THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PhD)

Genetic Factors Associated with the High Prevalence of Reduced High-density Lipoprotein Cholesterol Level in the Hungarian General and Roma populations: A Comparative Analysis

by Péter Pikó

Supervisor: Prof. Róza Ádány, MD, PhD, DSc

UNIVERSITY OF DEBRECEN

DOCTORAL SCHOOL OF HEALTH SCIENCES

DEBRECEN, 2021

Table of Contents

List of Abbreviations

ABCA1 - ATP-binding cassette transporter A-1

ABCG1 - ATP-binding cassette subfamily G member 1

ANGPTL4 - Angiopoietin like 4

APOA-1 - Apolipoprotein A-1

APOB - Apolipoprotein B

BMI - Body mass index

CE - Cholesteryl esters

CETP - Cholesteryl ester transfer protein

CVD - Cardiovascular disease

FADS2 - Fatty acid desaturase 2

FC - Free cholesterol

GALNT2 - Polypeptide N-acetylgalactosaminyltransferase 2

GPMSSP - Hungarian General Practitioners' Morbidity Sentinel Stations Programme

GRS - Genetic risk score (unweighted)

GWAS - Genome-wide association studies

H - Haplotype

HDL-C - High-density lipoprotein cholesterol

HG - Hungarian general

HMGCR - HMG-CoA Reductase

HR – Hungarian Roma

HWE - Hardy–Weinberg equilibrium

IDL - Intermediate-density lipoproteins

KCTD10 - Potassium channel tetramerization domain containing 10

LCAT - Lecithin cholesterol acyltransferase

LD - Linkage disequilibrium

LDL - Low-density lipoprotein

LDL-R - Low-density lipoprotein receptor

LIPC - Hepatic triglyceride lipase

LIPG - Endothelial lipase

LPL - Lipoprotein lipase

SCARB1 - Scavenger Receptor Class B Member 1

SNP - Single-nucleotide polymorphism

SRB1 - Scavenger receptor class B type 1

TG – Triglyceride

TRL - Triglyceride-rich lipoproteins

VLDL - Very-low-density lipoproteins

WWOX - WW Domain Containing Oxidoreductase

wGRS – Genetic risk score (weighted)

1. Introduction

Studying ethnicity-related biological, environmental, social, and psychological attributes have always been an essential component of multidisciplinary scientific studies related to the prevention and targeted intervention of diseases, with a special emphasis on the ones with high prevalence among ethnic groups [1].

1.1. Roma ethnicity

The Roma population, which constitutes the largest ethnic minority in Europe, is one of the main topics of ethnicity-based researches. It is estimated that 10-12 million Roma are scattered across the European continent [2], indicating an accumulation in Central, Eastern and Southern European countries (mainly Bulgaria, Hungary, Slovakia, Romania, and North Macedonia) [3]. The Roma are often concentrated in severely deprived regions [4], segregated colonies characterised by most unfavourable environmental conditions [5]. According to the latest census (2011), about 3.2% of the total population is Roma in Hungary; however, their estimated representation is much higher and is reaching some 8.5% of the total population of the country [6, 7]. This discrepancy stems from the fact that due to the widespread prejudices they face, many Roma people do not declare their ethnic origin [8]. Although studies on the Roma population face a number of challenges in terms of both data collection and methodology [8, 9], the available data strongly suggest that Roma populations suffer from poor health, lower life expectancy and barriers in access to healthcare [10-13].

1.2. Possible genetic causes in the background of high prevalence of CVD morbidity and mortality among Roma

Comparative studies of the risk profile of Roma adults indicate that the Roma population carries a significantly higher risk burden of cardiovascular diseases (CVDs) than the majority population, regardless of the country where they live [14-20]. Although some studies have also examined the possible role of genetic factors in relation to individual metabolic traits [21-23], the contribution of environmental and genetic factors to the increased risk of CVDs has not been determined in any of those studies.

The studies of which results are summarized of the present thesis are based on the findings of a previous health examination survey carried out by our research team, which estimated and compared the prevalence of metabolic syndrome (which is the most robust indicator of the risk of many non-communicable diseases) and its components on representative random samples of the Hungarian general population and Roma living in segregated settlements [24]. It was found that the prevalence of metabolic syndrome and its components, especially reduced plasma High-density lipoprotein cholesterol (HDL-C) level (normal values are <1.03 mmol/l in males and <1.29 mmol/l in females) or treated lipid disorders were significantly more frequent in all age groups of the Hungarian Roma population compared to the Hungarian general population. For more details on the age-specific prevalence of it see Figure 1.

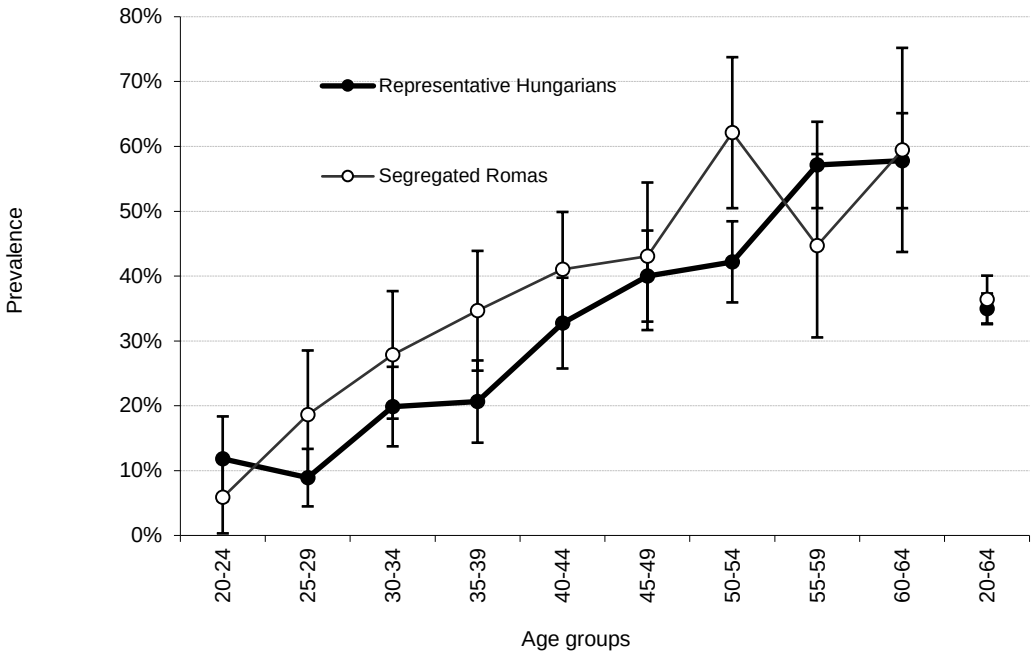


Figure 1. Age-specific prevalence of reduced HDL-C concentration or treated lipid disorders in 20-64-years old females' samples representative for Hungary and for segregated Roma colonies in Hajdú-Bihar and Szabolcs-Szatmár-Bereg counties

Source: Kosa, Z., et al., Prevalence of metabolic syndrome among Roma: a comparative health examination survey in Hungary. Eur J Public Health, 2015. 25(2): p. 299-304.

A low level of plasma HDL-C was identified as an independent risk factor for coronary artery disease in the 1980s, and has been one of the most important risk factors used by clinicians to assess the present or future presence of high cardiovascular risk ever since [25]. HDL-C metabolism has a strong genetic basis since 40-60% of the estimated interindividual variation in serum concentration is regulated at a genetic level [26, 27].

Based on the above results and the high degree of consanguinity in the Roma population [28, 29]; the prevalence of endogamy among the Roma, i.e. the prevalence of first cousin couples is 16 times higher among them than that in the Hungarian population [30], it is reasonable to assume that ethnic differences exist not only in the prevalence of reduced HDL-C levels but also in the prevalence of related gene polymorphisms.

1.3. Some key players in lipid metabolism

1.3.1. High-density lipoprotein cholesterol

High-density lipoprotein (HDL) is one of the five major groups of lipoproteins, which are complex particles composed of multiple proteins which transport lipids (including cholesterol, phospholipids, and triglycerides) in the bloodstream. HDL is mainly secreted by the liver and the small intestines but it is the liver that secretes ~70–80% of the HDL to total HDL-C present in the plasma.

It is a well-known fact that HDL-C levels in the blood show an inverse relationship with the incidence of cardiovascular diseases. In addition to reverse cholesterol transport, HDL is also known to have other health effects (anti-thrombotic [31], anti-inflammatory [32], anti-oxidative [33], and pro-vasodilatory [34]). Therefore, its measurable level in blood plasma serves as an overall indicator of an individual's risk to different non-communicable diseases. For more details on the role of HDL-C, see Figure 2.

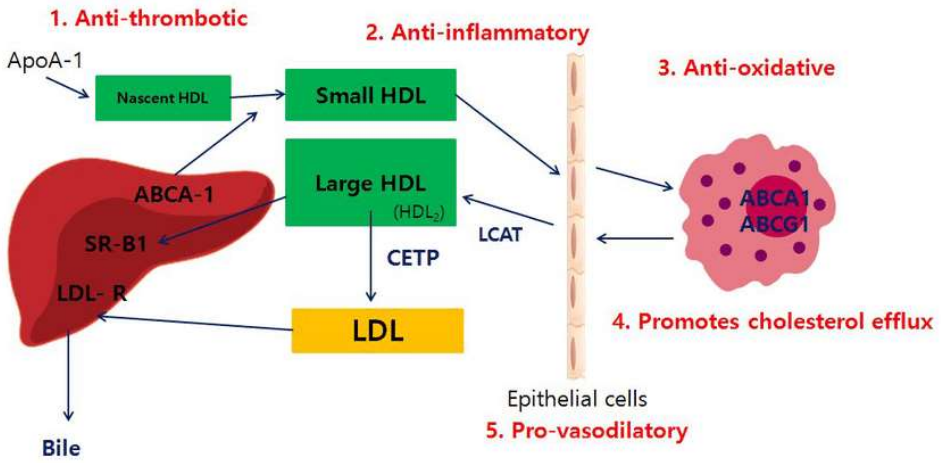


Figure 2. The role of high-density lipoprotein cholesterol and its involvement in the reverse cholesterol transport (the process of cholesterol movement from the extrahepatic tissues back to the liver).

ABCA-1, ATP-binding cassette transporter A-1; ABCG1, ATP-binding cassette subfamily G member 1; ApoA-1, apolipoprotein A-1; CETP, cholesteryl ester transfer protein; HDL, high-

density lipoprotein; LCAT, lecithin cholesterol acyltransferase; LDL, low-density lipoprotein; LDL-R, low-density lipoprotein receptor; SRB-1, scavenger receptor class B type 1.

Source: Ahn, N. and K. Kim, *High-density lipoprotein cholesterol (HDL-C) in cardiovascular disease: effect of exercise training.* Integr Med Res, 2016. **5**(3): p. 212-215. [35]

Many genes and their products are involved in the synthesis of HDL and the processes it regulates. Mutations in these genes and/or in the DNA segments that regulate them can affect the plasma levels of HDL-C and, consequently, the health of the individual. Among the many genes, cholesteryl ester transfer protein (*CETP*) and hepatic lipase (*LIPC*) should be highlighted as key contributors to the regulation of plasma HDL-C level.

1.3.2. Cholesteryl ester transfer protein (CETP)

The CETP promotes the transfer of cholesteryl esters from HDL-C (considered as antiatherogenic effect) to apolipoprotein B containing lipoproteins (considered as proatherogenic effect), including very-low-density lipoproteins, intermediate-density lipoproteins (IDL), and low-density lipoproteins (LDL) [36]. The reduced level of CETP is associated with increased HDL-C and decreased LDL-C levels, which is a typically antiatherogenic lipid profile [36]. The favourable lipid-profile associated with decreased level of CETP has been in the focus of studies for more than 30 years [37], and current studies are trying to find different ways of reducing CETP level by different drugs (e.g.: CETP inhibitors; [38]). For more details on the role of CETP, see Figure 3.

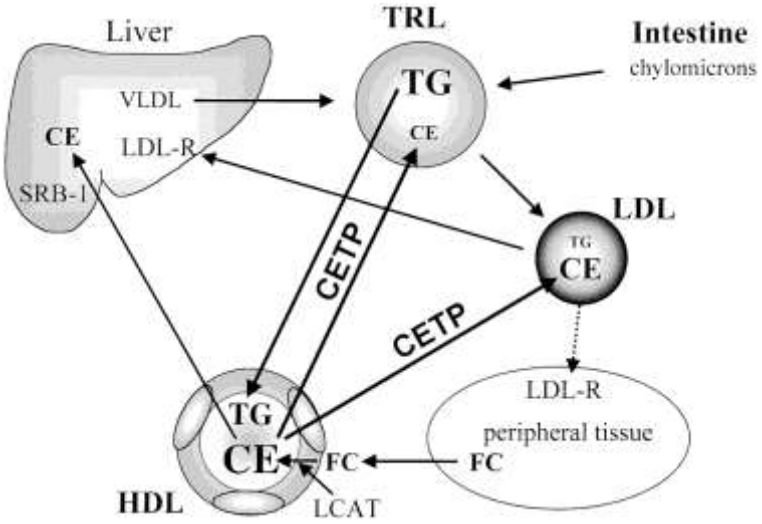


Figure 3. The role of cholesteryl ester transfer protein (CETP) in plasma lipid transport.

CE, cholesteryl esters; FC, free cholesterol; HDL, high-density lipoprotein; LDL, low-density lipoprotein; LDL-R, low-density lioprotein receptor; TG, triglyceride; TRL, triglyceride-rich lipoproteins; SRB-1, scavenger receptor class B type 1; VLDL, very-low-density lipoproteins.

Source: Barter, P.J., et al., *Cholesteryl ester transfer protein: a novel target for raising HDL and inhibiting atherosclerosis.* Arterioscler Thromb Vasc Biol, 2003. **23**(2): p. 160-7. [36]

1.4.2 Hepatic lipase also named as Lipase C (LIPC)

Hepatic lipase is a form of lipase with dual functions as triglyceride/phospholipid hydrolase and ligand/bridging factor for receptor-mediated lipoprotein uptake. Hepatic lipase is coded by the LIPC gene and is expressed mainly in hepatocytes and endothelial cells. It plays an important role in triglyceride hydrolysis in the blood by maintaining steady levels of IDL, HDL and LDL [39]. Hepatic lipase and ABCL1 protein also help the transfer of free fatty acids from plaques in the arteries to HDL. This process creates HDL-3 [40]. For more details on the role of hepatic lipase, see Figure 4.

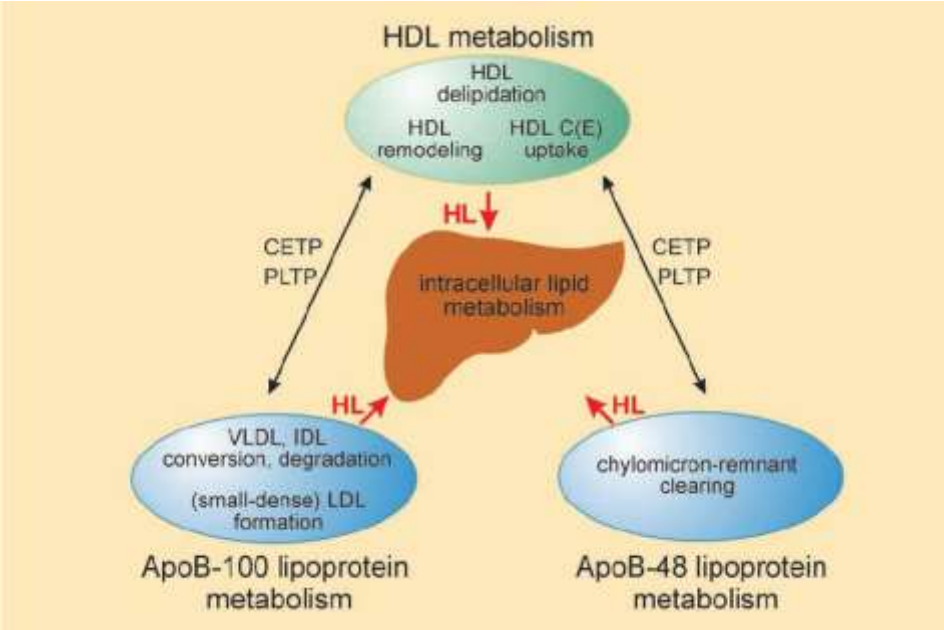


Figure 4. Role of hepatic lipase (HL) in lipid metabolism.

ApoB-48, apolipoprotein B-48; ApoB-100, apolipoprotein B-100; CETP, cholesteryl ester transfer protein; HDL, High-density lipoprotein; HDL C, High-density lipoprotein cholesterol; IDL, intermediate-density lipoproteins; LDL, low-density lipoproteins; PLTP, phospholipid transfer protein; VLDL, very-low-density lipoproteins

Source: Jansen, H., A.J. Verhoeven, and E.J. Sijbrands, Hepatic lipase: a pro- or anti-atherogenic protein? J Lipid Res, 2002. 43(9): p. 1352-62.

1.4. Genetic risk score calculations

Genetic variations describe the differences occurring among individuals. When a specific nucleotide at a particular position in the genome differs from the reference nucleotide and it is present in a sufficiently large fraction of the population (at least 1%), it is called a single nucleotide polymorphism (SNP). SNPs are the most common genetic variations that exist in the human genome, and a typical person differs from the reference human genome at 4-5 millions of these variations [41] some of which have a real effect on different traits (lipid metabolism, carbohydrate metabolism, pathways involved in blood pressure regulation, etc.) while others do not. The individual risk alleles associated with these traits have only a small effect on them, with most genotype relative risks being

in the range of 1.1–2.0; consequently, the predictive value of a single SNP is very low, in general.

Calculating the polygenic/genetic risk score (GRS) is a very effective way of estimating the cumulative effect of selected SNPs on a biological outcome [42]. However, in order to do these calculations we need independently published genome-wide association studies (GWAS) which estimate the full impact of SNPs [43]. In addition to the many benefits of GRS, one of the critical points against it is that the impact of SNPs has been described mainly on populations of European descent and may not be applicable to different ethnic groups with non-European origin such as the Roma.

1.5. Applicability of SNPs described in European populations to a population with non-European origin

The effect of SNPs associated with lipid metabolism is lesser known in non-European populations. Most genetic studies dealing with ethnic minorities underperform as a result of relatively small sample sizes, what is more, most of the genotyping platforms are designed to have optimally cover the genetic variants previously identified in European populations. Therefore, it cannot be excluded that the directions and/or magnitudes of the effect of SNPs on HDL-C concentration differ significantly in non-European populations when compared to those obtained in populations of European origin [44-46].

According to the NHGRI-EBI Catalogue of published genome-wide association studies, the majority of GWAS for HDL-C level were performed in populations of European descent [47-71], some of them in other populations [72-91], and none of them in the Roma.

A comprehensive research on the effect of SNPs discovered in different populations is needed that would explain either the generalizability or the lack of them across populations before applying them in GRS calculations for the Roma population. The received generalization data will show whether the SNPs identified by GWAS are simply tagSNPs or they are more likely to be true functional SNPs [92] across the different populations.

1.6. Haplotype analysis

Haplotype analysis is a tool for arranging alleles/SNPs that are inherited together on the same genes/chromosomes. Haplotyping involves an identification process where the recombination events are located between different markers. In order to avoid multicollinearity, it is enough to examine one of the SNPs with a linkage disequilibrium (LD) above 0.8 during the GRS calculation. Nevertheless, haplotype analysis provides an opportunity to examine the combined effect of these SNPs on the observed trait and what is more, it can also be considered an extension of the GRS calculation.

1.7. Aims

The aim of our studies was to define whether genetic susceptibility contributes to the higher prevalence of reduced HDL-C level among Roma in addition to the effects of unfavourable environmental factors (e.g. unhealthy nutrition, smoking, alcohol consumption) being more common among them [93, 94]. Our hypothesis on the involvement of genetic factors in the background of the observed phenomenon was tested in four stages:

The first stage was a structured literature search to identify polymorphisms that are thought to contribute to the occurrence/development of low HDL-C level. From the

identified SNPs, the most relevant ones for the study were selected and genotyped in the Hungarian general and Roma populations.

The second stage was determining whether the effect of genotyped SNPs described in European-descent population-based studies can be transferred to the Roma population (with Asian origin) [95, 96].

In the third stage, if the applicability of SNPs in Roma populations is demonstrated, unweighted and weighted genetic risk score values were calculated for both study populations. Biostatistical methods were used to examine whether the calculated GRSs show a correlation with HDL-C level (as continuous or binary variables as outcome) in both populations or not. If a significant association of GRSs with HDL-C level is confirmed, it became possible to compare the genetic burden of the study populations.

In the fourth stage, haplotype analysis was used to characterize the combined effect of SNPs in the same gene on HDL-C level and to identify the unique co-occurrence of SNPs in the Hungarian general and Roma populations.

The answers to our questions have important implications for the design and implementation of - hopefully effective - interventions that, in the presence of genetically determined susceptibility, may prevent the clinical manifestation of the elevated cardiovascular disease risk resulting from reduced HDL-C level in vulnerable groups.

2. Materials and methods

2.1. Sample populations

The studies described in the present thesis involved 757 randomly selected Hungarian Roma individuals living in segregated colonies in North-East Hungary and 1783 individuals selected also at random from the Hungarian general population.

2.1.1. Sample of Roma living in segregated colonies

Roma participants were enrolled from two Northeast Hungarian counties (Hajdú-Bihar and Szabolcs-Szatmár-Bereg) where the majority of their segregations are found by using stratified multistage sampling [24].

Briefly, segregated colonies of the Roma were identified by field workers with Roma origins within the framework of a project of the Hungarian Ministry of Environmental Protection and the University of Debrecen [5]. The ethnic affiliation of individuals was assessed by self-declaration. Only those segregated colonies were included in the study which consisted of more than 100 individuals. Sixty-four such colonies were identified and 40 of them were selected randomly for the study (25 from Hajdú-Bihar and 15 from Szabolcs-Szatmár-Bereg county). For each colony, 25 households were randomly selected, and for each household, all individuals aged 20 years or more were identified and one of them was randomly selected to participate in the study. After 3 GPs refused to participate, a total of 925 people from the practices of 37 GPs formed the final sample population (22 GPs in Hajdú-Bihar county (22X25 individuals) and 15 GPs in Szabolcs-Szatmár-Bereg county (15X25 individuals).

As a part of this health examination survey, medical histories and socio-demographic characteristics (age, sex, marital status, education, financial situation, number of people

living in a household and employment) were recorded and physical examinations were carried out (weight, height, waist circumference, blood pressure measurements) for each participant. Medical history information was completed by GP. This included data collection on the occurrence of known hypertension, known carbohydrate metabolism disorder, known lipid metabolism disorder, known obesity, and known metabolic syndrome. As well as co-morbidities based on hospital or specialist diagnosis for stroke, peripheral vascular disease, acute myocardial infarction, ischemic heart disease, heart failure, diabetic nephropathy, asthma, AV block, sick sinus syndrome, and kidney disease. Blood pressure measurements were performed twice on the same day with a standard mercury sphygmomanometer at intervals of at least 3 minutes. The average of the two measurements was used in the analysis. Blood samples were taken for laboratory (among others serum triglyceride, HDL-cholesterol, glucose levels were determined) and genotype investigations.

In 757 of the 925 planned sample collections, informed consent was signed, the questionnaire was completed and performing the physical examination. This represents an 82% turnout.

2.1.2. Hungarian general population

The Hungarian General Practitioners' Morbidity Sentinel Stations Programme (GPMSSP) was established in 1998 by the School of Public Health, University of Debrecen and the National Public Health and Medical Officer Service [97]. The main purpose of the GPMSSP is to monitor the incidence and prevalence of non-communicable chronic diseases (e.g.: lipid disorders, diabetes mellitus, ischaemic heart disease, hypertension, stroke, acute myocardial infarction, liver cirrhosis and malignancies of the colon and rectum, breast, prostate, cervix, and the respiratory tract) of great public health importance.

At the start of the program, samples were collected from four counties (Hajdú-Bihar, Győr-Moson-Sopron, Szabolcs-Szatmár-Bereg and Zala counties), which were later supplemented by four more counties (Komárom-Esztergom and Bács-Kiskun, Baranya, and Heves counties). The source population of the sample collection was selected from the Hungarian citizens aged 20-69 in 59 participating GP practice. Samples were randomly selected from the files of residents in the catchment area and 2006 individuals were selected at random, proportional to the size of the practice, and stratified for age and gender. The resulting sample population can be considered representative of the Hungarian general population on the bases of age, gender and geographical distribution.

The study followed a methodology similar to that of the Roma one: in health examination surveys [98], medical histories and socio-demographic characteristics were recorded and physical examinations were carried out (weight, height, waist circumference, blood pressure measurements) for each participant. Blood samples were taken for laboratory tests (among them serum triglyceride, HDL-C, fasting glucose levels were measured) and genotype investigations.

In 1783 of the 2006 planned sample collections, informed consent was signed, the questionnaire was completed and performing the physical examination. This represents an 89% turnout.

2.2. Ethical approval

All procedures performed in the studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments.

This study was approved by the Ethical Committee of the University of Debrecen, Medical Health Sciences Centre (reference No. 2462-2006) and by the Ethical Committee

of the Hungarian Scientific Council on Health (reference Nos. NKFP/1/0003/2005; 8907-O/2011-EKU,).

2.3. DNA isolation

DNA was isolated by using a MagNA Pure LC system (Roche Diagnostics, Basel, Switzerland) with a MagNA Pure LC DNA Isolation Kit–Large Volume according to the manufacturer's instructions. Extracted DNA was eluted in 200 µl MagNA Pure LC DNA Isolation Kit-Large Volume elution buffer.

2.4. Selection of SNPs

A systematic literature review on the PubMed, HuGE Navigator and Ensembl databases was conducted to identify SNPs most strongly associated with HDL synthesis and cholesterol transport using different combinations of the following keywords and terms: high-density lipoprotein cholesterol, cholesterol transport and synthesis, single nucleotide polymorphism, candidate gene, and meta-analysis. The literature search identified 130 SNPs.

Those studies were used to identify the relevant ones for the study which were original researches conducted on human samples and which applied either the candidate gene or the GWAS approach since both of them result in susceptible/protective alleles. The references of the selected articles and reviews were also examined to identify additional related studies. SNPs were selected from the above studies only if they were found to be consistently associated with plasma HDL-C level in samples with a biostatistically acceptable size. 76 SNPs met the criteria described above and were included in the further investigation.

The 76 SNPs were listed by priority (effect and association strength on HDL-C level) and the first 33 were selected for genotyping. During the assay design, a pool of 23 SNPs was created for genotyping by the service provider (Mutation Analysis Core Facility of the Karolinska University Hospital, Sweden). For more details on the SNP selection see Figure 5.

Figure 5. Flowchart of the SNP selection process.

Genotyping was successful for 21 SNPs which then became the subjects of our further studies. For 18 of the 21 genotyped SNPs, properly documented estimated effect (weight value) was available for comparative analysis. To avoid multicollinearity, only one SNP is selected from each identified LD block. For the list and role of SNPs considered in the study see Table 1.

Table 1. List and role of SNPs considered in the study

No.	SNP	Effect comp.	GRS	wGRS	Hapl. analy.	Excl.	Gene	References
1	rs4149268	X	X	X	-	-	*ABCA1*	[99] [100]
2	rs1883025	-	-	-	-	X		[100] [101]
3	rs2967605	-	-	-	-	X	*ANGPTL4*	[102]
4	rs964184	-	-	-	-	X	*APOA1*	[102] [103]
5	rs693	X	X	-	-	-	*APOB*	[104] [105] [106]
6	rs6754295	-	-	-	-	X		[107]
7	rs3764261	-	-	-	-	X	*CETP*	[107] [108]
8	rs1800775	-	-	-	-	X		[109] [110]
9	rs5882	X	X	X	X	-		[111] [112]
10	rs1532624 [a]	X	X	-	X	-		[113] [114] [115] [116]
11	rs173539	-	-	-	-	X		[114] [115] [117]
12	rs9989419	X	X	X	X	-		[107] [118] [110]
13	rs7499892	X	X	X	X	-		[114] [111] [119]
14	rs1864163	-	-	-	-	X		[100] [118]
15	rs708272[a]	X	X	X	X			[112]
16	rs174570	-	-	-	-	X	*FADS2*	[120]
17	rs4846914[d]	X	X	-	-	-	*GALNT2*	[111] [113] [121] [122] [123] [124]
18	rs2144300[d]	X	X	X				[99] [115]
19	rs3846662	X	X	-	-	-	*HMGCR*	[125]
20	rs2338104	X	X	X	-	-	*KCTD10*	[113] [122]
21	rs1800588[c]	X	X	-	X	-	*LIPC*	[111] [114] [115] [121]
22	rs10468017[b]	X	X	-	X	-		[99] [113] [122] [64]
23	rs4775041[b]	X	X	X	X	-		[99] [100] [111]
24	rs1532085	X	X	X	X	-		[113] [126]
25	rs1077834[c]	X	X	X	X	-		[114]
26	rs2070895[c]	X	X	-	X	-		[114] [119]
27	rs4939883	X	X	X	-	-	*LIPG*	[99] [122]
28	rs2000813	X	X	X	-	-		[114]
29	rs328	X	X	X	-	-	*LPL*	[111] [113] [114] [121]
30	rs2083637	-	-	-	-	X		[107]
31	rs331	-	-	-	-	X		[107] [127]
32	rs5888	-	-	-	-	X	*SCARB1*	[128]
33	rs2548861	X	X	-	-	-	*WWOX*	[129]

[a]LD block 1; [b]LD block2; [c]LD block3; [d]LD block 4

Effect. comp., effect comparison analysis; GRS, unweighted genetic risk score; wGRS, weighted genetic risk score; Hapl. analy., haplotype analysis; Excl., excluded from further study; ABCA1, ATP-binding cassette transporter ABCA1; ANGPTL4, Angiopoietin Like 4; APOA1, ATP-binding cassette transporter ABCA1; APOB, Apolipoprotein B; CETP, Cholesteryl ester transfer protein; FADS2, Fatty acid desaturase 2; GALNT2, Polypeptide N-acetylgalactosaminyl transferase 2; HMGCR, HMG-CoA Reductase; KCTD10, Potassium channel tetramerization domain containing 10; LIPC, Hepatic lipase; LIPG, Endothelial lipase; LPL, Lipoprotein lipase; SCARB1, Scavenger Receptor Class B Member 1; WWOX, WW Domain Containing Oxidoreductase

2.5. Genotyping

Genotyping was performed on a MassARRAY platform (Sequenom Inc., San Diego, CA, USA) with iPLEX Gold chemistry by a service provider (Mutation Analysis Core Facility of the Karolinska University Hospital, Stockholm, Sweden). Validation, concordance analysis and quality control were conducted by the service provider according to their protocols.

2.6. Adequate sample size and statistical power calculation

The adequate sample size for the study groups was computed using Online Sample Size Estimator online calculator (http://osse.bii.a-star.edu.sg/calculation1.php), assuming a power of 80%, an alpha-level of 0.05 for the case-control ratio of 1:2.5 (Hungarian Roma: Hungarian general samples ratio). The allele frequencies for CEU (Utah Residents with Northern and Western Ancestry) and for GIH (Gujarati Indian from Houston, Texas) populations were taken from phase 3 of the 1000 Genome project as an example and they were applied for sample size estimation, considering the fact that the Hungarian Roma population arrived to the Balkans from North-India [95]. Power estimation was performed for the study populations by the Quanto Software Version 1.2.4 [130]. For more details on adequate sample size and result of power calculation see Table 3.

2.7. Comparison of the individual effects of SNPs on HDL-C level in the Hungarian general and Roma populations

The individual effects of SNPs (size and direction) on HDL-C level were estimated and compared between the HG and HR populations; furthermore, the data obtained from the available literature on the estimated effect of SNPs in independent European populations were compared to our results. To test for the effects of genotypes on the

quantitative trait (HDL-C level), linear regression analyses were performed and all results were adjusted for relevant covariates (age, sex, and BMI). Transferability of results obtained was defined as having no significant differences in the effect of SNPs on HDL-C level between the two study groups. This definition is based on the assumption that the effect in the Hungarian general population is consistent with the results previously obtained from populations with European ancestry.

The nominal level of significance ($p<0.05$) was applied in case of each comparison. The use of nominal significance for replication and for generalization on non-European populations has been widely applied by some large consortium and is considered an acceptable criterion for publication [45, 131].

2.8. Calculation of GRS and wGRS values

Unweighted and weighted genetic risk scores (GRS and wGRS) were calculated to define the combined effect of several HDL-C related SNPs. Individuals without full genotypic data were excluded from the studies.

In the GRS, each person was assigned a score based on the number of risk alleles carried. Thus, risk allele homozygotes were coded as genotype "2", heterozygotes as genotype "1", while "0" indicated the absence of the risk allele. When the effect allele was reported to be protective, the coding was "0" for effect allele homozygotes and "2" for reverse allele homozygotes [132]. By using these codes, a simple count score (unweighted) was calculated as it is described by equation (1) in which G_i is the number of the risk alleles for the i^{th} SNP. This model sums up all risk alleles of all loci as a summary score assuming that all alleles have the same effect.

$$GRS = \sum_{i=1}^{I} G_i \quad (1)$$

In a weighted approach, rather than giving equal weight to all SNPs, we used their effects described in the literature selected as weight scores, and those with a greater impact on the outcome contribute more to the wGRS. Weights were derived from the risk coefficient for each loci based on the reported beta values of previous HDL-C associated studies [56, 61, 69, 133-136]. The estimated effect on HDL-C level was available for 18 SNPs in 7 genes (*ABCA1* gene: rs4149268; *CETP* gene: rs1532624, rs5882, rs708272, rs7499892, and rs9989419; *GALNT2* gene: rs2144300 and rs4846914; *KCTD10* gene: rs2338104; *LIPC* gene: rs10468017, rs1077834, rs1532085, rs1800588, rs2070895, and rs4775041; *LIPG* gene: rs2000813 and rs4939883; *LPL* gene: rs328) based on the result of literature research. The calculation of the weighted genetic risk score is described by equation (2). In this weighted genetic risk score, weights (w_{β_i}) were derived from the risk effect for each SNP based on relative effect size determined (in beta value) by previous studies. These weights (w_{β_i}) were multiplied by 0, 1 or 2 according to the number of risk alleles carried by each person (X_i) [132, 137, 138].

$$wGRS = \sum_{i=1}^{I} w_{\beta\_i} X_i \quad (2)$$

Two-sided t tests were applied to compare the distribution of GRSs between the study populations. To reveal if the association between genetic risk and ethnicity depends on the influence of factors relevant to HDL-C levels (sex, age, body mass index, systolic and diastolic blood pressure, fasting glucose level, as well as antihypertensive, antidiabetic, and lipid-lowering treatments)) multivariate linear regression analyses were used in which GRSs were the dependent variable, while the ethnicity of population (HR and HG), relevant covariates were considered as independent variables. Quintiles were generated

based on wGRS and the association of quintiles with changes in HDL-C levels and the prevalence of reduced HDL-C level was examined by trend analysis. The association of GRSs with reduced plasma HDL-C level (as binary outcome) and plasma HDL-C levels (as continuous outcome) were further evaluated by using univariate and multiple regression models separately in HR and HG subjects. In univariate regression analysis, GRS was used as an independent variable. Multiple regression analysis was used in 2 form, in the first, in addition to the GRS, age and gender were used as independent variables, while in the second, all relevant ones. The possibility of ethnic differences in association patterns were further confirmed by analysing study populations (HR and HG) together in one multiple regression model by adding ethnicity and all other relevant factors (sex, age, body mass index, systolic and diastolic blood pressure, fasting glucose level, as well as antihypertensive, antidiabetic, and lipid-lowering treatments)as covariates. The applied thresholds for reduced serum HDL cholesterol level were <1.03 mmol/l for men and <1.29 mmol/l for women. Based on the consensus definition of the International Diabetes Federation for the metabolic syndrome [139], those individuals were considered to have reduced HDL-C level whose parameters were below the above threshold or who were treated with lipid abnormality.

2.9. Haplotype block analyses

The haplotype block analyses of SNPs were estimated by using the expectation maximization algorithm carried out by the SNPStats online tool [140].

To avoid effects resulting from ethnicity related factors (e.g. environment and cultural effects), the study populations were examined together in a combined population, and then ethnicity was used as a covariate in the statistical models. All models were adjusted

by relevant covariates (ethnicity, sex, age, BMI, systolic and diastolic blood pressure, fasting glucose level, antihypertensive, antidiabetic and lipid-lowering treatment).

Bonferroni corrected p-values were used for haplotype analyses. The number of independent SNPs for each gene (*CETP* and *LIPC*) was defined by using the SNPsnap web-based tool [141], which was 4 in case of the *CETP* gene and 2 for the *LIPC*. For haplotype analysis, the result was considered significant if the p-value was less than 0.0125 for haplotypes in *CETP* and less than 0.025 for haplotypes in the *LIPC* gene according to the Bonferroni correction method.

2.10. Statistical analyses

Statistical tests were conducted with Stata (version 13) and PLINK (version 1.07) softwares. Mann-Whitney U test was used to compare the age, BMI and HDL-C level of the populations. Prevalence data were compared by $\chi 2$ test. Shapiro-Wilk test was applied for testing the normality of data. When necessary, dependent quantitative variables were transformed by using a two-step approach suggested by Templeton to reduce the effect of non-normality [142]. The Bonferroni adjustment was applied when genetic risk calculation was being performed ($p<0.0023$).

The existence of Hardy-Weinberg equilibrium (HWE) and the differences of allele frequencies for all SNP variants between the Hungarian general and Hungarian Roma populations were evaluated with $\chi 2$ test. Linkage disequilibrium between polymorphisms was analysed by using Haploview software (version 4.2). Generally, the conventional p threshold of 0.05 was applied.

3. Results

3.1. Characteristics of the study populations

Only those subjects were considered whose genotype data were available in addition to clinical and demographic data. In total, 616 Hungarian Roma (HR) and 1497 Hungarian general (HG) samples were included in the baseline comparison. The proportion of males in Roma samples was significantly lower (HG: 47.3% vs. HR: 38.3%, $p<0.001$). The age distribution of Roma subjects was shifted towards the younger age groups and deviated from the HG one (HG: 44.16 vs. HR: 40.30, $p<0.001$). There were significant differences in mean plasma HDL-C levels (HG: 1.43 mmol/L vs. HR: 1.21 mmol/L, $p<0.001$), but the distribution of BMI values kg/m^{2}did not differ significantly (HG: 27.43 kg/m^2 vs. HR: 27.497 kg/m^2; $p=0.898$). The proportion of subjects with reduced plasma HDL-C level was significantly higher among the Roma compared to the Hungarian general population (HG: 28.20% vs. HR: 53.00%, $p<0.001$). Lipid-lowering (HG: 13.90% vs. HR: 10.80%, p=0.050) and antihypertensive treatment (HG: 29.90% vs. HR: 25.40%, p=0.042) was more common in the Hungarian general population than in the Roma. There was no difference in the prevalence of antidiabetic treatment between the study populations. For more details on the demographic characteristics and laboratory data of the study populations, see Table 2.

Table 2. Characteristics of the study populations.

	Hungarian general (N=1497)	Roma (N=616)	p-value
	Mean (95% CI)		
Age (year)	44.16 (43.53 - 44.78)	40.30 (39.39 - 41.21)	<0.001
BMI (kg/m^2)	27.43 (27.15 - 27.70)	27.47 (26.64 - 28.30)	0.898
Systolic blood pressure (mmHg)	126.81 (125.95 - 127.66)	125.21 (123.67 - 126.76)	0.059
Diastolic blood pressure (mmHg)	80.26 (79.79 - 80.72)	78.43 (77.63 - 79.23)	<0.001
Fasting glucose level (mmol/L)	4.83 (4.74 - 4.92)	5.44 (5.29 - 5.59)	<0.001
HDL-C level (mmol/L)	1.43 (1.40 - 1.45)	1.21 (1.18 - 1.24)	<0.001
	Prevalence (%)		p-value
Sex (female/male)	52.7/47.3	61.7/38.3	<0.001
Lipid-lowering treatment	13.90	10.80	0.050
Antidiabetic treatment	5.40	5.20	0.901
Antihypertensive treatment	29.90	25.40	0.042
Reduced HDL-C level[a]	28.20	53.00	<0.001

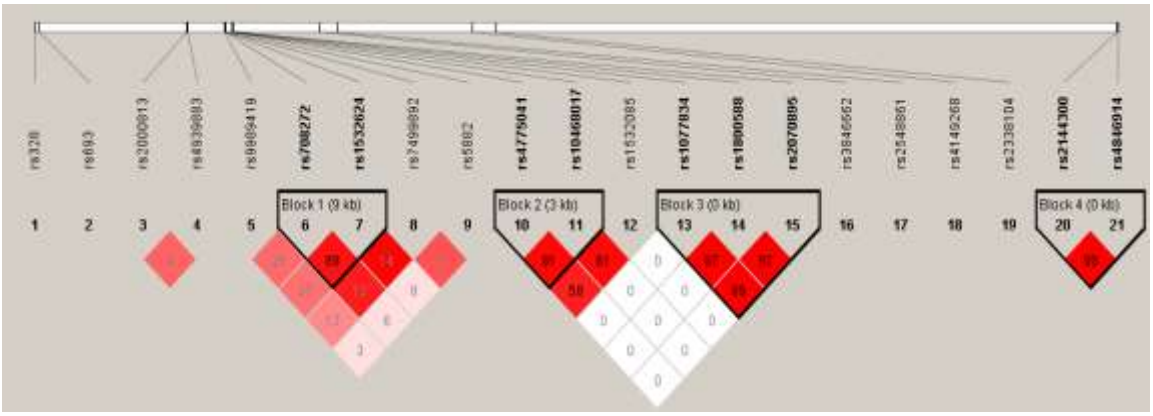
[a]Reduced HDL-C levels: <1.03 mmol/L in male and <1.29 mmol/L in female.

3.2. Results of linkage analysis

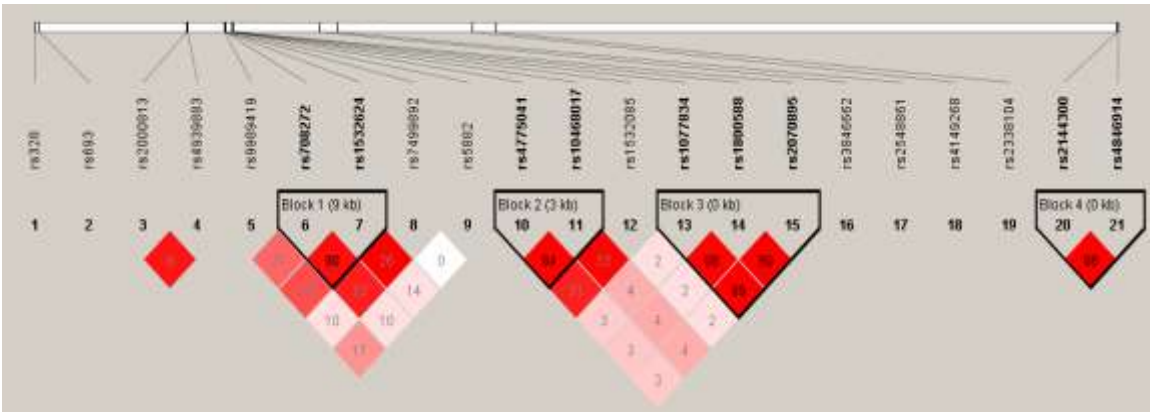
The SNPs of the *CETP* gene (rs708272 - rs1532624), the SNPs of the *LIPC* gene (rs4775041-rs104688017 and rs1077843-rs1800588-rs2070895) and rs2144300-rs4846914 SNPs of the *GALNT2* gene were in complete linkage disequilibrium (for more details, see Figure 6). Therefore, from among these genes, only rs708272 of *CETP*, rs4775041 and rs1077843 of the *LIPC* gene and rs2144300 of the *GALNT2* gene were taken into consideration in the calculation of the weighted genetic risk score, resulting in a wGRS composed of 13 SNPs.

Figure 6. Haplotype block organization of 21 SNPs related to high-density lipoprotein cholesterol level on the LD maps for the Hungarian general (1A) and Roma (1B) populations

1A



1B



Linkage analyses were performed separately in the study populations. According to the LD map generated by Haploview, there are four haplotype blocks (outlined in a bold black line) consisting of variants that occur in high LD. The blocks were formed by the SNPs of the *CETP*, *LIPC* and *GALNT2* genes. The numbers above the map show the rs-numbers of SNPs. The colour scheme is a standard Haploview colour scheme (white D′ < 1 and LOD < 2, shades of pink/red: D′ < 1 and LOD ≥ 2, and bright red D′ = 1 and LOD ≥ 2). Numbers in squares are D′ values.

3.3. Comparison of allele frequencies

Twenty-one SNPs were tested to determine whether the observed genotype frequencies were consistent with the Hardy-Weinberg equilibrium in both study populations. No significant deviation from HWE was observed. Differences between the HR and HG populations remained significant for 12 SNPs after the multiple test correction.

Six protective alleles consisting SNPs with a HDL-C level-increasing effect (rs1532085, rs1800588, rs2070895, rs2000813, rs4939883, and rs328) were more frequent in the Roma population, and three alleles (rs4149268, rs2144300, and rs38466629) were more prevalent in the HG one. Out of 7 susceptibility SNPs, four (rs7499892, rs4846914, rs2338104, and rs2548861) were more frequent in the Roma population and two of them (rs5882 and rs9989419) were more common in the Hungarian general one. More details on allele frequencies in the study populations, which were calculated on the basis of the obtained genotype distributions, are shown in Table 3.

Table 3. The calculated adequate sample size with the statistical power and comparison of protective and susceptibility allele frequencies (%) in the Hungarian general (HG) and Roma populations.

Genes	SNPs	Allele	Adequate sample size for HG population	Adequate sample size for Roma population	Power	HG population (N=1497)	Roma population (N=616)	p value
Protective allele			N=			Frequencies (%)		
ABCA1	**rs4149268**	C	242	97	0.999	**62.19**	46.67	**<0.001**
APOB	rs693	A	99	39	0.999	46.66	44.32	0.103
CETP	rs1532624	A	1905	762	0.726	44.15	45.05	0.568
GALNT2	**rs2144300**	T	169	68	0.999	**58.48**	43.99	**<0.001**
HMGCR	**rs3846662**	G	127	51	0.999	**45.22**	38.72	**<0.001**
LIPC	rs10468017	T	642	257	0.999	25.85	25.81	0.849
LIPC	rs1077834	C	5321	2129	0.334	23.61	26.38	0.137
LIPC	**rs1532085**	A	839	336	0.972	34.50	**49.19**	**<0.001**
LIPC	**rs1800588**	T	142	57	0.999	21.44	**24.68**	**0.021**
LIPC	**rs2070895**	A	130	52	0.999	22.14	**24.76**	**0.049**
LIPC	rs4775041	C	844	338	0.971	27.19	26.79	0.771
LIPG	**rs2000813**	T	2409	964	0.625	31.00	**38.64**	**<0.001**
LIPG	**rs4939883**	C	917	367	0.959	83.97	**89.45**	**<0.001**
LPL	**rs328**	G	30050	12020	0.094	10.12	**13.88**	**<0.001**
Susceptibility allele			N=		Power	Frequencies (%)		p value
CETP	**rs5882**	A	4044	1618	0.420	**69.17**	57.06	**<0.001**
CETP	rs708272	G	1359	544	0.859	56.35	53.98	0.155
CETP	**rs7499892**	T	43129	17252	0.078	15.33	**23.46**	**<0.001**
CETP	**rs9989419**	A	66209	26484	0.064	**37.54**	33.93	**0.023**
GALNT2	**rs4846914**	G	153	61	0.999	40.18	**54.55**	**<0.001**
KCTD10	**rs2338104**	C	125	50	0.999	39.35	**49.27**	**<0.001**
WWOX	**rs2548861**	T	188	75	0.999	38.24	**46.10**	**<0.001**

3.4. Comparison of individual effects (size and direction) of SNPs on HDL-C level between Hungarian general and Roma populations

Eleven SNPs (located in the *ABCA1*, *CETP*, *LIPC*, *LPL* and *KCTD10* genes) were associated significantly with HDL-C level in the HG population, while only 4 SNPs (in the *CETP* and *LIPC* genes) showed a significant effect on HDL-C level among Roma. Three of these 4 SNPs (rs1532624, rs708272 and rs7499892) are identical to those we identified in the *CETP* gene in the HG population (rs1532624, rs708272, rs7499892 and rs9989419), while rs2000813 in the *LIPC* gene showed a nominally significant association with HDL-C level only in the Roma population. More details on the adjusted (age, sex, and BMI) linear associations between the individual SNPs and HDL-C levels are shown separately for the study populations in Table 4. As opposed to this, 8 of the 11 SNPs were found to be associated with HDL-C level only in the Hungarian general population. Six of these genetic variants - the *LIPC* (rs10468017, rs1077834, rs1800588 and rs2070895), *ABCA1* (rs4149268) and *KCDT10* (rs2338104) genes - showed only a nominally significant association with HDL-C level. In general, significant associations were more likely to be observed among the Hungarian general participants (with European origin) than among the Roma participants (with South Asian origin). Power analyses in the Roma group demonstrated decreased statistical power to detect the reported effects, which were largely attributable to its lower sample size (N_{HG}=1497 vs. N_{HR}=616).

The associations of the rs1532624, rs708272 and rs7499892 variants in CETP were consistent both in magnitude and direction, while the effect of three SNPs (rs693 in APOB, rs2000813 in LIPG, and rs2548861 in WWOX) were opposite in direction between the Hungarian general and Roma populations. Stata software provides the ability

to test for a difference in association with a quantitative trait between study populations by comparing the two regression coefficients. In the comparison analyses of SNP effects (measured as β values in the regression analyses) between the two groups, the p-values for heterogeneity ranged between 0.092 and 0.964 for most of the SNPs (18 of the 21 examined), and only nominally significant differences, ranging between 0.024 and 0.032, were found only for three SNPs (rs693 in APOB, rs2548861 in *WWOX* and rs9989419 in *CETP)***.**

Table 4. The effect alleles and association measures (β-value) of SNPs obtained from the literature and the results of linear regression analyses (adjusted by age, sex and BMI) for the Hungarian general and Roma populations

SNP (rs number)	Gene (short)	Effect allele	Beta values (mmol/l) from European populations	Hungarian general N=1497				Hungarian Roma N=616				p-value*
				β-value	95% CI	p-value	Power	β-value	95% CI	p-value	Power	
rs693	*APOB*	A	NA	-0.024	-0.050 - 0.001	0.063	0.690	0.030	-0.011 - 0.071	0.154	0.512	0.027
rs4149268	***ABCA1***	C	0.021 – 0.039	**0.030**	**0.004 - 0.056**	**0.027**	0.591	0.024	-0.015 - 0.064	0.225	0.221	0.828
rs1532624	***CETP***	C	0.075 – 0.122	**0.060**	**0.035 - 0.085**	**<0.001**	1.000	**0.075**	**0.033 - 0.117**	**<0.001**	0.999	0.535
rs5882	*CETP*	A	-0.034 – -0.016	-0.027	-0.055 - 0.001	0.052	0.857	-0.004	-0.044 - 0.036	0.854	0.058	0.336
rs708272	***CETP***	A	0.062 – 0.080	**0.066**	**0.041 - 0.091**	**<0.001**	1.000	**0.075**	**0.034 - 0.117**	**<0.001**	0.999	0.708
rs7499892	***CETP***	T	-0.177 – -0.078	**-0.123**	**-0.157 - -0.089**	**<0.001**	1.000	**-0.085**	**-0.133 - -0.038**	**<0.001**	0.961	0.190
rs9989419	***CETP***	A	-0.073 – -0.042	**-0.065**	**-0.091 - -0.039**	**<0.001**	0.999	-0.011	-0.053 - 0.032	0.621	0.086	**0.024**
rs2000813	***LIPG***	T	0.029 - 0.031	-0.001	-0.028 - 0.026	0.947	0.051	**0.041**	**0.001 - 0.080**	**0.044**	0.801	0.092
rs4939883	*LIPG*	C	0.036 – 0.048	0.015	-0.020 - 0.050	0.398	0.227	0.022	-0.044 - 0.089	0.511	0.260	0.838
rs10468017	***LIPC***	T	0.036 - 0.046	**0.034**	**0.005 - 0.062**	**0.022**	0.959	0.005	-0.040 - 0.050	0.836	0.064	0.296
rs1077834	***LIPC***	T	-0.043	**-0.032**	**-0.062 - -0.002**	**0.040**	0.519	-0.041	-0.089 - 0.007	0.097	0.432	0.765
rs1532085	*LIPC*	A	0.038 – 0.047	0.023	-0.003 - 0.049	0.087	0.710	0.024	-0.017 - 0.065	0.245	0.366	0.964
rs1800588	***LIPC***	T	0.036 – 0.044	**0.033**	**0.003 - 0.063**	**0.034**	0.957	0.039	-0.008 - 0.086	0.105	0.803	0.845
rs2070895	***LIPC***	A	0.042 – 0.054	**0.033**	**0.003 - 0.063**	**0.030**	0.960	0.039	-0.007 - 0.086	0.098	0.818	0.835
rs4775041	*LIPC*	C	0.034 – 0.036	0.020	-0.008 - 0.048	0.162	0.623	0.011	-0.034 - 0.055	0.638	0.126	0.731
rs3846662	*HMGCR*	G	NA	0.017	-0.009 - 0.043	0.199	0.403	0.023	-0.019 - 0.064	0.283	0.343	0.816
rs328	***LPL***	G	0.058 – 0.098	**0.074**	**0.033 - 0.116**	**<0.001**	1.000	0.056	-0.002 - 0.115	0.058	0.989	0.626
rs2144300	*GALNT2*	T	0.029 – 0.048	0.021	-0.005 - 0.047	0.117	0.341	0.014	-0.026 - 0.053	0.500	0.103	0.763
rs4846914	*GALNT2*	G	-0.030 – -0.016	-0.019	-0.046 - 0.007	0.148	0.305	-0.014	-0.053 - 0.026	0.504	0.102	0.807
rs2338104	***KCTD10***	C	-0.026 – -0.012	**-0.031**	**-0.056 - -0.006**	**0.016**	0.811	-0.036	-0.076 - 0.005	0.084	0.670	0.853
rs2548861	*WWOX*	T	NA	0.017	-0.009 - 0.043	0.197	0.337	-0.033	-0.074 - 0.007	0.106	0.574	0.032

Results in **bold** indicate significant associations between the SNP and HDL-C within a population.

*p-value for the differences in association with the quantitative traits among the study populations; values less than 0.025 can be considered as significant

NA: no available data

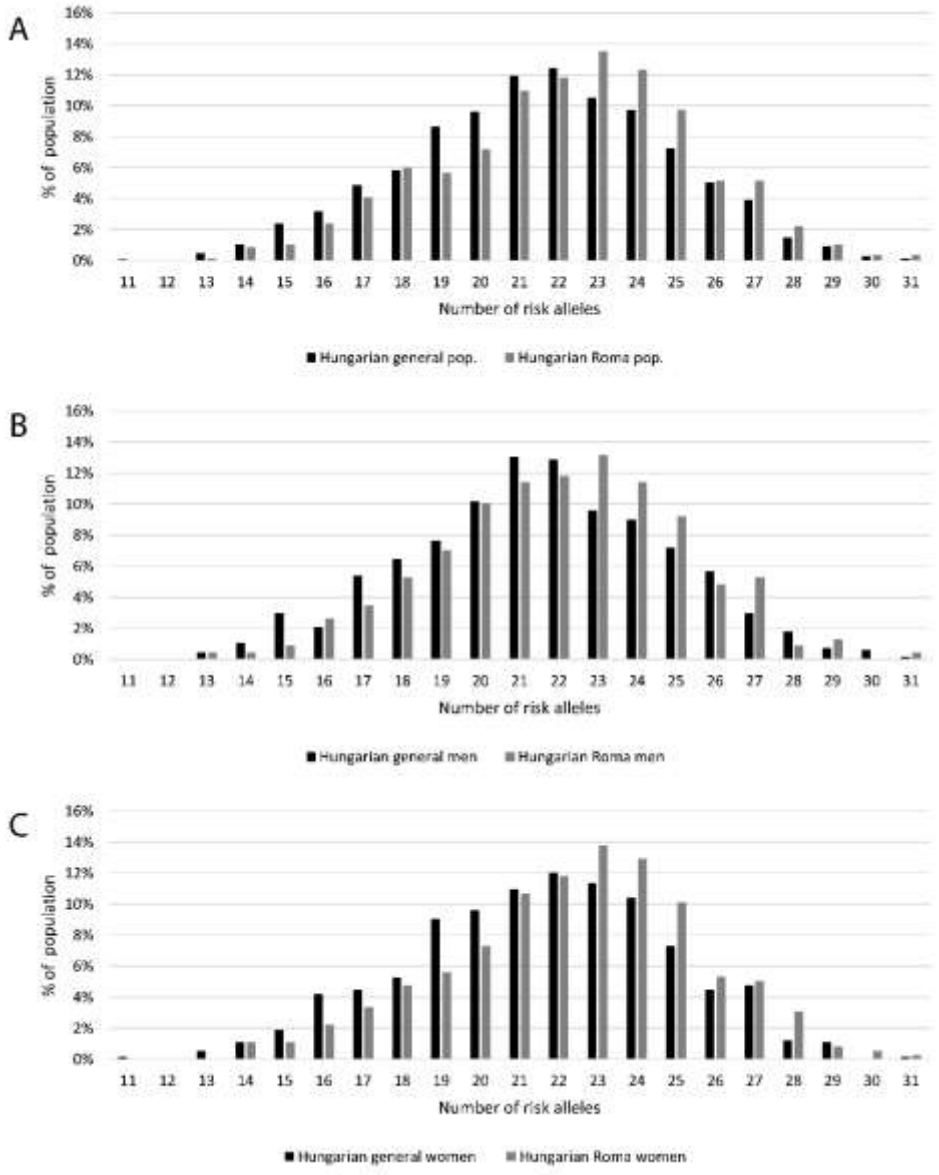
3.5. Comparison of our results with data obtained in other studies on European populations

The effect direction and size, expressed as β-values, of the SNPs obtained from our regression analyses were compared to the effect size measures from literature data on European populations (for more details, see Table 4). For the SNPs we examined, there was no measurable difference in the effect size and direction when compared with the literature data in any of the study populations.

3.6. Comparison of genetic risk scores between study populations

The unweighted GRS calculated on the basis of 21 SNPs ranged from 11 to 31 in the Hungarian general population and from 13 to 31 among the Roma. GRS data showed normal distribution in both study populations. The mean value of the score was 21.5±3.3 in the HG population and 22.2±3.2 in the Roma group. The distribution of GRS in case of the Roma population shows a strong shift to the right (higher risk) compared to the Hungarian general one and a strong, significant difference ($p<0.001$) was measured between the two groups using two-sided t-test (for more details, see Figure 7. A). The results showed a similar pattern when examined by gender (for more details, see Figure 7. B, and C), where individuals belonging to the Roma population in both sexes have, on average, a higher genetic burden than those belonging to the Hungarian general population (men: GRS_{HR} = 22.1±3.3 vs. GRS_{HG} = 21.5±3.1, $p<0.050$; women: GRS_{HR} = 22.3±3.4 vs. GRS_{HG} = 21.6±3.2, $p<0.001$; for more details, see Figure 7. B, and C).

Figure 7. Distribution of unweighted genetic risk scores based on 21 single nucleotide polymorphisms by study populations (A) and separately for men (B) and women (C) in both populations.



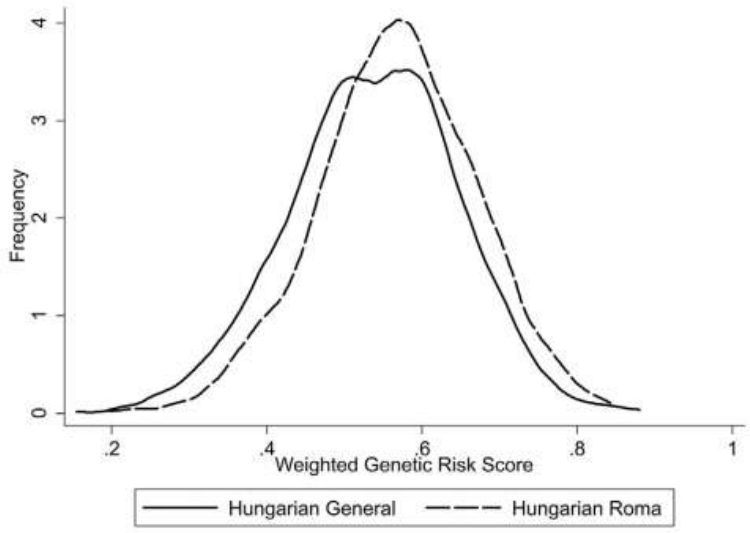
To calculate the weighted GRS, it is essential to know the individual effect of SNPs on HDL-C level. For more details on the estimated impact of each SNP based on literature data, see Table 5 [56, 57, 61, 69, 133-135, 143].

Table 5. SNPs considered in the weighted genetic risk scores, their genes, effect alleles and weighting numbers adopted from original publications

SNPs	Genes	Effect allele	Weighting number	Publications for effect alleles
rs4149268	*ABCA1*	C	0.039	[30]
rs5882	*CETP*	A	-0.016	[35]
rs708272	*CETP*	G	-0.062	[33]
rs7499892	*CETP*	T	-0.078	[32]
rs9989419	*CETP*	A	-0.042	[30]
rs2144300	*GALNT2*	T	0.048	[34]
rs2338104	*KCTD10*	C	-0.048	[30]
rs1077834	*LIPC*	C	0.043	[29]
rs1532085	*LIPC*	A	0.038	[31]
rs4775041	*LIPC*	C	0.036	[30]
rs2000813	*LIPG*	T	0.030	[29]
rs4939883	*LIPG*	C	0.039	[30]
rs328	*LPL*	G	0.057	[32]

The wGRS data were normally distributed in both study populations. The average wGRS value was 0.53±0.1 in the Hungarian general population, while this value was significantly higher in case of the Roma (0.57±0.1, $p<0.001$). For more details on the distribution curves of wGRS by study populations, see Figure 8.

Figure 8. Distribution curves of weighted genetic risk scores for the Hungarian general (HG) and Hungarian Roma (HR) populations. The distribution of wGRS in the HR population (dashed line) is right-shifted compared to the HG one (continuous line).



Both study populations were classified into quintiles based on wGRS. The first quintile included the genetically least vulnerable (wGRS: 0.15 - ≤0.30), while the 5th quintile included the most vulnerable individuals (wGRS: 0.75 – 0.88). The genetic risk of individuals in the third quintile is considered to be average. After this, it was examined how the distribution of individuals evolved between the quintiles in both populations.

Only half a percent of the Roma population belonged to the least vulnerable group compared to 1.83% in the Hungarian general population (p=0.025). Nineteen percent of the Hungarian general population and only 11% of the Roma one belonged to the lower-than-average genetic risk group. The 5th highest risk quintile included 2.61% of the HG population while 5.14% of the HR one (p=0.004). For more details on the distribution of study populations according to quintiles of wGRS, see Table 6.

Table 6. Distribution of study populations by wGRS quintiles

	Hungarian general population (%)	**Roma population (%)**	**p-value**
1st quintile of wGRS (0.15 - ≤0.30)	1.83	0.51	0.025
2nd quintile of wGRS (0.31 - ≤0.45)	17.18	10.45	<0.001
3rd quintile of wGRS (0.46 - <0.59)	48.38	49.14	0.756
4th quintile of wGRS (0.6 - ≤ 0.74)	30.00	34.76	0.037
5th quintile of wGRS (0.75 – 0.88)	2.61	5.14	0.004

3.7. Association of genetic risk scores with plasma HDL-C level

Based on descriptive statistics (see Table 7), there is a clear trend for both the average level of HDL-C and the prevalence of reduced HDL-C by wGRS-based quintiles in both populations. Nevertheless, regardless of the quintiles, lower average HDL-C levels as well as higher representations of reduced HDL-C levels were more common in the Roma population. For more details, see Table 7.

Table 7. The average HDL-C levels (in mmol/L) and proportion (in %) of subjects with reduced plasma HDL-C levels in the Hungarian general and Roma populations according to wGRS-based.

	1st quintile of wGRS (0.15-≤0.30)	**2nd quintile of wGRS (0.31-≤0.45)**	**3rd quintile of wGRS (0.46-<0.59)**	**4th quintile of wGRS (0.6 -≤0.74)**	**5th quintile of wGRS (0.75 – 0.88)**	
Hungarian general (Men;Women)	**N=26 (8;18)**	**N=241 (115;126)**	**N=681 (320;361)**	**N=417 (200;217)**	**N=36 (21;15)**	**p-values for trend**
Average HDL-C levels (mmol/L)	1.56	1.47	1.41	1.38	1.33	0.008
Prevalence of reduced HDL-C levels	19.23	24.17	29.25	29.98	35.29	0.072
Hungarian Roma (Men;Women)	**N=3 (1;2)**	**N=61 (28;33)**	**N=287 (112;175)**	**N=203 (76;127)**	**N=30 (11;19)**	**p-values for trend**
Average HDL-C levels (mmol/L)	1.26	1.24	1.23	1.2	1.09	0.067
Prevalence of reduced HDL-C levels (in %) quintiles	33.33	45.90	50.88	57.71	63.33	0.024

The unweighted GRS showed a significant correlation with HDL-C level in the unadjusted models in case of the study populations, where a one-unit rise in the GRS increased the risk of reduced plasma HDL-C levels (OR_{HG}=1.065, p=0.001; OR_{HR}=1.080, p=0.004). These results remained unchanged even after a correction just for age and sex (OR_{HG}=1.066, p<0.001; OR_{HR}=1.076, p=0.007) or for all relevant covariates (OR_{HG}=1.075, p<0.001; OR_{HR}=1.075, p=0.012) in linear regression analyses.

The weighted GRS showed a significant association with the outcome in both populations (OR_{HG}=3.411, p=0.029; OR_{HR}=10.385, p=0.005), and this correlation remained even after a correction for age and sex in logistic regression models (OR_{HG}=3.665, p=0.021; OR_{HR}=8.866, p=0.009). After correction for all relevant covariates, a similarly strong association between wGRS and reduced HDL-C cholesterol levels was measured in both populations (OR_{HG}=4.331, p=0.015; OR_{HR}=9.303, p=0.012).

In case of unweighted GRS, the estimated effect per allele was similar in both populations (about 6-7%), while in case of weighted GRS, a one-unit change showed a greater effect in the Roma population than in the Hungarian general one regardless of the number of covariates (OR_{HG}=3.411 vs. OR_{HR}=10.382; OR_{HG}=3.665 vs. OR_{HR}=8.866; OR_{HG}=4.331 vs. OR_{HR}=9.303). For more details on the association of GRSs and wGRSs with reduced plasma HDL-C levels see Table 8.

Table 8. Association of GRSs and wGRSs with reduced plasma HDL-C levels by study group.

	Hungarian general		**Roma**	
GRS	**OR (95% CI)**	**p value**	**OR (95% CI)**	**p value**
Unadjusted	1.065 (1.028 - 1.104)	0.001	1.080 (1.025 - 1.137)	0.004
Adjusted for age and sex	1.066 (1.028 - 1.105)	<0.001	1.076 (1.021 - 1.134)	0.007
Adjusted for all relevant covariates	1.075 (1.035 - 1.118)	<0.001	1.075 (1.016 - 1.137)	0.012
wGRS	**OR (95% CI)**	**p value**	**OR (95% CI)**	**p value**
Unadjusted	3.411 (1.132 - 10.274)	0.029	10.382 (2.045 - 52.713)	0.005
Adjusted for age and sex	3.665 (1.211 - 11.085)	0.021	8.866 (1.705 - 46.095)	0.009
Adjusted for all relevant covariates	4.331 (1.328 - 14.126)	0.015	9.303 (1.633 - 53.010)	0.012

The association of GRS and wGRS with reduced plasma HDL-C level (as binary outcome) were evaluated by unadjusted regression models and by regression models adjusted for relevant covariates (sex, age, body mass index, systolic and diastolic blood pressure, fasting glucose level, as well as antihypertensive, antidiabetic and lipid-lowering treatments) separately in Roma and HG subjects. In all models the HDL-C was the dependent variable, and the GRS/wGRS were the independent variables.

Similar associations of unweighted and weighted GRS with plasma HDL-C level (as a continuous variable) were observed in both study populations. Similar to the results described above, in this case, the change of one unit in the Roma population had a greater effect on the development of HDL-C levels in both GRS and wGRS. For more details on the effect of GRSs on HDL-C levels see Table 9.

Table 9. Association of GRS and wGRAS with plasma HDL-C level by study groups.

	Hungarian general		Roma	
GRS	**β (95% CI)**	**p value**	**β (95% CI)**	**p value**
Unadjusted	-0.011 (-0.017 - -0.004)	0.001	-0.013 (-0.022 - -0.003)	0.010
Adjusted for age and sex	-0.011 (-0.017 - -0.005)	<0.001	-0.013 (-0.023 - -0.004)	0.006
Adjusted for all relevant covariates	-0.011 (-0.017 - -0.006)	<0.001	-0.013 (-0.022 - -0.004)	0.005
wGRS	**β (95% CI)**	**p value**	**β (95% CI)**	**p value**
Unadjusted	-0.302 (-0.494 - -0.109)	0.002	-0.356 (-.655 - -0.057)	0.020
Adjusted for age and sex	-0.276 (-0.458 - -0.095)	0.003	-0.392 (-.686 - -0.097)	0.009
Adjusted for all relevant covariates	-0.295 (-0.465 - -0.126)	0.001	-0.358 (-0.637 - -0.080)	0.012

The association of GRS and wGRS with plasma HDL-C level (as binary outcome) were evaluated by unadjusted regression models and by regression models adjusted for relevant covariates (sex, age, body mass index, systolic and diastolic blood pressure, fasting glucose level, as well as antihypertensive, antidiabetic and lipid-lowering treatments) separately in Roma and HG subjects. In all models the HDL-C was the dependent variable, and the GRS/wGRS were the independent variables.

The effect of GRS and wGRS on reduced plasma HDL-C level remained significant when the two study populations were analysed together, i.e. ethnicity was used as a covariate in the regression models (GRS: OR=1.076, 95% CI: 1.042 - 1.111, p<0.001; wGRS: OR=6.087, 95% CI: 2.306 - 16.071, p<0.001).

In addition to GRSs, we were able to identify a number of factors that influence the development of reduced HDL-C levels in both populations. The risk-increasing factors were Roma ethnicity, female gender, increasing BMI, and increasing glucose levels. Older age has a protective effect. For more details on effect of GRSs and relevant covariates (sex, age, body mass index, systolic and diastolic blood pressure, fasting glucose level, as well as antihypertensive, antidiabetic, and lipid-lowering treatments) see Table 10.

Table 10. The association of reduced HDL-C level unweighted (A) and weighted (B) genetic risk scores adjusted for relevant covariates (sex, age, body mass index, systolic and diastolic blood pressure, fasting glucose level, as well as antihypertensive, antidiabetic, and lipid-lowering treatments)

A - Dependent variable: reduced plasma HDL-C level

Independent variables	OR (95% CI)	p-value
Ethnicity (HG as reference was used)	2.805 (2.215 - 3.552)	<0.001
Sex (male as reference was used)	1.533 (1.238 - 1.898)	0.006
Age	0.986 (0.976 - 0.996)	<0.001
BMI	1.119 (1.095 - 1.144)	<0.001
Systolic blood pressure	0.994 (0.984 - 1.004)	0.213
Diastolic blood pressure	0.993 (0.976 - 1.009)	0.384
Fasting glucose level	1.148 (1.031 - 1.278)	0.012
Lipid-lowering treatment	0.985 (0.702 - 1.382)	0.931
Antidiabetic treatment	1.254 (0.947 - 1.660)	0.115
Antihypertensive treatment	1.293 (0.755 - 2.213)	0.349
GRS	1.076 (1.042 - 1.111)	<0.001

B - Dependent variable: reduced plasma HDL-C level

Independent variables	OR (95% CI)	p-value
Ethnicity (HG as reference was used)	2.766 (2.184 - 3.504)	<0.001
Sex (male as reference was used)	1.547 (1.250 - 1.915)	<0.001
Age	0.986 (0.976 - 0.996)	0.007
BMI	1.118 (1.094 - 1.142)	<0.001
Systolic blood pressure	0.994 (0.984 - 1.003)	0.206
Diastolic blood pressure	0.993 (0.976 - 1.009)	0.382
Fasting glucose level	1.160 (1.042 - 1.291)	0.007
Lipid-lowering treatment	0.980 (0.699 - 1.375)	0.908
Antidiabetic treatment	1.262 (0.953 - 1.671)	0.104
Antihypertensive treatment	1.307 (0.763 - 2.238)	0.329
wGRS	6.087 (2.306 - 16.071)	<0.001

Regardless of the covariates used, both unweighted and weighted GRS showed a strong significant correlation with HDL-c levels (GRS: β=-0.012, $p<0.001$; wGRS: β=-0.332, $p<0.001$). In addition to the covariates affecting HDL-C levels described above, antidiabetic treatment also had a significant reducing effect on HDL-C levels.

Table 11. The association of HDL-C levels with unweighted (A) and weighted (B) genetic risk scores adjusted for relevant covariates (sex, age, body mass index, systolic and diastolic blood pressure, fasting glucose level, as well as antihypertensive, antidiabetic, and lipid-lowering treatments)

A - Dependent variable: plasma HDL-C level

Independent variables	β (95% CI)	p-value
Ethnicity (HG as reference was used)	-0.192 (-0.228 - -0.156)	<0.001
Sex (male as reference was used)	0.213 (0.181 - 0.245)	<0.001
Age	0.003 (0.002 - 0.005)	<0.001
BMI	-0.021 (-0.024 - -0.018)	<0.001
Systolic blood pressure	0.001 (0.000 - 0.002)	0.157
Diastolic blood pressure	0.003 (0.000 - 0.005)	0.044
Fasting glucose level	-0.027 (-0.043 - -0.011)	0.001
Lipid-lowering treatment	-0.003 (-0.055 - 0.049)	0.902
Antidiabetic treatment	-0.047 (-0.089 - -0.004)	0.032
Antihypertensive treatment	-0.074 (-0.158 - 0.009)	0.082
GRS	-0.012 (-0.017 - -0.007)	<0.001

B - Dependent variable: plasma HDL-C level

Independent variables	β (95% CI)	p-value
Ethnicity (HG as reference was used)	-0.189 (-0.226 - -0.153)	<0.001
Sex (male as reference was used)	0.211 (0.179 - 0.243)	<0.001
Age	0.003 (0.002 - 0.005)	<0.001
BMI	-0.021 (-0.024 - -0.018)	<0.001
Systolic blood pressure	0.001 (0.000 - 0.003)	0.148
Diastolic blood pressure	0.003 (0.000 - 0.005)	0.042
Fasting glucose level	-0.029 (-0.045 - -0.013)	<0.001
Lipid-lowering treatment	-0.002 (-0.054 - 0.050)	0.932
Antidiabetic treatment	-0.048 (-0.091 - -0.006)	0.026
Antihypertensive treatment	-0.075 (-0.159 - 0.009)	0.079
wGRS	-0.332 (-0.476 - -.0187)	<0.001

3.8. Haplotypes in the CETP and LIPC genes and their frequencies in the Hungarian general and Roma populations

Haplotype analysis involved different combinations of the 5 SNPs in the *CETP* (rs1532624, rs5882, rs708272, rs7499892 and rs9989419) and the 6 SNPs in the *LIPC* gene (rs10468017, rs1077834, rs1532085, rs1800588, rs2070895, rs4775041). For the results of structure analysis of linkage disequilibrium, see Figure 5.

We identified 10 haplotype blocks in the *CETP* and six in *LIPC* genes, the prevalence of which had been higher than 1% in the combined population (HR and HG together). A total 8 out of 10 in the haplotypes blocks in case of *CETP* (H1–H5 and H8–H10) and 4 out of 6 in *LIPC* (H1 and H4–H6) showed a significant difference in prevalence between the study populations (for more details, see Table 12 and 13). The $H8_{CETP}$ occurs almost exclusively in the Roma population (HR: 7.28% vs. HG: 0.14%; $p < 0.001$).

Table 12. The frequency of haplotypes in the *CETP* gene in the combined sample, as well as in the Roma (HR) and Hungarian general (HG) populations.

Haplotype	rs1532624	rs5882	rs708272	rs7499892	rs9989419	Frequency in the Combined Population	Frequency in HG Population	Frequency in HR Population	p-value
H1	**A**	**G**	**A**	**C**	**G**	20.57%	17.32%	**28.29%**	**<0.001**
H2	**A**	**A**	**A**	**C**	**G**	19.55%	**21.78%**	14.24%	**<0.001**
H3	**C**	**A**	**G**	**C**	**A**	13.98%	**15.05%**	11.45%	**0.015**
H4	**C**	**A**	**G**	**C**	**G**	13.69%	**14.79%**	11.09%	**0.015**
H5	**C**	**A**	**G**	**T**	**A**	12.63%	11.60%	**15.07%**	**0.019**
H6	C	G	G	C	A	5.55%	5.95%	4.60%	0.170
H7	A	A	A	C	A	2.77%	2.92%	2.43%	0.467
H8	**C**	**G**	**G**	**T**	**G**	2.26%	0.14%	**7.28%**	**<0.001**
H9	**C**	**G**	**G**	**C**	**G**	2.60%	**3.04%**	1.56%	**0.039**
H10	**C**	**A**	**G**	**T**	**G**	2.43%	**2.86%**	1.41%	**0.038**

Haplotypes with a significantly different frequency ($p < 0.05$) in the two populations and their frequency with the higher values are highlighted in bold.

Table 13. The frequency of haplotypes in the *LIPC* gene in the combined sample, as well as in Roma (HR) and Hungarian general (HG) populations.

Haplotype	rs10468017	rs1077834	rs1532085	rs1800588	rs2070895	rs4775041	Frequency in the Combined Population	Frequency in HG Population	Frequency in HR Population	p-value
H1	**C**	**T**	**G**	**C**	**G**	**G**	44.41%	**48.38%**	34.97%	**<0.001**
H2	T	T	A	C	G	C	21.08%	20.40%	22.70%	0.073
H3	C	C	G	T	A	G	13.81%	13.63%	14.25%	0.470
H4	**C**	**T**	**A**	**C**	**G**	**G**	10.20%	7.23%	**17.24%**	**<0.001**
H5	**T**	**C**	**A**	**C**	**G**	**G**	4.69%	**5.48%**	2.81%	**0.006**
H6	**C**	**C**	**A**	**T**	**A**	**G**	3.43%	2.08%	**6.63%**	**<0.001**

Haplotypes with significantly different frequency ($p < 0.05$) in the two populations and their frequency with the higher value are highlighted in bold.

3.9. Association of haplotypes in CETP and LIPC genes with HDL-C level in the combined study population

The most prevalent haplotype of the investigated two genes in the combined population ($H1_{CETP}$: AGACG and $H1_{LIPC}$: CTGCGG) was used as a reference for the comparative analysis on the relationship of haplotypes with HDL-C level.

$H3_{CETP}$ (β=−0.05, p=0.016 and OR=1.34, p=0.040) and $H8_{CETP}$ (β=−0.14 p=0.001 and OR=2.60, p=0.002) have at least a nominally significant lipid-lowering effect on the outcome. The prevalence of $H3_{CETP}$ in the Hungarian general population (HG: 15.05% vs. HR: 11.45%, p=0.015), and the prevalence of $H8_{CETP}$ in the Roma population (HG: 0.14% vs. HR: 7.28%, p<0.001) were found to be significantly higher in comparison with each other. For more details on the prevalence of haplotypes in the *CETP* gene, see Table 12, and on the effect of them on HDL-C level, see Table 14.

Table 14. The effect of haplotypes in the *CETP* gene on high-density lipoprotein cholesterol (HDL-C) levels in the combined study population (Hungarian Roma and Hungarian general together). The association was evaluated under adjusted models (ethnicity, sex, age, body mass index, systolic and diastolic blood pressure, fasting glucose level, as well as antihypertensive, antidiabetic and lipid-lowering treatments).

Haplotypes	β (95% CI)	p-value	OR (95% CI)	p-value
H1	reference	---	reference	---
H2	0.02 (−0.01 – 0.06)	0.220	0.88 (0.67 – 1.16)	0.370
H3	**−0.05 (−0.09 – −0.01)**	**0.016***	**1.34 (1.01 – 1.76)**	**0.040***
H4	−0.01 (−0.06 – 0.03)	0.470	0.91 (0.68 – 1.20)	0.490
H5	**−0.11 (−0.16 – −0.07)**	**<0.001****	**1.74 (1.32 – 2.30)**	**<0.001****
H6	0.01 (−0.06 – 0.07)	0.880	0.86 (0.55 – 1.34)	0.500
H7	0.04 (−0.05 – 0.13)	0.370	0.93 (0.50 – 1.72)	0.810
H8	**−0.14 (−0.22 – −0.06)**	**0.001****	**2.60 (1.43 – 4.72)**	**0.002****
H9	−0.07 (−0.17 – 0.03)	0.140	1.04 (0.50 – 2.13)	0.920
H10	−0.07 (−0.16 – 0.02)	0.130	1.80 (0.98 – 3.30)	0.058

At least nominally significant associations between haplotypes and lipid levels (cut-off for HDL-C: < 1.03 mmol/L in male and < 1.29 mmol/L in female) are highlighted in bold. * Significant p-values without Bonferroni correction. ** Significant p-values with Bonferroni correction.

The $H2_{LIPC}$ (β=0.05, p=0.003 and OR = 0.74, p=0.006) and $H3_{LIPC}$ (β=0.07, p=0.001) have a significant effect on HDL-C level as continuous and binary outcome, and their prevalence did not differ significantly between the study groups. $H5_{LIPC}$ (β=0.09, p=0.004; OR=0.50, p=0.005) was significantly associated with HDL-C level, and its prevalence was significantly higher in the HG population (HG: 5.48% vs. HR: 2.81%, p = 0.006). For more details on the prevalence of haplotypes in the *LIPC* gene, see Table 13, and on the effect of them on HDL-C levels, see Table 15.

Table 15. The effect of haplotypes in the *LIPC* gene on high-density lipoprotein cholesterol (HDL-C) in the combined study population (Hungarian Roma and Hungarian general together). The association was evaluated under adjusted models (ethnicity, sex, age, body mass index, systolic and diastolic blood pressure, fasting glucose level, as well as antihypertensive, antidiabetic and lipid-lowering treatments).

Haplotypes	β (95% CI)	p-value	OR (95% CI)	p-value
H1	reference	---	reference	---
H2	**0.05 (0.01–0.08)**	**0.004** **	**0.75 (0.60–0.92)**	**0.007** **
H3	**0.07 (0.03–0.11)**	**0.001** **	0.78 (0.59–1.03)	0.079
H4	0.03 (−0.01–0.07)	0.160	0.97 (0.73–1.29)	0.840
H5	**0.09 (0.03–0.15)**	**0.005** **	**0.50 (0.31–0.81)**	**0.005** **
H6	0.05 (−0.02–0.13)	0.170	0.91 (0.56–1.46)	0.680

At least nominally significant associations between haplotypes and lipid levels (cut-off for HDL-C: < 1.03 mmol/L in male and < 1.29 mmol/L in female) are highlighted in bold. * Significant *p*-value without Bonferroni correction. ** Significant *p*-values with Bonferroni correction.

4. Discussion

Low HDL-C level is one of the most important predictors of cardiometabolic diseases, especially atherosclerosis. It has long been a well-known epidemiological fact that an individual's HDL-C level is highly dependent on his or her age, sex, and lifestyle factors (characteristics of diet, level and type of physical activity, smoking, and alcohol consumption) [144]. Furthermore, a low socio-economic situation has also been shown to predict adverse changes in a number of cardiometabolic factors [145].

In addition to all these life-style related factors, the degree of genetic sustainability, as evidenced by numerous family and twin studies, must also be taken into consideration [27]. Candidate gene studies have identified a number of genes that encode proteins that are an integral part of HDL-C metabolism and, as a result, they can potentially effect an individual's plasma HDL-C concentrations [107, 146]. GWAS studies have provided evidence for the polygenic inheritance of dyslipidaemias, which means that many genes and the SNPs which they contain have combined effect on lipid traits and consequently on lipid levels in the blood [61, 71, 147]. Several studies have estimated the age- and sex-adjusted inheritance of reduced HDL-C level, including the Tehran Lipid and Glucose Study, a large-scale epidemiological investigation, which estimated the weight of genetic factors on lipid metabolism at 40% in general; among lipid components the genetic determination of the HDL-C level is the highest (46%) [27]. Reviews summarizing recent reports on the genetic background of lipid metabolism report more than 50 genes that are directly responsible for an individual's HDL cholesterol level [148] and plenty of SNPs are showed to be associated with HDL-C metabolism [149].

Significantly lower serum HDL-C levels were reported in the Roma population (independently from sex) compared to the general one [17, 150-154], and reduced HDL-

C levels were also found to be a lot more frequent among Roma children and adolescents as well [155].

These facts, together with the findings of our previous study which showed that average HDL-C level was significantly lower and the prevalence of reduced HDL cholesterol level was higher in all age groups of the Hungarian Roma population than in the corresponding age groups of the Hungarian general one [24] strongly suggest that in the Roma there are genetic factors behind this phenomenon.

The present study is the first one that examined in detail the existence of genetic causes underlying the high prevalence of the low HDL-C level among Roma. The SNPs in the current study are located in genes encoding proteins involved in HDL-C metabolism including lipid transfer. Twenty-one polymorphisms associated with HDL-C level were genotyped, and differences in the prevalence of 15 SNPs proved to be significant when the two study groups were compared with each other. The susceptibility alleles of rs5882 and rs9989419 (in the *CETP* gene) were more frequent among participants representing the Hungarian general population, whereas rs7499892 (in the *CETP* gene), rs4846914 (in the *GALNT2* gene), rs2338104 (in the *KCTD10* gene) and rs2548861 (in the *WWOX* gene) showed significantly higher frequency in the Roma population. Among the protective (i.e. with a HDL-C increasing effect) variants, the rs4149268 (in the *ABCA1* gene), rs2144300 (in the *GALNT2* gene) and rs3846662 (in the *HMGCR* gene) polymorphisms were observed more frequently in the HG population; while the alleles of rs1532085, rs1800588, and rs2070895 (in the *LIPC* gene), rs2000813 and rs4939883 (in the *LIPG* gene), and rs328 (in the *LPL* gene) showed significantly higher prevalence in the HR one.

The effect of SNPs, as estimated in independent studies, was validated and utilized to model and compare the genetic risk related to reduced plasma HDL-C level in the Hungarian Roma and general populations [156]. The individual effect of all twenty-one SNPs in 10 genes closely related to plasma HDL-C level was evaluated in both study populations. Out of the 21 SNPs, 12 showed a significant association with HDL-C level in at least one of the study populations - eight SNPs in the Hungarian general population and one among Roma, and three in both populations. All the three common SNPs (rs1532624, rs708272, and rs7499892) are located in the *CETP* gene. There was no significant difference between literature data and our results in any of the study populations. Thus, the studied SNPs are suitable for calculating genetic risk scores for both populations.

The genetic risk score was calculated in both populations in unweighted and weighted forms. For the unweighted GRS, all 21 SNPs were used, while only 13 were used to calculate the weighted GRS. In case of both unweighted (HR: 22.2 vs. HG: 21.5, $p<0.001$) and weighted (HR: 0.57 vs. HG: 0.53, $p<0.001$) GRS, the Roma population carried a significantly higher genetic burden compared to the Hungarian general one. Furthermore, it was demonstrated that both GRSs showed a significant association with HDL-C level regardless of the effect of confounding factors [157, 158].

Most of the SNPs with a significant effect on HDL-C level are found in 2 genes, *CETP* and *LIPC*; thus, haplotype analysis was performed using SNPs in *CETP* and *LIPC* genes in the study [159]. Three haplotypes in the *CETP* gene (H3, H5, and H8) showed at least a nominally significant association with decreased HDL-C level and 3 in the *LIPC* gene (H2, H3, and H5) with elevated HDL-C level. Haplotype 8 in the *CETP* gene, which has the greatest reducing effect on HDL-C level ($\beta=-0.14$, $p<0.001$), was found almost exclusively in the Roma population (HR: 7.28% vs. HG: 0.14%). The 0.14%

representation of this haplotype in the HG population is so low, that the possibility that the two persons showed to have the haplotype 8 might belong to the Roma ethnicity (in the HG population selected at random the estimated number of Roma is about 140).

The results show that the majority of the susceptibility alleles were accumulated in the Roma population, and they override the slight predominance of protective alleles among them. The comparison of unweighted and weighted genetic risk score distributions showed that the Hungarian Roma population carry a greater load of risk alleles compared to the Hungarian general one. The haplotype analysis confirmed that lower HDL-C levels among the Roma can be partly explained by the special coexistence of polymorphisms in the *CETP* gene.

Concerning the role of CETP in lipid metabolism it was already shown a quarter of a century ago that CETP is responsible for all cholesteryl ester (CE) and triglyceride (TG) transfer activity in human plasma [160] and after inspiring animal studies it was well documented in human randomized trials that CETP inhibition results in significant elevation in HDL-C level combined with strong decrease of LDL-C and apolipoprotein B100 levels [161]. Although the results of phase III studies on CETP inhibitors have not yet lived up to expectations [162], it is no doubt that decreased CETP activity due to genetic alterations strongly contributes to the development of disturbances in HDL-C metabolism resulting in low HDL-C and consequently increased risk to atherosclerosis [163].

Regarding LIPC, its role – even the question whether it is either a protective or proatherogenic agent – is a subject of intense speculations and discussions, studies on SNPs in the gene have presented very variable, sometimes inconsistent findings (see reviewed [164]). Anyhow it is generally accepted that LIPC has a crucial role in

converting large, TG-rich HDL_2 into small, dense HDL_3 and is a negative regulator of HDL-C level [25]. On the effects of different SNPs in the LIPC gene further studies are needed.

An obvious limitation of this study is that although the majority of the Roma population has accumulated in the region in which the samples were collected, this study population cannot be interpreted as a representative sample for the whole Hungarian Roma one. According to the study design, it is also important to note that the representative sample of the Hungarian general population included some Roma people, so it is possible that their inclusion resulted in a slight underestimation of the differences between the two study populations. However, as many Roma individuals are reluctant to define their ethnicity as Roma, it would be very difficult to overcome this limitation. Exposure to epigenetic factors, rare or structural variants, gene-environment, and gene-gene interactions were not taken into consideration, although it is well known that all of them can modify the effect of genetic risk on any traits.

Our analyses were corrected only for the major covariates; however, a number of behavioural factors (e.g., physical inactivity and unhealthy diet) can modify susceptibility to reduced HDL-C level, and, consequently, may result in the underestimation of differences in plasma HDL-C level between the Hungarian general and Roma populations to a certain extent. Considering the high rate of consanguinity, it is assumed that there are a number of founding mutations among the Roma which may affect HDL-C level. The founder mutations related to diseases which have been identified among them follow a Mendelian inheritance pattern. One of these, the R148X homozygous nonsense mutation in the NDRG1 gene, is associated with lower HDL-C level in men in a single study with a very small sample size [165]. However, based on the strong primary founder effect in

this ethnicity, it cannot be ruled out that founder mutations associated with multifactorial diseases may also exist even though they have not been identified yet.

In conclusion, we demonstrated that the vast majority of the effects of targeted SNPs identified in European populations could be replicated in Hungarian general and Roma populations. Consequently, effect size measures obtained from the literature can be used for risk estimation not only in case of the Hungarian general population (European origin) but also in case of the Roma one (South Asian origin). It was strongly suggested that the Roma population has a genetic susceptibility for reduced HDL-C level. Further genetic susceptibility studies on HDL-C level can utilize the effect size of SNPs identified in European populations to estimate the weight of genetic factors on this trait among Roma populations.

This is the first study carried out to investigate the possible genetic background of the high prevalence of reduced HDL-C level among the Roma. The SNPs investigated include those that are mainly involved in HDL-C metabolism. By using DNA samples from study groups that are representative of the Hungarian general and the Roma populations, it was possible to estimate and compare the genetic risk. Although the genetic variants in our study could explain differences in plasma HDL-C level and are stronger predictors of reduced plasma HDL-C levels in the Roma population compared to the Hungarian general population, it is still important to emphasise that further genetic research on minority populations is required. At present, there is no sufficient support for the clinical application of gene-based prediction models in lipid disturbances, yet, there are promising signs pointing towards its future applicability [166]. We must accept that careful clinical trial programmes are needed in order to determine which HDL-raising therapeutic interventions may indeed exert protective effect [167].

5. Összefoglalás

Számos kutatás igazolta, hogy a roma populáció esetében a csökkent HDL-koleszterin (HDL-C) szint kortól és nemtől függetlenül gyakrabban fordul elő, mint a többségi lakosságnál. Ezt a jelenséget egymástól függetlenül több ország roma populációjában is leírták. A vér HDL-koleszterin szintje kb. 50%-ban genetikailag meghatározott, a fennmaradó 50%-on környezeti és életmódbeli tényezők osztoznak. Kutatásunk célja a romák körében az általános populáció esetében észleltnél szignifikánsan gyakrabban tapasztalható alacsony HDL-C szint hátterében álló genetikai fogékonyság igazolása és a genetikai háttér tisztázásához adatok szolgáltatása.

Huszonegy – irodalmi adatok alapján a HDL-C metabolizmussal összefüggésbe hozható - polimorfizmus genotipizálása történt meg 1497 a magyar általános populációból, valamint 616 a telepszerű körülmények között élő roma populációból származó mintán. Tizenöt SNP allélfrekvenciája esetében volt szignifikáns eltérés a két vizsgált populáció között. Az SNP-k hatáserőssége jelentősen nem különbözött a két populáció esetében és jelentősen nem tért el az irodalmi adatoktól sem. A súlyozatlan és súlyozott genetikai rizikópontszám a roma minta esetében jobbra tolódást mutatott (a nagyobb genetikai kockázat irányába) a magyar általános populációval összehasonlítva. Regressziós elemzéssel sikerült igazolnunk, hogy a genetikai rizikó pontszám szignifikánsan összefügg a HDL-koleszterin szinttel mindkét populáció esetében. Számos a *CETP* gént érintő előnytelen mutáció felhalmozódása volt igazolható a roma populáció esetében.

Öt a *CETP* és hat a *LIPC* génben található SNP-re alapozva a *CETP* gén esetében három a HDL-C szintet szignifikánsan csökkentő, míg a *LIPC* gén esetében szintén három a HDL-C szintet szignifikánsan növelő haplotípust is sikerült azonosítanunk. A *CETP* génben található nyolcas haplotípus amely a legerősebb HDL-C csökkentő hatással bír, szinte kizárólag a roma populációban volt fellelhető.

Az eredmények alapján a roma és magyar általános lakosság HDL-koleszterin szintjében tapasztalható különbségek hátterében a genetikai meghatározottság egyértelműen kijelenthető, de nem elhanyagolható a környezeti és életmódbeli tényezők hatása sem.

Mindkét populáció esetében a HDL-koleszterin szint növelését célzó beavatkozásoknak figyelembe kell vennie az egyének genetikai hátterét a konvencionális életmódbeli és környezeti rizikótényezők mellett.

6. Summary

Numerous studies have shown that reduced high-density lipoprotein cholesterol levels (HDL-C) are more common in the Roma population, regardless of age and sex, than in the majority population. This observation has been described independently in the Roma population of several countries. HDL-C level is genetically determined in approximately 50%, but also strongly influenced by environmental and lifestyle factors. The aim of our research is to confirm the genetic susceptibility underlying high frequency of low HDL-C level among Roma, and to compare the results with that of the Hungarian general population.

Twenty-one polymorphisms were genotyped in 1497 samples from the Hungarian general population and 616 from the segregated Roma population living in Hungary. For allele frequencies of fifteen SNPs, there was a significant difference between the two populations studied. The effect of SNPs on HDL-C level did not differ significantly between the two populations and also did not differ significantly from literature data. The unweighted and weighted genetic risk scores in case of the Roma sample showed a shift to the right (towards a higher genetic risk) compared to the Hungarian general population. We were able to identify the accumulation of several harmful mutations affecting the *CETP* gene in the Roma population. By regression analysis we were able to demonstrate that the genetic risk scores were significantly associated with HDL-cholesterol level in both study populations.

Haplotype analysis was also performed using 5 SNPs in the *CETP* and 6 SNPs in the *LIPC* gene. We identified three haplotypes that significantly reduced HDL-C level in *CETP* and three haplotypes that increased HDL-C level in the *LIPC* gene. Haplotype 8 in the *CETP* gene had the greatest HDL-C lowering effect, and this haplotype could be found almost exclusively in the Roma population.

Based on the results, the differences in HDL-cholesterol level between the Roma and Hungarian general populations are strongly linked to genetic determinants, but the impact of environmental and lifestyle factors cannot be neglected either.

In case of both populations, in addition to conventional lifestyle and environmental risk factors, the genetic background of individuals should also be considered when interventions to increase HDL cholesterol level is considered.

7. Keywords

ethnicity, High-density lipoprotein cholesterol, Roma, single nucleotide polymorphism, genetic risk score, genetic susceptibility, haplotype analysis, CETP gene, LIPC gene

8. Kulcsszavak

etnikum, HDL-koleszterin, roma, egypontos nukleotid-polimorfizmus, genetikai rizikó pontszám, genetikai hajlam, haplotípus analízis, CETP gén, LIPC gén

9. Acknowledgements

First and foremost, I would like to express my sincere gratitude to my PhD supervisor, Prof. Róza Ádány for her help and guidance during my PhD work.

I owe many thanks to Szilvia Fiatal for her help and support. I would also like to thank all of my colleagues involved in the publications on which this thesis is based. I am grateful to Zsuzsa Tóth and Beáta Soltész for their tireless work in the laboratories. My thanks also go out to my colleagues in the MTA-DE Public Health Research Group and Department of Public Health and Epidemiology, Faculty of Medicine from whom I received a lot of support.

I would like to thank my family and friends for their support and patience.

This work was supported by the TÁMOP-4.2.2.AA-11/1/KONV-2012-0031 [IGEN HUNGARIAN), TÁMOP 4.2.1. B-09/1/KONV-2010-0007 and GINOP-2.3.2-15-2016-00005 projects which are co-financed by the European Union and the European Social Fund, and by the Hungarian Academy of Sciences [MTA 11003, 2006TKI227). This research was also supported by ÚNKP-19-3 New National Excellence Program of the Ministry for Innovation and Technology.

10. References

1. Race, E. and G. Genetics Working, *The use of racial, ethnic, and ancestral categories in human genetics research.* Am J Hum Genet, 2005. **77**(4): p. 519-32.
2. Liegeois, J.-P., *Roma, Gypsies, Travellers.* 1994, Croton-on-Hudson, NY 10520: Manhattan Publishing Company.
3. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee, and the Committee of the Regions: An EU Framework for National Roma Integration Strategies up to 2020*. 2011, European Commission.
4. McKee, M., R. Adany, and L. MacLehose, *Health status and trends in candidate countries*, in *Health Policy and European Union Enlargement*, M. McKee, L. MacLehose, and E. Nolte, Editors. 2004, Open University Press: Buckingham. p. 24-42.
5. Kosa, K., L. Darago, and R. Adany, *Environmental survey of segregated habitats of Roma in Hungary: a way to be empowering and reliable in minority research.* Eur J Public Health, 2011. **21**(4): p. 463-8.
6. *Population Census in Hungary in 2011*. 2011, Hungarian Central Statistical Office.
7. *Estimates of Roma population in European Countries*. 2014, Council of Europe.
8. Kosa, K. and R. Adany, *Studying vulnerable populations: lessons from the Roma minority.* Epidemiology, 2007. **18**(3): p. 290-9.
9. Voko, Z., 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): p. 455-60.
10. Adany, R., *Roma health is global ill health.* Eur J Public Health, 2014. **24**(5): p. 702-3.
11. Kuhlbrandt, C., et al., *An examination of Roma health insurance status in Central and Eastern Europe.* Eur J Public Health, 2014. **24**(5): p. 707-12.
12. Arora, V.S., C. Kuhlbrandt, and M. McKee, *An examination of unmet health needs as perceived by Roma in Central and Eastern Europe.* Eur J Public Health, 2016. **26**(5): p. 737-742.
13. Foldes, M.E. and A. Covaci, *Research on Roma health and access to healthcare: state of the art and future challenges.* Int J Public Health, 2012. **57**(1): p. 37-9.
14. Zeljko, H., et al., *Traditional CVD risk factors and socio-economic deprivation in Roma minority population of Croatia.* Coll Antropol, 2008. **32**(3): p. 667-76.
15. Dobranici, M., A. Buzea, and R. Popescu, *The cardiovascular risk factors of the Roma (gypsies) people in Central-Eastern Europe: a review of the published literature.* J Med Life, 2012. **5**(4): p. 382-9.
16. Zeljko, H.M., et al., *Age trends in prevalence of cardiovascular risk factors in Roma minority population of Croatia.* Econ Hum Biol, 2013. **11**(3): p. 326-36.
17. Weiss, E., et al., *Cardiovascular risk factors in a Roma sample population from Romania.* Rom J Intern Med, 2018. **56**(3): p. 193-202.
18. Hujova, Z., et al., *Cardiovascular risk predictors in central Slovakian Roma children and adolescents: regional differences.* Cent Eur J Public Health, 2010. **18**(3): p. 139-44.
19. Fedacko, J., et al., *Prevalence of cardiovascular risk factors in relation to metabolic syndrome in the Roma population compared with the non-Roma*

population in the eastern part of Slovakia. Cent Eur J Public Health, 2014. **22 Suppl**: p. S69-74.
20. Fedacko, J., et al., *Clinical and biochemical determinants of metabolic syndrome among Roma and non-Roma subjects in the eastern part of Slovakia.* Cent Eur J Public Health, 2014. **22 Suppl**: p. S75-80.
21. Adamkova, V., et al., *Genetic and biochemical characteristics in the Roma minority in the South Bohemia Region.* Neuro Endocrinol Lett, 2015. **36 Suppl 2**: p. 29-34.
22. Sumegi, K., et al., *Functional variants of lipid level modifier MLXIPL, GCKR, GALNT2, CILP2, ANGPTL3 and TRIB1 genes in healthy Roma and Hungarian populations.* Pathol Oncol Res, 2015. **21**(3): p. 743-9.
23. Zeljko, H.M., et al., *E2 allele of the apolipoprotein E gene polymorphism is predictive for obesity status in Roma minority population of Croatia.* Lipids Health Dis, 2011. **10**: p. 9.
24. Kosa, Z., et al., *Prevalence of metabolic syndrome among Roma: a comparative health examination survey in Hungary.* Eur J Public Health, 2015. **25**(2): p. 299-304.
25. Gordon, T., et al., *High density lipoprotein as a protective factor against coronary heart disease. The Framingham Study.* Am J Med, 1977. **62**(5): p. 707-14.
26. Qasim, A. and D.J. Rader, *Human genetics of variation in high-density lipoprotein cholesterol.* Curr Atheroscler Rep, 2006. **8**(3): p. 198-205.
27. Zarkesh, M., et al., *Heritability of the metabolic syndrome and its components in the Tehran Lipid and Glucose Study (TLGS).* Genet Res (Camb), 2012. **94**(6): p. 331-7.
28. Kalaydjieva, L., D. Gresham, and F. Calafell, *Genetic studies of the Roma (Gypsies): a review.* BMC Med Genet, 2001. **2**: p. 5.
29. Feher, A., et al., *Dopamine metabolism-related gene polymorphisms in Roma (Gypsy) and Hungarian populations.* J Genet, 2011. **90**(2): p. e72-5.
30. Assal, S., E. Susanszky, and A. Czeizel, *High consanguinity rate in Hungarian gipsy communities.* Acta Paediatr Hung, 1991. **31**(3): p. 299-304.
31. Mineo, C., et al., *Endothelial and antithrombotic actions of HDL.* Circ Res, 2006. **98**(11): p. 1352-64.
32. Navab, M., G.M. Anantharamaiah, and A.M. Fogelman, *The role of high-density lipoprotein in inflammation.* Trends Cardiovasc Med, 2005. **15**(4): p. 158-61.
33. Brites, F., et al., *Antioxidative activity of high-density lipoprotein (HDL): Mechanistic insights into potential clinical benefit.* BBA Clin, 2017. **8**: p. 66-77.
34. Kuvin, J.T., et al., *A novel mechanism for the beneficial vascular effects of high-density lipoprotein cholesterol: enhanced vasorelaxation and increased endothelial nitric oxide synthase expression.* Am Heart J, 2002. **144**(1): p. 165-72.
35. Ahn, N. and K. Kim, *High-density lipoprotein cholesterol (HDL-C) in cardiovascular disease: effect of exercise training.* Integr Med Res, 2016. **5**(3): p. 212-215.
36. Barter, P.J., et al., *Cholesteryl ester transfer protein: a novel target for raising HDL and inhibiting atherosclerosis.* Arterioscler Thromb Vasc Biol, 2003. **23**(2): p. 160-7.
37. Brown, M.L., et al., *Molecular basis of lipid transfer protein deficiency in a family with increased high-density lipoproteins.* Nature, 1989. **342**(6248): p. 448-51.

38. Tall, A.R., *CETP inhibitors to increase HDL cholesterol levels.* N Engl J Med, 2007. **356**(13): p. 1364-6.
39. Jansen, H., A.J. Verhoeven, and E.J. Sijbrands, *Hepatic lipase: a pro- or anti-atherogenic protein?* J Lipid Res, 2002. **43**(9): p. 1352-62.
40. Cedo, L., et al., *Human hepatic lipase overexpression in mice induces hepatic steatosis and obesity through promoting hepatic lipogenesis and white adipose tissue lipolysis and fatty acid uptake.* PLoS One, 2017. **12**(12): p. e0189834.
41. Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.
42. Burgess, S. and S.G. Thompson, *Use of allele scores as instrumental variables for Mendelian randomization.* Int J Epidemiol, 2013. **42**(4): p. 1134-44.
43. Evans, D.M., P.M. Visscher, and N.R. Wray, *Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk.* Hum Mol Genet, 2009. **18**(18): p. 3525-31.
44. Bryant, E.K., et al., *A multiethnic replication study of plasma lipoprotein levels-associated SNPs identified in recent GWAS.* PLoS One, 2013. **8**(5): p. e63469.
45. Dumitrescu, L., et al., *Genetic determinants of lipid traits in diverse populations from the population architecture using genomics and epidemiology (PAGE) study.* PLoS Genet, 2011. **7**(6): p. e1002138.
46. Chang, M.H., et al., *Racial/ethnic variation in the association of lipid-related genetic variants with blood lipids in the US adult population.* Circ Cardiovasc Genet, 2011. **4**(5): p. 523-33.
47. van Leeuwen, E.M., et al., *The challenges of genome-wide interaction studies: lessons to learn from the analysis of HDL blood levels.* PLoS One, 2014. **9**(10): p. e109290.
48. Wang, K., et al., *Pathway-Wide Association Study Implicates Multiple Sterol Transport and Metabolism Genes in HDL Cholesterol Regulation.* Front Genet, 2011. **2**: p. 41.
49. Ridker, P.M., et al., *Polymorphism in the CETP gene region, HDL cholesterol, and risk of future myocardial infarction: Genomewide analysis among 18 245 initially healthy women from the Women's Genome Health Study.* Circ Cardiovasc Genet, 2009. **2**(1): p. 26-33.
50. Diabetes Genetics Initiative of Broad Institute of, H., et al., *Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels.* Science, 2007. **316**(5829): p. 1331-6.
51. Southam, L., et al., *Whole genome sequencing and imputation in isolated populations identify genetic associations with medically-relevant complex traits.* Nat Commun, 2017. **8**: p. 15606.
52. An, P., et al., *Genome-wide association studies identified novel loci for non-high-density lipoprotein cholesterol and its postprandial lipemic response.* Hum Genet, 2014. **133**(7): p. 919-30.
53. Surakka, I., et al., *A genome-wide association study of monozygotic twin-pairs suggests a locus related to variability of serum high-density lipoprotein cholesterol.* Twin Res Hum Genet, 2012. **15**(6): p. 691-9.
54. Kraja, A.T., et al., *A bivariate genome-wide approach to metabolic syndrome: STAMPEED consortium.* Diabetes, 2011. **60**(4): p. 1329-39.
55. Sabatti, C., et al., *Genome-wide association analysis of metabolic traits in a birth cohort from a founder population.* Nat Genet, 2009. **41**(1): p. 35-46.

56. Heid, I.M., et al., *Genome-wide association analysis of high-density lipoprotein cholesterol in the population-based KORA study sheds new light on intergenic regions.* Circ Cardiovasc Genet, 2008. **1**(1): p. 10-20.
57. Kathiresan, S., et al., *Erratum: Corrigendum: Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans.* Nature Genetics, 2008. **40**(11): p. 1384-1384.
58. Keller, M., et al., *THOC5: a novel gene involved in HDL-cholesterol metabolism.* J Lipid Res, 2013. **54**(11): p. 3170-6.
59. Kathiresan, S., et al., *Common variants at 30 loci contribute to polygenic dyslipidemia.* Nat Genet, 2009. **41**(1): p. 56-65.
60. Aulchenko, Y.S., et al., *Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts.* Nat Genet, 2009. **41**(1): p. 47-55.
61. Willer, C.J., et al., *Newly identified loci that influence lipid concentrations and risk of coronary artery disease.* Nat Genet, 2008. **40**(2): p. 161-9.
62. Nagy, R., et al., *Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 Generation Scotland participants.* Genome Med, 2017. **9**(1): p. 23.
63. He, L., et al., *Pleiotropic Meta-Analyses of Longitudinal Studies Discover Novel Genetic Variants Associated with Age-Related Diseases.* Front Genet, 2016. **7**: p. 179.
64. Middelberg, R.P., et al., *Genetic variants in LPL, OASL and TOMM40/APOE-C1-C2-C4 genes are associated with multiple cardiovascular-related traits.* BMC Med Genet, 2011. **12**: p. 123.
65. Zemunik, T., et al., *Genome-wide association study of biochemical traits in Korcula Island, Croatia.* Croat Med J, 2009. **50**(1): p. 23-33.
66. Postmus, I., et al., *Meta-analysis of genome-wide association studies of HDL cholesterol response to statins.* J Med Genet, 2016. **53**(12): p. 835-845.
67. Surakka, I., et al., *The impact of low-frequency and rare variants on lipid levels.* Nat Genet, 2015. **47**(6): p. 589-97.
68. Smith, E.N., et al., *Longitudinal genome-wide association of cardiovascular disease risk factors in the Bogalusa heart study.* PLoS Genet, 2010. **6**(9): p. e1001094.
69. Teslovich, T.M., et al., *Biological, clinical and population relevance of 95 loci for blood lipids.* Nature, 2010. **466**(7307): p. 707-13.
70. Irvin, M.R., et al., *A genome-wide study of lipid response to fenofibrate in Caucasians: a combined analysis of the GOLDN and ACCORD studies.* Pharmacogenet Genomics, 2016. **26**(7): p. 324-33.
71. Willer, C.J., et al., *Discovery and refinement of loci associated with lipid levels.* Nat Genet, 2013. **45**(11): p. 1274-1283.
72. Hiura, Y., et al., *Identification of genetic markers associated with high-density lipoprotein-cholesterol by genome-wide screening in a Japanese population: the Suita study.* Circ J, 2009. **73**(6): p. 1119-26.
73. Zhu, Y., et al., *Susceptibility loci for metabolic syndrome and metabolic components identified in Han Chinese: a multi-stage genome-wide association study.* J Cell Mol Med, 2017. **21**(6): p. 1106-1116.
74. Kurano, M., et al., *Genome-wide association study of serum lipids confirms previously reported associations as well as new associations of common SNPs within PCSK7 gene with triglyceride.* J Hum Genet, 2016. **61**(5): p. 427-33.

75. Lu, X., et al., *Genetic Susceptibility to Lipid Levels and Lipid Change Over Time and Risk of Incident Hyperlipidemia in Chinese Populations.* Circ Cardiovasc Genet, 2016. **9**(1): p. 37-44.
76. Moore, C.B., et al., *Phenome-wide Association Study Relating Pretreatment Laboratory Parameters With Human Genetic Variants in AIDS Clinical Trials Group Protocols.* Open Forum Infect Dis, 2015. **2**(1): p. ofu113.
77. Ko, A., et al., *Amerindian-specific regions under positive selection harbour new lipid variants in Latinos.* Nat Commun, 2014. **5**: p. 3983.
78. Weissglas-Volkov, D., et al., *Genomic study in Mexicans identifies a new locus for triglycerides and refines European lipid loci.* J Med Genet, 2013. **50**(5): p. 298-308.
79. Wakil, S.M., et al., *A common variant association study reveals novel susceptibility loci for low HDL-cholesterol levels in ethnic Arabs.* Clin Genet, 2016. **90**(6): p. 518-525.
80. Below, J.E., et al., *Meta-analysis of lipid-traits in Hispanics identifies novel loci, population-specific effects, and tissue-specific enrichment of eQTLs.* Sci Rep, 2016. **6**: p. 19429.
81. Wu, Y., et al., *Genetic association with lipids in Filipinos: waist circumference modifies an APOA5 effect on triglyceride levels.* J Lipid Res, 2013. **54**(11): p. 3198-205.
82. Coram, M.A., et al., *Genome-wide characterization of shared and distinct genetic components that influence blood lipid levels in ethnically diverse human populations.* Am J Hum Genet, 2013. **92**(6): p. 904-16.
83. Kim, Y.J., et al., *Large-scale genome-wide association studies in East Asians identify new genetic loci influencing metabolic traits.* Nat Genet, 2011. **43**(10): p. 990-5.
84. Waterworth, D.M., et al., *Genetic variants influencing circulating lipid levels and risk of coronary artery disease.* Arterioscler Thromb Vasc Biol, 2010. **30**(11): p. 2264-76.
85. Chambers, J.C., et al., *Common genetic variation near MC4R is associated with waist circumference and insulin resistance.* Nat Genet, 2008. **40**(6): p. 716-8.
86. Spracklen, C.N., et al., *Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels.* Hum Mol Genet, 2017. **26**(9): p. 1770-1784.
87. van Leeuwen, E.M., et al., *Meta-analysis of 49 549 individuals imputed with the 1000 Genomes Project reveals an exonic damaging variant in ANGPTL4 determining fasting TG levels.* J Med Genet, 2016. **53**(7): p. 441-9.
88. Zhou, L., et al., *A genome wide association study identifies common variants associated with lipid levels in the Chinese population.* PLoS One, 2013. **8**(12): p. e82420.
89. Lettre, G., et al., *Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARe Project.* PLoS Genet, 2011. **7**(2): p. e1001300.
90. Kamatani, Y., et al., *Genome-wide association study of hematological and biochemical traits in a Japanese population.* Nat Genet, 2010. **42**(3): p. 210-5.
91. Lowe, J.K., et al., *Genome-wide association studies in an isolated founder population from the Pacific Island of Kosrae.* PLoS Genet, 2009. **5**(2): p. e1000365.

92. Rasmussen-Torvik, L.J., et al., *Fasting glucose GWAS candidate region analysis across ethnic groups in the Multiethnic Study of Atherosclerosis (MESA).* Genet Epidemiol, 2012. **36**(4): p. 384-91.
93. Paulik, E., et al., *Smoking behaviour and attitudes of Hungarian Roma and non-Roma population towards tobacco control policies.* Int J Public Health, 2011. **56**(5): p. 485-91.
94. Sárváry, A., Kósa, Zs., Jávorné Erdei, R., Gyulai, A., Takács, P., Sándor, J., Sárváry, A., Németh, Á., Halmai, R., Ádány, R., *Telepszerű körülmények között élő gyermekek egészségmagatartás Északkelet-Magyarországon {Article in Hungarian: Health behaviour of children living in colonies in North-Eastern Hungary}.* Népegészségügy, 2012. **90**(4): p. 230-244.
95. Gresham, D., et al., *Origins and divergence of the Roma (gypsies).* Am J Hum Genet, 2001. **69**(6): p. 1314-31.
96. Mendizabal, I., et al., *Reconstructing the Indian origin and dispersal of the European Roma: a maternal genetic perspective.* PLoS One, 2011. **6**(1): p. e15988.
97. Szeles, G., et al., *A preliminary evaluation of a health monitoring programme in Hungary.* Eur J Public Health, 2005. **15**(1): p. 26-32.
98. Jermendy, G., et al., *Prevalence rate of diabetes mellitus and impaired fasting glycemia in Hungary: cross-sectional study on nationally representative sample of people aged 20-69 years.* Croat Med J, 2010. **51**(2): p. 151-6.
99. Willer, C.J., et al., *Newly identified loci that influence lipid concentrations and risk of coronary artery disease.* Nature Genetics, 2008. **40**(2): p. 161-169.
100. Dumitrescu, L., et al., *Genetic Determinants of Lipid Traits in Diverse Populations from the Population Architecture using Genomics and Epidemiology (PAGE) Study.* Plos Genetics, 2011. **7**(6).
101. Cougnard-Gregoire, A., et al., *Elevated high-density lipoprotein cholesterol and age-related macular degeneration: the Alienor study.* PLoS One, 2014. **9**(3): p. e90973.
102. Weissglas-Volkov, D., et al., *Investigation of variants identified in caucasian genome-wide association studies for plasma high-density lipoprotein cholesterol and triglycerides levels in Mexican dyslipidemic study samples.* Circ Cardiovasc Genet, 2010. **3**(1): p. 31-8.
103. Kristiansson, K., et al., *Genome-Wide Screen for Metabolic Syndrome Susceptibility Loci Reveals Strong Lipid Gene Contribution But No Evidence for Common Genetic Basis for Clustering of Metabolic Syndrome Traits.* Circulation-Cardiovascular Genetics, 2012. **5**(2): p. 242-249.
104. Huggins, G.S., et al., *Do genetic modifiers of high-density lipoprotein cholesterol and triglyceride levels also modify their response to a lifestyle intervention in the setting of obesity and type-2 diabetes mellitus?: The Action for Health in Diabetes (Look AHEAD) study.* Circ Cardiovasc Genet, 2013. **6**(4): p. 391-9.
105. Park, M.H., et al., *Genetic loci associated with lipid concentrations and cardiovascular risk factors in the Korean population.* J Med Genet, 2011. **48**(1): p. 10-5.
106. Rudkowska, I., et al., *Gene-diet interactions on plasma lipid levels in the Inuit population.* Br J Nutr, 2013. **109**(5): p. 953-61.
107. Boes, E., et al., *Genetic-epidemiological evidence on genes associated with HDL cholesterol levels: a systematic in-depth review.* Exp Gerontol, 2009. **44**(3): p. 136-60.

108. Pare, G., et al., *Genetic analysis of 103 candidate genes for coronary artery disease and associated phenotypes in a founder population reveals a new association between endothelin-1 and high-density lipoprotein cholesterol.* American Journal of Human Genetics, 2007. **80**(4): p. 673-682.
109. Hamrefors, V., et al., *A gene score of nine LDL and HDL regulating genes is associated with fluvastatin-induced cholesterol changes in women.* Journal of Lipid Research, 2010. **51**(3): p. 625-634.
110. Kathiresan, S., et al., *A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study.* BMC Med Genet, 2007. **8 Suppl 1**: p. S17.
111. Asselbergs, F.W., et al., *Large-Scale Gene-Centric Meta-analysis across 32 Studies Identifies Multiple Lipid Loci.* American Journal of Human Genetics, 2012. **91**(5): p. 823-838.
112. Boekholdt, S.M. and J.F. Thompson, *Natural genetic variation as a tool in understanding the role of CETP in lipid levels and disease.* Journal of Lipid Research, 2003. **44**(6): p. 1080-1093.
113. Ronald, J., et al., *Analysis of recently identified dyslipidemia alleles reveals two loci that contribute to risk for carotid artery disease.* Lipids in Health and Disease, 2009. **8**.
114. Heid, I.M., et al., *Genome-Wide Association Analysis of High-Density Lipoprotein Cholesterol in the Population-Based KORA Study Sheds New Light on Intergenic Regions.* Circulation-Cardiovascular Genetics, 2008. **1**(1): p. 10-U65.
115. Tachmazidou, I., et al., *A rare functional cardioprotective APOC3 variant has risen in frequency in distinct population isolates.* Nature Communications, 2013. **4**.
116. Weissglas-Volkov, D., et al., *Genomic study in Mexicans identifies a new locus for triglycerides and refines European lipid loci.* Journal of Medical Genetics, 2013. **50**(5): p. 298-308.
117. Kraja, A.T., et al., *A Bivariate Genome-Wide Approach to Metabolic Syndrome STAMPEED Consortium.* Diabetes, 2011. **60**(4): p. 1329-1339.
118. Ridker, P.M., et al., *Polymorphism in the CETP Gene Region, HDL Cholesterol, and Risk of Future Myocardial Infarction Genomewide Analysis Among 18 245 Initially Healthy Women From the Women's Genome Health Study.* Circulation-Cardiovascular Genetics, 2009. **2**(1): p. 26-33.
119. Holmes, M.V., et al., *Mendelian randomization of blood lipids for coronary heart disease.* European Heart Journal, 2015. **36**(9): p. 539-+.
120. Lu, Y., et al., *Dietary n-3 and n-6 polyunsaturated fatty acid intake interacts with FADS1 genetic variation to affect total and HDL-cholesterol concentrations in the Doetinchem Cohort Study.* Am J Clin Nutr, 2010. **92**(1): p. 258-65.
121. Kathiresan, S., et al., *Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans (vol 40, pg 189, 2008).* Nature Genetics, 2008. **40**(11): p. 1384-1384.
122. Kathiresan, S., et al., *Common variants at 30 loci contribute to polygenic dyslipidemia.* Nature Genetics, 2009. **41**(1): p. 56-65.
123. Varga, T.V., et al., *Genetic Determinants of Long-Term Changes in Blood Lipid Concentrations: 10-Year Follow-Up of the GLACIER Study.* Plos Genetics, 2014. **10**(6).
124. Willer, C.J., et al., *Discovery and refinement of loci associated with lipid levels.* Nature Genetics, 2013. **45**(11): p. 1274-+.

125. Ma, L., et al., *Knowledge-driven analysis identifies a gene-gene interaction affecting high-density lipoprotein cholesterol levels in multi-ethnic populations.* PLoS Genet, 2012. **8**(5): p. e1002714.
126. Sabatti, C., et al., *Genome-wide association analysis of metabolic traits in a birth cohort from a founder population.* Nature Genetics, 2009. **41**(1): p. 35-46.
127. Lanktree, M.B., et al., *Replication of genetic associations with plasma lipoprotein traits in a multiethnic sample.* J Lipid Res, 2009. **50**(7): p. 1487-96.
128. Roberts, C.G., et al., *Variants in scavenger receptor class B type I gene are associated with HDL cholesterol levels in younger women.* Hum Hered, 2007. **64**(2): p. 107-13.
129. Lee, J.C., et al., *WW-domain-containing oxidoreductase is associated with low plasma HDL-C levels.* Am J Hum Genet, 2008. **83**(2): p. 180-92.
130. Gauderman, W.J. and J.M. Morrison, *Quanto 1.1: A computer program for power and sample size calculations for genetic epidemiology studies.* 2010.
131. Plenge, R.M., et al., *Recommendations for publication of genetic association studies in Arthritis & Rheumatism.* Arthritis Rheum, 2011. **63**(10): p. 2839-47.
132. Talmud, P.J., et al., *Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study.* BMJ, 2010. **340**: p. b4838.
133. Asselbergs, F.W., et al., *Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci.* Am J Hum Genet, 2012. **91**(5): p. 823-38.
134. Kathiresan, S., *Will cholesteryl ester transfer protein inhibition succeed primarily by lowering low-density lipoprotein cholesterol? Insights from human genetics and clinical trials.* J Am Coll Cardiol, 2012. **60**(20): p. 2049-52.
135. McDonough, C.W., et al., *Atenolol induced HDL-C change in the pharmacogenomic evaluation of antihypertensive responses (PEAR) study.* PLoS One, 2013. **8**(10): p. e76984.
136. Bustami, J., et al., *Cholesteryl ester transfer protein (CETP) I405V polymorphism and cardiovascular disease in eastern European Caucasians - a cross-sectional study.* BMC Geriatr, 2016. **16**: p. 144.
137. Sebastiani, P., N. Solovieff, and J.X. Sun, *Naive Bayesian Classifier and Genetic Risk Score for Genetic Risk Prediction of a Categorical Trait: Not so Different after all!* Front Genet, 2012. **3**: p. 26.
138. Johnson, T., *Efficient calculation for Multi-SNP Genetic Risk Score*, in *American Society of Human Genetics Annual Meeting*. 2012: San Francisco, November 6-10, .
139. *The IDF consensus worldwide definition of the metabolic syndrome*. 2006, International Diabetes Federation: Brussels, Belgium.
140. Sole, X., et al., *SNPStats: a web tool for the analysis of association studies.* Bioinformatics, 2006. **22**(15): p. 1928-9.
141. Pers, T.H., P. Timshel, and J.N. Hirschhorn, *SNPsnap: a Web-based tool for identification and annotation of matched SNPs.* Bioinformatics, 2015. **31**(3): p. 418-20.
142. Templeton, G.F., *A Two-Step Approach for Transforming Continuous Variables to Normal: Implications and Recommendations for IS Research.* Communications of the Association for Information Systems, 2011. **28**(1): p. 4.
143. Ronald, J., et al., *Analysis of recently identified dyslipidemia alleles reveals two loci that contribute to risk for carotid artery disease.* Lipids Health Dis, 2009. **8**: p. 52.
144. Singh, I.M., M.H. Shishehbor, and B.J. Ansell, *High-density lipoprotein as a therapeutic target: a systematic review.* JAMA, 2007. **298**(7): p. 786-98.

145. Elovainio, M., et al., *Socioeconomic differences in cardiometabolic factors: social causation or health-related selection? Evidence from the Whitehall II Cohort Study, 1991-2004.* Am J Epidemiol, 2011. **174**(7): p. 779-89.
146. Sviridov, D. and P.J. Nestel, *Genetic factors affecting HDL levels, structure, metabolism and function.* Curr Opin Lipidol, 2007. **18**(2): p. 157-63.
147. Chasman, D.I., et al., *Genetic loci associated with plasma concentration of low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, triglycerides, apolipoprotein A1, and Apolipoprotein B among 6382 white women in genome-wide analysis with replication.* Circ Cardiovasc Genet, 2008. **1**(1): p. 21-30.
148. Asselbergs, F.W., R.C. Lovering, and F. Drenos, *Progress in genetic association studies of plasma lipids.* Curr Opin Lipidol, 2013. **24**(2): p. 123-8.
149. Brunham, L.R. and M.R. Hayden, *Human genetics of HDL: Insight into particle metabolism and function.* Prog Lipid Res, 2015. **58**: p. 14-25.
150. Hubkova, B., et al., *Assessment of clinical biochemical parameters in Roma minority residing in eastern Slovakia compared with the majority population.* Cent Eur J Public Health, 2014. **22 Suppl**: p. S12-7.
151. Petrikova, J., et al., *Serum Uric Acid in Roma and Non-Roma-Its Correlation with Metabolic Syndrome and Other Variables.* Int J Environ Res Public Health, 2018. **15**(7).
152. Macejova, Z., et al., *The Roma Population Living in Segregated Settlements in Eastern Slovakia Has a Higher Prevalence of Metabolic Syndrome, Kidney Disease, Viral Hepatitis B and E, and Some Parasitic Diseases Compared to the Majority Population.* Int J Environ Res Public Health, 2020. **17**(9).
153. Babinska, I., et al., *Is the cardiovascular risk profile of people living in Roma settlements worse in comparison with the majority population in Slovakia?* Int J Public Health, 2013. **58**(3): p. 417-25.
154. Cook, B., et al., *Revisiting the evidence on health and health care disparities among the Roma: a systematic review 2003-2012.* Int J Public Health, 2013. **58**(6): p. 885-911.
155. Hujova, Z., J. Desatnikova, and D. Gabor, *Some cardiovascular risk factors in Gypsy children and adolescents from Central Slovakia.* Bratisl Lek Listy, 2009. **110**(4): p. 233-9.
156. Piko, P., et al., *Generalizability and applicability of results obtained from populations of European descent regarding the effect direction and size of HDL-C level-associated genetic variants to the Hungarian general and Roma populations.* Gene, 2019. **686**: p. 187-193.
157. Piko, P., et al., *Genetic factors exist behind the high prevalence of reduced high-density lipoprotein cholesterol levels in the Roma population.* Atherosclerosis, 2017. **263**: p. 119-126.
158. Pikó, P., et al., *Data to genetic risk assessment on high-density cholesterol level associated polimorphisms in Hungarian General and Roma populations.* Data in Brief, 2017. **"submitted"**.
159. Piko, P., et al., *The Effect of Haplotypes in the CETP and LIPC Genes on the Triglycerides to HDL-C Ratio and Its Components in the Roma and Hungarian General Populations.* Genes (Basel), 2020. **11**(1).
160. Tall, A., *Plasma lipid transfer proteins.* Annu Rev Biochem, 1995. **64**: p. 235-57.
161. Armitage, J., M.V. Holmes, and D. Preiss, *Cholesteryl Ester Transfer Protein Inhibition for Preventing Cardiovascular Events: JACC Review Topic of the Week.* J Am Coll Cardiol, 2019. **73**(4): p. 477-487.

162. Bochem, A.E., J.A. Kuivenhoven, and E.S. Stroes, *The promise of cholesteryl ester transfer protein (CETP) inhibition in the treatment of cardiovascular disease.* Curr Pharm Des, 2013. **19**(17): p. 3143-9.
163. Kardassis, D., et al., *Metabolism of HDL and its regulation.* Curr Med Chem, 2014. **21**(25): p. 2864-80.
164. Kobayashi, J., et al., *Hepatic Lipase: a Comprehensive View of its Role on Plasma Lipid and Lipoprotein Metabolism.* J Atheroscler Thromb, 2015. **22**(10): p. 1001-11.
165. Hunter, M., et al., *NDRG1 interacts with APO A-I and A-II and is a functional candidate for the HDL-C QTL on 8q24.* Biochem Biophys Res Commun, 2005. **332**(4): p. 982-92.
166. Blackett, P.R. and D.K. Sanghera, *Genetic determinants of cardiometabolic risk: a proposed model for phenotype association and interaction.* J Clin Lipidol, 2013. **7**(1): p. 65-81.
167. Landmesser, U., *High density lipoprotein - should we raise it?* Curr Vasc Pharmacol, 2012. **10**(6): p. 718-9.

11. Publications

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/6/2021.PL
Subject: PhD Publication List

Candidate: Péter Pikó
Doctoral School: Doctoral School of Health Sciences
MTMT ID: 10061786

List of publications related to the dissertation

1. **Pikó, P.**, Fiatal, S., Werissa, N. A., Begashaw, B., Rácz, G., Kósa, Z., Sándor, J., Ádány, R.: The Effect of Haplotypes in the CETP and LIPC Genes on the Triglycerides to HDL-C Ratio and Its Components in the Roma and Hungarian General Populations.
Genes. 11 (56), 1-13, 2020.
DOI: http://dx.doi.org/10.3390/genes11010056
IF: 3.759 (2019)

2. **Pikó, P.**, Fiatal, S., Kósa, Z., Sándor, J., Ádány, R.: Generalizability and applicability of results obtained from populations of European descent regarding the effect direction and size of HDL-C level-associated genetic variants to the Hungarian general and Roma populations.
Gene. 686, 187-193, 2019.
DOI: http://dx.doi.org/10.1016/j.gene.2018.11.067
IF: 2.984

3. **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. 263, 119-126, 2017.
IF: 4.467

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

List of other publications

4. Llanaj, E., **Pikó, P.**, Nagy, K., Rácz, G., Sándor, J., Kósa, Z., Fiatal, S., Ádány, R.: Applicability of Obesity-Related SNPs and their Effect Size Measures Defined on Populations with European Ancestry for Genetic Risk Estimation among Roma.
Genes. 11, 1-13, 2020.
DOI: http://dx.doi.org/10.3390/genes11050516
IF: 3.759 (2019)

5. Fiatal, S., **Pikó, P.**, Ádány, R.: Application of Single-Nucleotide Polymorphism-Related Risk Estimates in Identification of Increased Genetic Susceptibility to Non-communicable Diseases.
In: Personalised Health Care. Fostering Precision Medicine Advancements for Gaining Population Health Impact. / Stefania Boccia, Róza Ádány, Paolo Villari, Martina C. Cornel, Corrado De Vito, Roberta Pastorino, Springer Nature Switzerland AG, Cham, Switzerland, 22-26, 2020.

6. **Pikó, P.**, Werissa, N. A., Fiatal, S., Sándor, J., Ádány, R.: Impact of Genetic Factors on the Age of Onset for Type 2 Diabetes Mellitus in Addition to the Conventional Risk Factors.
JPM. 11 (1), 6-, 2020.
DOI: http://dx.doi.org/10.3390/jpm11010006
IF: 4.433 (2019)

7. Ádány, R., **Pikó, P.**, Fiatal, S., Kósa, Z., Sándor, J., Bíró, É., Kósa, K., Paragh, G., Bácsné Bába, É., Veres-Balajti, I., Bíró, K., Varga, O., Balázs, M.: 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. 17 (4833), 1-22, 2020.
DOI: http://dx.doi.org/10.3390/ijerph17134833
IF: 2.849 (2019)

8. Diószegi, J., **Pikó, P.**, Kósa, Z., Sándor, J., Llanaj, E., Ádány, R.: Taste and Food Preferences of the Hungarian Roma Population.
Front. Public Health. 8, 1-11, 2020.
DOI: https://doi.org/10.3389/fpubh.2020.00359
IF: 2.483 (2019)

9. Soltész, B., **Pikó, P.**, Sándor, J., Kósa, Z., Ádány, R., Fiatal, S.: The genetic risk for hypertension is lower among the Hungarian Roma population compared to the general population.
PLoS One. 15 (6), 1-17, 2020.
DOI: http://dx.doi.org/10.1371/journal.pone.0234547
IF: 2.74 (2019)

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

10. 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. 179, 37-44, 2019.
DOI: http://dx.doi.org/10.1016/j.thromres.2019.04.031
IF: 2.869

11. Werissa, N. A., **Pikó, P.**, Fiatal, S., Kósa, Z., Sándor, J., Ádány, R.: SNP-Based Genetic Risk Score Modeling Suggests No Increased Genetic Susceptibility of the Roma Population to Type 2 Diabetes Mellitus.
Genes. 10 (11), 1-16, 2019.
DOI: http://dx.doi.org/10.3390/genes10110942
IF: 3.759

12. **Pikó, P.**, Fiatal, S., Kósa, Z., Sándor, J., Ádány, R.: A csökkent HDL-koleszterin szint a roma lakosság körében tapasztalt magas prevalenciájában szerepet játszó genetikai tényezők.
Népegészségügy. 96 (1), 39-50, 2018.

13. van El, C. G., Baccolini, V., **Pikó, P.**, Cornel, M. C.: Stakeholder Views on Active Cascade Screening for Familial Hypercholesterolemia.
Healthcare. 6 (3), 1-11, 2018.
DOI: http://dx.doi.org/10.3390/healthcare6030108

14. **Pikó, P.**, Fiatal, S., Kósa, Z., Sándor, J., Ádány, R.: Data to genetic risk assessment on high-density cholesterol level associated polymorphisms in Hungarian general and Roma populations.
Data Brief. 14, 354-359, 2017.
DOI: http://dx.doi.org/10.1016/j.dib.2017.07.053

Total IF of journals (all publications): 34,102
Total IF of journals (publications related to the dissertation): 11,21

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.

07 January, 2021