Paradigm Academic Press
Journal of Innovations in Medical Research
ISSN 2788-7022
AUG. 2025 VOL.4, NO.4

Microscopic Effect of Tobacco on Human Health: How Tobacco Impacts Gene Expression Levels

Jingquan Shi[1]

[1] Claremont High School, Claremont, CA 91711, US

Correspondence: Jingquan Shi, Claremont High School, Claremont, CA 91711, US.

doi:10.63593/JIMR.2788-7022.2025.08.005

Abstract

Tobacco exposure is commonly linked to airway epithelium and lung diseases, yet its impact on the enzymes that catalyze adenosine-to-inosine (A-to-I) RNA editing remains unclear. The goal of this research is to compare the three gene expression levels, ADAR, ADARB1, ADARB2, between smokers and non-smoker, which play an important role in RNA editing and cellular regulation. RNA-sequence count data for smokers (362 samples) and non-smokers (635) were obtained from the GEO publication "Cigarette Smoking-Associated Isoform Switching and 3' UTR Lengthening Via Alternative Polyadenylation" (GSE171730). Data obtained was first applied with $\log_2$-transformation to dampen outliers, then compared with T-test. The result of the study shows that smoking had no significant effect on ADAR1, but produced robust up-regulation of ADARB1 and ADARB2. These results indicate that tobacco selectively enhances expression of two ADAR paralogs while sparing the ubiquitously expressed ADAR1, pointing to a targeted modulation of the RNA-editing machinery in smokers. Such selective induction may shift global A-to-I editing profiles and contribute to smoking-associated disease risk.

Keywords: smoking, RNA editing, ADAR enzymes, airway epithelium, gene expression, transcriptomics

1. Introduction

For decades, smoking has been associated with increased risk of developing various diseases, including lung and cardiovascular diseases. It has already been found that smoking is a main risk factor for the increase in prevalence of HIV, tuberculosis, SARS-CoV, and the current SARS-CoV-2 (Jiang C, Chen Q & Xie M., 2020). Moreover, the study "Cigarette smoke and adverse health effects: An overview of research trends and future needs" has shown its relationship with being the main cause of lung cancers, which caused about 124,730 lives to be lost per year, according to the Lung Cancer Research Foundation. Researchers are now going beyond the clinical results to see if smoking can cause any negative effect on the molecular level.

Many changes on the molecular level have been reported in the literature. In this study, RNA editing is specifically chosen and three gene expression levels, ADAR, ADARB1, ADARB2, as they are necessary in RNA editing, are focused on. The ADAR gene, also known as ADAR1, is discovered to edit and provide instructions for synthesizing RNA-specific protein, proteins that bind to RNA without strong sequence preferences. Meanwhile, ADARB1, also known as ADAR2, is discovered to edit the majority of the coding regions (neural editing) and serve the function of a catalyst. Lastly, ADARB2, also known as ADAR3, is found to be responsible for regulating mRNA translation in RNA editing processes, specifically in brain areas.

It is known that RNA editing has the ability to alter nucleotide sequences, which can strongly affect protein synthesis according to the study "RNA editing enzymes: structure, biological functions and applications". Since these changes are site-specific, RNA editing allows clear observations and comparisons between groups to be made. Based on findings from previous knowledge, this research aims to compare these three gene expression

levels between smokers and non-smokers by visualization of the data by boxplot and analysis with T-test to see if there is a significant difference between the 2 groups. By determining if there is a significant difference between the three gene levels of smokers and non-smokers, we can deepen our understanding on how smoking affects human bodies from a microscopic perspective. Since mRNA abundance of ADAR paralogs often correlates with global A-toI editing levels, hence by using the ADAR genes as a proxy for RNA editing levels, we can observe how RNA molecules are manipulated. In this case, we hypothesize that chronic smoking would alter the transcription in the ADAR-family genes in human airway epithelial tissues.

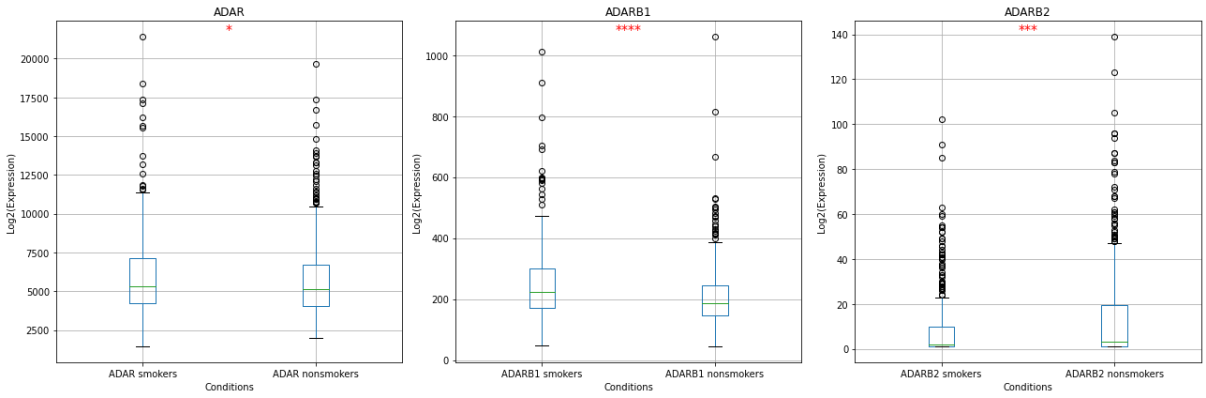
2. Methods

Data in this study, whole blood samples from 454 current and 767 former smokers using Illumina HiSeq 2000, was adopted from a GEO publication (GSE171730) (Xu, Zhonghui, et al., 2021). Firstly, the original data was obtained and log2 conversion was applied to it in order to reduce the effect of outliers on the following analysis. Next, the resulting data is converted into a boxplot in order to visualize the distribution of the three gene expression levels of smokers and non-smokers. In this study, a T-test was performed in order to distinguish the difference between smokers and non-smokers for the three gene expression levels: ADAR, ADARB1, ADARB2 respectively. Lastly, the p-value was calculated in order to see whether the comparison is statistically significant (if the p-value is small enough), and also compare the median values of the three gene expressions of both groups.

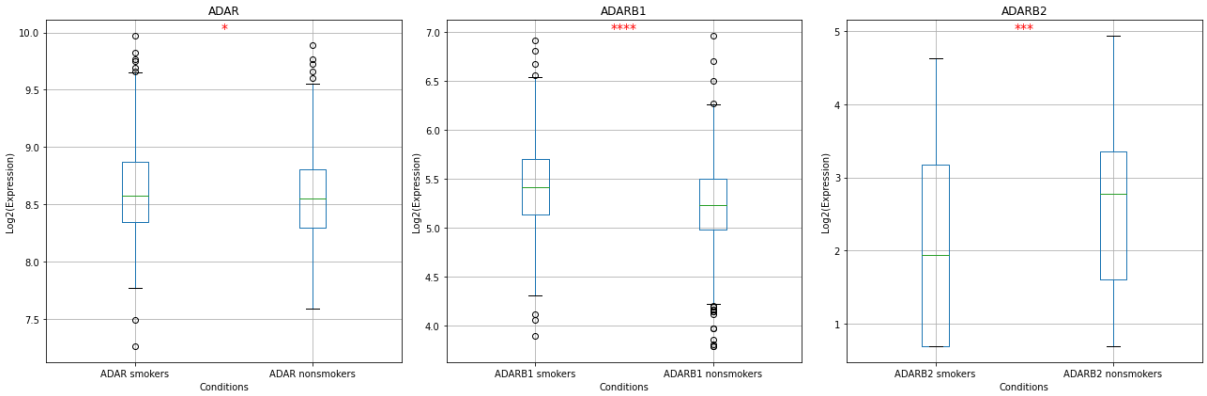
3. Results

The hypothesis for this study was that smoking would affect ADAR gene expression. Final computed results showed that for ADAR, the p-value ($p = 0.049$) was slightly less than 0.05 when using the original data, and ($p = 0.057$) was more than 0.05 after reducing the effect of outliers with log2 conversion. No significant difference was observed between the two groups, which means there was no significant difference between the gene expression of ADAR between Smokers and Non-smokers.

For ADARB1, there was a significant difference between two groups as the p-value ($p = 0$ for both original data and data with log 2 applied) was way less than 0.001 when using both data. For ADARB2, both groups were significantly different as the p-value was less than 0.001 using both data (original $p=0.00079$) (converted $p=0.00015$ after applying log 2). The low p-value showed that there was a high possibility smoking would affect and change the ADARB1 (higher in smokers) and ADARB2 (higher in nonsmokers) genes and there was a high possibility of a real difference, allowing the result to be accepted.



(a)



(b)

Figure 1. Expression levels of ADAR genes (a) original data on a box plot graph; (b) data after log2 on a box plot graph

* ADAR: p=0.057 > 0.05; **** ADARB1: p=0 < 0.001; *** ADARB2: p=0.00079 < 0.001.

(The smaller is the p-value the more significant the difference will be).

4. Discussion

From the result, it can be concluded that smoking is not a changing factor for gene expression of ADAR as it doesn't show a significant difference in the ADAR level between smokers and non-smokers. However, ADARB1 and ADARB2 can be more easily affected by smoking as it showed a significant difference in the ADARB1 and ADARB2 levels between smokers and non-smokers. In comparison to the study "Decrease in ADAR1 expression by exposure to cigarette smoke enhances susceptibility to oxidative stress" which shows results with more significant differences along with a potential therapeutic approach. Although its findings may appear stronger, the accuracy of the results is questionable since their data were derived from cell lines. Cells in cell lines can undergo genetic and functional changes over time since they divide and grow rapidly, resulting in a possibly inaccurate result. Moreover, the RNA editing levels were not assessed, which limits its insight into the functional consequences of ADAR expression changes. In contrast, the current study includes analysis across human samples, offering greater reliability. In the future, assay global A-to-I levels can be done to see how pervasive this RNA modification is distributed. In-vitro airway-organoid smoke exposure can also be tested to see the effect of smoking in respiratory systems and how A-to-I editing is affected by that.

References

Dailamy, A., Lyu, W., Nourreddine, S. et al., (2024). Charting and probing the activity of ADARs in human development and cell-fate specification. *Nat Commun*, *15*, 9818. https://doi.org/10.1038/s41467-024-53973-0

Jiang C, Chen Q, Xie M., (2020, Jul 14). Smoking increases the risk of infectious diseases: A narrative review. *Tob Induc Dis.*, *18*, 60. doi: 10.18332/tid/123845. PMID: 32765200; PMCID: PMC7398598. https://pmc.ncbi.nlm.nih.gov/articles/PMC7398598/

Saha SP, Bhalla DK, Whayne TF Jr, Gairola C., (2007, Fall). Cigarette smoke and adverse health effects: An overview of research trends and future needs. *Int J Angiol.*, *16*(3), 77-83. doi: 10.1055/s-0031-1278254. PMID: 22477297; PMCID: PMC2733016. https://pmc.ncbi.nlm.nih.gov/articles/PMC2733016/

Savva, Y.A., Rieder, L.E. & Reenan, R.A., (2012). The ADAR protein family. *Genome Biol.*, *13*, 252. https://doi.org/10.1186/gb-2012-13-12-252

Xu, Zhonghui, et al., (2021 Dec 01). Cigarette Smoking-Associated Isoform Switching and 3' UTR Lengthening via Alternative Polyadenylation. *Genomics*, *U.S. National Library of Medicine*, pubmed.ncbi.nlm.nih.gov/34763026/. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE171730

Zhang D, Zhu L, Gao Y, Wang Y, Li P., (2024, Mar 16). RNA editing enzymes: structure, biological functions and applications. *Cell Biosci.*, *14*(1), 34. doi: 10.1186/s13578-024-01216-6. PMID: 38493171; PMCID: PMC10944622. https://pmc.ncbi.nlm.nih.gov/articles/PMC10944622/