

SHORT THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PhD)

Determining Susceptibility to Smoking among Hungarians with a
Special Focus on the Roma Population and Gene Expression Analysis
in Coronary Artery Disease Patients: An Integrated Investigation

By Mohammed Abdulridha Merzah

Supervisor: Dr. Szilvia Fiatal



UNIVERSITY OF DEBRECEN
DOCTORAL SCHOOL OF HEALTH SCIENCE

DEBRECEN, 2023

Determining Susceptibility to Smoking among Hungarians with a Special Focus on the Roma Population and Gene Expression Analysis in Coronary Artery Disease Patients: An Integrated Investigation

By Mohammed Abdulridha Merzah, MPH

Supervisor: Dr. Szilvia Fiatal, PhD

Doctoral School of Health Science, University of Debrecen

Head of the defence board:

Prof. Dr. Endre Nagy, PhD, DSc

Reviewers:

Dr. Zsuzsanna Máté, PhD

Dr. Attila Borbély, PhD

Members of the defence board:

Prof. Dr. Edit Paulik, PhD

Dr. Éva Katona, PhD

The **PhD defence** takes place at Lecture Hall of Department of Internal Medicine, **Building ‘A’**, Faculty of Medicine, University of Debrecen on **February 21st, 2024, at 13:00**.

Introduction

Smoking has a profound worldwide impact, leading to more than 8 million deaths annually across the globe. If the current global smoking trend persists without alteration, it is estimated that the number of deaths attributed to tobacco-related illnesses could exceed 8 million annually by 2030. This ongoing prevalence of smoking poses a significant public health concern in Europe, remarkably reflected in its contribution to approximately 700,000 annual deaths. Notably, Hungary faces substantial challenges in this regard, as the prevalence of smoking remains high. In fact, 24.9% of individuals aged 15 and above in Hungary are daily smokers.

Significant variations in smoking habits are observed within the Hungarian population. Specifically, men and individuals aged 18-34 exhibit alarmingly high smoking rates, reaching as much as 35.6% within the latter group. The situation is further complicated by disparities in smoking prevalence among the Roma minority, particularly those living in segregated areas. These areas report even higher smoking rates compared to the general Hungarian population.

Given the prevalence of smoking in Hungary and its associated health risks, it is imperative to understand the underlying factors contributing to this issue. This becomes particularly crucial when considering the social isolation experienced by minorities residing in segregated areas and potential genetic predispositions that might contribute to unhealthy lifestyles and disease development. Consequently, identifying the genetic factors influencing smoking behaviors becomes crucial for practical interventions aimed at smoking cessation.

Moreover, smoking stands as a well-established and modifiable risk factor for coronary artery disease (CAD). Unfortunately, Hungary carries a disproportionate burden of smoking-related diseases, with roughly a quarter of its adult population reporting daily smoking in 2019. This statistic places Hungary as the third-highest country in Europe concerning smoking prevalence. By 2020, the prevalence of tobacco use escalated to 28%, surpassing the European average of

25%. Hungary's persistently high smoking rates are particularly concerning due to the nation's significant burden of coronary heart disease (CHD). CHD deaths account for 24.5% of all deaths in Hungary, according to the latest health report.

Despite the established epidemiological links between smoking and non-communicable diseases, such as CHD, a notable research gap persists in the literature concerning the impact of smoking on gene expression, particularly within the Hungarian population. This research project thus endeavors to address this gap by examining the global gene expression patterns associated with smoking in individuals with CAD. Doing so aims to shed light on the genetic foundations of smoking behavior and its potential implications for targeted interventions.

Aims and Objectives

The overarching goal of this study was to uncover the genetic factors contributing to elevated smoking rates within the Hungarian Roma (HR) population. Additionally, the study aimed to investigate variations in gene expression in the whole blood of patients with CAD, differentiating between smokers and non-smokers, in order to reveal potential connections between smoking habits and CAD susceptibility. The specific objectives were as follows:

1. To identify the underlying genetic factors associated with the high prevalence of smoking in the HR population.
2. To analyze gene expression patterns in the whole blood of both smokers and non-smokers among patients with CAD.

Material and Methods

Study 1: Roma Socioeconomic Status Has a Higher Impact on Smoking Behaviours than Genetic Susceptibility

The study employed a health survey to create a comprehensive database for comparative and associational research, aiming to understand the reasons behind the health challenges faced by the Roma population, particularly their increased susceptibility to cardiometabolic disorders. Adult participants from both the Hungarian general (HG) and HR populations, aged 20 to 64, underwent questionnaire-based, physical, and laboratory assessments in this cross-sectional study. Out of the initial 832 individuals (417 HG and 415 HR), 18 subjects were excluded due to missing anthropometric measurements.

The study focused on Single-Nucleotide Polymorphisms (SNPs) associated with smoking behaviours. A selection of 10 SNPs out of 33 was made based on published studies with sufficient sample sizes and consistent relevance to smoking intensity. DNA was isolated from blood samples and genotyped using the mass-array platform. Genetic Risk Scores (GRSs) were assigned to participants based on the number of risk alleles they carried for the selected SNPs. Both unweighted and weighted GRS were calculated using specific formulas. Smoking habits were categorized into four classes: never smokers, former smokers, moderate smokers, and heavy smokers.

Statistical analysis was conducted using various methods, including chi-square tests, linear regression, binary regression, and multinomial regression. The analysis aimed to assess associations between genotypes and smoking behaviours, predict smoking-related indicators, and examine the impact of factors like socioeconomic status, genetic risk scores, and more. Ethical approval was obtained for the study, and informed consent was secured from all participants.

Study 2: A Transcriptomic Analysis of Smoking-Induced Gene Expression Alterations in Coronary Artery Disease Patients

This single-centre cross-sectional study was conducted at the University of Debrecen Medical Centre, Division of Cardiology and Cardiac Surgery between November 2021 and October 2022. It included 61 patients meeting specific criteria with CAD. Blood samples were collected from participants for various assays, including cotinine measurement and RNA isolation. RNA sequencing libraries were prepared, and RNA sequencing was performed on Illumina NextSeq 500. Cotinine levels were used to differentiate non-smokers from active smokers. Clinical and demographic data were gathered, including age, sex, smoking status, medication, blood pressure, body mass index (BMI), and laboratory results. Quantitative coronary angiography was employed to assess the extent of coronary stenosis. Nicotine dependence among smokers was evaluated using the Fagerstrom test for nicotine dependency.

Statistical analyses were performed using R software for descriptive statistics and Strand NGS for differential gene expression analysis. Quality sequencing data was controlled using FastQC, and alignment to the human reference genome was performed. Differentially expressed genes were identified using the DESeq algorithm. Gene ontology (GO) analysis was conducted using CytoScape and ClueGo applications to identify significant pathways associated with differentially expressed genes.

The study was ethically approved, and participants provided informed consent. This research aimed to shed light on factors related to coronary artery disease, particularly genetic expression and related biological processes.

Results

Study 1: Roma Socioeconomic Status Has a Higher Impact on Smoking Behaviours than Genetic Susceptibility

Of 814 participants, around 35.9% were male, 36.3% were HR, and 25% were over 50. HR people had higher smoking rates (65.2%) than HG (32.8%). Significant differences in smoking habits ($p\text{-value} < 0.0001$) were observed. HR females were twice as likely to smoke as HG females, with an odds ratio of 2.1. Significant disparities in SES ($p\text{-value} < 0.0001$), with more HR individuals in the lower middle SES category (78.6% vs. 60.2% of HG individuals), were observed.

In regards to the association of genotype with qualitative smoking behaviours, the additive models of rs2036534, rs578776, rs16969968, rs6517442, and rs2235186 significantly differed between HG and HR populations based on smoking status (smokers vs. nonsmokers). A high prevalence of both homozygous and heterozygous risk alleles was observed among HR smokers for all significant SNPs. For the quantitative smoking behaviours, HG individuals who carried either homozygous or heterozygous of the risk allele of six SNPs (rs10490162, rs2036534, rs6517442, rs16969968, rs2235186, and rs578776) exhibit higher smoking rates. However, the same model of only four SNPs (rs2673931, rs2235186, rs578776, and rs578776) was linked to higher CPD among the HR population. Furthermore, HR people had a 33% higher likelihood of CPD than HG people ($p < 0.001$). Regarding age at smoking initiation, the HR individuals were likelier to initiate tobacco at a younger age (standardized $\beta = -0.23$, $p < 0.001$). Nonetheless, while gender showed a significant association with CPD, no significant association was found between gender and age at smoking initiation.

The GRS median and IQR were similar for HR and HG individuals, with a median of 9.0 (IQR: 7.0-10). However, when wGRS were considered, HR individuals had a slightly higher median

(5.2; IQR: 4-6) than HG individuals (4.9; IQR: 3-5), with a p-value of 0.02. Consequently, the wGRS distribution skewed towards higher values among HR individuals.

A binary regression test showed that individuals who smoked in both populations exhibited a lower SES than those who never smoked, with a negative β coefficient of -0.04 ($p=0.02$) and -0.01($p=0.00$) for HR and HG, respectively. The results of a multinomial regression analysis demonstrated an inverse association between SES and smoking behaviours, with a statistical significant of $p<0.001$. After adjusting for sex, age, BMI, and genetic factors, it was observed that HR individuals had a higher likelihood of being smokers compared to HR individuals. In conclusion, the analyses revealed that SES had a stronger relationship to smoking behaviours than genetic factors.

Study 2: A Transcriptomic Analysis of Smoking-Induced Gene Expression Alterations in Coronary Artery Disease Patients

Sixty-one subjects were recruited in this study, with one-half being smokers (50.8%, $n=31$). Smokers males were higher than non-smokers (37.7% vs 19.7%) in contrast to female gender (29.5% vs 13.1%). A significant difference was identified regarding the median age between smokers and non-smokers, with a lower median age among smokers (60 vs. 75 years, $p<0.001$). The percentage of smokers with a familial CAD background was considerably higher than non-smokers.

The initial analysis revealed 502 genes with significant expression differences, indicated by a p-value of <0.05 . Up on adjusting the fold change (FC) threshold to ≥ 1.5 , only 13 genes remained significant. After removing outlier samples by applying principle component analysis and clustering, forty-four samples remained for further analyses. A hundred and twenty genes were identified with an FC of ≥ 1.5 and a p-value of <0.05 . Of them, fifteen genes were overexpressed, and the rest were downregulated. Significantly, there was a substantial

upregulation of 7-fold of gene *GPR15*, with a p-value of <0.001 . Conversely, gene *KRT1* demonstrated a 3-fold downregulation, as evidenced by a p-value of 0.006. On the other hand, when an FC of ≥ 2.0 was used with an adjusted p-value of <0.05 , a total of 31 differentially expressed genes (DEGs) retained their significance. Of them, five genes were upregulated, and 26 genes were downregulated.

Moreover, 27 DEGs were identified when comparing individuals with high, moderate, and low-to-moderate levels of ND to those with low levels of ND. Remarkably, across all comparison groups, a consistent pattern of overexpression of nine genes (*ASB16*, *ADPGK-AS1*, *RN7SL2*, *IGKV1-12*, *SPTLC1P2*, *SCGB3A1*, *UBE2SP2*, *ZNF696*, and *USP9Y*) was observed.

Discussion

This thesis thoroughly examined the factors contributing to the high smoking rates within the Hungarian population and the impact of smoking on gene expression in CAD patients. Two rigorous observational studies were conducted using primary data. The first study identified a significant correlation between smoking behaviour and SES among HR individuals, despite the absence of evidence for genetic predisposition. The second study yielded valuable insights into the effects of cigarette smoking on gene expression in peripheral blood and its relation to smoking-induced CAD.

Giving significance to the first aim of this thesis, minorities living in segregated areas experience social isolation and a genetic predisposition that could heighten their likelihood of engaging in unhealthy lifestyles or developing certain diseases. Identifying genetic factors underlying smoking behaviours is crucial to enable the implementation of novel interventions for smoking cessation. Our findings revealed no evidence of a genetic predisposition to smoking behaviours among HR individuals; however, a negative inverse link was identified between SES and smoking behaviours.

Comparable medians of GRSs were observed in both populations. When considering the wGRS, HR individuals had a slightly higher score than HG people. However, there is no significant relationship between GRSs and smoking behaviours. In contrast, ethnicity and SES revealed a significant correlation. Despite the equivalence of GRSs in the study populations, heavy smokers still exhibited higher risk scores. This observation could be attributed to the selection of specific SNPs, most of which are linked to nicotine dependence, a characteristic more commonly found among heavy smokers. Therefore, further studies incorporating a more comprehensive range of SNPs associated with smoking behaviours are necessary to obtain a comprehensive understanding.

Our findings indicate no evidence of genetic vulnerability toward smoking behaviours among the HR population. Instead, the higher prevalence among those individuals residing in economically disadvantaged segregated areas appears to be potentially influenced by environmental and cultural factors. Prior studies have revealed a considerable link between cigarette smoking and sociodemographic factors, with a higher smoking rate observed among people of low socioeconomic status. Across Europe, smoking tends to be more prevalent among unemployed or self-employed individuals. Manual employees have the highest smoking rates among employed individuals, with 40% being smokers. Similarly, a strong correlation exists between tobacco use and lower levels of educational achievement. These trends are evident within Roma communities, as the EU health report on the Roma population indicates. HR exhibit a higher prevalence of smoking, with a tendency to initiate smoking at a younger age and a propensity to smoke more CPD than the non-Roma population.

Furthermore, SES emerged as a prominent factor that could influence the health and lifestyles of HR individuals. Living in economic difficulties increases the likelihood of smoking and reduces the prospect of smoking cessation. Our study revealed that living in a lower SES increases the likelihood of being a smoker, consistent with findings by others. Among HR

individuals, 65% were smokers, with 77% living in the lower-middle categories of SES. Additionally, 50% of the HR smokers were identified as heavy smokers. These findings highlight a strong association between smoking behaviours and SES, indicating that the prevalence of heavy smoking habits increases as the economic status of HR individuals decreases. In essence, the lower the economic status in which HR individuals reside, the greater the prevalence of heavy smoking habits. Consequently, by elucidating the impacts of segregation, it is possible to classify intervention and disease prevention strategies as alternative implications. Public health endeavours should prioritize the identification of social and cultural barriers that hinder the reduction of smoking rates.

Several policies have successfully reduced the prevalence rate of smoking among the HG population however have proven ineffective among HR individuals. A qualitative study in Slovenia shed light on smoking as a fundamental component of Roma's ethnic, cultural, and individual identity. The study revealed that older family members introduced cigarettes to younger generations, considering it a part of the natural "growing up" process. Therefore, it is crucial to prioritize qualitative research to comprehend the barriers hindering smoking reduction among HR individuals. Although many HR people oppose endorsing tobacco control policies, it is vital to thoroughly evaluate the extent to which Roma people feel marginalized and discriminated against, particularly concerning the action of public authorities.

However, there needs to be more research conducted to investigate the effects of smoking on blood gene expression in individuals with CAD in Hungary. Therefore, the second part of this thesis aimed to analyse the gene expression in the blood samples of CAD patients, focusing on their smoking habits. The results revealed a notable correlation between smoking and the overexpression of genes associated with immune response and oxygen-binding activities. In addition, we observed a significant difference in the expression of 120 DEGs between smokers and non-smokers among CAD patients.

Linked of upregulated genes to CAD and smoking

Atherosclerosis is the main underlying cause of CAD, where various inflammatory processes play a significant role. Our findings exposed fifteen genes that exhibited increased activity, all known to be contributed to the control of immune responses. Among these, six genes were specifically associated with immunoglobulin receptors or secretions, namely *IGKV3-11*, *IGHA2*, *IGLC1*, *IGLV3-21*, *IGLL5*, and *JCHAIN*. In general, antibodies, also known as immunoglobulins, are a group of glycoproteins produced by B-lymphocytes. They act as receptors, and upon binding to specific antigens, they stimulate B-lymphocyte replication and differentiation, producing plasma cells that secrete immunoglobulins. These secreted immunoglobulins are vital in mediating the effector phase of humoral immunity. Evidence suggests that humoral immunity plays a role in modulating the processes involved in plaque development. Consistent with our findings, a study has demonstrated that smoking has distinct effects on the concentrations of IgM and IgA, resulting in elevated levels of these antibodies in the bloodstream.

Additionally, another study has identified that measuring the plasma level of *IGHA2*, which encodes IgA, could be a valuable biomarker for detecting early-stage atherosclerosis. Similarly, the expression levels of *CD28*, *TRAT1*, *TRAV13-1*, *LEF1*, and other genes have been associated with smoking and smoking-related atherosclerosis. These genes are centrally involved in activating adaptive immunity, particularly T-cell activation, which significantly contributes to the formation of atherosclerotic lesions.

It has been discovered that the G protein-coupled receptor 15 (*GPR15*) is a critical regulator of T cell trafficking during systemic inflammation. Recent research has found that increased expression of *GPR15* is linked to cardiovascular disorders and plays a role in the negative effects of smoking. This analysis revealed that *GPR15* expression was significantly higher in

smokers than non-smokers, indicating a defensive immune response by the host to counteract the effects of smoking. Additionally, smoking is associated with the upregulation of C-C chemokine receptor type 7 protein (encoded by the *CCR7* gene), which facilitates T cell trafficking in patients with coronary artery disease.

The *ITGA6* gene, which belongs to the integrin family, has a vital function in various biological processes such as cell proliferation, invasion, and adhesion. Elevated expression of *ITGA6* has been observed to have a notable impact on macrophages, contributing significantly to the development of atheroma, a fatty deposit in the arterial wall. Scott and Palmer have suggested that smoking affects the levels of circulating adhesion molecules. Smoking can result in increased levels of soluble adhesion molecules, indicating of ongoing inflammatory processes that play a crucial role in the formation of atheroma.

The *RCAN3*, a coding protein, is believed to have a role in calcium-mediated signaling pathways. No specific studies have investigated the relationship between *RCAN3*, CAD, or smoking, except for one study that indicates a negative association between DNA methylation and the risk of heart attack. However, because *RCAN3* is involved in processes mediated by calcium, a recent study has proposed a new mechanism in which nicotine induces intracellular calcium, leading to increased calcification of atherosclerotic plaques. Furthermore, two other genes, *TRAV4*, which encodes the T cell receptor alpha variable 4, and *JCHAIN*, which encodes the Immunoglobulin J Chain, have been linked to adenocarcinoma. However, their connection to CAD or smoking has yet to be studied. Considering their involvement in adaptive immunity, the overexpression observed in this analysis may be related to their role in the immune response.

Pathways of the downregulated genes

Rather than exploring the extensive catalogue of more than a hundred genes that have been suppressed, this section chooses to focus on the corresponding GO terms. Our findings indicate

important molecular functions such as oxygen transport, oxygen binding, haemoglobin interaction, binding to organic acids, and binding to haem. A total of six genes, namely *HBQ1*, *HBA2*, *HBA1*, *HBB*, *HBM*, and *HBD*, were identified as exhibiting these molecular functions in common. The haemoglobin subunit genes delta (*HBD*) and beta (*HBB*) have a vital role in ensuring the effective operation of haemoglobin, which is responsible for the transportation of oxygen within the bloodstream. The capacity of haemoglobin to effectively transport oxygen may be hindered by smoking, leading to a reduced capacity. This effect can be attributed to the suppression or downregulation of these genes. Alternatively, haemoglobin subunit $\alpha 2$ (*HBA2*) is a closely related variant of $\alpha 1$ (*HBA1*), which itself is an important variant of theta1 (*HBQ1*). These genes have been associated with erythrocyte function, specifically carbon dioxide uptake and oxygen delivery. Consequently, smoking may cause a decrease in oxygen availability among patients with CAD.

Moreover, the downregulation of alpha haemoglobin stabilizing protein (*AHSP*), a small protein responsible for regulating the stability and folding of the alpha haemoglobin (α Hb) subunit, could potentially be linked to impaired oxygen transportation in CAD patients as a consequence of smoking. Similarly, the downregulation of *SLC4A1*, a member of the soluble carrier family 4, has been observed to impact haemoglobin-binding activity. A study has indicated that oxidative stress (OS) can influence the expression of *SLC4A1*, and these effects could have adverse implications for diseases associated with OS, such as CAD. With the growing body of evidence connecting smoking and OS, this correlation could shed light on the downregulation of the *SLC4A1* gene observed in smoker CAD patients. Significantly, the presence of carbon dioxide and other harmful chemicals in cigarettes amplifies the adverse health consequences of smoking. Prolonged exposure to these chemicals can result in OS, which disrupts the regular expression of genes and causes the downregulation of genes involved in vital physiological processes, including oxygen and haem binding or carrier activity.

The *ALAS2* gene encodes a protein called 5'-Aminolevulinate Synthase 2, which is involved in initiating the biosynthesis of haem. Recent evidence suggests that this protein may mediate, linking smoking with IL6 and CRP, two inflammatory markers, and exhibiting an inverse association. The reduced expression of this gene could be attributed to smoking-related inflammation. Likewise, *PTGDS* is a gene responsible for encoding a protein engaged in synthesising of prostaglandin D. It plays a role in regulating the relaxation and contraction of smooth muscle cells and platelet aggregation. Consistent with our findings, research studies have indicated that the *PTGDS* gene is downregulated in individuals who smoke and is associated with both smoking and the development of plaques. Furthermore, a mouse model has provided evidence of the involvement of *PTGDS* in promoting thrombosis. Ultimately, our analysis revealed that these genes (*ALAS2*, *HBA1*, *HBA2*, *HBB*, *HBQ1*, *HBD*, *HBM*, and *PTGDS*) are involved in pathways associated with organic acid-binding. Further comprehensive studies are needed to examine their potential influence on the development and advancement of CAD among individuals who smoke. Understanding the intricate connection between smoking, CAD, and organic acid binding is crucial for unravelling the underlying mechanisms and identifying potential therapeutic targets.

Despite the cross-sectional design, the findings can enhance our understanding of the connection between genetic factors and smoking behaviours from one side and between cigarette consumption and CAD from another, thereby supporting previous epidemiological observations. It is essential to recognize that cross-sectional studies primarily provide a snapshot of associations at a specific moment and do not account for other confounding factors or changes over time. Therefore, further research, such as longitudinal or cohort studies, is necessary better to understand the causal relationship between smoking and CAD. Nonetheless, the results of this study are valuable as they contribute to the existing body of knowledge and reinforce the previously observed epidemiological patterns. They emphasize the significant

public health initiatives to reduce cigarette consumption and promote smoking cessation, considering that smoking is a significant risk factor for developing CAD and other cardiovascular diseases.

Limitation

One significant limitation to be reported is that we classified heavy smokers based on the number of CPDs; they were presumed to be nicotine dependent. Alternatively, relying on reliable measurement tools such as the Fagerstrom Test for Nicotine Dependence (FND), Nicotine Dependence Syndrome Scale, or the Wisconsin Inventory of Smoking Dependence would provide valuable insights into the level of nicotine dependence among participants. Another limitation was the inadequate representation of the overall HR population with the selected sample. This inadequacy arose from certain participants' reluctance to self-identify as HR, leading to the possibility of including HR individuals within the low-risk (HG) group. This misclassification could have resulted in underestimating the outcomes related to genetic susceptibility, thus potentially impacting the overall findings.

The second part of this dissertation had certain limitations. One significant limitation was related to the referral and admission process, which took longer to complete than the time it takes for plasma cotinine (a marker of smoking) to degrade by half, approximately 16-17 hours. Consequently, most smokers would have undetectable levels of cotinine within that timeframe. Self-reported smoking status and the Fagerström standard for nicotine dependence were utilized to address this limitation.

Furthermore, the inclusion of patients with coronary stenosis of $\geq 50\%$ of the luminal diameter of the affected artery/is was determined based on angiography, which relies on the cardiologist's interpretation. While this approach has been commonly employed in previous studies as the accepted means of diagnosing CAD, a more quantitative assessment of CAD diagnosis could be achieved using invasive intravascular ultrasound methods.

Atherosclerosis is a complex disease that involves interactions between circulating blood and the endothelium. Therefore, relying solely on blood circulation may only capture some of the intricate aspects of the pathophysiology of atherogenesis. Consequently, it may be more suitable to utilize intracoronary sampling as a more appropriate approach.

Conclusions

In conclusion, this thesis aimed to explore the genetic characteristics and determinants associated with smoking behaviours, with a focus on the HR populations, as well as the impact on gene expression in CAD patients with different smoking statuses. However, the initial hypothesis regarding genetic vulnerability among HR individuals towards smoking behaviours was disproven. The findings revealed a higher frequency and early initiation of smoking among HR individuals, but genetics did not significantly influence these behaviours. Socioeconomic factors demonstrated an inverse relationship with smoking behaviours. Future research should prioritize investigating non-genetic factors, particularly socioeconomic determinants, to better understand their influence on smoking behaviours, which can inform tobacco control efforts. Moreover, a significant difference in gene expression was observed among CAD patients based on their smoking status. This thesis unveiled novel overexpressed genes, including *RCAN3*, *TRAV4*, and *JCHAIN*, suggesting their potential involvement among CAD smokers. These findings could contribute to understanding smoking-related health risks and inform targeted interventions. There is an increasing need for blood biomarkers in diagnosing CAD to provide a more precise diagnostic approach in addition to relying solely on angiography and imaging techniques. Further research on the identified genes could be valuable in discovering blood markers and investigating potential targeted therapies, leading to personalized medicine approaches. This approach could enhance treatment effectiveness and patient outcomes by aligning prescribed medications with the individual gene expression profiles.

New contributions to academic knowledge

Study 1: Roma Socioeconomic Status Has a Higher Impact on Smoking Behaviour than Genetic Susceptibility

- ***Higher Smoking Frequency and Early Initiation was observed among HR Individuals***

HR individuals exhibited a significant 33% ($p < 0.001$) increase in smoking frequency compared to HG individuals, along with a strong tendency to initiate smoking at a younger age (standardized $\beta = -0.23$, $p < 0.001$).

- ***High prevalence of smoking behaviours and nicotine dependence in HR individuals is not genetically influenced***

The median values of the GRS did not show any significant difference, although the wGRS was slightly higher among Roma individuals (5.2 vs. 4.9; $p = 0.02$).

- ***An inverse relationship was found between SES and smoking behaviours***

Smokers tend to have lower SES than those who have never smoked, as indicated by ($\beta_{HR} = -0.039$, $p = 0.023$; $\beta_{HG} = -0.010$, $p = 0.049$).

Study 2: A Transcriptomic Analysis of Smoking-Induced Gene Expression Alterations in Coronary Artery Disease Patients

- ***120 DEGs were identified between SM and NSM among CAD patients***

A total of 120 genes, with 15 genes upregulated and 105 genes downregulated, were identified based on $FC \geq 1.5$ and a p -value < 0.05 .

- ***Upregulated genes were related to immune response pathways***

GO analysis showed that over-expressed genes were significantly annotated to immune response pathways ($FDR \leq 0.03$).

- ***Downregulated genes were associated to oxygen and haem binding activities***

Downregulated genes exhibited significant annotations with pathways involving oxygen and haem binding or activity ($FDR \leq 0.03$).

- ***Unexplored overexpression of JCHAIN, RCAN3, and TRAV4 genes was identified among smokers***

Upregulation of *JCHAIN*, *RCAN3*, and *TRAV4* genes were identified between SM and NSM with an FC of 2.1, 1.5, and 1.6, respectively, where FDR was ≤ 0.003 .

Summary

This thesis thoroughly examined the factors contributing to the high smoking rates within the Hungarian population and the impact of smoking on gene expression in CAD patients. Two rigorous observational studies were conducted using primary data. The first study identified a significant correlation between smoking behaviour and SES among HR individuals, despite the lack of evidence for genetic predisposition. The second study yielded valuable insights into the effects of cigarette smoking on gene expression in peripheral blood and its relation to smoking-induced CAD.

This thesis was based on two observational studies. The first was a comprehensive health survey with 814 participants comprising 412 HG and 402 HR. The study compared risk allele frequencies using GRS and wGRS based on ten SNPs. HR individuals exhibited significantly higher smoking frequency (33% higher) and tended to start smoking at a younger age (standardized $\beta = -0.23$, $p < 0.001$). However, despite the high prevalence of smoking behaviours and nicotine dependence among HR individuals, the study found no significant genetic influence, as median GRS values did not differ significantly, although wGRS was slightly higher among Roma individuals (5.2 vs 4.9; $p = 0.02$). Additionally, an inverse relationship between SES and smoking behaviours was observed, with smokers showing lower SES ($\beta_{HR} = -0.039$, $p = 0.023$; $\beta_{HG} = -0.010$, $p = 0.049$). The second analysis was based on 61 CAD patients with a median age of 67 years, and only 44 subjects were included for further analyses. Between SM and NSM, 120 DEGs (15 upregulated and 105 downregulated) were identified with a $FC \geq 1.5$ and a p -value < 0.05 . GO analyses revealed that the upregulated genes were related to immune response pathways ($FDR \leq 0.03$). In contrast, downregulated genes exhibited significant annotations with pathways involving oxygen and haem binding or activity ($FDR \leq 0.03$).

Our first analysis provides evidence of higher smoking frequency and early initiation among HR individuals, but these behaviours are not genetically influenced. The inverse relationship between SES and smoking behaviours highlights the importance of considering socioeconomic factors when studying and addressing smoking habits. On the other hand, the second study offers valuable insights into the potential effects of smoking on gene expression in whole blood and their connection to smoking-related CAD. The previously unexplored overexpression of *RCAN3*, *TRAV4*, and *JCHAIN* genes indicates a possible involvement in CAD among individuals who smoke.

Acknowledgement

First and foremost, I would like to express my sincere gratitude to my kind supervisor, **Dr. Szilvia Fiatal**, for her unwavering support and guidance throughout my Ph.D. journey. I am also deeply grateful to my co-authors, particularly **Prof. Sandor Janos**, **Dr. Szilard Poliska**, **Dr. Laszlo Balogh**, and **Dr. Istvan Szasz**, for their valuable contributions.

I would like to extend my heartfelt appreciation to my **loving family**, including **my parents, wife, brothers, and sisters** for their continuous support during my studies. Their unwavering encouragement has been instrumental in my academic pursuits.

Furthermore, I would like to acknowledge the support received from **Tempus Public Foundation**, which has played a significant role in facilitating my research endeavours.

Funding

The funding for these studies was provided by various sources, including the Stipendium Hungaricum scholarship, the European Union through the European Regional Development Fund (GINOP-2.3.2-15-2016-00005), the Hungarian Academy of Sciences (MTA11010 and TK2016-78), and the National Research, Development and Innovation Office of the Hungarian Ministry of Innovation and Technology (grant No OTKA 135784).



Registry number:
Subject:

DEENK/449/2023.PL
PhD Publication List

Candidate: Mohammed Merzah
Doctoral School: Doctoral School of Health Sciences
MTMT ID: 10086224

List of publications related to the dissertation

1. **Merzah, M.**, Póliska, S., Balogh, L., Sándor, J., Szász, I., Natae, S., Fiatal, S.: A Transcriptomic Analysis of Smoking-Induced Gene Expression Alterations in Coronary Artery Disease Patients.
Int. J. Mol. Sci. 24 (18), 1-14, 2023.
DOI: <http://dx.doi.org/10.3390/ijms241813920>
IF: 5.6 (2022)
2. **Merzah, M.**, Kósa, Z., Sándor, J., Natae, S., Pikó, P., Ádány, R., Fiatal, S.: Roma Socioeconomic Status Has a Higher Impact on Smoking Behaviour than Genetic Susceptibility.
Int. J. Environ. Res. Public Health. 18 (6), 1-12, 2021.
DOI: <http://dx.doi.org/10.3390/ijerph18063206>
IF: 4.614

List of other publications

3. Natae, S., **Merzah, M.**, Sándor, J., Ádány, R., Bereczky, Z., Fiatal, S.: A combination of strongly associated prothrombotic single nucleotide polymorphisms could efficiently predict venous thrombosis risk.
Front. Cardiovasc. Med. 10, 1-11, 2023.
DOI: <http://dx.doi.org/10.3389/fcvm.2023.1224462>
IF: 3.6 (2022)
4. **Merzah, M.**, Sulaiman, D., Karim, A. A., Khalil, M. E., Gupta, S., Almuzaini, Y., Hashemi, S., Mathew, S., Khatoon, S., Hoque, M. B.: A systematic review and meta-analysis on the prevalence and impact of coronary artery disease in hospitalized COVID-19 patients.
Heliyon. 9 (9), 1-12, 2023.
DOI: <http://dx.doi.org/10.1016/j.heliyon.2023.e19493>
IF: 4 (2022)





5. Hashemi, S., Ghazanfari, F., **Merzah, M.**, Rezaei, M., Astaraki, P., Birjandi, M.: Relationship between job stress and work-related quality of life among emergency medical technicians: a cross-sectional study.
BMJ Open. 13 (6), 1-7, 2023.
DOI: <http://dx.doi.org/10.1136/bmjopen-2022-066744>
IF: 2.9 (2022)
6. Natae, S., Kósa, Z., Sándor, J., **Merzah, M.**, Bereczky, Z., Pikó, P., Ádány, R., Fiatal, S.: The Higher Prevalence of Venous Thromboembolism in the Hungarian Roma Population Could Be Due to Elevated Genetic Risk and Stronger Gene-Environmental Interactions.
Front. Cardiovasc. Med. 8, 1-13, 2021.
DOI: <http://dx.doi.org/10.3389/fcvm.2021.647416>
IF: 5.846
7. **Merzah, M.**, Pikó, P., Ádány, R., Fiatal, S.: Roma socioeconomic status has higher impact on smoking behaviour than genetic susceptibility.
Eur. J. Public Health. 30 (Suppl5), 321-322, 2020.
DOI: <http://dx.doi.org/10.1093/eurpub/ckaa165.874>
IF: 3.367

Total IF of journals (all publications): 29,927

Total IF of journals (publications related to the dissertation): 10,214

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

02 October, 2023

