DISSERTATION 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

UNIVERSITY OF DEBRECEN
DOCTORAL SCHOOL OF HEALTH SCIENCE

DEBRECEN, 2023

DISSERTATION 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

Table of Contents

1 Introduction ... 1

2 Aims and Objectives: ... 3

3 Literature Review ... 4

3.1 Smoking ... 4

3.2 Genetic susceptibility to smoking ... 5

3.3 Coronary artery disease (CAD) ... 6

3.4 Smoking as a risk factor of CAD ... 6

3.5 Genetic susceptibility to CAD ... 7

3.6 Rational of the study ... 8

4 Material and Methods ... 10

4.1 Part I ... 10

4.1.1 Study design and population ... 10

4.1.2 Single-nucleotide polymorphism (SNPs) selection ... 10

4.1.3 DNA isolation and genotyping ... 10

4.1.4 Genetic risk scores ... 11

4.1.5 Smoking phenotypes ... 12

4.1.6 Hardy-Weinberg Equilibrium (HWE) ... 12

4.1.7 Socioeconomic status (SES) ... 12

4.1.8 Genetic models ... 12

4.1.9 Statistical analysis ... 13

4.1.10 Ethical approval ... 14

4.2 Part II ... 15

4.2.1 Study design and population ... 15

4.2.2 Blood samples collection ... 15

4.2.3 RNA isolation, library preparation, and sequencing ... 16

4.2.4 Cotinine assay ... 16

4.2.5 Clinical and demographical data ... 16

4.2.6 Statistical analysis ... 17

5 Results ... 19

5.1 Part I ... 19

5.1.1 Characteristics of the study populations ... 19

5.1.2 Risk allele frequencies ... 20

5.1.3 Association of genotype with quantitative smoking behaviours (cigarettes smoked per day and age of smoking initiation) ... 21

5.1.4 Association of genotype with qualitative smoking behaviours 24
5.1.5 Comparison of genetic risk scores between HR and HG Individuals 26
5.1.6 Assessing potential risk factors for quantitative smoking behaviours (cigarettes smoked per day and age at smoking initiation) 26
5.1.7 Assessing potential risk factors for qualitative smoking behaviours 28
5.2 Part II 30
5.2.1 Characteristics of the study populations 30
5.2.2 Differentially expressed genes (DEGs) between smokers and non-smokers..... 31
5.2.3 Gene Ontology and pathway analyses of the DEGs 37
6 Discussion 38
7 Limitation 45
8 Conclusions 47
9 New contributions to academic knowledge 48
10 Summary 50
Keywords 51
11 References 52
12 Acknowledgement 69
13 Funding 69
14 Appendices I
Appendix A: Supplementary Tables I
Appendix B: Supplementary Figures XII

List of Tables

Table 4-1: Regression models for analysing smoking behaviours and related factors 14

Table 5-1: Descriptive characteristics of the study populations.. 19

Table 5-2: Differences in risk allele frequencies between study populations 20

Table 5-3: Association of genotype with smoking status: additive model.............................. 25

Table 5-4: Regression model on cigarette per day in both populations 27

Table 5-5: Regression model on age at smoking imitation in both populations 27

Table 5-6: Association of SES with smoking status by study group. 28

Table 5-7: Association of smoking behaviours with SES and other variables........................ 29

Table 5-8: Demographic characteristics of study participants (N=61) 30

Table 5-9: The top 10 upregulated and downregulated genes (FC≥1.5, N=44)...................... 32

Table 5-10: Differentially expressed genes with a fold change of ≥2 (N=44)........................ 34

Appendices:

Table 14-1: Hardy-Weinberg equilibrium for Hungarian Roma...I

Table 14-2: Smoking status among both populations based on gender I

Table 14-3: Differences in risk allele frequencies between study populations II

Table 14-4: Genotype by smoking behaviours in general population.................................... III

Table 14-5: Genotype by smoking behaviours in the Roma population IV

Table 14-6: Differentially expressed genes with a fold change of ≥1.5 (N=61)...................... V

Table 14-7: Demographic characteristics of study participants (N=44)................................. VI

Table 14-8: Differentially expressed genes among smokers based on Nicotine Dependence

(N=31) VII

Table 14-9: Gene ontology of the 15 upregulated genes (N=44)..IX

Table 14-10: Gene ontology of the 105 downregulated genes (N=44) X

List of Figures

Figure 5-1: Cigarette per day (CPD) by genotype in the study populations. 21

Figure 5-2: CPD by genotype in HG (a1,2), (b1,2) in HR. 22

Figure 5-3: Age at smoking initiation by genotype in the study populations. 23

Figure 5-4: Age at smoking initiation by genotype.. 23

Figure 5-5: Frequency distributions of GRS and wGRS based on populations...................... 26

Figure 5-6: The heatmap of the DEGs with a fold change of ≥1.5 (N=44) 33

Figure 5-7: The heatmap of the differentially expressed genes with a fold change of ≥2 36

Figure 5-8: Gene ontology analysis of the most significant pathways................................... 37

Appendices:

Figure 14-1: The heatmap of the DEGs among smokers based on Nicotine Dependence (n=31) **..**XII

Figure 14-2: GO analysis of the DEGs among smokers categorized based on Nicotine Dependence XIII

List of Abbreviation:

BMI	Body mass index
CAD	Coronary artery disease
CAG	Coronary angiogram
CHD	Coronary heart disease
CI	Confidence interval
CPD	Cigarette per day
CVD	Cardiovascular disease
DALYs	Disability-adjusted life-years
DEGs	Differentially expressed genes
FC	Fold change
FDR	False discovery rate
FSM	Former smoker
FTND	Fagerstrom test for nicotine dependency
GO	Gene ontology
GRS	Genetic risk score
GWAS	Genome-wide association studies
HDL-C	High-density lipoprotein cholesterol
HG	Hungarian general
HR	Hungarian Roma
HSM	Heavy smoker
HWE	Hardy-Weinberg equilibrium
IQR	Interquartile range
LDL-C	Low-density lipoprotein cholesterol
MSM	Moderate smoker
ND	Nicotine dependence
NSM	Non-smokers
OR	Odd ratio
OS	Oxidative stress

QCA	Quantitative coronary angiography
RIN	RNA integrity number
SES	Socioeconomic status
SM	Smokers
SNP	Single nucleotide polymorphism
WBCs	White blood cells
wGRS	Weighted genetic risk score
WHO	World Health Organization

1 Introduction

Smoking has a profound global impact, contributing to over 8 million deaths each year globally[1]. If the current trend of smoking worldwide persists without alteration, the number of deaths caused by tobacco-related illnesses is estimated to exceed 8 million per year by 2030[2]. Smoking remains a significant public health concern in Europe, contributing to approximately 700,000 deaths annually[3]. In Hungary, smoking prevalence is high, with 24.9% of the population aged 15 and over engaging in daily smoking[4]. Moreover, significant disparities exist within the Hungarian population, with higher daily smoking prevalence observed among men and individuals aged 18-34 exhibit the highest rates, representing 35.6%[4]. Additionally, the Roma minority living in segregated areas is reported to have higher smoking rates compared to the general Hungarian population[5-7].

The alarming prevalence of smoking in Hungary and its association with various health risks necessitates a comprehensive understanding of its underlying factors. Addressing this issue becomes crucial in light of the social isolation experienced by minorities in segregated areas[8,9] and their potential genetic predisposition that may contribute to unhealthy lifestyles and disease development[9,10]. Therefore, identifying the genetic factors influencing smoking behaviours becomes imperative for implementing effective interventions aimed at smoking cessation.

Smoking is a well-established and modifiable risk factor for coronary artery disease (CAD)[11,12]. Unfortunately, Hungary bears a heavy burden of smoking-related diseases, with approximately a quarter of the adult population reporting daily smoking in 2019[13]. This rate makes Hungary as the third-highest country in Europe regarding smoking prevalence. In 2020, the prevalence of tobacco use was 28%, higher than the European average of 25%[13]. The continued high smoking rates are concerning since Hungary has a significant burden of coronary heart disease

(CHDs). According to the latest health report, CHD deaths make up 24.5% of all deaths in the country[13,14].

Despite the epidemiological findings linking smoking to non-communicable diseases, including CHDs[12,15,16], a research gap in the literature still exist regarding the effect of smoking on gene expression, specifically among the Hungarian population. Therefore, the second section of this dissertation aimed to explore the global gene expression patterns associated with smoking among CAD patients, providing insights into the genetic underpinnings of smoking behaviour and its potential implications for targeted interventions.

2 Aims and objectives:

1- To detect the possible genetic background of the high prevalence of smoking in the Hungarian Roma (HR) population

a. Determine the extent of smoking behaviours among HR compared to the Hungarian general (HG) population.

b. Comparing genetic profiles of smokers (SM) and non-smokers (NSM) within the HR and HG populations.

c. Identify genetic susceptibility to smoking using genetic risk score (GRS) and weighted genetic risk score (wGRS) of previously selected single nucleotide polymorphisms (SNPs).

d. Identify significant associations and quantify the strength of the relationship between genetic factors and smoking behaviours

e. Identify significant associations and quantify the strength of the relationship between other possible risk factors and smoking behaviours.

2- To analyse the gene expression in the whole blood of smokers and non-smokers among CAD patients.

a. Identify potential differences in gene expression associated with smoking status among individuals with CAD.

b. Gain insights into the molecular mechanisms and pathways related to smoking and CAD through gene ontology (GO) analyses.

3 Literature review

3.1 Smoking

Smoking has a profound global impact, contributing to over 8 million deaths from direct tobacco use each year globally[1]. However, among non-smokers, an additional 1.2 million deaths were attributed to second-hand smoke[1]. Unless there is a change in the ongoing smoking patterns, the number of deaths caused by tobacco-related illnesses is estimated to exceed 8 million per year by 2030[2]. In 2020 according to WHO, the proportion of men smokers were 4.7 times higher than women (36.7% vs 7.8%)[1]. According to a systematic analysis of 204 countries between 1990 and 2019, tobacco consumption is the third most significant risk factor for disability-adjusted life-years (DALYs)[17]. Tobacco smoking has caused 7.69 million deaths and 200 million DALYs, representing 13.6% of all human deaths and 7.89% of all DALYs[18].

Smoking remains a significant public health concern in Europe, contributing to approximately 700,000 deaths annually[3]. In Hungary, smoking prevalence is high, with 24.9% of the population aged 15 and over engaging in daily smoking[4]. Moreover, significant disparities exist within the Hungarian people, with higher daily smoking rates observed among men, and individuals aged 18-34 exhibit the highest rates, representing 35.6%[4]. Additionally, the Roma minority living in segregated areas was reported to have higher smoking rates compared to the general Hungarian population[5-7]. A study aimed to assess the smoking rate and support tobacco policies among HR revealed that the influence of ethnicity on attitudes towards tobacco control was partially elucidated by the Roma population's elevated smoking prevalence and comparatively lower educational attainment[7]. Numerous studies declared that smoking prevalence among HR individuals was higher than in HG people, estimated at 41-72%. This proportion was about five times higher than the rate among HG individuals[19-21]. Remarkably, a significant proportion of HR women, approximately 51.1%, are smokers and unlikely to quit

smoking. Notably, 70.3% of these women continued smoking during pregnancy, reflecting substantial resistance to cessation efforts[22,23].

Despite the implementation of tobacco control policies, smoking prevalence in Hungary remain high, particularly among vulnerable populations such as youth, low-income individuals, and those with lower levels of education[20,24]. Ongoing efforts are needed to further reduce smoking rates and improve public health outcomes in Hungary. Smoking cessation interventions and public education campaigns can play an essential role in reducing smoking rates and improving public health in Hungary.

3.2 Genetic susceptibility to smoking

Genetic predisposition to smoking behaviours might exist, as indicated by multiple genome-wide association studies (GWAS)[25-30]. Several susceptibility loci have been identified for nicotine dependence (ND) among African American and Chinese Han smokers[27,28]. A meta-analysis of 14 independent GWAS revealed a significant association between ND phenotypes and loci on the following genes: *CHRNA4, THSD7B, RBFOX1, RELN, IREB2, HYKK, PSMA, 4TTFCP2L1, COPB2, CHRNA5, CHRNA3, CHRNB4, INO80C, CHRNA4,* and *ZNF804A*[29]. A recent meta-analysis on decent European smokers identified over 1,250 significant variants on chromosomes 1, 4, 5, 9, 15, and 19 associated with various smoking phenotypes[31]. Studies involving twins and families have indicated that smoking addiction does not solely result from a single gene, but rather involves several genes contributing to increased vulnerability to nicotine addiction, revealing a complex genetic framework[32]. Knowledge is growing on the role of genes that play crucial roles in the development of ND. Recent analysis with a substantial sample size of 3.4 million individuals of 4 major global ancestries identified 3,823 variants associated with smoking phenotypes and alcohol consumption[33]. Among these, 2,486 distinct genetic variants and 1,346 loci were associated with smoking initiation. Additionally, 33 loci (comprising 39 variants), 140 loci (including 243 variants), and 128 loci (consisting of 206

variants) were found to be associated with the age of smoking onset, number of cigarettes per day, and smoking cessation, respectively[33].

3.3 Coronary artery disease (CAD)

CVD is the main cause of death in Europe[34,35], including CAD[36,37]. CAD is the most common form of CVDs. also called ischemic heart disease. In Europe, approximately one-fifth (20%) of all deaths can be attributed to CAD, making it the most prevalent form of CVD. The latest data reveals that CAD accounts for 16% of premature mortality in women under 75 years and 18% in men[38]. Numerous modifiable risk factors could lead to the development of CAD, including certain behaviours[39]. It is widely agreed upon that adopting a healthier lifestyle, such as quitting smoking, maintaining a healthy diet, and engaging in regular physical activity, can significantly reduce the incidence and mortality rates of CHD[40-43].

3.4 Smoking as a risk factor of CAD

A systematic review of 29 Mendelian randomized studies demonstrated a significant association between genetic predisposition to smoking, such as smoking initiation or lifetime smoking, and an increased risk of developing circulatory system diseases[44]. The estimated odds ratio (OR) for the risk of developing CAD associated with genetic predisposition to smoking was 1.31 (confidence interval (CI): 1.26-1.36)[44].

Smoking is a well-established and modifiable risk factor for CAD[11,12]. Unfortunately, Hungary bears a heavy burden of smoking-related diseases, with approximately a quarter of the adult population reporting daily smoking in 2019[13]. This rate makes Hungary the third highest country in Europe regarding smoking prevalence. In 2020, the prevalence of tobacco use was 28%, higher than the European average of 25%[13]. The continued high smoking rates are concerning since Hungary has a significant burden of CHD. According to the latest health report, coronary heart disease deaths make up 24.5% of all deaths in the country[13,14].

A study establishes a clear correlation between smoking intensity and a decline in heart functions[45]. Prolonged smoking over a period of ten years can lead to the development of ischemic cardiomyopathy characterized by a thicker, heavier, and weaker heart muscle, resulting in reduced systolic function compared to non-smokers[45]. Moreover, the chemical compounds in cigarettes can potentially harm the endothelium of arteries, causing the accumulation of fatty materials and forming atherosclerosis. The latter results in the development of arterial plaques, which may lead to restricted blood flow and ischemia, at the same time increasing the risk of severe conditions such as heart attack and stroke[12,46]. By eliminating risk factors such as smoking, it is possible to prevent atherosclerosis[47], as the primary constituents of cigarettes can affect the expression of several genes in the endothelial cells of arteries[48]. CAD or atherosclerosis is a complex disease involving multiple types of cells present in the peripheral blood, including B-cells, natural killer cells, T-cells, white blood cells (WBCs), platelets and circulating stem cells[49,50].

3.5 Genetic susceptibility to CAD

Not only are hyperlipidaemia, hypertension, obesity, and unhealthy lifestyle known to increase the risk of developing CAD[51], but it has also been discovered that CAD phenotypes are influenced by a combination of genetic and environmental factors[52-54]. Recent publications have highlighted the significant role of genetics on CAD risk. It is estimated that 40-60% of CAD cases can be attributed to genetic factors[51]. Individuals with a family history of CAD are approximately 1.5 times more likely to develop the disease compared to those without such a family history[55,56]. Numerous genes and genetic variations that correlate with an elevated susceptibility to CAD have been discovered through genetic research. These genes play a role in diverse biological activities connected to inflammation, lipid metabolism, blood clotting, and vascular function[57-61].

Understanding the genetic basis of CAD can have implications for disease prevention, risk assessment, and personalized treatment approaches[62]. Genetic testing and counselling may benefit for individuals with a family history of CAD or those who wish to assess their genetic predisposition to the disease[63,64]. However, it's important to note that genetic testing for CAD is still evolving, and the interpretation of results requires careful consideration in conjunction with other clinical and lifestyle factors[41,42].

3.6 Rational of the study

Genetic factors have been implicated in influencing smoking behaviours, as indicated by numerous GWAS, suggesting that specific genetic backgrounds may contribute to smoking tendencies[25,65]. Previous research has consistently provided strong evidence linking the higher prevalence of smoking behaviours and nicotine dependence to a genetic predisposition[26,66,67]. Additionally, earlier studies have demonstrated the prevalence of numerous risk alleles associated with specific non-communicable diseases among the Roma population[68-70]. These findings support the idea that a complex interplay of inheritable factors shapes the health status of the Roma community.

Other risk factors, such as SES, have also been identified as contributing to the high smoking rate among Roma individuals[71]. SES has been found to strongly influence the health of the Roma population and is associated with a higher prevalence of unhealthy behaviours, including smoking at a younger age[19,71].

Given these considerations, it is crucial to prioritize studies that focus on investigating specific diseases and health-related behaviours to enhance the well-being of the Roma minority population. Such studies can help identify specific genetic factors associated with smoking behaviours and inform targeted interventions and prevention strategies. Therefore, the first part

of this dissertation was conducted to highlight the genetic predispositions and other potential determinants contributing to the high smoking rate among Roma individuals.

Correspondingly, genetic studies have significantly contributed to advancing our understanding of the underlying pathophysiological mechanisms in CAD and identifying novel treatment targets[56,72-74]. While the association between smoking and CAD is well-established[12,15,16,46], there remains a need for extensive research that investigates the interplay between genetic factors and traditional risk factors, particularly smoking. Moreover, the impact of smoking on the blood transcriptome among CAD patients has yet to be previously studied. Therefore, the second study in this doctoral thesis aimed to address this gap in the literature by examining the global gene expression in the whole blood of SM and NSM among CAD patients. The findings can reinforce the existing epidemiological evidence and underscore the importance of further exploring the identified genes' functional implications and therapeutic potential.

4 Material and Methods

4.1 Part I

4.1.1 Study design and population

A health survey was previously created to build a complex database for comparison and association research to better understand the causes of the feeble health of Roma people, particularly the high burden of cardiometabolic disorders. Adult (20–64 years) Hungarian general (HG) and the Hungarian Roma (HR) individuals who participated in this cross-sectional study were subjected to questionnaire-based, physical, and laboratory examinations. A total of 832 people (417 HG and 415 HR) were recruited in this study[75]. Due to lacking anthropometric measurements, 18 subjects (13 HR and 5 HG) from the current research were disqualified in order to maintain a complete record database.

4.1.2 Single-nucleotide polymorphism (SNPs) selection

The process of SNPs selection was described previously[76]. An intensive literature search on PubMed was carried out, and a total of thirty-three SNPs were chosen. The selected criteria were that SNPs should be extracted from published studies with an adequate sample size and demonstrate a consistency relevant to the smoking intensity and a strong influence on smoking habits. A total of 10 SNPs were selected, while others were omitted for various reasons, including the lack of estimated effect size from genome-level association studies or to avoid collinearity (when two or more SNPs were strongly linked or correlated that they provide redundant information).

4.1.3 DNA isolation and genotyping

As reported previously[75], the DNA isolation and genotyping technique were disclosed. In brief, the DNA preparation was done from EDTA-anticoagulation blood samples on the day of sample collection. DNA was extracted from plasma samples using the MagNA Pure LC DNA Isolation Kit–Large Volume (Roche Diagnostics, Basel, Switzerland) following the

manufacturer's instructions. The genotyping took place on the mass-array platform (Sequenom Inc., San Diego, California, USA), with an iPLEX gold chemistry from a service provider for the mutation analysis. The validation and analysis of the concordant and quality controls were carried out following the provider's protocol (MAF, Karolinska Institute, Solna, Sweden).

4.1.4 Genetic risk scores

Each participant was assigned a score based on the presence of risk alleles they carried. A score of "0" indicated the absence of any risk alleles, a score of "1" was assigned to individuals who had one risk allele (heterozygote), and a score of "2" was given to individuals who carried two risk alleles in a homozygous state. Using these assigned scores, both weighted and unweighted genetic risk scores were calculated. *Equation (1)* was utilized to determine the unweighted genetic risk score (GRS), considering the number of risk alleles (G_i) associated with each SNP for each participant. Subsequently, the scores of all ten SNPs were combined for each person, assuming that all SNPs have a similar influence.

$$GRS = \sum_{i=1}^{i} G_i \tag{1}$$

Using the ORs of the risk allele of each SNP, obtained from previous studies[25,77-81], the weighted genetic risk scores (wGRS) were calculated. The calculation of the wGRS was performed using the *equation (2)*, where (W_{OR}) represents the risk coefficients assigned to each risk allele based on their previously determined relative effect sizes. Additionally, (X_i) indicates the number of effective alleles each individual carries.

$$wGRS = \sum_{i=1}^{i} (W_{OR}\, X_i) \tag{2}$$

4.1.5 Smoking phenotypes

Four categories were used to classify *qualitative smoking habits*: non-smokers (NSM) are individuals who have smoked fewer than 100 cigarettes in their whole life. Former smokers (FSM) were participants who had smoked 5 cigarettes per day (CPD) or more than 100 cigarettes during their lives but had abstained from smoking for at least a year. Participants who smoked 6 CPD but less than 20 CPD were classified as moderate smokers (MSM), whereas those who smoked more than 20 CPD were classified as heavy smokers (HSM)[24]. For further analysis, smokers were defined as former, moderate, and heavy smokers together. The study also considered *quantitative smoking behaviour* (CPD and age upon smoking initiation).

4.1.6 Hardy-Weinberg Equilibrium (HWE)

Hardy-Weinberg Equilibrium (HWE) was employed to evaluate the genotype distribution deviation for each population separately. Due to the failure of rs3762611 to follow HWE (Appendix: ***Table 14-1***) and the poor clustering of rs4105144, both SNPs were excluded from further analysis.

4.1.7 Socioeconomic status (SES)

The SES of participants was determined using the modified Kuppuswamy Scale 2019, a tool for classifying SES based on education level, occupation, and monthly family income. The criteria for calculating SES were previously described elsewhere[82]. Based on their SES, participants were divided into five categories: higher, upper-middle, lower-middle, upper-lower, and lower.

4.1.8 Genetic models

Three primary genetic models were employed to analyse the inheritance patterns of *smoking behaviours (quantitative and qualitative)*. The *Dominant Model* was used to ascertain whether one allele masked the effects of another. In cases where a dominant allele was present, it expressed its association with smoking behaviours. The *Recessive Model* was also applied to

determine whether smoking behaviours require two recessive alleles for expression. Additionally, the *Additive Model* was used to investigate whether the combination of two alleles contributed to smoking behaviours.

4.1.9 Statistical analysis

The data analysis was performed using Plink v.1.9[83] and the Statistical Packages for Social Sciences (SPSS v.25, IBM Corporation: Armonk, NY, USA)[84]. A chi-square test was employed to assess the associations between genotypes and smoking phenotypes (*quantitative* and *qualitative*) (Section 4.1.5), and other demographic variables. Linear regression analysis was employed to predict the most suitable indicator for quantifying smoking behaviours. ***Model I*** was employed for predicting CPD, while ***Model II*** was used for predicting the age at which individuals initiated smoking. In order to determine the optimal predictor of smoking status (smoker vs. non-smoker), binary regression models were conducted. ***Model III*** included SES, sex, age, body mass index (BMI), and genetic risk score (GRS) as independent variables. In contrast, ***Model IV*** included weighted genetic risk score (wGRS) as an additional independent variable. In addition, multinomial regression analysis was conducted to predict smoking behaviours, categorized as NSM, FSM, MSM, and HSM. The dependent variable in this analysis was the smoking behaviour category. Independent variables included SES, sex, age, ethnicity, BMI, and either GRS in ***Model V*** or wGRS in ***Model VI.*** A Benjamini-Hochberg multiple correction test was applied to control the false discovery rate (FDR) and compute the modified significant results. In the following table, a summary of the regression models employed in our study to analyse various aspects of smoking behaviour and related factors is presented.

Table 4-1: Regression models for analysing smoking behaviours and related factors

Model	Prediction	Regression Type	Independent Variables
I	CPD	Linear	SES, sex, age, BMI, GRS
II	Age at Smoking Initiation	Linear	SES, sex, age, BMI, GRS
III	Smoking Status (SM vs. NSM)	Binary	SES, sex, age, BMI, GRS
IV	Smoking Status (SM vs. NSM)	Binary	SES, sex, age, BMI, wGRS
V	Smoking Behaviours	Multinomial	SES, sex, age, ethnicity, BMI, GRS
VI	Smoking Behaviours	Multinomial	SES, sex, age, ethnicity, BMI, wGRS

CPD= cigarette per day; SM= smokers; NSM= non-smokers; SES= socioeconomic status; BMI= body mass index; GRS= genetic risk scores; wGRS= weighted genetic risk scores.

4.1.10 Ethical approval

The protocol was approved by the Hungarian Scientific Council on Health Committee (61327-2017/EKU). Formal informed consent was obtained from each individual, indicating their agreement to be part of this study.

4.2 Part II

4.2.1 Study design and population

This study was a single-centre study with a cross-sectional design, carried out at the University of Debrecen Medical Centre, Division of Cardiology and Cardiac Surgery. The study was conducted between November 2021 and October 2022. The Cardiology and Cardiac Surgery Clinic provides health services to a population of around 2.5 million individuals residing in the Eastern region of Hungary[85].

Sixty-one patients meeting specific inclusion criteria were recruited in this study. The inclusion criteria comprised adult patients who underwent their initial coronary angiogram (CAG) and were diagnosed with coronary lesions characterized by a more than 50% diameter reduction in the affected artery/arteries. Patients with hypertension, obesity, diabetes mellitus, previous coronary revascularization, or a luminal diameter less than 50% of the original diameter of the affected artery were excluded.

All participants provided informed written consent, and the research ethics committee of the University of Debrecen approved the study (protocol No.: 5797-2021).

4.2.2 Blood samples collection

Blood samples were collected from all participants using EDTA tubes for cotinine assay and laboratory tests, and PAXgene tubes (PreAnalytiX, Switzerland; reference number: 762165) for RNA isolation. The acquired PAXgene tubes were subjected to a two-hour incubation at room temperature post-sampling, followed by storage at -80°C until required. Plasma was promptly separated into aliquots within 30 minutes of blood collection using EDTA tubes and stored at -80°C for subsequent analysis.

4.2.3 RNA isolation, library preparation, and sequencing

Following the manufacturer's instructions, total RNA was extracted from blood samples using the PAXgene Blood RNA Kit (PreAnalytix, Switzerland). RNA quality was assessed using the Eukaryotic total RNA Nano assay on the Agilent 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA)[86]. The mean RNA integrity number (RIN) of all isolated RNA was 8.5 (SD: 0.6); these RIN values indicate good RNA integrity for subsequent library preparation[87]. RNA sequencing libraries were prepared using the NEBNext Ultra II RNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA) following the manufacturer's instructions. Briefly, ribosomal RNA was depleted using the NEBNext rRNA Depletion Kit (Human) according to the manufacturer's instructions. The remaining RNA was fragmented using the NEBNext Magnesium RNA Fragmentation Module, followed by cDNA synthesis using random hexamer primers and reverse transcriptase. The cDNA fragments were subjected to end repair, A-tailing, and adapter ligation using the NEBNext Ultra II DNA Library Prep Kit for Illumina. PCR amplification was performed to enrich adapter-ligated fragments, and the resulting libraries were purified using AMPure XP beads (Beckman Coulter, Indianapolis, IN, USA). RNA sequencing was performed on an Illumina NextSeq 500 system using a single-end sequencing protocol with a read length of 75-cycle.

4.2.4 Cotinine assay

The cotinine levels were evaluated using a specific ELISA kit designed for cotinine measurement, following the instructions provided by the manufacturer (Cloud-Clone Corp., Wuhan, China; Reference number: CET058Ge). Non-smokers were identified by a cotinine level of <1 ng/mL, while active smokers had a cotinine concentration of 10 ng/mL or higher[88].

4.2.5 Clinical and demographical data

Relevant demographic information, including age, sex, smoking status, nicotine dependence, anticoagulation medication, blood pressure, BMI, and laboratory data such as total cholesterol,

HDL-C, LDL-C, and triglyceride, were collected from every participant. A quantitative coronary angiography (QCA) was utilized to assess each study participant's extent of coronary stenosis. Patients with a luminal diameter of less than 50% of the original diameter were strictly excluded. Participants with stenosis ranging from 50% to 70% were categorized as having moderate stenosis, while those with stenosis greater than 70% were classified as having severe stenosis[89]. The level of nicotine dependence (ND) among smokers was examined using the FTND. The FTND consists of 6 items with a possible scale of 1-10, where a score of ≥ 8 indicates high ND, 5-7 indicates moderate ND, 3-4 indicates low-to-moderate ND, and ≤ 2 indicates low ND[90].

4.2.6 Statistical analysis

The MD-Anderson Bioinformatics website, which facilitates the computation of sample sizes for genomic experiments[91], was utilized to determine the appropriate sample size for this study. For detecting a 1.5-fold differences with a two-tailed alpha level of 0.05 and a power of 0.8, a sample size of 60 patients was required.

Descriptive statistics were conducted to analyse continuous and numerical variables. Measures of central tendency such as mean, standard deviation, and median were employed as appropriate. Relative frequencies or absolute values were used for representation categorical variables. All statistical analyses were conducted using the R software package.

FastQC was used to assess the quality control of all throughput sequencing data[92]. After trimming the sequence ends, the remaining high-quality reads underwent alignment to the human reference genome (GRCh38) through the utilization of the HISTAT2 algorithm. This process resulted in the generation of BAM files. Afterward, the analysis was performed using the StrandNGS software (accessible at www.strand-ngs.com). DESeq algorithm was utilized for normalization. In order to identify differentially expressed genes between groups, a Moderated-t-test was conducted.

GO analysis was performed using the CytoScape (v3.4) and ClueGo (v2.3.5) applications. The GO Biological Process Database and the list of differentially expressed genes were utilized to identify relevant pathways with statistical significance. A two-sided hypergeometric test with FDR correction at a significance level of <0.05 was applied to determine significance.

5 Results

5.1 Part I

5.1.1 Characteristics of the study populations

The demographical data of participants were summarized using descriptive statistics. Males represent 35.9% of the combined study population; 36.3% of them were HR people. One-quarter of the study subjects were aged ≥50 years. Regarding the smoking status, most of the HR were smokers (n=262, 65.2%), while 32.8% (n=135) of the HG individuals were reported to be smokers. Among the study populations, a statistically significant distinction was found regarding smoking status (p-value < 0.0001). HR females were twice more likely to be smokers than HG females, with an OR of 2.1 (95% CI: 1.5 - 2.9) (See Appendix: ***Table 14-2***). The study population exhibits a substantial SES variation, as evidenced by a significant p-value of less than 0.0001. Most HR people live at and under the lower middle SES compared to HG people (78.6% vs. 59.2%). Details on participants' characteristics are presented in ***Table 5-1***.

Table 5-1: Descriptive characteristics of the study populations

Variables		**Population** HG (N=412) n (%)	HR (N=402) n (%)	**χ2**	***p* value**
Sex	Male	186 (45.1)	106 (26.4)	28.69	**<0.001**
	Female	226 (54.9)	296 (73.6)		
Age (years)	20-29	70 (16.9)	79 (19.3)	9.19	0.10
	30-39	76 (18.4)	81 (19.8)		
	40-49	117 (28.4)	103 (25.1)		
	50-59	91 (22.2)	95 (23.2)		
	≥ 60	58 (14.1)	44 (10.8)		
Smoking Status	Smoker	135 (32.8)	262 (65.2)	86.49	**<0.001**
	Non-smoker	277 (67.2)	140 (34.8)		
Socioeconomic Status (SES)	Lower	0 (0)	21 (5.2)	26.98	**<0.001**
	Upper lower	69 (16.7)	106 (26.4)		
	Lower middle	175 (42.5)	189 (47.0)		
	Upper middle	150 (36.4)	86 (21.4)		
	Upper	18 (4.4)	0 (0)		

Bold font highlights significant differences between populations. HG= Hungarian general; HR= Hungarian Roma, N= total population size; n= number of observations; SES=socioeconomic status. This table adapted from *Merzah et al., 2021*.

5.1.2 Risk allele frequencies

The differences in the risk allele frequencies of the selected SNPs among HG and HR populations were conducted using chi-square (χ2) test. At an alpha level of <0.05, significant differences were identified in the frequencies of rs2673931, rs6517442, and rs2235186 (0.01, 0.03, and 0.01, respectively); while rs578776 was significant at <0.0001 level (Appendix: ***Table 14-3***). Upon adjustment and elimination of missing genotypes, all significant SNPs remained unchanged. In contrast, when applying the same analyses for each gender separately, the frequency of only rs2673931 remained significant among males with a p-value of <0.05 after adjusting. However, only the frequency of rs578776 among females remained significant after adjusting (p< 0.0001). See ***Table 5-2*** for more details.

Table 5-2: Differences in risk allele frequencies between study populations

SNP	Risk Allele	Frequency		χ2	*p* value	OR	ADJ-*p*[1]
		HR	HG				
A: Males							
rs10490162	T	0.90	0.91	0.22	0.63	0.87	0.73
rs2673931	**T**	**0.50**	**0.62**	**7.55**	**0.01**	**0.61**	**0.04**
rs4142041	G	0.34	0.40	2.04	0.15	0.77	0.25
rs2036534	T	0.75	0.78	0.46	0.49	0.87	0.67
rs16969968	A	0.34	0.34	0.01	0.94	0.99	0.95
rs578776	G	0.62	0.71	5.35	**0.02**	0.65	0.06
rs6517442	C	0.39	0.30	4.42	**0.03**	1.48	0.07
rs2235186	A	0.44	0.30	5.88	**0.01**	1.88	0.06
B: Females							
rs10490162	T	0.91	0.88	3.76	0.05	1.49	0.08
rs2673931	T	0.55	0.62	4.49	**0.03**	0.76	0.08
rs4142041	G	0.36	0.39	0.77	0.38	0.89	038
rs2036534	T	0.77	0.82	4.11	**0.04**	0.72	0.08
rs16969968	A	0.33	0.37	1.62	0.20	0.84	0.23
rs578776	**G**	**0.59**	**0.76**	**32.28**	**<0.001**	**0.45**	**<0.001**
rs6517442	C	0.34	0.29	2.53	0.11	1.25	0.15
rs2235186	A	0.39	0.33	4.72	**0.03**	1.34	0.08

Bold fonts highlight significant differences. Legend: [1] adjusted p-value. HG= Hungarian general; HR= Hungarian Roma, N= total population size; n= number of observations.

5.1.3 Association of genotype with quantitative smoking behaviours (cigarettes smoked per day and age of smoking initiation)

An independent sample-t test was applied to assess the difference in the mean of *CPD* among populations. The analysis revealed a significant variation in the CPD (p=0.002), where the mean was (12±13) in HR people and (5±9) in the HG population. Comparison based on gender revealed that HR males (M=14, SD=7) consumed approximately twice as much CPD compared to HG males (M=6, SD=5). On the other hand, HG females consumed roughly one-third of the CPD compared to their counterparts from the HR population. The CPD for HG females averaged around 4 and 12 for HR females (4±4 vs. 12±5).

In regards to genotype and CPD, the dominant homozygous of the risk allele of rs16969968-A, rs2235186-A, rs6517442-C, rs578776-G, and rs4142041-G was related to the higher CPD among study populations. Additionally, the additive-dominant models of rs2673931-T and rs10490162-T were associated with the high consumption of CPD in the study populations (***Figure 5-1***).

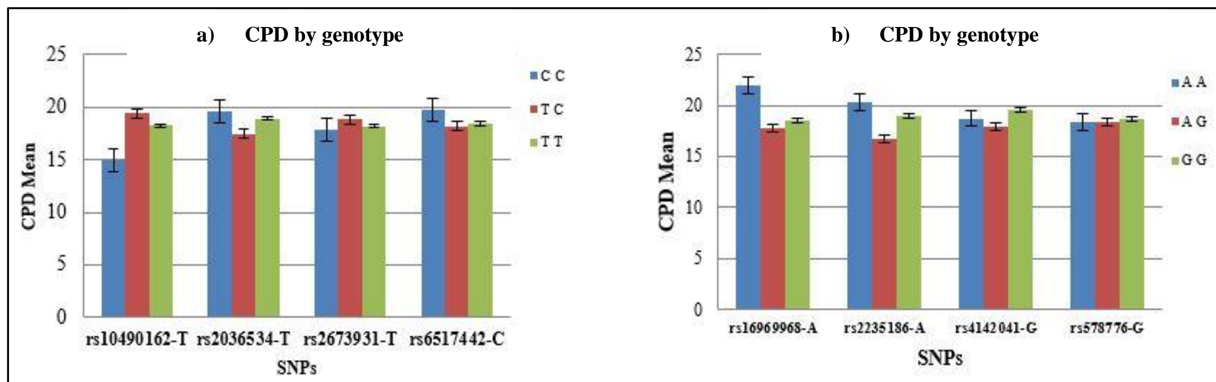

Figure 5-1: Cigarette per day (CPD) by genotype in the study populations.
Risk allele is written beside each SNP

The relationship between CPD and genotype varied when examining study population separately. Out of 8 SNPs, 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, by using the same model only four SNPs

(rs2673931, rs2235186, rs578776, and rs578776) were linked to higher CPD in the HR population (***Figure 5-2***).

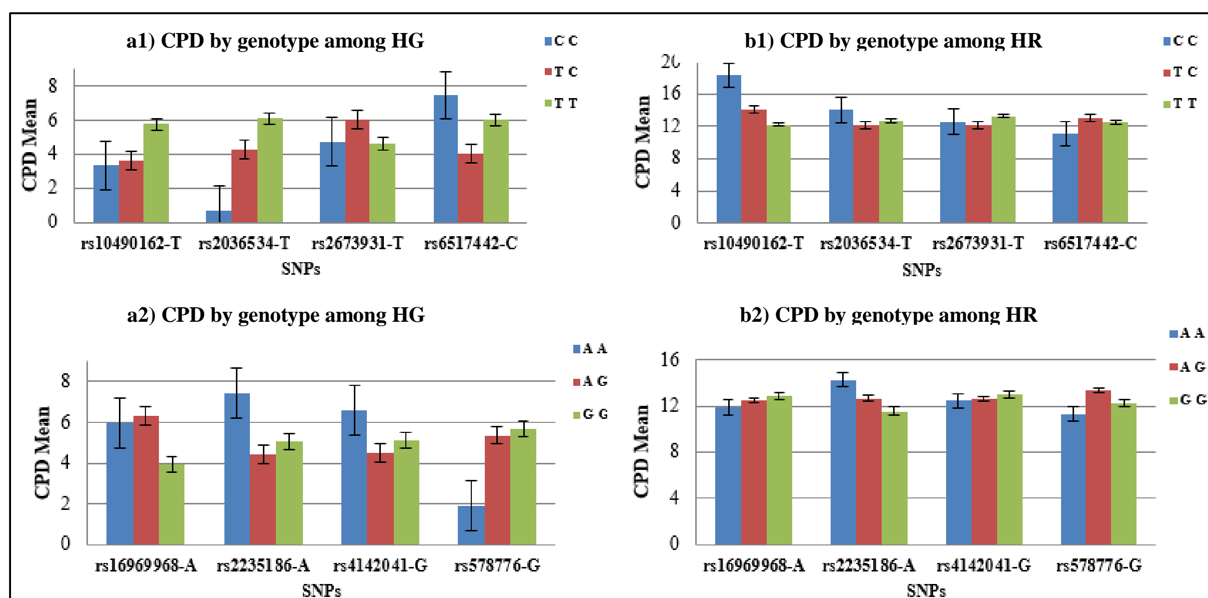

Figure 5-2: CPD by genotype in HG (a1,2), (b1,2) in HR.
Risk allele is written beside each SNP; HG= Hungarian general; HR= Hungarian Roma

Regarding *age at smoking initiation*, although the average age at smoking initiation was lower among HR compared to HG, 16±1.774 years and 17±0.308 years, respectively; our analysis revealed no significant differences in the mean age at smoking initiation between study groups. In general, individuals of the study subjects who carried a dominant homozygous allele of rs16969968, rs2235186, and rs4142041 were at higher risk of consuming cigarettes at an early age. In contrast, individuals who carried the dominant homozygous or heterozygous risk allele of rs10490162 and rs6517442 were found to initiate smoking at an early age of life (***Figure 5-3***).

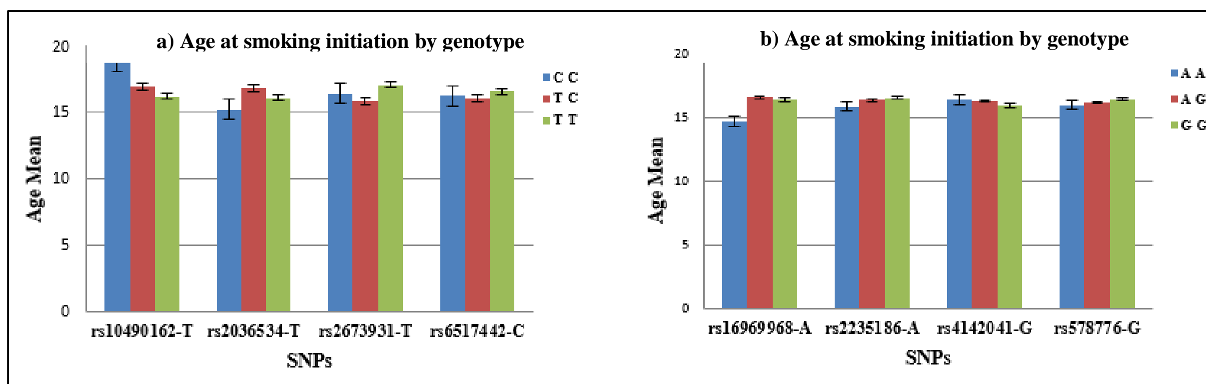

Figure 5-3: Age at smoking initiation by genotype in the study populations.
Risk allele is written beside each SNP

Additionally, the homozygous dominant model of the risk allele of rs6517442 was related to the tendency of smoking initiation at a younger age among HR individuals, with an average age of initiation being around 15±0.8 years. Similarly, HG individuals with a homozygous dominant of the risk allele at rs16969968 tended to start smoking at an early age, with an average age of 15.6±0.07 years. Among HR individuals, those with either dominant homozygous or heterozygous risk alleles for all SNPs, except for rs2673931-T, tended to start smoking at a young age. However, in the case of HG individuals, the early age of smoking initiation was associated with carrying homozygous risk alleles for six SNPs. See ***Figure 5-4*** for more details.

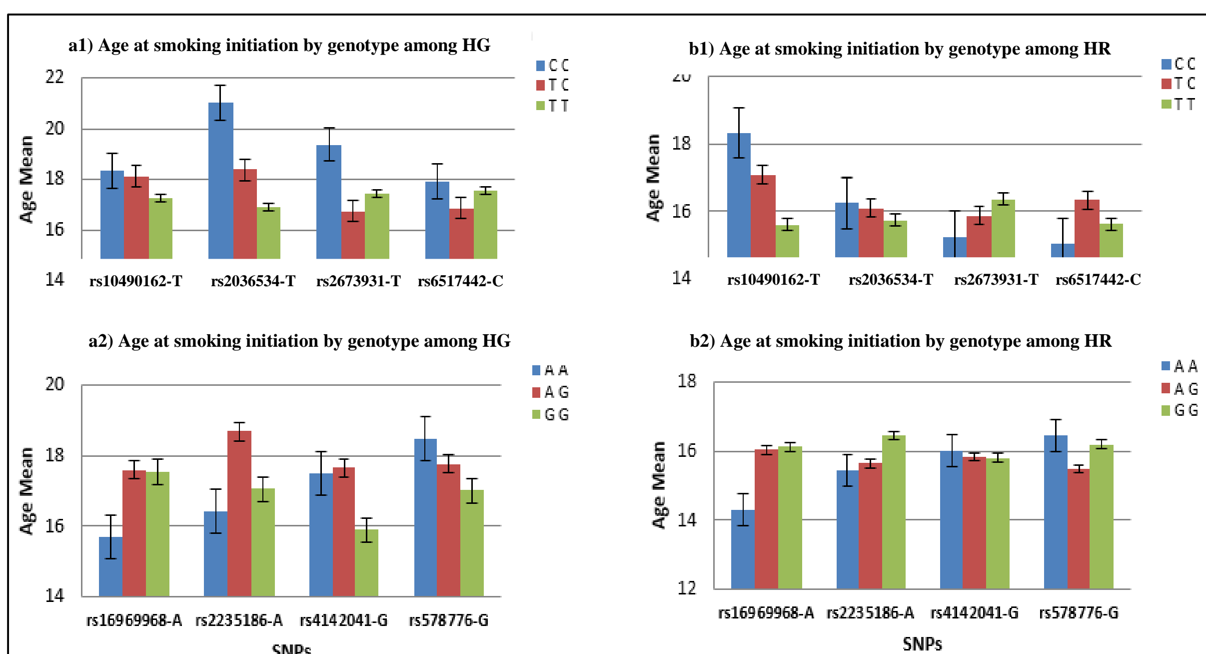

Figure 5-4: Age at smoking initiation by genotype.
Risk allele is written beside each SNP; HG= Hungarian general; HR= Hungarian Roma

5.1.4 Association of genotype with qualitative smoking behaviours

In order to identify the association between genotype and smoking status, a chi-square test was conducted on the study samples. This test was also conducted independently within each population to investigate the relationship between genotype and smoking behaviours. The additive models of rs2036534, rs578776, rs16969968, rs6517442, and rs2235186 were significantly differed between HG and HR populations according to smoking status (SM vs. NSM). A high prevalence of both homozygous and heterozygous risk alleles was observed among HR smokers for all significant SNPs (***Table 5-3***). The additive model revealed no significant SNP in HR when study population's smoking behaviours were investigated separately. However, in HG individuals, the only significant result ($p<0.001$) was observed for the additive model of rs2235186 with smoking behaviours (Appendix: ***Table 14-4*** and ***Table 14-5***).

Table 5-3: Association of genotype with smoking status: additive model

SNPs	Gene	Genotype	HG (n=412)		HR (n=402)		*p* value
			SM %	**NSM %**	**SM %**	**NSM %**	
rs2036534-T	*AGPHD1*	C C	0.2	3.2	4.1	2.3	**0.04**
		C T	9.6	23.8	24.2	12.0	
		T T	22.8	40.4	36.9	20.6	
rs578776-G	*CHRNA3*	A A	0.7	5.7	11.6	7.5	**<0.001**
		G A	13.9	25.4	30.2	12.9	
		G G	18.4	35.8	23.2	14.7	
rs16969968-A	*CHRNA5*	A A	3.4	7.3	6.4	5.6	**0.03**
		G A	18.8	30.3	28.5	13.7	
		G G	10.5	29.6	30.3	15.5	
rs4142041-G	*CTNNA3*	A A	13.4	21.8	28.1	15.1	0.30
		A G	15.6	36.1	28.8	15.8	
		G G	4.2	8.9	7.9	4.3	
rs6517442-C	*KCNJ6*	C C	3.7	4.9	6.3	5.1	**0.01**
		T C	10.5	31.1	32.2	14.5	
		T T	18.6	31.3	26.6	15.2	
rs10490162-T	*NRXN1*	C C	0.5	1.0	0.8	0.0	0.27
		T C	4.4	14.7	11.0	5.9	
		T T	27.9	51.5	53.2	29.2	
rs2235186-A	*MAOA*	A A	7.1	11.3	15.0	7.4	**<0.001**
		A G	9.3	16.9	26.4	10.7	
		G G	16.2	39.2	23.9	16.8	
rs2673931-T	*TRPC7*	C C	5.2	8.8	13.2	7.1	0.12
		C T	16.2	31.9	33.0	18.5	
		T T	11.3	26.5	19.0	9.1	

Bold font highlights significant differences. SM= smokers; NSM= non-smokers. Risk allele is written beside each SNP.

5.1.5 Comparison of genetic risk scores between HR and HG individuals

The median and interquartile ranges (IQRs) of GRSs were similar among HR and HG individuals, with a median value of 9, a p-value of 0.618, and an IQR of 7-10. However, when considering the wGRSs, the median was marginally higher for HR individuals (median=5.2) than HG individuals (median=4.9), with a p-value of 0.02. ***Figure 5-5*** shows the distribution of the GRSs among the study population. Among HG individuals, the IQR of wGRS was slightly lower than that of HG individuals (3-5 and 4-6, respectively). As a result, the distribution of wGRS was shifted towards higher values among HR individuals.

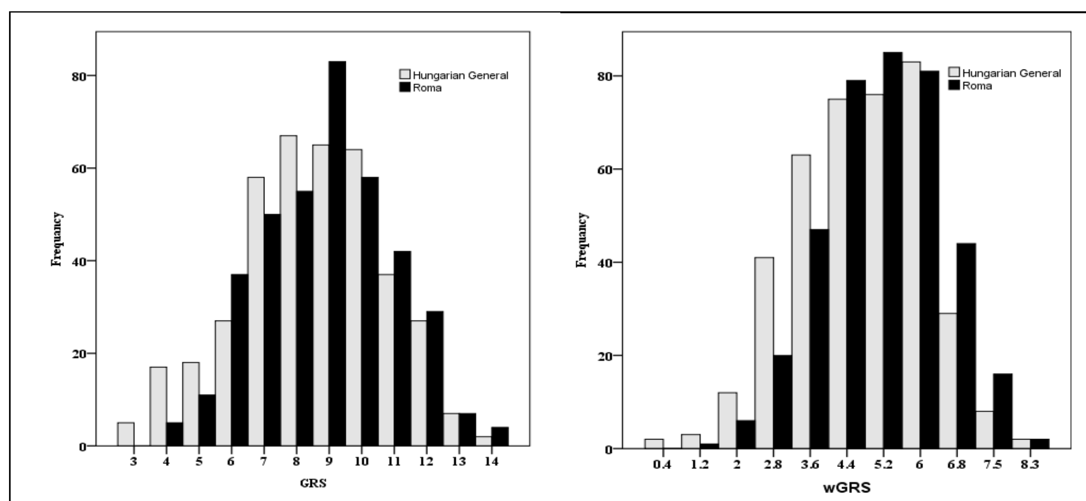

Figure 5-5: Frequency distributions of GRS and wGRS based on populations

5.1.6 Assessing potential risk factors for quantitative smoking behaviours (cigarettes smoked per day and age at smoking initiation)

Linear regression was applied to identify the best factor that might predict the risk of high smoked CPD and early age at smoking initiation. The dependent variable was CPD (***Model I***) and age at smoking imitation (***Model II***), where GRSs, population, SES, gender, age, and BMI were set as independent variables for both models. Results of ***Model I*** indicate that HR people had a 33% higher likelihood of smoking than HG people (p<0.001). Regarding age at smoking

initiation, the HR individuals were likelier to initiate tobacco at a younger age (standardized ß= -0.23, p<0.001). Furthermore, the analyses revealed that individuals who engaged in heavier smoking and began smoking at a younger age exhibit low BMI. Nonetheless, while sex showed a significant association with CPD, no significant association was found between sex and age at smoking initiation (**Table 5-4** and **Table 5-5**).

Table 5-4: Regression model on cigarette per day in both populations

Variables	Cigarette per day (CPD)							
	Model I[a]				Model I[b]			
	β	Standardized β	*t*	*p* value	β	Standardized β	*t*	*p* value
GRSs	-0.04	-0.01	-0.22	0.83	0.06	0.01	0.22	0.83
Population	7.67	0.33	9.28	**0.00**	7.63	0.33	9.17	**0.00**
SES	-0.06	-0.02	-0.46	0.65	-0.06	-0.02	-0.47	0.64
Gender	-2.45	-0.10	-2.93	**0.01**	-2.45	-0.10	-2.92	**0.01**
Age	-0.29	-0.03	-0.95	0.34	-0.30	-0.03	-0.97	0.33
BMI	-1.18	-0.10	-2.90	**0.01**	-1.18	-0.10	-2.89	**0.01**

Bold font highlights significant result. Legend: [a] CPD was set as a dependent variable while GRS, population, SES, gender, age and BMI were set as independent variables; [b] CPD was set as a dependent variable while wGRS, population, SES, gender, age and BMI were set as independent variables. Hungarian General was set as a reference for population. R^2=0.113.

Table 5-5: Regression model on age at smoking imitation in both populations

Variables	Age at smoking initiation							
	Model II[a]				Model II[b]			
	β	Standardized β	*t*	*p* value	β	Standardized β	*t*	*p* value
GRSs	0.07	0.02	0.46	0.65	0.03	0.01	0.12	0.90
Population	-4.05	-0.23	-6.17	**0.00**	-4.07	-0.23	-6.16	**0.00**
SES	0.02	0.01	0.20	0.84	0.02	0.01	0.21	0.83
Gender	0.03	0.00	0.05	0.96	0.03	0.00	0.05	0.96
Age	0.20	0.03	0.81	0.42	0.20	0.03	0.83	0.41
BMI	-0.70	-0.08	-2.19	**0.03**	-0.70	-0.08	-2.18	**0.03**

Bold font highlights significant result. Legend: [a] age at smoking initiation was set as a dependent variable while GRS, population, SES, gender, age and BMI were set as independent variables; [b] age at smoking initiation was set as a dependent variable while wGRS, population, SES, gender, age and BMI were set as independent variables. Hungarian General was set as a reference for population. R^2= 0.051.

5.1.7 Assessing potential risk factors for qualitative smoking behaviours

Further analyses used a binary regression test where the smoking status was applied as a dependent variable. The independent variables were SES, GRS (***Model III -Table 5-6 /A***), wGRS (***Model IV - Table 5-6/B***), age, sex, and BMI. Results of both models demonstrated that SES was the most influential predictor of smoking status, exhibiting a significant p-value of <0.05 across the studied populations. Individuals who smoked exhibited a lower SES in both populations, 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.

Table 5-6: Association of SES with smoking status by study group.

Variables	**A**					
	Hungarian Roma (*n*=402)			**Hungarian General (*n*=412)**		
	β	**95% *CI***	***p* value**	**β**	**95% *CI***	***p* value**
SES[a]	-0.04	0.02-0.03	**0.02**	-0.04	0.04-0.06	**0.05**
GRS	-0.01	-0.04-0.03	0.15	0.03	-0.03-0.09	0.30
Sex	-0.03	-0.12-0.24	0.61	-0.24	-0.59-0.12	0.19
Age	-0.06	-0.09-0.04	0.27	0.06	-0.05-0.17	0.27
BMI	-0.15	-0.26-0.03	**0.01**	-0.12	-0.31-0.07	0.20
			B			
SES[a]	-0.04	0.05-0.07	**0.02**	-0.01	0.06-0.07	**0.05**
wGRS	-0.16	-0.20-0.09	0.09	0.07	-0.04-0.17	0.16
Sex	-0.03	-0.20-0.63	0.36	-0.26	-0.07-0.08	0.13
Age	-0.06	-0.23-0.07	0.29	0.07	-0.05-0.19	0.23
BMI	-0.15	-0.44-$^{-}$0.08	**0.01**	-0.15	-0.35-0.05	0.13

Bold font means significant. R^2= 0.22 (**A**), R^2= 0.23 (**B**); smoking status (smokers and non-smokers) was fixed as a dependent variable. Legend: a = socioeconomic status. Adjusted regression analysis was used to evaluate the association where the model was adjusted for SES, sex, age, BMI and GRS (**A**) and with wGRS (**B**).

Additionally, a multinomial regression analysis was performed with two different models to predict the most influential factor of smoking behaviours. ***Model V*** employed GRSs, while ***Model VI*** incorporated the wGRSs. With HG designated as a reference category, the tested independent variables of both models included SES, age, sex, BMI, and population. Both models disclosed that SES was a significant predictor of smoking behaviours. The results demonstrated an inverse association between SES and smoking behaviours, with a statistical

significance 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 (***Table 5-7***). In conclusion, the analyses revealed that SES had a stronger relationship to smoking behaviours than genetic factors.

Table 5-7: Association of smoking behaviours with SES and other variables

Smoking Behaviours	Variables	Model V			Model VI		
		OR	**95% *CI***	***p* value**	**OR**	**95% *CI***	***p* value**
Former Smoker	GRSs	1.03	0.70-1.52	0.87	1.01	0.76-1.52	0.68
	SES=upper lower	1.01	1.00-1.09	**<0.01**	1.34	1.29-1.95	**<0.01**
	SES=lower middle	1.06	1.04-1.22	**<0.01**	1.48	1.05-1.70	**<0.01**
	SES=upper middle	1.36	1.35-1.75	**<0.01**	1.56	1.45-1. 60	**<0.01**
	Age	1.27	0.89-1.83	0.19	1.28	0.89-1.83	0.19
	BMI	0.85	0.54-1.34	0.48	0.85	0.53-1.34	0.48
	[Sex=Male]	0.54	0.17-1.68	0.29	0.54	0.17-1.68	0.29
	[Population=HG]	0.29	0.09-1.40	**0.01**	0.29	0.27-1.27	**0.02**
Moderate Smoker	GRSs	1.03	0.87-1.22	0.76	1.03	0.82-1.11	0.51
	SES=upper lower	1.99	1.89-2.56	**<0.01**	1.27	1.11-1.28	**<0.01**
	SES=lower middle	1.25	1.15-1.78	**<0.01**	1.10	1.09-1.62	**<0.01**
	SES=upper middle	0.39	0.29-0.85	**<0.01**	0.63	0.63-0.96	**<0.01**
	Age	1.06	0.90-1.25	0.46	0.90	0.91-1.25	0.44
	BMI	0.75	0.60-0.92	**0.01**	0.60	0.60-0.92	**0.01**
	[Sex=Male]	0.85	0.54-1.34	0.49	0.85	0.54-1.34	0.49
	[Population=HG]	0.38	0.29-4.02	**<0.01**	0.54	0.24-4.13	**<0.01**
Heavy Smoker	GRSs	1.03	0.88-1.20	0.74	1.05	0.91-1.19	0.53
	SES=upper lower	1.98	1.16-2.18	**<0.01**	1.66	1.23-1.69	**<0.01**
	SES=lower middle	1.11	1.10-1.86	**<0.01**	1.89	1.71-1.91	**<0.01**
	SES=upper middle	0.89	0.66-1.02	**<0.01**	0.96	0.68-0.99	**<0.01**
	Age	0.98	0.85-1.137	0.82	0.98	0.85-1.14	0.82
	BMI	0.73	0.61-0.88	**0.01**	0.73	0.61-0.88	**0.01**
	[Sex=Male]	1.62	1.10-2.40	**0.02**	1.62	1.09-2.39	**0.02**
	[Population=HG]	0.15	0.11-4.99	**<0.01**	0.15	0.11-4.93	**<0.01**

Bold font highlights significant results. Non-smoker was set as a reference. SES= socioeconomic status. Model V= using genetic risk score (GRS) with SES, BMI, age, sex, and population as independent variables. Model VI= using genetic risk score (wGRS) with SES, BMI, age, sex, and population as independent variables. SES= upper was set as a reference category; the lower category of SES was removed from the table, as it was considered as a redundant because no single HG subject was indicated in this category. The R2 of the association was 0.176

5.2 Part II

5.2.1 Characteristics of the study populations

Sixty-one subjects were recruited in this study, with one-half being smokers (50.8%, n=31). The frequency of male smokers (37.7%, n=23) exceeded that of male non-smokers (19.7%, n=12). However, among females, the proportion of non-smokers (29.5%, n=18) was greater than that of smokers (13.1%, n=8). 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. ***Table 5-8*** expressed other details on subjects' characteristics.

Table 5-8: Demographic characteristics of study participants (N=61)

Variables	Total	Smoking status		*p* value
		Smokers	Non-smokers	
Gender, n (%)				
Male	**35 (57.4)**	**23 (74.2)**	**12 (40.0)**	**0.007**
Female	**26 (42.6)**	**8 (25.8)**	**18 (60.0)**	
Age, median (range) years	**67 (28-88)**	**60 (28-75)**	**75 (54-88)**	**<0.001**
Cholesterol (mmol/L), mean (SD)	5.0 (1.4)	5.2 (1.5)	4.8 (1.4)	NS
Triglyceride (mmol/L), mean (SD)	1.6 (2.0)	1.9 (2.7)	1.4 (0.7)	NS
HDL-C (mmol/L), mean (SD)	1.3 (0.6)	1.4 (0.7)	1.3 (0.3)	NS
LDL-C (mmol/L), mean (SD)	3.0 (1.2)	3.1 (1.3)	2.9 (1.2)	NS
Number of affected arteries, median (range)	2 (1-3)	2 (1-3)	2 (1-3)	NS
Family history of CAD, n (%)	**20 (32.8)**	**14 (45.2)**	**6 (20.0)**	**0.04**
Anticoagulation medication, n (%)	61 (100)	31 (100)	30 (100)	NS
BMI, mean (SD)	26.0 (3.1)	25.3 (3.4)	26.8 (2.5)	NS
Blood pressure (mmHg), mean (SD)				
Systolic	125.9 (22.4)	120.8 (21.9)	131.3 (22.0)	NS
Diastolic	73.3 (12.0)	71.9 (13.5)	74.8 (10.4)	NS
Plasma cotinine ng/mL, mean (SD)	0.3 (0.1)	0.4 (0.2)	0.2 (0.1)	NS
Angiography findings, n (%)				
>2 arteries stenosis	45 (73.8)	20 (64.5)	25 (83.3)	NS
LAD	50 (82.0)	24 (77.4)	26 (86.6)	NS
LM	10 (16.4)	6 (19.4)	4 (13.3)	NS
CX	**34 (55.7)**	**13 (42.0)**	**21 (70.0)**	**0.03**
RCA	35 (57.4)	17 (54.8)	18 (60.0)	NS

LAD= left anterior descending artery; LM= left main coronary artery; CX= circumflex artery; n= number of subject; NS= non-significant; RCA= right coronary artery; SD= standard deviation. Bold text means significant.

5.2.2 Differentially expressed genes (DEGs) between smokers and non-smokers

To identify and compare genes that exhibited a differential expression between smokers and non-smokers, moderated-t test was performed. The initial analysis revealed 502 genes with significant expression differences with a p-value of <0.05. Up on adjusting the fold change (FC) threshold to ≥1.5, only 13 genes remained significant. Appendix: ***Table 14-6*** provides further details on those genes. After removing outlier samples by applying principle component analysis and clustering, forty-four samples remained for further analyses (Appendix: ***Table 14-7***). 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. The top 10 genes that displayed significant upregulation and downregulation, along with their corresponding FCs and p-values, were shown in ***Table 5-9***. 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.

Table 5-9: The top 10 upregulated and downregulated genes (FC≥1.5, N=44)

No.	Gene Symbol	FC	*p* value*	Description
Upregulated genes				
1	*GPR15*	7.5	1.4E-9	G protein-coupled receptor 15 [Source:HGNC Symbol;Acc:HGNC:4469]
2	*IGHA2*	2.6	0.034	immunoglobulin heavy constant alpha 2 (A2m marker) [Source:HGNC Symbol;Acc:HGNC:5479]
3	*IGLV3-21*	2.4	0.030	immunoglobulin lambda variable 3-21 [Source:HGNC Symbol;Acc:HGNC:5905]
4	*IGKV3-11*	2.2	0.006	immunoglobulin kappa variable 3-11 [Source:HGNC Symbol;Acc:HGNC:5815]
5	*JCHAIN*	2.1	0.037	joining chain of multimeric IgA and IgM [Source:HGNC Symbol;Acc:HGNC:5713]
6	*IGLC1*	2.0	0.044	immunoglobulin lambda constant 1 [Source:HGNC Symbol;Acc:HGNC:5855]
7	*IGLL5*	1.9	0.046	immunoglobulin lambda like polypeptide 5 [Source:HGNC Symbol;Acc:HGNC:38476]
8	*LEF1*	1.7	0.001	lymphoid enhancer binding factor 1 [Source:HGNC Symbol;Acc:HGNC:6551]
9	*CCR7*	1.6	0.030	C-C motif chemokine receptor 7 [Source:HGNC Symbol;Acc:HGNC:1608]
10	*CD28*	1.6	0.002	CD28 molecule [Source:HGNC Symbol;Acc:HGNC:1653]
Downregulated genes				
1	*KRT1*	-2.9	0.006	keratin 1 [Source:HGNC Symbol;Acc:HGNC:6412]
2	*NEAT1_3*	-2.7	0.005	Nuclear enriched abundant transcript 1 conserved region 3 [Source:RFAM;Acc:RF01957]
3	*PDZK1IP1*	-2.6	0.044	PDZK1 interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:16887]
4	*PLVAP*	-2.6	8.1E-6	plasmalemma vesicle associated protein [Source:HGNC Symbol;Acc:HGNC:13635]
5	*CA1*	-2.4	7.2E-4	carbonic anhydrase 1 [Source:HGNC Symbol;Acc:HGNC:1368]
6	*VWCE*	-2.4	0.036	von Willebrand factor C and EGF domains [Source:HGNC Symbol;Acc:HGNC:26487]
7	*KLF1*	-2.4	8.1E-4	Kruppel like factor 1 [Source:HGNC Symbol;Acc:HGNC:6345]
8	*HBM*	-2.3	2.3E-4	hemoglobin subunit mu [Source:HGNC Symbol;Acc:HGNC:4826]
9	*AHSP*	-2.3	0.016	alpha hemoglobin stabilizing protein [Source:HGNC Symbol;Acc:HGNC:18075]
10	*EPB42*	-2.2	0.006	erythrocyte membrane protein band 4.2 [Source:HGNC Symbol;Acc:HGNC:3381]

FC= fold change; negative sign means overexpressed in non-smokers. * corrected p-value.

A visual representation of the 120 DEGs was expressed using a heatmap. ***Figure 5-6*** illustrates the clustering of the DEGs based on study groups.

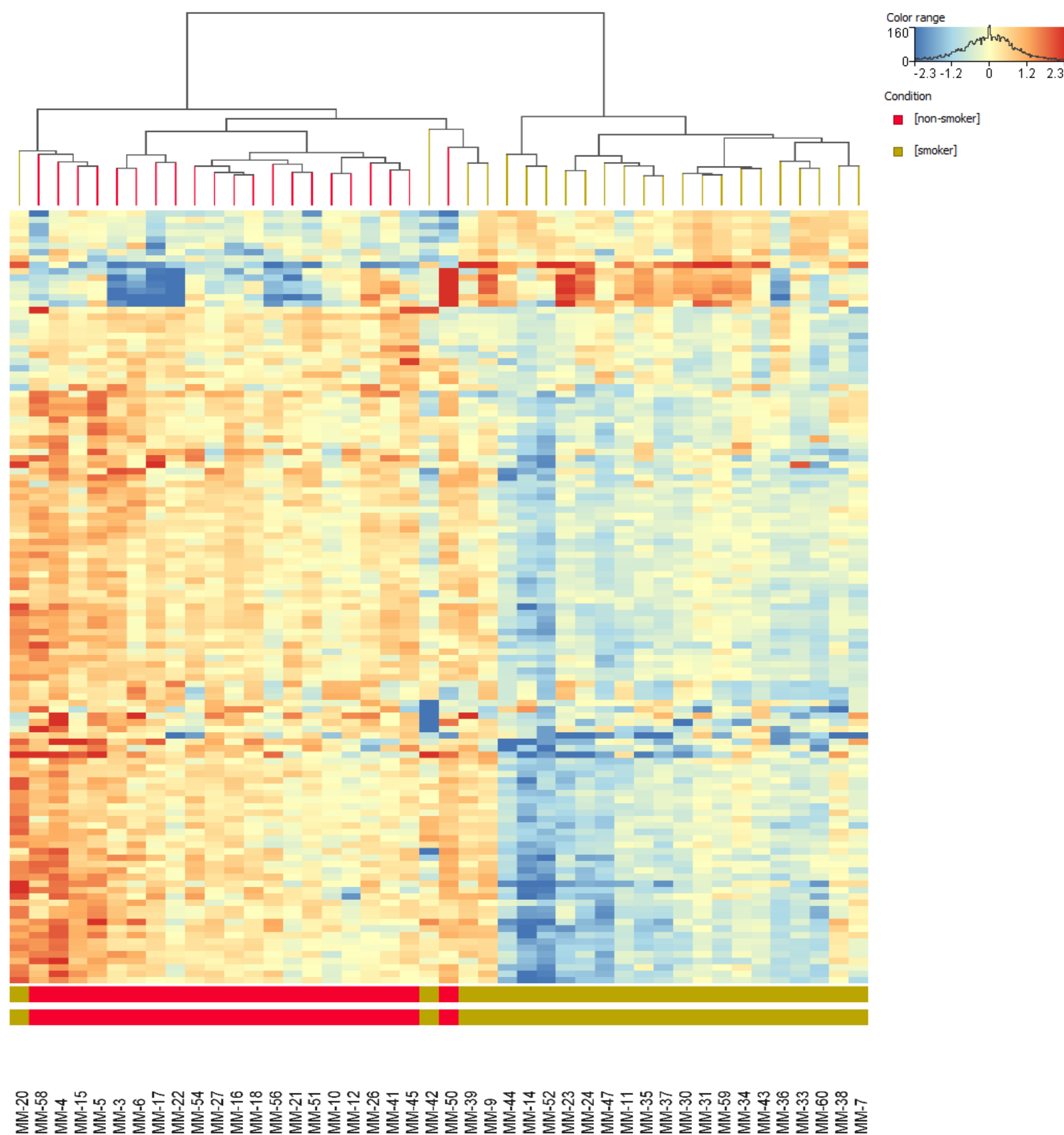

Figure 5-6: The heatmap of the DEGs with a fold change of ≥1.5 (N=44)

Additionally, when an FC of ≥2.0 was used with an adjusted p-value of <0.05, a total of 31 DEGs retained their significance. Of them, five genes were upregulated, and 26 genes were downregulated, as indicated in ***Table 5-10***.

Table 5-10: Differentially expressed genes with a fold change of ≥2.0 (N=44)

No.	Gene Symbol	FC	*p* value*	Description
Upregulated genes				
1	*GPR15*	7.5	1.4E-09	G protein-coupled receptor 15 [Source:HGNC Symbol;Acc:HGNC:4469]
2	*IGHA2*	2.6	2.8E-02	immunoglobulin heavy constant alpha 2 (A2m marker) [Source:HGNC Symbol;Acc:HGNC:5479]
3	*IGKV3-11*	2.2	6.2E-03	immunoglobulin kappa variable 3-11 [Source:HGNC Symbol;Acc:HGNC:5815]
4	*IGLV3-21*	2.4	3.0E-02	immunoglobulin lambda variable 3-21 [Source:HGNC Symbol;Acc:HGNC:5905]
5	*JCHAIN*	2.1	3.8E-02	joining chain of multimeric IgA and IgM [Source:HGNC Symbol;Acc:HGNC:5713]
Downregulated genes				
1	*ABALON*	-2.1	8.6E-04	apoptotic BCL2L1-antisense long non-coding RNA [Source:HGNC Symbol;Acc:HGNC:49667]
2	*AHSP*	-2.3	6.1E-03	alpha hemoglobin stabilizing protein [Source:HGNC Symbol;Acc:HGNC:18075]
3	*ALAS2*	-2.2	1.6E-02	5'-aminolevulinate synthase 2 [Source:HGNC Symbol;Acc:HGNC:397]
4	*BCL2L1*	-2.0	1.0E-03	BCL2 like 1 [Source:HGNC Symbol;Acc:HGNC:992]
5	*CA1*	-2.4	3.6E-02	carbonic anhydrase 1 [Source:HGNC Symbol;Acc:HGNC:1368]
6	*EPB42*	-2.2	3.5E-04	erythrocyte membrane protein band 4.2 [Source:HGNC Symbol;Acc:HGNC:3381]
7	*HBD*	-2.0	1.7E-03	hemoglobin subunit delta [Source:HGNC Symbol;Acc:HGNC:4829]
8	*HBM*	-2.3	1.6E-02	hemoglobin subunit mu [Source:HGNC Symbol;Acc:HGNC:4826]
9	*IFIT1B*	-2.2	1.5E-03	interferon induced protein with tetratricopeptide repeats 1B [Source:HGNC Symbol;Acc:HGNC:23442]
10	*KLC3*	-2.2	1.8E-03	kinesin light chain 3 [Source:HGNC Symbol;Acc:HGNC:20717]
11	*KLF1*	-2.4	2.4E-04	Kruppel like factor 1 [Source:HGNC Symbol;Acc:HGNC:6345]
12	*KRT1*	-2.9	5.8E-03	keratin 1 [Source:HGNC Symbol;Acc:HGNC:6412]
13	*MRC2*	-2.2	3.6E-04	mannose receptor C type 2 [Source:HGNC Symbol;Acc:HGNC:16875]
14	*MYL4*	-2.2	3.6E-04	myosin light chain 4 [Source:HGNC Symbol;Acc:HGNC:7585]
15	*NEAT1_3*	-2.7	4.40E-02	Nuclear enriched abundant transcript 1 conserved region 3 [Source:RFAM;Acc:RF01957]
16	*OSBP2*	-2.2	7.1E-04	oxysterol binding protein 2 [Source:HGNC Symbol;Acc:HGNC:8504]
17	*PDZK1IP1*	-2.6	2.8E-06	PDZK1 interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:16887]

18	*PLVAP*	-2.6	7.3E-04	plasmalemma vesicle associated protein [Source:HGNC Symbol;Acc:HGNC:13635]
19	*RUNDC3A*	-2.2	2.1E-04	RUN domain containing 3A [Source:HGNC Symbol;Acc:HGNC:16984]
20	*SELENBP1*	-2.2	2.3E-03	selenium binding protein 1 [Source:HGNC Symbol;Acc:HGNC:10719]
21	*SHISA4*	-2.2	1.9E-02	shisa family member 4 [Source:HGNC Symbol;Acc:HGNC:27139]
22	*SLC4A1*	-2.1	1.8E-03	solute carrier family 4 member 1 (Diego blood group) [Source:HGNC Symbol;Acc:HGNC:11027]
23	*SNCA*	-2.2	4.7E-03	synuclein alpha [Source:HGNC Symbol;Acc:HGNC:11138]
24	*TNS1*	-2.1	1.4E-03	tensin 1 [Source:HGNC Symbol;Acc:HGNC:11973]
25	*TRIM58*	-2.0	6.9E-03	tripartite motif containing 58 [Source:HGNC Symbol;Acc:HGNC:24150]
26	*VWCE*	-2.4	8.1E-04	von Willebrand factor C and EGF domains [Source:HGNC Symbol;Acc:HGNC:26487]

FC= fold change between smoker and nonsmokers; negative sign means overexpressed in nonsmokers. * corrected p-value.

A heatmap was generated to visualize the expression pattern of the 31 DEGs. The resulting heatmap vividly represented how the DEGs were grouped based on their expression levels (***Figure 5-7***).

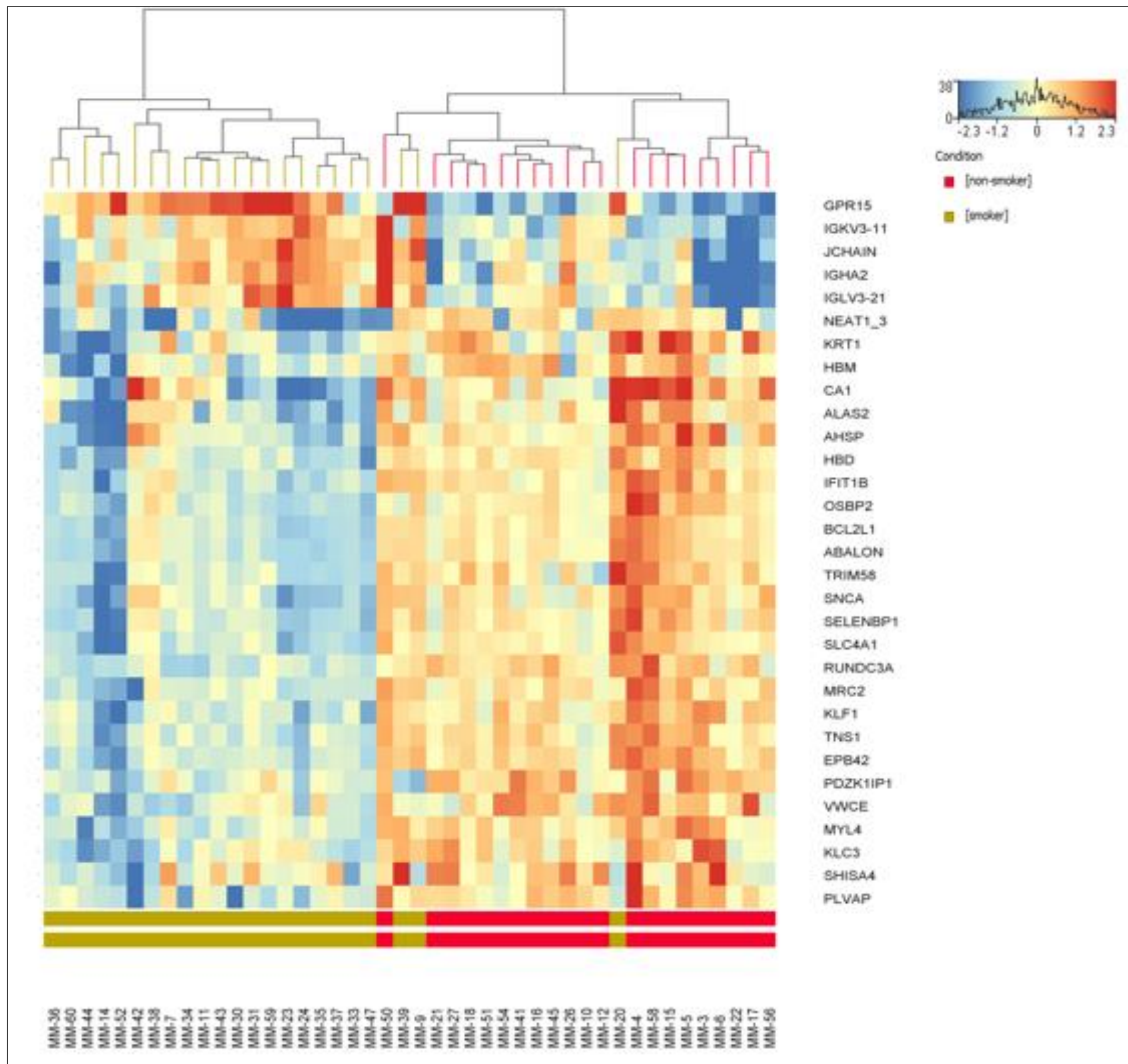

Figure 5-7: The heatmap of the differentially expressed genes with a fold change of ≥2

Among smoker subjects, subsequent analyses were performed using the Kruskal-Wallis test to identify DEGs based on the level of ND. By applying an FC of ≥1.5 and a p-value of <0.05, 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 (Appendix: ***Table 14-8*** and Figure 14-1).

5.2.3 Gene ontology and pathway analyses of the DEGs

The GO analyses demonstrated that the upregulated genes were primarily engaged in immune repose and inflammatory pathways. In contrast, the downregulated genes were significantly related to metabolic processes of respiratory burst and reactive oxygen species. See ***Figure 5-8*** for more details. ***Table 14-9*** displays the annotated GO functions of the upregulated genes, where three cellular components and 18 biological processes were annotated, with an adjusted p-value of <0.05. For the downregulated genes, five cellular components, five molecular functions, and 14 biological processes were identified (Appendix: ***Table 14-10***).

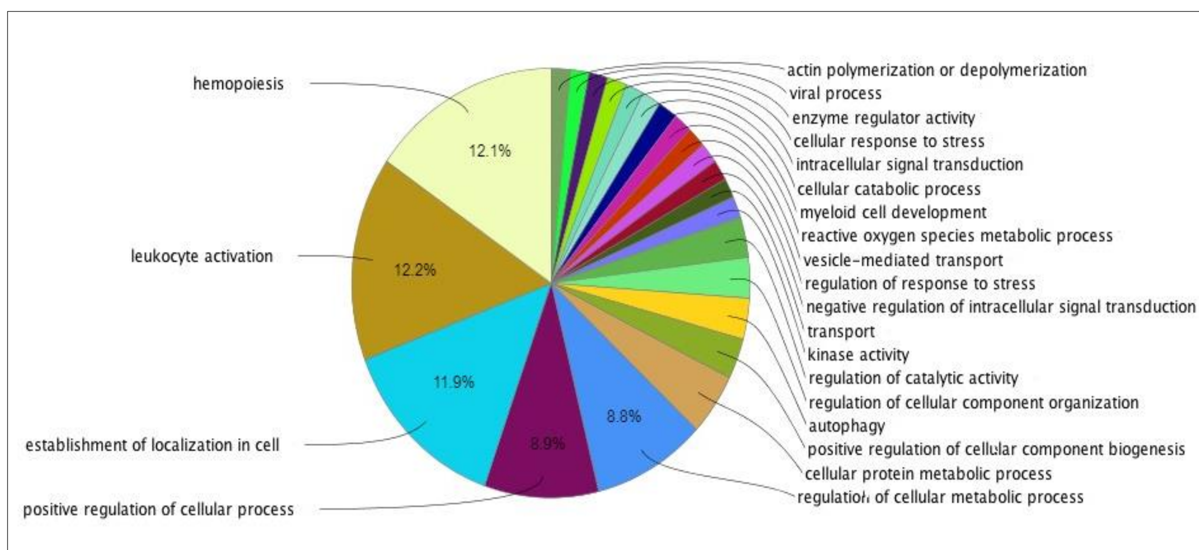

Figure 5-8: Gene ontology analysis of the most significant pathways

Among smokers, the GO analyses revealed that the upregulated genes were primarily linked to immune response activities. On the other hand, the downregulated genes were annotated with activities related to receptor regulation, inhibition, and antagonism (Appendix: ***Figure 14-2****)*.

6 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 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[8,9] and a genetic predisposition that could heighten their likelihood of engaging in unhealthy lifestyles or developing certain diseases[9,10]. 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[93]. 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 in 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[71, 94] . 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[71]. 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[71,95].

Furthermore, SES emerged as a prominent factor that could influence the health and lifestyles of HR individuals[20,96,97]. Living in economic difficulties increases the likelihood of smoking and reduces the prospect of smoking cessation[22,98,99.]. Our study revealed that living in a lower SES increases the likelihood of being a smoker, consistent with findings by others[6,20,71,97,99]. 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[7]. 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[100]. 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[100], 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.

Atherosclerosis is the main underlying cause of CAD[101], where various inflammatory processes play a significant role[102,103,104]. 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[105]. They act as receptors, and upon binding to specific antigens, they stimulate B-lymphocyte replication and differentiation, producing plasma cells that secrete immunoglobulins[106]. 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[107]. 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[108,109].

Additionally, another study has identified that measuring the plasma level of protein *IGHA2* could be a valuable biomarker for detecting early-stage atherosclerosis[110]. Similarly, the expression levels of *CD28, TRAT1, TRAV13-1, LEF1*, and other genes have been associated with smoking and smoking-related atherosclerosis[111-114]. These genes are centrally involved in activating adaptive immunity, particularly T-cell activation, which significantly contributes to the formation of atherosclerotic lesions[107,115-117.]

It has been discovered that the G protein-coupled receptor 15 (*GPR15*) is a critical regulator of T cell trafficking during systemic inflammation[118]. Recent research has found that increased expression of *GPR15* is linked to cardiovascular disorders and plays a role in the negative effects of smoking[119-124]. 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) [125-127], which facilitates T cell trafficking in patients with CAD[128-131].

The *ITGA6* gene, which belongs to the integrin family, has a vital function in various biological processes such as cell proliferation, invasion, and adhesion[132]. 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[133]. *Scott and Palmer* have suggested that smoking affects the levels of circulating adhesion molecules[134]. 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[45,134,135] .

The *RCAN3*, a coding protein, is believed to have a role in calcium-mediated signalling pathways[136]. 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[137]. 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[138]. Furthermore, two other genes, *TRAV4*[139], which encodes the T cell receptor alpha variable 4, and *JCHAIN*[140], which encodes the Immunoglobulin J Chain, have been linked to adenocarcinoma[141,142]. 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.

Having explored the implications of gene overexpression in the previous section, we now pivot our focus to the converse phenomenon—gene downregulation. Instead of delving into the extensive list of over a hundred genes that have been downregulated, this section opts to concentrate on the associated 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[143]. The capacity of haemoglobin to effectively transport oxygen may be hindered by smoking[144-146], leading to a reduced capacity. This effect can be attributed to the suppression or downregulation of these genes. Alternatively, haemoglobin subunit α2 (*HBA2*) is a closely

related variant of α1 (*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[147-149].

Moreover, the downregulation of alpha haemoglobin stabilizing protein (*AHSP*), a small protein responsible for regulating the stability and folding of the alpha haemoglobin (αHb) subunit[150], 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[151], 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[152, 153, 154]. Prolonged exposure to these chemicals can result in OS[153] , 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[155, 156].

The *ALAS2* gene encodes a protein called 5'-Aminolevulinate Synthase 2[157], which is involved in initiating the biosynthesis of haem[158]. Recent evidence indicates that this protein might act as a mediator, linking smoking with IL6 and CRP—two inflammatory markers— and exhibiting an inverse association[159]. 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 synthesizing of prostaglandin D[160]. 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[161,162]. Furthermore, a mouse model has provided evidence of the involvement of *PTGDS* in promoting thrombosis [162-164]. 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.

7 Limitation

One significant limitation of the first part of the dissertation to be reported is that we classified heavy smokers based on the number of CPDs; they were presumed to be nicotine dependent[97,165,166]. Alternatively, relying on reliable measurement tools such as the Fagerstrom Test for Nicotine Dependence (FTND)[167], Nicotine Dependence Syndrome Scale[168], or the Wisconsin Inventory of Smoking Dependence[169] 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[170]. Consequently, most smokers would have undetectable levels of cotinine within that timeframe. Self-reported smoking status and the Fagerström standard for nicotine dependence[90] were utilized to address this limitation.

Furthermore, the inclusion of patients with coronary stenosis of ≥50% of the luminal diameter of the affected artery 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 intravascular ultrasound.

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.

8 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.

9 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% ($p <0.001$) higher 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 ($\beta_{HR} = -0.039$, $p = 0.023$; $\beta_{HG} = -0.010$, $p = 0.049$).

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 ($FDR \leq 0.03$).

- ***Formerly 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.

10 Summary

This thesis investigates the genetic characteristics and determinants linked to smoking behaviours and explores the influence of smoking on gene expression in CAD patients. Studying these health-related behaviours can improve the well-being of the Roma minority population. Additionally, the impact of smoking on the blood transcriptome in Hungarian CAD patients remains understudied. These findings would be valuable in strengthening existing epidemiological evidence, identifying relevant determinants, and informing targeted interventions.

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 data of 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 fold change (FC) ≥ 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.

Keywords

Socioeconomic status, smoking behaviours, genetic susceptibility, Roma, Hungarian population, gene expression, coronary artery disease, next-generation sequencing, integrated investigation, transcriptomic analysis.

11 References

1. Tobacco. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/tobacco#:~:text=Tobacco kills more than 8,- and middle-income countries. Published 2022. Accessed June 27, 2023.
2. Organization WH. *WHO Report on the Global Tobacco Epidemic, 2017: Monitoring Tobacco Use and Prevention Policies*. World Health Organization; 2017.
3. Tobacco Overview. European Commission. https://health.ec.europa.eu/tobacco/overview_en. Accessed June 27, 2023.
4. Smoking habits 2019. Hungarian General Statistical Office. https://www.ksh.hu/docs/hun/xftp/idoszaki/elef/dohanyzas_2019/index.html. Published 2019. Accessed June 27, 2023.
5. Fiatal S, Tóth R, Moravcsik-Kornyicki Á, et al. High prevalence of smoking in the Roma population seems to have no genetic background. *Nicotine Tob Res*. 2016;18(12):2260-2267. doi:10.1093/ntr/ntw161
6. Kósa Z, Széles G, Kardos L, et al. A comparative health survey of the inhabitants of Roma settlements in Hungary. *Am J Public Health*. 2007;97(5):853-859. doi:10.2105/AJPH.2005.072173
7. Paulik E, Nagymajtényi L, Douglas Easterling, Rogers T. Smoking behaviour and attitudes of Hungarian Roma and non- Roma population towards tobacco control policies. *Int J Public Heal*. 2011;56(5). doi:10.1007/s00038-011-0252-x.Smoking
8. States M, Union E. *Roma Health Report:Health Status of the Roma Population; Data Collection in the Member States of the European Union.*; 2014.
9. Orton L, Cuevas RA De, Stojanovski K, et al. Roma populations and health inequalities : a new perspective. *Int J Hum RIGHTS Healthc*. 2019;12(5):319-327. doi:10.1108/IJHRH-01-2019-0004
10. Lau AY, Ridge M, Lau AY, Ridge M. Addressing the impact of social exclusion on mental health in Gypsy , Roma , and Traveller communities. *Ment Heal Soc Incl*. 2011;15(3):129-137. doi:10.1108/20428301111165717
11. McEvoy JW, Blaha MJ, DeFilippis AP, et al. Cigarette Smoking and Cardiovascular Events: Role of Inflammation and Subclinical Atherosclerosis: The Multi-Ethnic Study

of Atherosclerosis (MESA). *Arter Thromb Vasc Biol*. 2015;35(3):700-709. doi:10.1161/ATVBAHA.114.304562.Cigarette

12. Ning Ding, Yingying Sang, MSa, Jingsha Chen SHB, Kalbaugh CA, Salameh MJ, et al. Cigarette Smoking, Smoking Cessation, and Long-term Risk of 3 Major Atherosclerotic Diseases. *J Am Coll Cardiol*. 2019;74(4):498-507. doi:10.1016/j.jacc.2019.05.049.Cigarette
13. OECD, European Observatory on Health Systems and Policies. State of Health in the EU - Hungary: Country Health Profile 2021. 2021:24.
14. Nasr N, Soltész B, Sándor J, Ádány R, Fiatal S. Comparison of Genetic Susceptibility to Coronary Heart Disease in the Hungarian Populations : Risk Prediction Models for Coronary Heart Disease. *Genes (Basel)*. 2023;14(1033). doi:https://doi.org/10.3390/genes14051033
15. Banks E, Joshy G, Korda RJ, et al. Tobacco smoking and risk of 36 cardiovascular disease subtypes : fatal and non-fatal outcomes in a large prospective Australian study. *BMC Med*. 2019;17(128):1-18.
16. Khoramdad M, Leila AV. Association between passive smoking and cardiovascular disease : A systematic review and meta-analysis. *Int Union Biochem Mol Biol*. 2019;72:677–686. doi:10.1002/iub.2207
17. Global burden of 87 risk factors in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet (London, England)*. 2020;396(10258):1223-1249. doi:10.1016/S0140-6736(20)30752-2
18. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990-2019: a systematic analysis from the Global Burden of Disease Study 2019. *Lancet (London, England)*. 2021;397(10292):2337-2360. doi:10.1016/S0140-6736(21)01169-7
19. Cse P, Ne R, Ko K, Ko Z, Sze G. Does socioeconomic status fully mediate the effect of ethnicity on the health of Roma people in Hungary ? *J Epidemiol Community Heal*. 2003;63:455-460. doi:10.1136/jech.2008.079715
20. Vokó Z, Csépe P, Németh R, et al. Does socioeconomic status fully mediate the effect of ethnicity on the health of Roma people in Hungary? *J Epidemiol Community Health*. 2009;63(6):455-460. doi:10.1136/jech.2008.079715

21. Loring B. Tobacco and inequities> Guidance for addressing inequities in tobacco-related harm. In: *World Health Organization*. ; 2014.

22. Peter B, Andrea G, Ildiko R, L FK. Continued Smoking Versus Spontaneous Quitting Among Pregnant Women Living in A High Risk Environment. *Cent Eur J Public Health*. 2018;26(3):164-170. doi:10.21101/cejph.a5048

23. Hajdu T, Kertesi G, Kézdi G. *Health Differences at Birth between Roma and Non-Roma Children in Hungary: Long-Run Trends and Decomposition*. Vol 45.; 2019. doi:10.1111/padr.12276

24. Sándor J, Kósa Z, Boruzs K, Boros J, Tokaji I. The decade of Roma Inclusion : did it make a difference to health and use of health care services ? *Int J Public Health*. 2017;62(7):803-815. doi:10.1007/s00038-017-0954-9

25. Caporaso N, Gu F, Chatterjee N, et al. Genome-Wide and Candidate Gene Association Study of Cigarette Smoking Behaviors. *PLoS One*. 2009;4(2). doi:10.1371/journal.pone.0004653

26. Ma JZ, Payne TJ, Lou X, Zhang D, Dupont RT, Elston RC. Genome-wide linkage scan for nicotine dependence in European Americans and its converging results with African Americans in the Mid-South Tobacco Family sample. *Mol Psychiatry*. 2008;13:407-416. doi:10.1038/sj.mp.4002038

27. Li MD, Payne TJ, Ma JZ, et al. A genomewide search finds major susceptibility loci for nicotine dependence on chromosome 10 in African Americans. *Am J Hum Genet*. 2006;79(4):745-751. doi:10.1086/508208

28. Li M, Chen Y, Yao J, et al. Genome-Wide Association Study of Smoking Behavior Traits in a Chinese Han Population. *Front Psychiatry*. 2020;11(September):1-10. doi:10.3389/fpsyt.2020.564239

29. Chen J, Loukola A, Gillespie NA, et al. Genome-Wide Meta-Analyses of FTND and TTFC Phenotypes. *Nicotine Tob Res*. 2020;22(6):900-909. doi:10.1093/ntr/ntz099

30. Xu K, Li B, McGinnis KA, et al. Genome-wide association study of smoking trajectory and meta-analysis of smoking status in 842,000 individuals. *Nat Commun*. 2020;11(1). doi:10.1038/s41467-020-18489-3

31. Buchwald J, Chenoweth MJ, Palviainen T, et al. Genome-wide association meta-analysis of nicotine metabolism and cigarette consumption measures in smokers of

European descent. *Mol Psychiatry*. 2021;26(6):2212-2223. doi:10.1038/s41380-020-0702-z

32. Davies GE, Soundy TJ. The genetics of smoking and nicotine addiction. *S D Med*. 2009;Spec No(June):43-49.

33. Saunders GRB, Wang X, Chen F, et al. Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature*. 2022;612(7941):720-724. doi:10.1038/s41586-022-05477-4

34. Townsend N, Kazakiewicz D, Lucy Wright F, et al. Epidemiology of cardiovascular disease in Europe. *Nat Rev Cardiol*. 2022;19(2):133-143. doi:10.1038/s41569-021-00607-3

35. Timmis A, Vardas P, Townsend N, et al. European Society of Cardiology: cardiovascular disease statistics 2021: Executive Summary. *Eur Hear J - Qual Care Clin Outcomes*. 2022;8(4):377-382. doi:10.1093/ehjqcco/qcac014

36. Timmis A, Townsend N, Gale CP, et al. European society of cardiology: Cardiovascular disease statistics 2019. *Eur Heart J*. 2020;41(1):12-85. doi:10.1093/eurheartj/ehz859

37. MedTech Europe. Cardiovascular Disease in Europe: A Call to Action 2019-2024. European Reserach Group. https://www.medtecheurope.org/wp-content/uploads/2019/10/MedTech-Europe-Cardiovascular-Disease-in-Europe-Call-to-Action-August2020.pdf. Published 2019. Accessed July 4, 2023.

38. Daponte-Codina A, Knox EC, Mateo-Rodriguez I, et al. Gender and Social Inequalities in Awareness of Coronary Artery Disease in European Countries. *Int J Environ Res Public Health*. 2022;19(3):1-14. doi:10.3390/ijerph19031388

39. Yusuf S, Joseph P, Rangarajan S, et al. Modifiable risk factors, cardiovascular disease, and mortality in 155 722 individuals from 21 high-income, middle-income, and low-income countries (PURE): a prospective cohort study. *Lancet (London, England)*. 2020;395(10226):795-808. doi:10.1016/S0140-6736(19)32008-2

40. Oertelt-Prigione S, Seeland U, Kendel F, et al. Cardiovascular risk factor distribution and subjective risk estimation in urban women - The BEFRI Study: A randomized cross-sectional study. *BMC Med*. 2015;13(1). doi:10.1186/s12916-015-0304-9

41. Ramirez FD, Chen Y, Di Santo P, Simard T, Motazedian P, Hibbert B. Association

between self-reported potentially modifiable cardiac risk factors and perceived need to improve physical health: A population-based study. *J Am Heart Assoc*. 2017;6(5):13-15. doi:10.1161/JAHA.117.005491

42. Thakkar J, Heeley EL, Chalmers J, Chow CK. Inaccurate risk perceptions contribute to treatment gaps in secondary prevention of cardiovascular disease. *Intern Med J*. 2016;46(3):339-346. doi:10.1111/imj.12982

43. Nehme Z, Andrew E, Bernard S, et al. Impact of a public awareness campaign on out-of-hospital cardiac arrest incidence and mortality rates. *Eur Heart J*. 2017;38(21):1666-1673. doi:10.1093/eurheartj/ehw500

44. Larsson SC, Burgess S. Appraising the causal role of smoking in multiple diseases: A systematic review and meta-analysis of Mendelian randomization studies. *eBioMedicine*. 2022;82:104154. doi:10.1016/j.ebiom.2022.104154

45. Holt E, Skaarup KG, Lassen MCH, et al. The effects of smoking on cardiac structure and function in a general population. *Eur Heart J*. 2022;43(Supp. 2):2022.

46. Li J, Liu S, Cao G, et al. Nicotine induces endothelial dysfunction and promotes atherosclerosis via GTPCH1. *J Cell Mol Med*. 2018;22(June):5406-5417. doi:10.1111/jcmm.13812

47. McEvoy JW, BAO1 KN, DeFilippis AP, et al. The Relationship of Cigarette Smoking with Inflammation and Subclinical Vascular Disease: The Multi-Ethnic Study of Atherosclerosis. *Arter Thromb Vasc Biol*. 2015;35(4):1002-1010. doi:10.1161/ATVBAHA.114.304960.The

48. Pittilo RM. Cigarette smoking , endothelial injury and cardiovascular disease. *Cardiovasc Pathol*. 2000:219-230.

49. Fan H, Hegde PS. The Transcriptome in Blood : Challenges and Solutions for Robust Expression Profiling. *Curr Mol Med*. 2005;5:3-10.

50. Hafiz M, Abdullah N, Othman Z, et al. Peripheral blood gene expression profile of atherosclerotic coronary artery disease in patients of different ethnicity in Malaysia. *J Cardiol*. 2012;60(3):192-203. doi:10.1016/j.jjcc.2012.05.009

51. Ralapanawa U, Sivakanesan R. Epidemiology and the magnitude of coronary artery disease and acute coronary syndrome: A narrative review. *J Epidemiol Glob Health*. 2021;11(2):169-177. doi:10.2991/JEGH.K.201217.001

52. Ozaki K, Tanaka T. Molecular genetics of coronary artery disease. *J Hum Genet*. 2016;61(1):71-77. doi:10.1038/jhg.2015.70

53. Duan S, Luo X, Dong C. Identification of susceptibility modules for coronary artery disease using a genome wide integrated network analysis. *Gene*. 2013;531(2):347-354. doi:10.1016/j.gene.2013.08.059

54. Amit VK, Sekar K. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat Rev Genet*. 2017;18(6):331-344. doi:10.1038/nrg.2016.160.Genetics

55. Wang H, Liu Z, Shao J, et al. Pathogenesis of premature coronary artery disease: Focus on risk factors and genetic variants. *Genes Dis*. 2022;9(2):370-380. doi:10.1016/j.gendis.2020.11.003

56. Kessler T, Schunkert H. Coronary Artery Disease Genetics Enlightened by Genome-Wide Association Studies. *JACC Basic to Transl Sci*. 2021;6(7):610-623. doi:10.1016/j.jacbts.2021.04.001

57. Aragam KG, Jiang T, Goel A, et al. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat Genet*. 2022;54(12):1803-1815. doi:10.1038/s41588-022-01233-6

58. Cadby G, Giles C, Melton PE, et al. Comprehensive genetic analysis of the human lipidome identifies loci associated with lipid homeostasis with links to coronary artery disease. *Nat Commun*. 2022;13(1):1-17. doi:10.1038/s41467-022-30875-7

59. Van Meurs JBJ, Pare G, Schwartz SM, et al. Common genetic loci influencing plasma homocysteine concentrations and their effect on risk of coronary artery disease. *Am J Clin Nutr*. 2013;98(3):668-676. doi:10.3945/ajcn.112.044545

60. Ganopoulou M, Moysiadis T, Gounaris A, et al. Single Nucleotide Polymorphisms' Causal Structure Robustness within Coronary Artery Disease Patients. *Biology (Basel)*. 2023;12(5). doi:10.3390/biology12050709

61. Nikpay M, Goel A, Won H-H, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015;47(10):1121-1130. doi:10.1038/ng.3396

62. Assimes TL, Roberts R. Genetics: Implications for Prevention and Management of Coronary Artery Disease. *J Am Coll Cardiol*. 2016;68(25):2797-2818. doi:10.1016/j.jacc.2016.10.039

63. Klarin D, Natarajan P. Clinical utility of polygenic risk scores for coronary artery disease. *Nat Rev Cardiol*. 2022;19(5):291-301. doi:10.1038/s41569-021-00638-w

64. Rotter JI, Lin HJ. An Outbreak of Polygenic Scores for Coronary Artery Disease. *J Am Coll Cardiol*. 2020;75(22):2781-2784. doi:10.1016/j.jacc.2020.04.054

65. David SP, Hamidovic A, Chen GK, et al. Genome-wide meta-analyses of smoking behaviors in African Americans. *Transl Psychiatry*. 2012;2(5):e119. doi:10.1038/tp.2012.41

66. Erzurumluoglu AM, Liu M, Jackson VE, et al. Meta-analysis of up to 622,409 individuals identifies 40 novel smoking behaviour associated genetic loci. *Mol Psychiatry*. 2020;25(10):2392-2409. doi:10.1038/s41380-018-0313-0

67. Verhagen M, Kleinjan M, Engels RC. A systematic review of the A118G (Asn40Asp) variant of OPRM1 in relation to smoking initiation, nicotine dependence and smoking cessation. *Pharmacogenomics*. 2012;13(8):917-933. doi:10.2217/pgs.12.76

68. Ko Z, Fiatal S. The genetic risk for hypertension is lower among the Hungarian Roma population compared to the general population. *PLoS One*. 2020;15(6):1-17. doi:10.1371/journal.pone.0234547

69. Pikó P, Fiatal S, Kósa Z, Sándor J, Ádány R. Genetic factors exist behind the high prevalence of reduced high-density lipoprotein cholesterol levels in the Roma population. *Atherosclerosis*. 2017;263:119-126. doi:10.1016/j.atherosclerosis.2017.05.028

70. Fiatal S, Pikó P, Kósa Z, Sándor J, Ádány R. Genetic profiling revealed an increased risk of venous thrombosis in the Hungarian Roma population. *Thromb Res*. 2019;179:37-44. doi:10.1016/j.thromres.2019.04.031

71. Hiscock R, Bauld L, Amos A, Fidler JA, Munaf M. Socioeconomic status and smoking : a review. *Ann NY Acad Sci*. 2012;1248:107-123. doi:10.1111/j.1749-6632.2011.06202.x

72. Zhifen Chen, Schunkert H. Genetics of coronary artery disease in the post-GWAS era. *J Intern Med*. 2021;290:980–992. doi:https://doi.org/10.1111/ joim.13362

73. Butnariu LI, Florea L, Badescu MC, Țarcă E, Costache I-I, Gorduza EV. Etiologic Puzzle of Coronary Artery Disease: How Important Is Genetic Component? *Life (Basel, Switzerland)*. 2022;12(6). doi:10.3390/life12060865

74. McPherson R, Tybjaerg-Hansen A. Genetics of Coronary Artery Disease. *Circ Res*. 2016;118(4):564-578. doi:10.1161/CIRCRESAHA.115.306566

75. Ádány R, Pikó P, Fiatal S, et al. Prevalence of insulin resistance in the Hungarian general and roma populations as defined by using data generated in a complex health (Interview and examination) survey. *Int J Environ Res Public Health*. 2020;17(13):1-22. doi:10.3390/ijerph17134833

76. Fiatal S, Tóth R, Moravcsik-kornyicki Á, et al. Original investigation High Prevalence of Smoking in the Roma Population Seems to Have No Genetic Background. *Nicotine Tob Res*. 2016;18(12):2260-2267. doi:10.1093/ntr/ntw161

77. Bierut LJ, Madden PAF, Breslau N, et al. Novel Genes Identified in a High-Density Genome Wide Association Study for Nicotine Dependence. *Hum Mol Genet*. 2007;16(1):24-35. doi:10.1093/hmg/ddl441

78. Liu JZ, Tozzi F, Waterworth DM, et al. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet*. 2010;42(5). doi:10.1038/ng.572

79. Thorgeirsson TE, Gudbjartsson DF, Surakka I, et al. Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat Genet*. 2010;42(5):448-453. doi:10.1038/ng.573

80. Furberg H, Kim Y, Dackor J, et al. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet*. 2010;42(5):441-447. doi:10.1038/ng.571

81. Saccone SF, Hinrichs AL, Saccone NL, et al. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet*. 2007;16(1):36-49. doi:10.1093/hmg/ddl438

82. Saleem SM. Modified Kuppuswamy socioeconomic scale updated for the year 2019. *Indian J Forensic Community Med*. 2019;6(1):1-3. doi:10.18231/2394-6776.2019.0001

83. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW DM& SP. PLINK: a toolset for whole-genome association and population-based linkage analysis. 2007. doi:81

84. Statistics IBMS. IBM SPSS Statistics for Windows, version 25. 2017.

85. University of Debrecen. Department of Cardiology and Cardiac Surgery.

https://unideb.hu/en/department-cardiology-and-cardiac-surgery-cardiac-surgery. Published 2021. Accessed March 13, 2023.

86. Bioanalyzer. Agilent 2100 Bioanalyzer System: 2100 Expert Software User's Guide. © *Agil Technol Inc 2000-2020*. 2020;2000-2020:1-252. https://www.agilent.com/cs/library/usermanuals/public/2100_Bioanalyzer_Expert_USR.pdf.

87. Mueller O, Lightfoot S, Schroeder A. RNA Integrity Number (RIN). *Agil Technol*. 2016:1-8.

88. Hukkanen J, Jacob P, Benowitz NL. Metabolism and disposition kinetics of nicotine. *Pharmacol Rev*. 2005;57(1):79-115. doi:10.1124/pr.57.1.3

89. Garcia-Garcia HM, Eugène P. McFadden AF, Mehran R, et al. Standardized End Point Definitions for Coronary Intervention Trials. *Circulation*. 2018;137:2635-2650. doi:10.1161/CIRCULATIONAHA.117.029289

90. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. *Br J Addict*. 1991;86(9):1119-1127. doi:10.1111/j.1360-0443.1991.tb01879.x

91. Sample Size for Microarray Experiments. https://bioinformatics.mdanderson.org/MicroarraySampleSize/. Accessed September 15, 2020.

92. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010.

93. Stevens VL, Bierut LJ, Talbot JT, et al. Nicotinic receptor gene variants influence susceptibility to heavy smoking. *Cancer Epidemiol Biomarkers Prev*. 2008;17(12):3517-3525. doi:10.1158/1055-9965.EPI-08-0585

94. Rosicova K, Geckova AM, Dijk JP Van, Rosic M, Zezula I, Groothoff JW. Socioeconomic indicators and ethnicity as determinants of regional mortality rates in Slovakia. *Int J Public Health*. 2009;54:274-282. doi:10.1007/s00038-009-7108-7

95. Rosicova K, Reijneveld SA, Geckova AM, et al. Inequalities in mortality by socioeconomic factors and Roma ethnicity in the two biggest cities in Slovakia : a multilevel analysis. *Int J Equity Health*. 2015;14(123):1-10. doi:10.1186/s12939-015-0262-z

96. Giuseppe G, Giulia C, Valentina M, et al. Socioeconomic and regional inequalities in smoking cessation in Italy, 2014-2017 2019. 2019:275-285. doi:10.19191/EP19.4.P275.078

97. Babinská I, Gecková AM, Jarčuška P, et al. Does the population living in Roma settlements differ in physical activity, smoking and alcohol consumption from the majority population in Slovakia? *Cent Eur J Public Health*. 2014;22(88):S22-S27. doi:10.21101/cejph.a3897

98. Roberts B, Chikovani I, Makhashvili N, Patel V, Mckee M. Tobacco Use and Nicotine Dependence among Conflict-Affected Men in the Republic of Georgia. *Int J Environ Res Public Health*. 2013;592:2185-2197. doi:10.3390/ijerph10062185

99. Attila S, Zsigmond K, Erdei JR, et al. Socioeconomic Status, Health Related Behaviour, And Self-Rated Health of Children Living in Roma Settlements in Hungary. *Cent Eur J Public Health*. 2019;27(1):24-31. doi:10.21101/cejph.a4726

100. Petek D, Pavlič DR, Švab I, Lolić D. Attitudes of Roma toward Smoking : Qualitative Study in Slovenia. *Croat Med J*. 2006;47(2):344-347.

101. Frostegård J. Immunity, atherosclerosis and cardiovascular disease. *BMC Med*. 2013;11(117):1-8. http://www.biomedcentral.com/1741-7015/11/117%0Apapers3://publication/uuid/D0C88CBA-A315-4DF2-8BF5-CD34F24146B2.

102. Lorenzatti A, Servato ML. Role of anti-inflammatory interventions in coronary artery disease: Understanding the Canakinumab Anti-inflammatory Thrombosis Outcomes Study (CANTOS). *Eur Cardiol Rev* . 2018;13(1):38-41. doi:10.15420/ecr.2018.11.1

103. Zoccali C, Mallamaci F. Innate Immunity System in Patients With Cardiovascular and Kidney Disease. *Circ Res*. 2023;132(8):915-932. doi:10.1161/CIRCRESAHA.122.321749

104. Guo X, Ma L. Inflammation in coronary artery disease-clinical implications of novel HDL-cholesterol-related inflammatory parameters as predictors. *Coron Artery Dis*. 2023;34(1):66-77. doi:10.1097/MCA.0000000000001198

105. Justiz Vaillant AA, Jamal Z, Patel P, Ramphul K. Immunoglobulin. In: Treasure Island (FL); 2023.

106. Jr HWS, Cavacini L. Structure and Function of Immunoglobulins. *J Allergy Clin*

Immunol. 2010;125(202):S41-S52. doi:10.1016/j.jaci.2009.09.046.

107. Porsch F, Mallat Z, Binder CJ. Humoral immunity in atherosclerosis and myocardial infarction : from B cells to antibodies. *Cardiovasc Res*. 2021;117:2544-2562. doi:10.1093/cvr/cvab285

108. Tarbiah N, Todd I, Tighe PJ, Fairclough LC. Cigarette smoking differentially affects immunoglobulin class levels in serum and saliva: An investigation and review. *Basic Clin Pharmacol Toxicol*. 2019;125(5):474-483. doi:10.1111/bcpt.13278

109. Khan SR, van der Burgh AC, Peeters RP, van Hagen PM, Dalm VASH, Chaker L. Determinants of Serum Immunoglobulin Levels: A Systematic Review and Meta-Analysis. *Front Immunol*. 2021;12(April):1-15. doi:10.3389/fimmu.2021.664526

110. Núñez E, Fuster V, Gómez-Serrano M, et al. Unbiased plasma proteomics discovery of biomarkers for improved detection of subclinical atherosclerosis. *eBioMedicine*. 2022;76:1-14. doi:10.1016/j.ebiom.2022.103874

111. Piaggeschi G, Rolla S, Rossi N, et al. Immune Trait Shifts in Association With Tobacco Smoking : A Study in Healthy Women. *Front Immunol*. 2021;12(March):1-12. doi:10.3389/fimmu.2021.637974

112. Ewald DA, Malajian D, Krueger JG, et al. Meta-analysis derived atopic dermatitis (MADAD) transcriptome defines a robust AD signature highlighting the involvement of atherosclerosis and lipid metabolism pathways. *BMC Med Genomics*. 2015:1-15. doi:10.1186/s12920-015-0133-x

113. Wang J-F, Huang Y, Lu S-F, et al. Comparative study of gene expression profiles rooted in acute myocardial infarction and ischemic/reperfusion rat models. *Am J Cardiovasc Dis*. 2020;10(2):84-100.

114. Vink JM, Jansen R, Brooks A, et al. Differential gene expression patterns between smokers and non-smokers: cause or consequence? *Addict Biol*. 2017;22(2):550-560. doi:10.1111/adb.12322

115. Xing S, Li F, Zeng Z, et al. Tcf1 and Lef1 transcription factors establish CD8(+) T cell identity through intrinsic HDAC activity. *Nat Immunol*. 2016;17(6):695-703. doi:10.1038/ni.3456

116. Kundu S, Freiberg MS, Tracy RP, et al. Circulating T Cells and Cardiovascular Risk in People With and Without HIV Infection. *J Am Coll Cardiol*. 2022;80(17):1633-1644.

doi:10.1016/j.jacc.2022.08.756

117. Chowdhury RR, D'addabbo J, Huang X, et al. Human Coronary Plaque T Cells Are Clonal and Cross-React to Virus and Self. *Circ Res*. 2022;130(10):1510-1530. doi:10.1161/CIRCRESAHA.121.320090

118. Bauer M, Hackermüller J, Schor J, et al. Specific induction of the unique GPR15 expression in heterogeneous blood lymphocytes by tobacco smoking. *Biomarkers*. 2019;24(3):217-224. doi:10.1080/1354750X.2018.1539769

119. Haase T, Müller C, Stoffers B, et al. G Protein-Coupled Receptor 15 Expression Is Associated with Myocardial Infarction. *Int J Mol Sci*. 2023;24(1). doi:10.3390/ijms24010180

120. Andersen AM, Lei MK, Beach SRH, Philibert RA, Sinha S, Colgan JD. Cigarette and cannabis smoking effects on GPR15+ helper T cell levels in peripheral blood: Relationships with epigenetic biomarkers. *Genes (Basel)*. 2020;11(2):1-19. doi:10.3390/genes11020149

121. Andersen AM, Lei M, Beach SRH, Philibert RA. Inflammatory biomarker relationships with helper T cell GPR15 expression and cannabis and tobacco smoking. *J Psychosom Res*. 2021;141. doi:10.1016/j.jpsychores.2020.110326.Inflammatory

122. Stoffers B, Bacmeister L, Braeuninger H, et al. GPR15-mediated T cell recruitment during acute viral myocarditis is associated with improved virus elimination and outcome. *Cardiovasc Res*. 2022;118(Supplement_1):70164. doi:10.1093/cvr/cvac066.127

123. Bauer M. The role of GPR15 function in blood and vasculature. *Int J Mol Sci*. 2021;22(19). doi:10.3390/ijms221910824

124. Zhao J, Liu X, Xu J, et al. Elevated Expression and Activation of GPR15 in Immune Cells in Graves' Disease. *Biomolecules*. 2022;12(12):1-10. doi:10.3390/biom12121899

125. Grievink HW, Smit V, Huisman BW, et al. Cardiovascular risk factors: The effects of ageing and smoking on the immune system, an observational clinical study. *Front Immunol*. 2022;13(September):1-14. doi:10.3389/fimmu.2022.968815

126. Zhang JF, Li Y, Zhang AZ, He QQ, Du YC, Cao W. Expression and pathological significance of CC chemokine receptor 7 and its ligands in the airway of asthmatic rats exposed to cigarette smoke. *J Thorac Dis*. 2018;10(9):5459-5467.

doi:10.21037/jtd.2018.08.124

127. Cao W, Du YC, Li Y, Wu XJ, Zhang AZ. Effect of cigarette smoke exposure on the expression of CCR7 and levels of Th1/Th2 cytokines in asthmatic rats. *Zhonghua Yi Xue Za Zhi*. 2018;98(28):2264-2268. doi:10.3760/cma.j.issn.0376-2491.2018.28.011

128. Wang C, Song C, Liu Q, et al. Gene Expression Analysis Suggests Immunological Changes of Peripheral Blood Monocytes in the Progression of Patients With Coronary Artery Disease. *Front Genet*. 2021;12(March):1-10. doi:10.3389/fgene.2021.641117

129. Kohli K, Pillarisetty VG, Kim TS. Key chemokines direct migration of immune cells in solid tumors. *Cancer Gene Ther*. 2022;29(1):10-21. doi:10.1038/s41417-021-00303-x

130. Collu R, Post JM, Scherma M, et al. Altered brain levels of arachidonic acid-derived inflammatory eicosanoids in a rodent model of anorexia nervosa. *Biochim Biophys Acta - Mol Cell Biol Lipids*. 2020;1865(4). doi:10.1016/j.bbalip.2019.158578

131. Reiss Y, Proudfoot AE, Power CA, Campbell JJ, Butcher EC. CC chemokine receptor (CCR)4 and the CCR10 ligand cutaneous T cell-attracting chemokine (CTACK) in lymphocyte trafficking to inflamed skin. *J Exp Med*. 2001;194(10):1541-1547. doi:10.1084/jem.194.10.1541

132. Ma C, Lu T, He Y, et al. Comprehensive analysis of autophagy-related gene expression profiles identified five gene biomarkers associated with immune infiltration and advanced plaques in carotid atherosclerosis. *Orphanet J Rare Dis*. 2023;18(1):66. doi:10.1186/s13023-023-02660-2

133. Xia Y, Brewer A, Bell JT. DNA methylation signatures of incident coronary heart disease: findings from epigenome-wide association studies. *Clin Epigenetics*. 2021;13(1):1-16. doi:10.1186/s13148-021-01175-6

134. Scott D, Palmer R. The influence of tobacco smoking on adhesion molecule profiles. *Tob Induc Dis*. 2002;1(1):7-25. doi:10.1201/b13288-6

135. Lee J, Cooke JP. The role of nicotine in the pathogenesis of atherosclerosis. *Atherosclerosis*. 2011;215(2):281-283. doi:10.1016/j.atherosclerosis.2011.01.003

136. Mulero MC, Aubareda A, Schlüter A, Pérez-Riba M. RCAN3, a novel calcineurin inhibitor that down-regulates NFAT-dependent cytokine gene expression. *Biochim Biophys Acta - Mol Cell Res*. 2007;1773(3):330-341. doi:10.1016/j.bbamcr.2006.12.007

137. Agha G, Mendelson MM, Ward-Caviness CK, et al. Blood Leukocyte DNA Methylation Predicts Risk of Future Myocardial Infarction and Coronary Heart Disease. *Circulation*. 2019;140(8):645-657. doi:10.1161/CIRCULATIONAHA.118.039357

138. Petsophonsakul P, Burgmaier M, Willems B, et al. Nicotine promotes vascular calcification via intracellular Ca2+-mediated, Nox5-induced oxidative stress, and extracellular vesicle release in vascular smooth muscle cells. *Cardiovasc Res*. 2022;118(9):2196-2210. doi:10.1093/cvr/cvab244

139. TRAV4 Gene - T Cell Receptor Alpha Variable 4. GeneCards. https://www.genecards.org/cgi-bin/carddisp.pl?gene=TRAV4. Published 2023. Accessed July 6, 2023.

140. JCHAIN Gene - Joining Chain Of Multimeric IgA And IgM. GeneCrads. https://www.genecards.org/cgi-bin/carddisp.pl?gene=JCHAIN. Published 2023. Accessed July 6, 2023.

141. Li Y, Fu W, Geng Z, et al. A pan-cancer analysis of the oncogenic role of ribonucleotide reductase subunit M2 in human tumors. *PeerJ*. 2022;10:1-29. doi:10.7717/PEERJ.14432

142. Slizhikova DK, Zinovyeva M V., Kuzmin D V., et al. Decreased expression of the human immunoglobulin J-chain gene in squamous cell cancer and adenocarcinoma of the lungs. *Mol Biol*. 2007;41(4):594-600. doi:10.1134/S0026893307040115

143. Marengo-Rowe AJ. Structure-function relations of human hemoglobins. *Proc Baylor Univ Med Cent*. 2006;19(3):239-245. doi:10.1080/08998280.2006.11928171

144. Thriveni R, Manshi P, Ramesh D V., Rachel B, Byatnal A, Kempwade P. Effects of smoking on hemoglobin and erythrocytes sedimentation rate and its association with ABO blood groups. *J Indian Acad Oral Med Radiol*. 2020;32(2):103-106. doi:10.4103/jiaomr.jiaomr_30_20

145. Kahar F, Wikandari RJ, Irnawati I, Penmaley MS. The Effect of Cigarette Smoking Duration on Hemoglobin Level Measured with Cyanmethemoglobin Method. *Indones J Med Lab Sci Technol*. 2022;4(2):157-167. doi:10.33086/ijmlst.v4i2.2674

146. Elisia I, Lam V, Cho B, et al. The effect of smoking on chronic inflammation, immune function and blood cell composition. *Sci Rep*. 2020;10(1):1-16. doi:10.1038/s41598-

020-76556-7

147. Salehi N, Janjani P, Tadbiri H, Rozbahani M, Jalilian M. Effect of cigarette smoking on coronary arteries and pattern and severity of coronary artery disease: a review. *J Int Med Res*. 2021;49(12). doi:10.1177/03000605211059893

148. Hammond D, Fong GT, McNeill A, Borland R, Cummings KM. Effectiveness of cigarette warning labels in informing smokers about the risks of smoking: Findings from the International Tobacco Control (ITC) Four Country Survey. *Tob Control*. 2006;15(SUPPL. 3):19-25. doi:10.1136/tc.2005.012294

149. Rietbrock N, Kunkel S, Worner W, Eyer P. Oxygen-dissociation kinetics in the blood of smokers and non-smokers: interaction between oxygen and carbon monoxide at the hemoglobin molecule. *NaunynSchmiedebergs Arch Pharmacol*. 1992;345:123-128. http://springerlink.metapress.com/openurl.asp?genre=article&id=doi:10.1007/BF00169200.

150. AHSP Gene - Alpha Hemoglobin Stabilizing Protein. GeneCards. https://www.genecards.org/cgi-bin/carddisp.pl?gene=AHSP. Published 2023. Accessed July 10, 2023.

151. Remigante A, Spinelli S, Pusch M, et al. Role of SLC4 and SLC26 solute carriers during oxidative stress. *Acta Physiol*. 2022;235(1):1-20. doi:10.1111/apha.13796

152. Kamceva G, Arsova-Sarafinovska Z, Ruskovska T, Zdravkovska M, Kamceva-Panova L, Stikova E. Cigarette smoking and oxidative stress in patients with coronary artery disease. *Open Access Maced J Med Sci*. 2016;4(4):636-640. doi:10.3889/oamjms.2016.117

153. Caliri AW, Tommasi S, Besaratinia A. Relationships among smoking, oxidative stress, inflammation, macromolecular damage, and cancer. *Mutat Res - Rev Mutat Res*. 2021;787:108365. doi:10.1016/j.mrrev.2021.108365

154. Metta S, Basalingappa DR, Uppala S, Mitta G. Erythrocyte Antioxidant Defenses Against Cigarette Smoking in Ischemic Heart Disease. *J Clin Diagn Res*. 2015;9(6):BC08-11. doi:10.7860/JCDR/2015/12237.6128

155. Aksoy S, Cam N, Gurkan U, et al. Oxidative stress and severity of coronary artery disease in young smokers with acute myocardial infarction. *Cardiol J*. 2012;19(4):381-386. doi:10.5603/cj.2012.0069

156. Revin V V., Gromova N V., Revina ES, et al. The influence of oxidative stress and natural antioxidants on morphometric parameters of red blood cells, the hemoglobin oxygen binding capacity, and the activity of antioxidant enzymes. *Biomed Res Int*. 2019;2019(2003). doi:10.1155/2019/2109269

157. ALAS2 Gene - 5'-Aminolevulinate Synthase 2. GeneCards. https://www.genecards.org/cgi-bin/carddisp.pl?gene=ALAS2. Published 2023. Accessed July 10, 2023.

158. Cox TC, Sadlon TJ, Schwarz QP, et al. The major splice variant of human 5-aminolevulinate synthase-2 contributes significantly to erythroid heme biosynthesis. *Int J Biochem Cell Biol*. 2004;36:281-295. doi:10.1016/S1357-2725(03)00246-2

159. Guauque-Olarte S, Gaudreault N, Piché MÈ, et al. The transcriptome of human epicardial, mediastinal and subcutaneous adipose tissues in men with coronary artery disease. *PLoS One*. 2011;6(5). doi:10.1371/journal.pone.0019908

160. PTGDS Gene - Prostaglandin D2 Synthase. GeneCards. https://www.genecards.org/cgi-bin/carddisp.pl?gene=PTGDS. Published 2023. Accessed July 10, 2023.

161. Beccacece L, Abondio P, Bini C, Pelotti S. The Link between Prostanoids and Cardiovascular Diseases. *Int J Mol Sci*. 2023;24(4193). doi:10.3390/ijms24044193

162. Verdugo RA, Zeller T, Rotival M, et al. Graphical Modeling of Gene Expression in Monocytes Suggests Molecular Mechanisms Explaining Increased Atherosclerosis in Smokers. *PLoS One*. 2013;8(1). doi:10.1371/journal.pone.0050888

163. Zhao S, Cheng CK, Zhang C-L, Huang Y. Interplay Between Oxidative Stress, Cyclooxygenases, and Prostanoids in Cardiovascular Diseases. *Antioxid Redox Signal*. 2021;34(10):784-799. doi:10.1089/ars.2020.8105

164. Arehart E, Stitham J, Asselbergs FW, et al. Acceleration of cardiovascular disease by a dysfunctional prostacyclin receptor mutation: potential implications for cyclooxygenase-2 inhibition. *Circ Res*. 2008;102(8):986-993. doi:10.1161/CIRCRESAHA.107.165936

165. Heffnera JL, Anthenelli RM, Adler CM, Strakowski SM, Beavers J, DelBello MP. Prevalence and correlates of heavy smoking and nicotine dependence in adolescents with bipolar and cannabis use disorders. *Psychiatry Res*. 2013;210(3):857-862.

doi:10.1016/j.psychres.2013.04.010.Prevalence

166. Lau AYH, Ridge M. Addressing the impact of social exclusion on mental health in Gypsy, Roma, and Traveller communities. *Ment Heal Soc Incl*. 2011;15(3):129-137. doi:10.1108/20428301111165717

167. Fagerström KO, Kunze M, Schoberberger R, et al. Nicotine dependence versus smoking prevalence: Comparisons among countries and categories of smokers. *Tob Control*. 1996;5(1):52-56. doi:10.1136/tc.5.1.52

168. Shiffman S, Waters AJ, Hickcox M. The Nicotine Dependence Syndrome Scale: A multidimensional measure of nicotine dependence. *Nicotine Tob Res*. 2004;6(2):327-348. doi:10.1080/1462220042000202481

169. Piasecki TM, Piper ME, Baker TB. Refining the tobacco dependence phenotype using the Wisconsin inventory of smoking dependence motives: II. Evidence from a laboratory self-administration assay. *J Abnorm Psychol*. 2010;119(3):513-523. doi:10.1037/a0020235

170. Benowitz NL, Jacob P. Metabolism of nicotine to cotinine studied by a dual stable isotope method. *Clin Pharmacol Ther*. 1994;56(5):483-493. doi:10.1038/clpt.1994.169

12 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. Sándor Janos***, ***Dr. Szilárd Poliska***, ***Dr. Laszló Balogh***, and ***Dr. István Szász***, 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.

13 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).

14 Appendices

Appendix A: Supplementary Tables

Table 14-1: Hardy-Weinberg equilibrium for Hungarian Roma

CHR	SNP	Gene	A1	A2	GENO	*p* value
2	rs10490162	NRXN1	C	T	3/66/327	1
4	rs3762611	GABRB1	A	G	21/105/272	**0.01**
5	rs2673931	TRPC7	C	T	80/207/112	0.42
10	rs4142041	CTNNA3	G	A	50/176/171	0.66
15	rs2036534	AGPHD1	C	T	25/143/230	0.68
15	rs16969968	CHRNA5	A	G	48/166/184	0.26
15	rs578776	CHRNA3	A	G	74/171/148	0.06
21	rs6517442	KCNJ6	C	T	47/186/166	0.74
23	rs2235186	MAOA	A	G	43/147/105	0.54

CHR= Chromosome, SNP= single nucleotide polymorphism, A1= minor allele; A2=major allele

Table 14-2: Smoking status among both populations based on gender

Gender	HR		HG		OR	95% *CI*
	SM %	NSM %	SM %	NSM %		
Males	16.5	9.8	15.8	29.3	1.89	1.59 - 2.26
Females	48.7	25.0	17.0	37.9	2.05	1.48 - 2.86

SM= smokers; NSM= non-smokers

Table 14-3: Differences in risk allele frequencies between study populations

SNP	Gene	Risk Allele	Frequency		$\chi 2$	*p* value	OR	ADJ-*p**
			HR	HG				
rs10490162	NRXN1	T	0.91	0.89	1.52	0.22	1.23	0.25
rs2673931	**TRPC7**	**T**	**0.54**	**0.62**	**10.24**	**0.00**	**0.72**	**0.01**
rs4142041	CTNNA3	G	0.35	0.39	2.37	0.12	0.85	0.17
rs2036534	AGPHD1	T	0.77	0.80	3.05	0.08	0.81	0.13
rs16969968	CHRNA5	A	0.33	0.36	1.06	0.30	0.89	0.30
rs578776	**CHRNA3**	**G**	**0.60**	**0.74**	**35.27**	**<0.001**	**0.52**	**<0.001**
rs6517442	**KCNJ6**	**C**	**0.35**	**0.29**	**5.68**	**0.02**	**1.29**	**0.03**
rs2235186	**MAOA**	**A**	**0.40**	**0.32**	**9.24**	**0.00**	**1.43**	**0.01**

Bold font highlights significant differences; * adjustable p-value

Table 14-4: Genotype by smoking behaviours in general population

SNPs	Genotype	Smoking Behaviors				*p* value
		HSM[a]	MSM[b]	FSM[c]	NSM[d]	
rs10490162-T	C C	0.0	0.5	0.0	1.0	0.46
	T C	1.2	2.2	0.0	15.6	
	T T	10.8	11.2	1.7	55.7	
rs16969968-A	A A	1.7	1.0	0.0	8.0	0.10
	G A	7.6	7.6	1.2	32.9	
	G G	2.7	5.4	0.5	31.5	
rs2036534-T	C C	0.0	0.2	0.0	3.2	0.14
	C T	2.9	3.9	0.7	25.7	
	T T	9.0	9.5	1.0	43.8	
rs2235186-A	A A	3.9	2.2	0.0	12.5	**0.00**
	A G	2.0	4.2	1.5	18.6	
	G G	5.9	7.6	0.2	41.6	
rs2673931-T	C C	1.0	3.2	0.2	9.6	0.14
	C T	7.6	5.1	1.0	34.6	
	T T	3.2	5.6	0.5	28.4	
rs4142041-G	A A	5.9	5.2	0.2	23.7	0.48
	A G	4.7	6.9	1.2	39.0	
	G G	1.5	2.0	0.2	9.4	
rs578776-G	A A	0.2	0.5	0.0	5.7	0.55
	G A	4.7	5.5	1.2	28.0	
	G G	7.2	7.9	0.5	38.5	
rs6517442-C	C C	1.0	1.7	0.0	5.9	0.05
	T C	3.7	5.1	0.0	32.7	
	T T	7.3	7.1	1.7	33.9	

Legend: [a] Heavy Smoker, [b] Moderate Smoker, [c] Former Smoker,[d] Non-smoker. Bold font means statistically significant at (0.001) level. Risk allele is written besides each SNP.

Table 14-5: Genotype by smoking behaviours in the Roma population

SNPs	Genotype	Smoking Behaviors				*p* value
		HSM[a]	MSM[b]	FSM[c]	NSM[d]	
rs10490162-T	C C	0.5	0.3	0.0	0.0	0.76
	T C	6.8	2.8	0.8	6.3	
	T T	30.8	16.2	2.8	32.8	
rs16969968-A	A A	3.5	2.3	0.3	6.0	0.41
	G A	15.8	8.8	2.3	14.8	
	G G	19.1	8.0	1.3	17.8	
rs2036534-T	C C	3.0	0.8	0.0	2.5	0.88
	C T	14.3	7.3	1.0	13.3	
	T T	20.9	11.1	2.8	23.1	
rs2235186-A	A A	10.5	2.8	1.0	7.8	0.36
	A G	14.3	8.0	1.5	13.3	
	G G	13.5	8.3	1.3	17.8	
rs2673931-T	C C	8.3	3.5	1.0	7.3	0.89
	C T	18.5	10.0	2.0	21.3	
	T T	11.5	5.5	0.8	10.3	
rs4142041-G	A A	15.9	8.3	2.3	16.6	0.42
	A G	16.1	9.6	1.0	17.6	
	G G	6.0	1.0	0.5	5.0	
rs578776-G	A A	7.1	2.5	0.8	8.4	0.53
	G A	18.3	8.7	1.0	15.5	
	G G	13.0	7.6	1.8	15.3	
rs6517442-C	C C	4.5	1.3	0.5	5.5	0.44
	T C	17.5	10.0	2.5	16.5	
	T T	16.3	7.8	0.8	16.8	

Legend: [a] Heavy Smoker, [b] Moderate Smoker, [c] Former Smoker, [d] Non-smoker. Risk allele is written besides each SNP.

Table 14-6: Differentially expressed genes with a fold change of ≥1.5 (N=61)

No.	Gene Symbol	FC	*p* value	Description
Upregulated genes				
1	*GPR15*	6.0	5.4E-12	G protein-coupled receptor 15 [Source:HGNC Symbol;Acc:HGNC:4469]
2	*IFI44*	1.8	1.8E-02	interferon induced protein 44 [Source:HGNC Symbol;Acc:HGNC:16938]
3	*IGKV3-11*	1.6	1.6E-02	immunoglobulin kappa variable 3-11 [Source:HGNC Symbol;Acc:HGNC:5815]
4	*IGLV1-47*	1.5	2.3E-02	immunoglobulin lambda variable 1-47 [Source:HGNC Symbol;Acc:HGNC:5880]
5	*IGLV2-23*	1.6	4.9E-02	immunoglobulin lambda variable 2-23 [Source:HGNC Symbol;Acc:HGNC:5890]
6	*IGLV3-21*	1.7	3.2E-02	immunoglobulin lambda variable 3-21 [Source:HGNC Symbol;Acc:HGNC:5905]
Downregulated genes				
1	*CHI3L1*	-1.5	4.5E-02	chitinase 3 like 1 [Source:HGNC Symbol;Acc:HGNC:1932]
2	*EPB42*	-1.5	3.9E-02	erythrocyte membrane protein band 4.2 [Source:HGNC Symbol;Acc:HGNC:3381]
3	*HBQ1*	-1.5	4.7E-02	hemoglobin subunit theta 1 [Source:HGNC Symbol;Acc:HGNC:4833]
4	*KLF1*	-1.6	1.9E-02	Kruppel like factor 1 [Source:HGNC Symbol;Acc:HGNC:6345]
5	*PDZK1IP1*	-1.7	3.5E-03	PDZK1 interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:16887]
6	*PTGDS*	-1.6	1.5E-02	prostaglandin D2 synthase [Source:HGNC Symbol;Acc:HGNC:9592]
7	*RETN*	-1.5	3.6E-02	resistin [Source:HGNC Symbol;Acc:HGNC:20389]

FC= fold change between smokers and nonsmokers.

Table 14-7: Demographic characteristics of study participants (N=44)

Variables	Total	Smoking status		*p* value
		Smokers	Nonsmokers	
Gender, n (%)				
Male	**18 (63.6)**	**18 (81.8)**	**10 (45.5)**	**<0.001**
Female	**16 (36.4)**	**4 (18.2)**	**12 (54.5)**	
Age, median (range) years	**68 (32-88)**	**61 (32-75)**	**75 (54-88)**	**<0.001**
Cholesterol (mmol/L), mean (SD)	5.0 (1.4)	5.3 (1.3)	4.7 (1.4)	NS
Triglyceride (mmol/L), mean (SD)	1.5 (0.6)	1.5 (0.5)	1.5 (0.7)	NS
HDL-C (mmol/L), mean (SD)	1.3 (0.6)	1.4 (0.8)	1.2 (0.2)	NS
LDL-C (mmol/L), mean (SD)	3.0 (1.2)	3.1 (1.2)	2.8 (1.3)	NS
Number of affected arteries, median (range)	2 (1-3)	2 (1-3)	2 (1-3)	NS
Family history of CAD, n (%)	**12 (27.3)**	**10 (22.7)**	**2 (9.1)**	**0.02**
Anticoagulation medication, n (%)	44 (100)	22 (100)	22 (100)	NS
BMI, mean (SD)	26.3 (3.0)	25.6 (3.1)	27.0 (2.8)	NS
Blood pressure (mmHg), mean (SD)				
Systolic	126.1 (22.5)	120.6 (23.9)	131.6 (20.1)	NS
Diastolic	72.8 (11.4)	71.5 (13.3)	74.1 (9.4)	NS
Plasma cotinine ng/mL, mean (SD)	2.8 (1.6)	3.0 (1.5)	2.5(1.7)	NS
Angiography findings, n (%)				
>2 arteries stenosis	36 (81.8)	18 (81.8)	18 (81.8)	NS
LAD	37 (84.1)	18 (81.8)	19 (86.4)	NS
LM	7 (15.9)	5 (22.7)	2 (9.1)	NS
CX	26 (59.1)	11 (50.0)	15 (68.2)	NS
RCA	26 (59.1)	14 (63.6)	12 (54.5)	NS

LAD= left anterior descending artery; LM= left main coronary artery; CX= circumflex artery; n= number of subject; NS= non-significant; RCA= right coronary artery; SD= standard deviation. Bold text means significant.

Table 14-8: Differentially expressed genes among smokers based on nicotine dependence (N=31)

Gene Symbol	p	FC ([high] vs [low])	Regulation ([high] vs [low])	FC ([low moderate] vs [low])	Regulation ([low moderate] vs [low])	FC ([moderate] vs [low])	Regulation ([moderate] vs [low])	Description
ADPGK-AS1	0.03	1.6	up	1.5	up	1.2	up	ADPGK antisense RNA 1 [Source:HGNC Symbol;Acc:HGNC:44144]
ASB16	0.01	1.5	up	1.3	up	1.1	up	ankyrin repeat and SOCS box containing 16 [Source:HGNC Symbol;Acc:HGNC:19768]
HRH4	0.04	1.6	up	-1.5	down	-1.6	down	histamine receptor H4 [Source:HGNC Symbol;Acc:HGNC:17383]
HCG4P5	0.01	1.6	up	-1.1	down	-2.3	down	HLA complex group 4 pseudogene 5 [Source:HGNC Symbol;Acc:HGNC:22925]
IGKV1-12	0.02	9.1	up	1.8	up	4.2	up	immunoglobulin kappa variable 1-12 [Source:HGNC Symbol;Acc:HGNC:5730]
HLA-A	0.01	1.8	up	1.7	up	-2.2	down	major histocompatibility complex, class I, A [Source:HGNC Symbol;Acc:HGNC:4931]
RN7SL2	0.04	2.4	up	2.0	up	1.4	up	RNA, 7SL, cytoplasmic 2 [Source:HGNC Symbol;Acc:HGNC:23134]
SCGB3A1	0.01	30.2	up	3.7	up	1.6	up	secretoglobin family 3A member 1 [Source:HGNC Symbol;Acc:HGNC:18384]
SPTLC1P2	0.03	2.4	up	1.6	up	1.3	up	serine palmitoyltransferase long chain base subunit 1 pseudogene 2 [Source:HGNC Symbol;Acc:HGNC:39669]
SLC29A1	0.03	1.5	up	-1.2	down	-1.7	down	solute carrier family 29 member 1 (Augustine blood group) [Source:HGNC Symbol;Acc:HGNC:11003]
TRAJ33	0.04	1.8	up	-1.9	down	-1.5	down	T cell receptor alpha joining 33 [Source:HGNC Symbol;Acc:HGNC:12063]
UBE2SP2	0.01	1.6	up	2.0	up	1.3	up	ubiquitin conjugating enzyme E2 S pseudogene 2 [Source:HGNC Symbol;Acc:HGNC:32196]
USP9Y	0.04	19.0	up	3.9	up	3.6	up	ubiquitin specific peptidase 9, Y-linked [Source:HGNC Symbol;Acc:HGNC:12633]
ZNF696	0.01	1.9	up	1.1	up	1.4	up	zinc finger protein 696 [Source:HGNC Symbol;Acc:HGNC:25872]
ARMC10P1	0.04	-2.5	down	-1.2	down	1.0	up	armadillo repeat containing 10 pseudogene 1 [Source:HGNC Symbol;Acc:HGNC:43646]
CACYBPP2	0.03	-2.8	down	-1.9	down	-1.6	down	calcyclin binding protein pseudogene 2 [Source:HGNC Symbol;Acc:HGNC:45123]
CCT5P1	0.02	-1.9	down	-1.2	down	-1.4	down	chaperonin containing TCP1 subunit 5 pseudogene 1 [Source:HGNC Symbol;Acc:HGNC:35135]
LYPD2	0.01	-2.0	down	-4.3	down	-8.5	down	LY6/PLAUR domain containing 2 [Source:HGNC Symbol;Acc:HGNC:25215]

MTATP8P2	0.02	-3.1	down	-1.4	down	-1.1	down	mitochondrially encoded ATP synthase 8 pseudogene 2 [Source:HGNC Symbol;Acc:HGNC:44572]
MTCO3P12	0.02	-1.7	down	1.0	up	-1.2	down	mitochondrially encoded cytochrome c oxidase III pseudogene 12 [Source:HGNC Symbol;Acc:HGNC:52042]
MTRNR2L12	0.01	-1.8	down	1.2	up	-1.1	down	MT-RNR2-like 12 [Source:HGNC Symbol;Acc:HGNC:37169]
MTRNR2L8	0.02	-2.7	down	-1.1	down	-1.3	down	MT-RNR2-like 8 [Source:HGNC Symbol;Acc:HGNC:37165]
RPL14P3	0.03	-1.9	down	-1.3	down	1.0	up	ribosomal protein L14 pseudogene 3 [Source:HGNC Symbol;Acc:HGNC:36638]
RPS18	0.01	-2.3	down	-2.1	down	1.1	up	ribosomal protein S18 [Source:HGNC Symbol;Acc:HGNC:10401]
RPS3AP47	0.04	-2.1	down	-1.8	down	1.0	up	ribosomal protein S3a pseudogene 47 [Source:HGNC Symbol;Acc:HGNC:36744]
SNORD58B	0.02	-1.7	down	-1.6	down	1.1	up	small nucleolar RNA, C/D box 58B [Source:HGNC Symbol;Acc:HGNC:10209]
TRGV4	0.04	-1.6	down	-2.7	down	-2.2	down	T cell receptor gamma variable 4 [Source:HGNC Symbol;Acc:HGNC:12289]

FC= fold change

Table 14-9: Gene ontology of the 15 upregulated genes (N=44)

Term ID	Term description	Strength	FDR	Matching proteins in our network (labels)
Biological process				
GO:0050900	Leukocyte migration	1.54	0.0018	*CCR7,IGJ,GPR15,ITGA6,IGLL5*
GO:0002376	Immune system process	0.85	0.0035	*CCR7,IGJ,LEF1,GPR15,TRAT1,CD28,ITGA6,IGLL5*
GO:0016477	Cell migration	1.16	0.0035	*CCR7,IGJ,LEF1,GPR15,ITGA6,IGLL5*
GO:0022409	Positive regulation of cell-cell adhesion	1.5	0.0099	*CCR7,LEF1,CD28,ITGA6*
GO:0051251	Positive regulation of lymphocyte activation	1.48	0.0099	*CCR7,LEF1,CD28,IGLL5*
GO:0002250	Adaptive immune response	1.44	0.012	*IGJ,LEF1,TRAT1,IGLL5*
GO:0045060	Negative thymic T cell selection	2.6	0.0151	*CCR7,CD28*
GO:0002863	Positive regulation of inflammatory response to antigenic stimulus	2.56	0.0156	*CCR7,CD28*
GO:0006955	Immune response	0.91	0.0157	*CCR7,IGJ,LEF1,TRAT1,CD28,IGLL5*
GO:0002684	Positive regulation of immune system process	1.06	0.0192	*CCR7,LEF1,TRAT1,CD28,IGLL5*
GO:0030217	T cell differentiation	1.66	0.0192	*CCR7,LEF1,CD28*
GO:0050862	Positive regulation of T cell receptor signaling pathway	2.46	0.0192	*CCR7,TRAT1*
GO:0044419	Interspecies interaction between organisms	0.84	0.0292	*CCR7,IGJ,LEF1,GPR15,CD28,IGLL5*
GO:0051897	Positive regulation of protein kinase B signaling	1.58	0.0292	*CCR7,TRAT1,CD28*
GO:0032673	Regulation of interleukin-4 production	2.18	0.0372	*LEF1,CD28*
GO:0050851	Antigen receptor-mediated signaling pathway	1.52	0.0372	*TRAT1,CD28,IGLL5*
GO:0050870	Positive regulation of T cell activation	1.49	0.0403	*CCR7,LEF1,CD28*
GO:0007165	Signal transduction	0.55	0.0407	*CCR7,LEF1,GPR15,TRAT1,CD28,RCAN3,ITGA6,IGLL5*
GO:0050778	Positive regulation of immune response	1.16	0.0407	*CCR7,TRAT1,CD28,IGLL5*
GO:0050900	Leukocyte migration	1.54	0.0018	
Cellular Component				
GO:0009897	External side of plasma membrane	1.42	0.0115	*CCR7,CD28,ITGA6,IGLL5*
GO:0042571	Immunoglobulin complex, circulating	2.79	0.0115	*IGJ,IGLL5*
GO:0098636	Protein complex involved in cell adhesion	2.09	0.0422	*CD28,ITGA6*

GO= gene ontology; FDR= false discovery rate.

Table 14-10: Gene ontology of the 105 downregulated genes (N=44)

Term ID	Term description	Strength	FDR	Matching proteins in our network (labels)
Biological process				
GO:0030218	Erythrocyte differentiation	1.37	3.03E-06	*SLC4A1,KLF1,TAL1,EPB42,AHSP,ALAS2,FAM210B,GATA1,DMTN*
GO:0042744	Hydrogen peroxide catabolic process	1.73	3.03E-06	*HBQ1,HBA2,HBA1,HBB,HBM,HBD,GPX1*
GO:0048821	Erythrocyte development	1.64	3.03E-06	*SLC4A1,TAL1,EPB42,ALAS2,FAM210B,GATA1,DMTN*
GO:0061515	Myeloid cell development	1.45	3.03E-06	*SLC4A1,TAL1,EPB42,ALAS2,TSPAN2,FAM210B,GATA1,DMTN*
GO:0030099	Myeloid cell differentiation	0.98	0.00016	*SLC4A1,KLF1,TAL1,EPB42,AHSP,ALAS2,TSPAN2,FAM210B,GATA1,DMTN*
GO:0098869	Cellular oxidant detoxification	1.22	0.00042	*HBQ1,HBA2,HBA1,HBB,HBM,HBD,GPX1*
GO:0015701	Bicarbonate transport	1.39	0.0023	*HBA2,SLC4A1,HBA1,HBB,CA1*
GO:0045639	Positive regulation of myeloid cell differentiation	1.11	0.0071	*TAL1,FAXDC2,TESC,TRIM58,FAM210B,GATA1*
GO:0002520	Immune system development	0.63	0.0087	*SLC4A1,LYL1,KLF1,TAL1,EPB42,AHSP,GLRX5,ALAS2,TSPAN2,POLL,FAM210B ,GATA1,DMTN*
GO:0030097	Hemopoiesis	0.65	0.0109	*SLC4A1,LYL1,KLF1,TAL1,EPB42,AHSP,GLRX5,ALAS2,TSPAN2,FAM210B,GATA 1,DMTN*
GO:0045648	Positive regulation of erythrocyte differentiation	1.42	0.0144	*TAL1,TRIM58,FAM210B,GATA1*
GO:0042168	Heme metabolic process	1.41	0.0154	*BLVRB,HMBS,ALAS2,SLC25A39*
GO:0020027	Hemoglobin metabolic process	1.76	0.0194	*EPB42,AHSP,ALAS2*
GO:0043249	Erythrocyte maturation	1.63	0.038	*TAL1,EPB42,FAM210B*
Molecular function				
GO:0005344	Oxygen carrier activity	1.96	1.06E-06	*HBQ1,HBA2,HBA1,HBB,HBM,HBD*
GO:0019825	Oxygen binding	1.55	6.60E-05	*HBQ1,HBA2,HBA1,HBB,HBM,HBD*
GO:0030492	Hemoglobin binding	1.98	0.00034	*SLC4A1,AHSP,HBB,HBD*
GO:0043177	Organic acid binding	0.87	0.0129	*HBQ1,HBA2,HBA1,ALAS2,HBB,HBM,PTGDS,HBD*
GO:0020037	Heme binding	0.98	0.0327	*HBQ1,HBA2,HBA1,HBB,HBM,HBD*
Cellular Component				
GO:0005833	Hemoglobin complex	2.09	3.19E-09	*HBQ1,HBA2,AHSP,HBA1,HBB,HBM,HBD*
GO:0031838	Haptoglobin-hemoglobin complex	2.06	8.98E-08	*HBQ1,HBA2,HBA1,HBB,HBM,HBD*
GO:0030863	Cortical cytoskeleton	1.12	0.0047	*GYPC,TMOD1,SLC4A1,EPB42,DMTN,SPTB*
GO:0014731	Spectrin-associated cytoskeleton	1.85	0.009	*ANK1,DMTN,SPTB*
GO:0072562	Blood microparticle	1.05	0.009	*HBA2,KRT1,SLC4A1,HBA1,HBB,HBD*

GO:0005829	Cytosol	0.23	0.0435	*STRADB,HBQ1,YBX3,HBA2,KRT1,TMOD1,GMPR,BLVRB,ANK1,FBXO7,HMBS, ARHGEF40,UBXN6,RILP,BCL2L1,UBB,AHSP,HBA1,CARM1,OSBP2,HBB,TESC, SNCA,MYL4,RBM38,HBM,GUK1,TAGLN2,IFIT1B,HBD,HAGH,SELENBP1,PHO SPHO1,GPX1,RUNDC3A,BAG1,DMTN,CA1,SPTB,FKBP8,SMOX*
GO:0071682	Endocytic vesicle lumen	1.5	0.0435	*HBA2,HBA1,HBB*

GO= gene ontology; FDR= false discovery rate.

Appendix B: Supplementary Figures

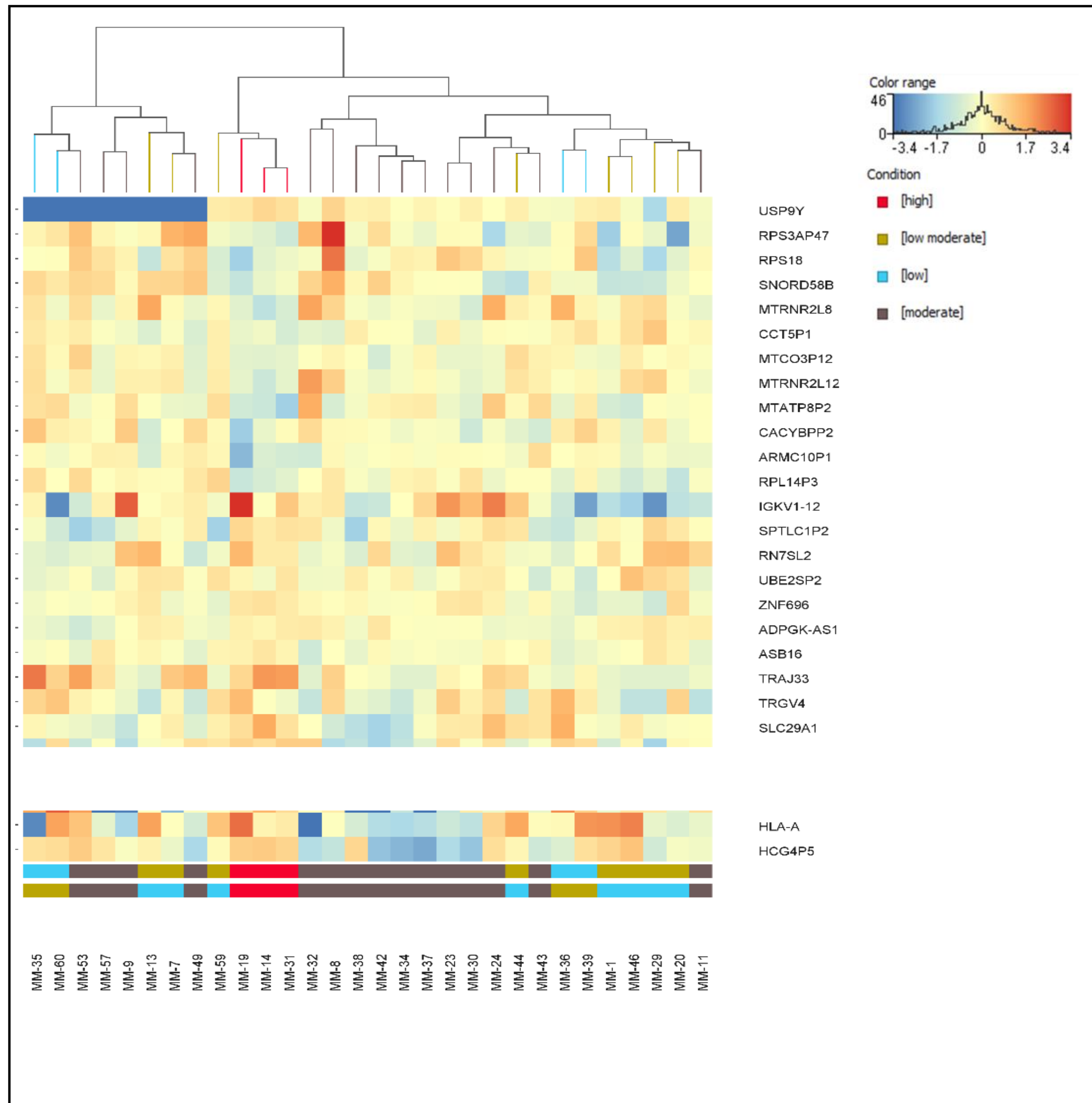

Figure 14-1: The heatmap of the DEGs among smokers based on nicotine dependence (n=31)

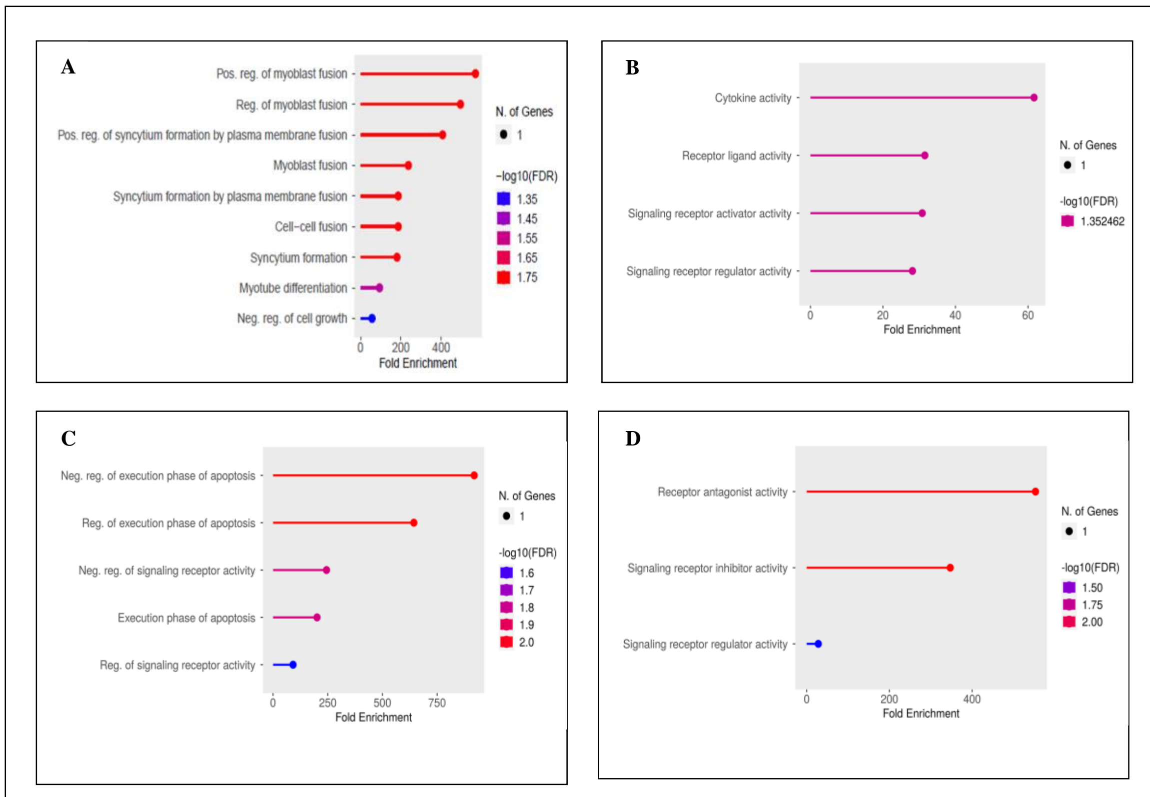

Figure 14-2: GO analysis of the DEGs among smokers categorized based on nicotine dependence

The DEGs were identified by comparing all ND categories to the low ND group. The analysis revealed the following: (A) Significant biological processes associated with the upregulated genes, (B) Significant molecular functions linked to the upregulated genes, (C) Significant biological processes related to the downregulated genes, and (D) Significant molecular functions associated with the downregulated genes.

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/449/2023.PL
Subject: 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)

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

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