DOI 10.18699/SBB-2022-445

Preliminary bioinformatic analysis to optimize the testing of the *SLC26A4* pathogenic variants involved in hearing loss

Danilchenko V.Yu.[1, 2]*, Zytsar M.V.[1], Maslova E.A.[1, 2], Posukh O.L.[1, 2]

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

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

* *danilchenko_valeri@mail.ru*

Key words: hearing loss, gene *SLC26A4*, pathogenic variants, diagnostics

Motivation and Aim: Over 5 % of the world's human population affects by hearing loss (HL), and more than a half of all HL cases due to genetic causes. To date, 134 genes are causally implicated in nonsyndromic HL with different types of inheritance (Hereditary Hearing Loss Homepage: https://hereditaryhearingloss.org).

Pathogenic variants in the *SLC26A4* (7q22.3, OMIM 605646) gene are one of the common causes of hereditary HL in many populations. Mutations in the *SLC26A4* gene cause non-syndromic recessive deafness (DFNB4, OMIM 600791) and Pendred syndrome (PDS, OMIM 274600) which combines sensorineural HL and goiter. The *SLC26A4* gene encodes pendrin, an anion transporter protein, which is mostly expressed in tissues of inner ear, thyroid and kidneys. Numerous studies revealed a diverse spectrum of *SLC26A4* mutations and their varying prevalence in patients from different regions of the world. The region- or ethno-specific prevalence of the *SLC26A4* mutations has not yet been fully elucidated.

SLC26A4 is relatively large gene covering about 57 kb of genomic DNA. The canonical transcript ENST00000644269.2 (ENSEMBL: https://www.ensembl.org), 4737 bp long, includes 20 coding exons within 2343-bp cDNA. Screening of pathogenic variants in the *SLC26A4* gene is an important part of molecular genetic testing for HL. Sequential analysis of the *SLC26A4* gene in a particular patient is continued until two recessive pathogenic *SLC26A4* variants are detected and, therefore, diagnosis could be made. A large size of *SLC26A4* predetermines the difficulties of its whole mutational analysis, which can be a labor-intensive, a time-consuming, and an expensive, especially in large patient cohorts.

Methods and Algorithms: Our study aims to optimize the testing of the *SLC26A4* gene by preliminary selection of the most diagnostically important regions of this gene for their priority analysis by direct sequencing.

To solve this task, we performed a thorough bioinformatic analysis of variations in the *SLC26A4* sequence based on the data from the Deafness Variation Database (DVD: https://deafnessvariationdatabase.org/gene/*SLC26A4*) [1]. The DVD, integrating all available genetic, genomic, and clinical data together with expert curation, generates the classification of variants: benign (B), likely benign (LB), likely pathogenic (LP), pathogenic (P), and variant of unknown significance (VUS) in many genes implicated in HL (including *SLC26A4*).

Results: To date, in total, 8647 different *SLC26A4* variants are reported in the DVD. The proportion of P and LP variants (PLP variants) is 7.0, 32.4, and 1.4 % in whole sequence, in coding and intronic regions of *SLC26A4* respectively. We selected all PLP variants within the *SLC26A4* coding and intronic regions (n = 603), and analyzed their distribution across the *SLC26A4* sequence. Among PLP variants, the common molecular

alterations are the missense variants, followed by the variants affecting splicing, and the frameshift variants. The number of PLP variants across the coding exons of *SLC26A4* varies substantially, as well as the physical size of the exons. For selection of exons with the highest load of PLP variants, the number of these variants was normalized by exon size. Ten "hot" exons had the variation rates higher than a mean (median value = 21.39) of a number of PLP variants per an exon size. Among PLP variants located in the *SLC26A4* intronic regions, the variants recognized by the DVD as the splice donor variants, are the most frequent followed by the splice acceptor variants.

In order to identify the *SLC26A4* regions with the highest load of PLP variants, we combined the data on PLP variants located in coding exons with those in adjacent intronic regions and revealed ten *SLC26A4* regions with the highest diagnostic value (exons 4, 6, 10, 11, 13-17, 19 and their flanking regions). Mutational analysis of these *SLC26A4* sequence regions can provide the diagnostic rate of 61.9 % (373/603) of all currently known PLP variants in the *SLC26A4* sequence.

Observed uneven prevalence of PLP variants in the *SLC26A4* gene in different populations or geographical regions could be influenced by ethnicity of the studied patients, history of a population to which they belong, as well as by a variable sensitivity of diagnostics methods applied in a particular group of patients. The accumulation of specific PLP variants in certain populations can be the result of founder effect as it was evidenced for c.2168A>G (p.His723Arg) in Japanese and Koreans and c.919-2A>G in Chinese.

Conclusion: The bioinformatic analysis of the distribution of PLP variants across *SLC26A4* gene sequence revealed the most diagnostically important regions of the *SLC26A4* gene (exons 4, 6, 10, 11, 13–17, 19 with flanking intronic regions), sequencing of which can provide the diagnostic rate of 61.9 % of all currently known PLP variants in the *SLC26A4* sequence. This approach can be used as an initial effective diagnostic testing in samples of patients of unknown ethnicity, or as a subsequent step after targeted testing of already known ethno- or region-specific *SLC26A4* pathogenic variants.

Acknowledgements: This work was supported by the RSF grant No. 21-75-00030 (to M.V.Z., E.A.M., and V.Y.D.) and by the Ministry of Education and Science of Russian Federation (grant No. FSUS-2020-0040) (to O.L.P.).

References

1. Azaiez H., Booth K.T., Ephraim S.S., Crone B., Black-Ziegelbein E.A., Marini R.J., Shearer A.E., Sloan-Heggen C.M., Kolbe D., Casavant T., Schnieders M.J., Nishimura C., Braun T., Smith R.J.H. Genomic landscape and mutational signatures of deafness-associated genes. *Am J Hum Genet.* 2018;103(4):484-497.