Roche, USA) according to the SeqCap EZ Library SR User's Guide Version 5.1 following the recommended protocols. Whole exome sequencing (WES) was performed using Illumina HiSeq 2500 platform (San Diego, CA, USA) on 86 samples from a group of elderly individuals without established diagnoses of neurodegenerative and mental diseases. A SeqCap EZ MedExome Enrichment Kit (Roche, USA) was used for exome sequencing. The total volume of nucleotide sequenced sequences is 47 million bp and includes exons, introns, 5'- and 3'-untranslated regulatory regions. The whole sequenced data aligned to the human reference genome (GRCh38) using the BWA-mem algorithm, version 0.7.10. Further, Genome Analysis Toolkit v.3.3 (GATK; Broad Institute) was used for data quality assurance as well as variant discovery. Variant characterization, including filtering, annotation, classification, prioritization and inheritance pattern analysis, was performed using the SnpEff toolbox and proprietary algorithms in the R environment. To elucidate the novel genetic variants associated with the variability of cognitive functions in the elderly we performed an exome-wide association study (EWAS). The compared groups comprised samples of individuals with the highest MoCA test values versus a sample of individuals with low test values. Fisher's exact test was chosen to compare the samples. The analysis was performed in the R environment using own script. The String and Gene Ontology databases were used for pathway and disease association enrichment analysis.

Results: In total, 310138 variants were found as a result of exome sequencing of 86 samples: 257711 were single nucleotide variants (SNV), 52427 are insertion-deletion polymorphisms (INDEL). An analysis of the distribution of variants by localization in the gene showed that 24.0 % of the detected variants are localized in introns, 12.7 % – in exons, 10.7 % – in upstream regions, 8.1 % – in downstream regions, 1.5 % – in 3'-

untranslated regions, 0.9 % – in 5'-untranslated regions, 1.3 % – in intron splicing sites. The average density of variants by exome was 1 variant per 9.98 bp. The array of variants for further analysis amounted to 28449 markers after the removal of singletons, selection of variants by frequency in the sample (at least 5 %), and additional filtering by the quality of genotypes (DP > 10 in each sample). Of these, 14528 were localized in exons, including 6473 non-synonymous mutations. After correction for multiple testing, only 4 markers were reached the exome-wide significance threshold: rs71148121 in *PKD*1, rs9898911 in *ATAD5*, rs8069561 and rs692529 in *DNAH17*.

Conclusion: Previously, associations with cognitive abilities, memory, Alzheimer's disease and other neurodegenerative diseases were not found for these markers. Analysis of these genes using the String and Gene Ontology databases showed that these genes are involved in processes associated with a change in the state or activity of the cell as a result of a stimulus indicating DNA damage due to environmental influences or metabolic errors (*ATAD5*), processes associated with cell localization (*PKD1*), and with microtubule processes, including microtubule-based movement (*DNAH17*).

Acknowledgements: The study is supported by The Research Institute of Medical Genetics, Tomsk National Research Medical Center (122020200083-8).