doi 10.18699/bgrs2024-9.1-24

# Identification of *SDHx* variants with gene expression model in head and neck paragangliomas

Snezhkina A.V.[1]*, Kobelyatskaya A.A.[1], Ayupova A.F.[1], Fedorova M.S.[1], Pavlov V.S.[1], Lantsova M.S.[1], Kuzovkina N.S.[1], Pudova E.A.[1], Kalinin D.V.[2], Kudryavtseva A.V.[1]

[1] *Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia*

[2] *Vishnevsky Institute of Surgery, Ministry of Health of the Russian Federation, Moscow, Russia*

** leftger@rambler.ru*

**Key words:** Head and neck paragangliomas; *SDHx* variants; gene expression; prediction model

*Motivation and Aim*: Head and neck paraganglioma (HNPGL) is a rare neuroendocrine neoplasm with a high rate of hereditary predisposition. Both germline and somatic mutations in the *SDHx* genes occur in over than 40 % of HNPGLs. Variants in the *SDHB* and *SDHD* genes are associated with an increased risk of metastatic and multifocal tumor forms, respectively. The objective of the study is to identify mRNA gene expression changes associated with *SDHx* variants and develop an expression model for mutation detection.

*Methods and Algorithms*: Whole-transcriptome sequencing data obtained on an Illumina platform for HNPGLs were utilized for differential gene expression analysis between *SDHx*-mutated and non-mutated tumors. Based on counts per million (CPM) values, a fully connected neural network (FCNN) was constructed using Keras, Tensorflow, and KerasR libraries. The obtained expression model was trained and tested both on a study cohort and the TCGA-PCPG RNA-Seq data. Subsequently, the model was validated on an extended tumor set (approximately 200 FFPE samples) using quantitative PCR (qPCR). The work was performed using the equipment of the EIMB RAS "Genome" center (http://www.eimb.ru/rus/ckp/ccu_genome_c.php).

*Results*: A two-step gene expression model for the identification of variants in *SDHx* genes using qPCR has been developed. At the initial step, the model defines variants in *SDHx* genes in general based on the mRNA expression of four genes – *CEP104*, *EYA4*, *FGD5*, and *NFRKB*. The first gene set demonstrated the following metrics in the test: sensitivity – 0.87, specificity – 0.94, accuracy – 0.90, and area under the curve (AUC) – 0.90. In the second step, the model predicts *SDHB* and *SDHD* variants using the expression mRNA data of *CEP104*, *EYA4*, *FGD5*, and *TERC* genes, and the following metrics were observed in the test: sensitivity – 1, specificity – 0.77, accuracy – 0.89, and AUC – 0.87.

*Conclusion*: The most effective method for identifying variants in the *SDHx* genes is genetic testing with targeted sequencing or whole-exome sequencing. However, these methods are expensive and time consuming. In this study, we created and validated an expression model for the prediction of *SDHx* variants in HNPGLs that can be used in routine clinical practice. In addition, the model enables the identification of *SDHB* and *SDHD* variants, which is very important for HNPGL management.

*Funding*: The study is supported by a grant from the Russian Science Foundation (No. 24-14-00439).