Identification of mRNA expression profiles and their characterization in age-related hearing loss

Kanglun Jiang*, Tan Wang†, Xinsheng Huang, Li Gao*

Department of Otorhinolaryngology-Head and Neck Surgery, Zhongshan Hospital, Fudan University, Shanghai 200030, China

Article Info

Abstract

Age-related hearing loss (ARHL), is a pervasive health problem worldwide. ARHL seriously affects the quality of life and reportedly leads to social isolation and dementia in the elderly. ARHL is caused by the degeneration or disorders of cochlear hair cells and auditory neurons. Numerous studies have verified that genetic factors contributed to this impairment, however, the mechanism behind remains unclear. In this study, we analyzed an mRNA expression dataset (GSE49543) from the GEO database. Differentially expressed genes (DEGs) between young control mice and presbycusis mice were analyzed using limma in R and weighted gene co-expression network analysis (WGCNA) methods. Functional enrichment analyses of the DEGs were conducted with the clusterProfiler R package and the results were visualized using ggplot2 R package. The STRING database was used for the construction of the protein-protein interaction (PPI) network of the screened DEGs. Two machine learning algorithms LASSO and SVM-RFE were used to screen the hub genes. We identified 54 DEGs in presbycusis using limma and WGCNA. DEGs were associated with the synaptic vesicle cycle, distal axon, neurotransmitter transmembrane transporter activity in GO analysis, and alcoholic liver disease, pertussis, lysosome pathway according to KEGG analyses. PPI network analysis identified three significant modules. Five hub genes (CLEC4D, MS4A7, CTSS, LAPTM5, ALOX5AP) were screened by LASSO and SVM-RFE. These hub genes were highly expressed in presbycusis mice compared with young control mice. We screened DEGs and identified hub genes involved in ARHL development, which might provide novel clues to understanding the molecular basis of ARHL.

Keywords: Age-related hearing loss, Bioinformatics, Differentially expressed genes, Machine learning.

1. Introduction

Age-related hearing loss (ARHL), or presbycusis, refers to progressive, bilateral symmetrical, sensorineural hearing loss mainly at higher frequency, and is the most prevalent sensory impairment in the aging population [1]. It is estimated by the WHO that around there will be over 500 million ARHL patients by 2025 [2]. Although ARHL is not life-threatening, it can affect the communication and social activity and life quality are decreased [3-5]. The treatment with cochlear implants or hearing aids is the main option for ARHL patients currently. Therefore, it is essential to deepen the understanding of the molecular basis and explore promising therapeutic targets for ARHL.

Hair cell loss, spiral ganglion neuron degeneration and stria vascularis atrophy are the primary causes of ARHL [6]. Based on the sites of abnormality, the ARHL is divided into four categories including sensory presbycusis, neural presbycusis, metabolic presbycusis, and conductive cochlear loss [7]. The hearing aids can replace the amplification and compression that the outer hair cells, and the sensory-motor cells of the cochlea fail to provide. Thus, patient perception of weak sounds can be improved in quiet conditions. However, the application of hearing aids is still limited by background noise conditions, degeneration of spiral ganglion neurons in the elderly and the high costs [8]. Thus, exploration of the potential therapeutic targets might contribute to ARHL prevention and treatment.

With the development of bioinformatics and gene chips, bioinformatics analysis has attracted increasing attention in the biomedical field for the potential of exploring biomarkers in diverse diseases. Previous studies have explored the potential biomarkers and potential targets of ARHL. For example, Liu et al. have proposed that hair cell aging is related to autophagy, and oxidative stress, and Sod1, Sirt6, Jund, and Cbx3 are related targets for ARHL therapy [9]. Peng et al. have reported that four aging-related genes (Ywhag, Mapre2, Fgf1, Acss2) are differentially expressed in ARHL and can be used for the construction of the prognostic model [10]. However, the molecular mechanisms in ARHL are not fully understood, and more efforts should be made to provide basis for the therapeutic intervention.

Our study intended to elucidate the expression pattern of DEGs and identify hub genes involved in ARHL based on bioinformatics analysis. We explored the gene expression pattern in ARHL in GSE49543 dataset using

* Corresponding author.
E-mail address: drgaoli@126.com (L. Gao).
# These authors contributed equally
Doi: http://dx.doi.org/10.14715/cmb/2024.70.4.40
the R limma package combined with WGCNA. The PPI network of the DEGs was constructed and the machine learning algorithms were applied for the selection of the hub genes. The findings of our work might provide clues for understanding the molecular basis of ARHL.

2. Materials and Methods

2.1. Data collection and processing
GeneChip data were downloaded from the GSE49543 dataset [11] from GEO database. The GSE49543 dataset contains 41 samples, with 9 young control CBA mouse cochlea samples and 32 ARHL mouse cochlea samples, and the annotation platform was GPL339 [MOE430A] Affymetrix Mouse Expression 430A Array.

2.2. Identification of differentially expressed genes (DEGs)
The DEGs between ARHL and young mice were screened using the limma package in R (version 4.2.1) in GSE49543 dataset and with logFold Change (FC) > 0.7, and P value < 0.05 as the threshold value. The statistical significance was calculated using Wilcoxon rank sum test. The results were visualized as heatmaps and volcano plots using the ggplot2 package.

2.3. Weighted gene co-expression network analysis (WGCNA)
WGCNA was used for the exploration of gene expression correlation [12]. The co-expression network of all genes in the GSE49543 dataset was constructed, and for the estimation of the network connection, adjacency was evaluated using the “soft” thresholding power (β) derived from co-expression similarity. Then the weighted adjacency matrix was converted into a topological overlap matrix, and its clustering tree structure is built using a hierarchical clustering method. Gene modules were identified by classifying genes with same expression pattern based on weighted correlation coefficient. The gene modules were represented by different clustering tree branches, and indicated by different colors. The modules with significant differences between the young-control and presbycusis group were then confirmed and genes in the differential modules were obtained.

The genes in the MEyellowgreen and MElightcyan modules and DEGs identified using limma in GSE49543 dataset were intersected and visualized by Venn diagram, and finally, the DEGs between presbycusis and young mice were screened for further analysis.

2.4. Functional enrichment analyses
Gene Ontology (GO) and Kyoto Genome Encyclopedia (KEGG) enrichment analyses of the screened DEGs were performed for the exploration of the biological functions and signalings implicated in ARHL progression using the clusterProfiler R package and ggplot2 R package.

2.5. Hub gene identification by machine learning
For the analyses of screened DEGs in presbycusis, two machine learning algorithms, Least Absolute Shrinkage and Selection Operator (LASSO) [13] using glmnet package and Support Vector Machine-Recursive Feature Elimination (SVM-RFE) [14] using e1071 package were performed. Finally, the results of the two machine learning algorithms were intersected and shown by Venn diagram.

2.6. Construction of the PPI networks
The STRING database (https://string-db.org/) was applied for the exploration of the interaction between target proteins and constructing the PPI networks. Cytoscape (version 3.7.2) was used for the evaluation and visualization of the networks. The clustering analysis was conducted using the MCODE function.

3. Results
3.1. Identification of DEGs in ARHL
The RMA-normalized microarray data in aging CBA mice were downloaded from the GSE49543 dataset. Totally 129 DEGs were in presbycusis mice compared with young control mice. The top 100 DEGs between ARHL mouse cochlea samples and young control samples were identified and shown in the heatmap (Fig. 1A). Fig. 1B presented the volcano plot of the DEGs screened by limma. For WGCNA network construction, totally 41 samples in the GSE49543 dataset were included, and we identified 17 modules on the thresholding power of β = 6 (Fig. 1C). We found that the yellow-green (coefficient=0.83, p=1e-04) and light-cyan modules (coefficient=-0.61, p=0.02) showed the highest correlation with ARHL, and were used as the pivotal modules for further analysis (Fig. 1D). We then evaluated the correlation between module membership and gene significance for age in the two selected modules. The results showed a positive correlation between the module membership in yellow-green modules and gene significance for age (Cor=0.7, p=3.9e-09) (Fig. 1E), and also between the module membership in light-
cyan module and gene significance for age (Cor=0.46, p=5.3e-18) (Fig. 1F). As shown in Figure 1G, totally 172 DEGs in the two modules were selected by WGCNA, and 129 DEGs were selected by limma in GSE49543 dataset, and 54 genes were identified in the intersection.

3.2. Functional enrichment analysis of the DEGs in presbycusis

GO and KEGG enrichment analyses of the DEGs between presbycusis and young mice were conducted, and the results indicated that these selected DEGs were closely related to transmission of neural signals such as synapse organization, synaptic transmission as well as neurotransmitter transport in biological processes (Fig. 2A); neuron, axon, and collagen in cellular component (Fig. 2B); glycosaminoglycan binding and in molecular function (Fig. 2C). KEGG analysis showed that the age-related DEGs were associated with the infectious diseases, IL-17 signaling pathway as well as synaptic vesicle cycle (Fig. 2D).

3.3. Protein-protein interaction (PPI) network of the DEGs

The PPI network analysis of the 54 DEGs was constructed on the STRING database. (Fig. 3A). The Cytoscape software was used to calculate the PPI network via the degree and the isolated nodes were removed. Furthermore, clustering analysis was conducted using the MCODE, and three significant gene clusters were obtained (Fig. 3B-D).

3.4. Screening of the hub genes based on machine learning

Two machine learning algorithms LASSO and SVM-RFE were used to screen the hub genes. Firstly, 5 genes were selected by LASSO regression algorithm from the DEGs (Fig. 4A). Next, 16 genes were identified by the SVM-RFE algorithm (Fig. 4B). Moreover, as shown in the Venn diagram, totally 5 intersecting genes were selected and confirmed as the hub genes, namely CLEC4D, MS4A7, CTSS, LAPTM5, and ALOX5AP. Moreover, the hub gene expression pattern in presbycusis was explored, and all these five hub genes were significantly upregulated in presbycusis mice in comparison with the young control mice in the GSE49543 dataset, which might be associated with the presbycusis progression (Fig. 4D-H).

4. Discussion

ARHL, also known as presbycusis, is a prevalent neurodegenerative disorder as well as a common chronic medical condition in the elderly [4]. Increasing evidence has revealed that the risk of cognitive impairment or decline increases in ARHL patients, which considerably affects public health and quality of life of patients. In the current study, we explored the potential genomic changes induced by age-related hearing loss based on the bioinformatics analysis. Totally, 54 DEGs were screened based on the WGCNA and limma in the GSE49543 dataset. Five
Bioinformatic identification of hub genes in ARHL.

5. Conclusion

In conclusion, 54 DEGs between ARHL and young control mice were identified, and 5 hub genes including Clec4d, Ms4a7, Cts5, Laptm5, and Alox5ap upregulated in ARHL were screened, which might provide clues for the anti-ARHL treatment.

Conflict of interests

The authors declare no competing interests.

Consent for publications

The author read and approved the final manuscript for publication.

Ethics approval and consent to participate

Not applicable.

Informed consent

Not applicable.

Availability of data and material

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors’ contributions

GL contributed to the study conception and design. Software, data collection and analysis were performed by JK, WT and HX. The first draft of the manuscript was written by JK. All authors commented on previous versions of the manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (82000980), Shanghai Science and Technology Committee Foundation (21Y31900500) and Foundation of Shanghai Municipal Commission of Economy and Informatization (23SHS05000).
Acknowledgements
Not applicable.

References


