



ELSEVIER

Contents lists available at ScienceDirect

## Data in Brief

journal homepage: [www.elsevier.com/locate/dib](http://www.elsevier.com/locate/dib)

## Data Article

# Data to genetic risk assessment on high-density cholesterol level associated polymorphisms in Hungarian general and Roma populations



Péter Pikó<sup>a,b</sup>, Szilvia Fialat<sup>b,c</sup>, Zsigmond Kósa<sup>d</sup>,  
János Sándor<sup>b,c</sup>, Róza Ádány<sup>a,b,c,\*</sup>

<sup>a</sup> MTA-DE Public Health Research Group of the Hungarian Academy of Sciences, Faculty of Public Health, University of Debrecen, Debrecen 4028, Hungary

<sup>b</sup> Department of Preventive Medicine, Faculty of Public Health, University of Debrecen, Debrecen 4028, Hungary

<sup>c</sup> WHO Collaborating Centre on Vulnerability and Health, Department of Preventive Medicine, Faculty of Public Health, University of Debrecen, Debrecen 4028, Hungary

<sup>d</sup> Department of Health Visitor Methodology and Public Health, Faculty of Health, University of Debrecen, Nyíregyháza 4400, Hungary

## ARTICLE INFO

## Article history:

Received 24 May 2017

Received in revised form

6 July 2017

Accepted 21 July 2017

Available online 26 July 2017

## Keywords:

Single nucleotide polymorphism

Genetic susceptibility

Genetic risk score

High-density lipoprotein cholesterol

Roma population

## ABSTRACT

Data obtained by genotyping single nucleotide polymorphisms (SNPs) related to high-density lipoprotein cholesterol (HDL-C) levels were utilized in Genetic Risk Score [unweighted (GRS) and weighted (wGRS)] computation on Hungarian general and Roma populations. The selection process of the SNPs as well as the results obtained are published in our research article (Piko et al., 2017) [1]. Linkage analyses were performed by study groups. Study populations were stratified by quintiles of weighted Genetic Risk Score. Multivariate linear regression analyses were performed using Genetic Risk Scores and HDL-C levels as dependent variables; and ethnicity, sex and age as independent variables. The study subjects were categorized into quintiles according their wGRS values. Associations of Genetic Risk Scores with plasma HDL-C levels (as a continuous variable) were observed in both populations. Finally, the two populations were merged and analyzed together by multivariate logistic regression where reduced

DOI of original article: <http://dx.doi.org/10.1016/j.atherosclerosis.2017.05.028>

\* Correspondence to: Kassai út 26., Debrecen 4028, Hungary. Fax: +36 52 512 769.

E-mail address: [adany.roza@sph.unideb.hu](mailto:adany.roza@sph.unideb.hu) (R. Ádány).

<http://dx.doi.org/10.1016/j.dib.2017.07.053>

2352-3409/© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

plasma HDL-C level was the dependent variable; while ethnicity, age and sex were the independent ones.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Specifications Table

Subject area	<i>Biology</i>
More specific subject area	<i>Molecular genetics, Public health genomics</i>
Type of data	<i>Figure, Table</i>
How data was acquired	<i>Survey, Blood sample collection, MassARRAY platform (Sequenom Inc., San Diego, CA, USA) with iPLEX Gold chemistry</i>
Data format	<i>Analyzed</i>
Experimental factors	<i>Genomic DNA from peripheral blood was isolated</i>
Experimental features	<i>Genotyping method of SNPs was based on MALDI-TOF (Matrix Assisted Laser Desorption-Ionisation-Time Of Flight) analysis, performed on MassARRAY Platform.</i>
Data source location	<i>Debrecen, Hungary, Latitude: 47.544062, 21° 38' 25'' E &amp; Longitude: 21.64283, 47° 32' 33'' N</i>
Data accessibility	<i>Data are presented in this article; DNA sample and raw data are available for further analyses in collaborative studies</i>

## Value of the data

- Several studies describe the health status of the Roma, which constitutes the largest ethnic minority in Europe however studies focusing on their genetic predisposition to common chronic diseases are scarce.
- Genetic background of atherosclerosis among Roma as well as the general Hungarian population can be studied separately or in international cohort.
- Genetic risk score assessment can be further utilized to compare susceptibility to reduced HDL-C level among different population groups.

## 1. Data

Distribution of SNPs related to HDL-C level were analysed in the Hungarian Roma and general populations and weighted Genetic Risk Scores were defined and used to categorize the population into quintiles.

## 2. Experimental design, materials and methods

### 2.1. Subjects

Study involved subjects of samples investigated during recent cross sectional surveys [2,3]. The Roma sample is representative to the Roma population living settlements in North-East Hungary in terms of age and sex and includes 646 individuals (Roma). The “General” sample consisting of 1542

individuals representative for the Hungarian general population in terms of geographic, age and sex distributions.

## 2.2. DNA extraction

DNA was isolated 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.

## 2.3. SNP selection

A systematic literature review on the PubMed, HuGE Navigator and Ensembl databases was conducted to identify SNPs most strongly associated with HDL-C metabolism (Table 1 and Fig. 1). The selection process of the SNPs is demonstrated in detail in our research article [1].

## 2.4. Genotyping

Genotyping was performed on a MassARRAY platform (Sequenom Inc., San Diego, CA, USA) with iPLEX Gold chemistry. Validation, concordance analysis and quality control were conducted by the facility according to their protocols.

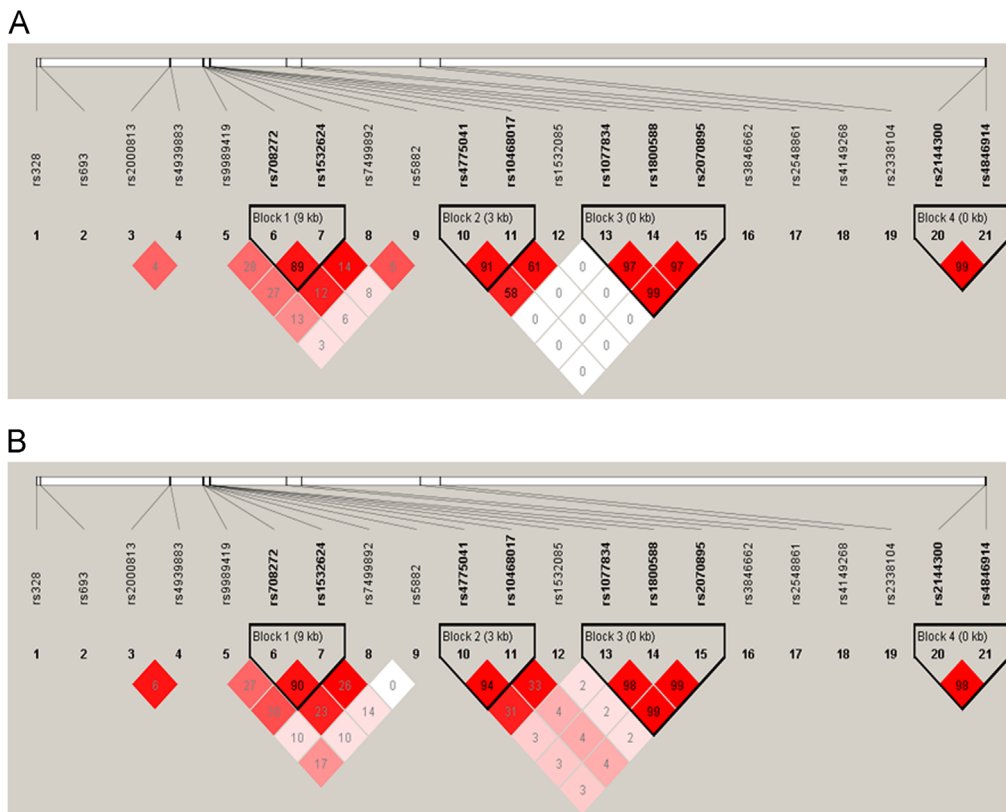
## 2.5. Statistical analyses

Two-sided t tests were used to compare the distribution of genetic risk scores in populations. To reveal the association between genetic risk, serum HDL-C level and ethnicity several statistical models were used (Tables 2–6).

**Table 1**

List of the SNPs which were involved in the research.

Nearest Gene	Gene (short)	SNP (rs number)	Chromosome
Apolipoprotein B	APOB	rs693	2
ATP-binding cassette transporter ABCA1	ABCA1	rs4149268	9
Cholesteryl ester transfer protein	CETP	rs1532624	16
Cholesteryl ester transfer protein	CETP	rs5882	16
Cholesteryl ester transfer protein	CETP	rs708272	16
Cholesteryl ester transfer protein	CETP	rs7499892	16
Cholesteryl ester transfer protein	CETP	rs9989419	16
Endothelial lipase	LIPG	rs2000813	18
Endothelial lipase	LIPG	rs4939883	18
Hepatic lipase	LIPC	rs10468017	15
Hepatic lipase	LIPC	rs1077834	15
Hepatic lipase	LIPC	rs1532085	15
Hepatic lipase	LIPC	rs1800588	15
Hepatic lipase	LIPC	rs2070895	15
Hepatic lipase	LIPC	rs4775041	15
HMG-CoA Reductase	HMGCR	rs3846662	5
Lipoprotein lipase	LPL	rs328	8
Polypeptide N-acetylgalactosaminyltransferase 2	GALNT2	rs2144300	1
Polypeptide N-acetylgalactosaminyltransferase 2	GALNT2	rs4846914	1
Potassium channel tetramerization domain containing 10	KCTD10	rs2338104	12
WW Domain Containing Oxidoreductase	WWOX	rs2548861	16



**Fig. 1.** Haplotype block organization of SNPs related to high-density lipoprotein cholesterol level on the LD maps for the Hungarian general (A) and Roma (B) populations. 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 are 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  $LOD \geq 2$ , and bright red  $D' = 1$  and  $LOD \geq 2$ ). Numbers in squares are  $D'$  values.

**Table 2**  
Distribution of study populations by wGRS quintiles.

	Hungarian general population (%)	Hungarian Roma population (%)	<i>p</i> -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	34.76	0.037
5th quintile of wGRS (0.75–0.88)	2.61	5.14	0.004

**Table 3**

Output of multiple regression models using unweighted and weighted genetic risk scores as dependent variable and ethnicity, age and sex as independent variables.

Dependent variable: GRS		R Square=0.009	
Independent variables	Coefficient	p-value	$\beta$
Ethnicity (Roma vs. General)	0.667	< 0.001	0.092
Sex women vs. men)	0.106	0.477	0.016
Age	-0.0003	0.068	-0.001
$\beta$ : relative strength of predictors			
Dependent variable: wGRS		R Square=0.017	
Independent variables	Coefficient	p-value	$\beta$
Ethnicity (Roma vs. General)	0.029	< 0.001	0.125
Sex (women vs. men)	-0.001	0.774	-0.006
Age	-0.0002	0.202	-0.028
$\beta$ : relative strength of predictors			

Multivariate regression analysis using age, sex as covariates did not change the inference neither for the GRS nor for wGRS.

**Table 4**

Proportion of subjects with reduced plasma HDL-C level in the General and Roma populations according to wGRS quintiles.

	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)	p-values for trend
<b>General (Men; Women)</b>	<b>N=26 (8;18)</b>	<b>N=241 (115;126)</b>	<b>N= 681 (320;361)</b>	<b>N=417 (200;217)</b>	<b>N= 36 (21;15)</b>	
Average HDL-C level (mmol/l)	1.56	1.47	1.41	1.38	1.33	0.021
Reduced plasma HDL-C (%)	11.54	23.77	27.8	28.5	31.43	0.083
<b>Roma (Men; Women)</b>	<b>N= 3 (1;2)</b>	<b>N=61 (28;33)</b>	<b>N=287 (112;175)</b>	<b>N=203 (76;127)</b>	<b>N= 30 (11;19)</b>	
Average HDL-C level (mmol/l)	1.26	1.24	1.23	1.2	1.09	0.076
Reduced plasma HDL-C (%)	33.33	44.26	49.83	52.22	56.67	0.054

**Table 5**

Association of GRSs with plasma HDL-C<sup>a</sup> level by study groups.

	Hungarian General		Hungarian Roma	
	$\beta$ (95% CI)	p-value	$\beta$ (95% CI)	p-value
<b>GRS</b>				
<b>Model I</b>	-0.01 (-0.018 to -0.003)	0.004	-0.013 (-0.023 to -0.003)	0.011
<b>Model II</b>	-0.011 (-0.018 to -0.004)	0.003	-0.013 (-0.023 to -0.003)	0.009
<b>wGRS</b>				
<b>Model III</b>	-0.243 (-0.466 to -0.020)	0.033	-0.318 (-0.633 to -0.002)	0.049
<b>Model IV</b>	-0.205 (-0.420 to 0.101)	0.062	-0.336 (-0.651 to -0.21)	0.036

The association of GRS and wGRS with plasma HDL-C level were evaluated under unadjusted regression models (Model I and III) and under regression models adjusted for age and sex (Model II and IV) separately in Roma and general subjects. In all models the HDL-C was the dependent variable, the GRS/wGRS were the independent variables.

95% CI: 95% confidence interval

<sup>a</sup> HDL-C values were non-normally distributed and were transformed using a two-step approach suggested by Templeton [4].

**Table 6**

The association of HDL-C level with genetic risk scores adjusted by ethnicity, sex and age.

Dependent variable: reduced plasma HDL-C level		R Square=0.046
<b>Independent variables</b>	<b>OR (95% CI)</b>	<b>p-value</b>
<b>Genetic constitution defined by GRS</b>	1.07 (1.04–3.31)	< 0.001
<b>Ethnicity (Roma vs. General)</b>	2.70 (2.19–3.31)	< 0.001
<b>Sex (women vs. men)</b>	0.99 (0.81–1.20)	0.942
<b>Age</b>	1.00 (0.99–1.01)	0.393
Dependent variable: reduced plasma HDL-C level		R Square=0.042
<b>Independent variables</b>	<b>OR (95% CI)</b>	<b>p-value</b>
<b>Genetic constitution defined by wGRS</b>	3.89 (1.56–9.69)	0.004
<b>Ethnicity (Roma vs. General)</b>	2.69 (2.19–3.31)	< 0.001
<b>Sex (women vs. men)</b>	1.00 (0.83–1.21)	0.993
<b>Age</b>	1.00 (0.99–1.01)	0.353

OR: odds ratio.

## Ethical approval

All procedures performed in 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 or comparable ethical standards.

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 and 8907-O/2011-EKU).

This article does not contain any studies with animals performed by any of the authors.

## Acknowledgements

The project was co-financed by the European Union under the European Social Fund (TÁMOP 4.2.1. B-09/1/KONV-2010-0007 and TÁMOP 4.2.2.A-11/1/KONV-2012-0031) and European Regional Development Fund (GINOP-2.3.2-15-2016-00005), as well as by the Hungarian Academy of Sciences (MTA11010 and TK2016-78).

## Transparency document. Supplementary material

Transparency data document associated with this article can be found in the online version at doi: <http://dx.doi.org/10.1016/j.atherosclerosis.2017.05.028>.

## Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi: <http://dx.doi.org/10.1016/j.atherosclerosis.2017.05.028>.

## References

- [1] P. Píró, S. Fiatal, Z. Kosa, J. Sandor, R. Adany, Genetic factors exist behind the high prevalence of reduced high-density lipoprotein cholesterol levels in the Roma population, *Atherosclerosis* 263 (2017) 119–126.
- [2] E. Szigethy, G. Szeles, A. Horvath, T. Hidvegi, G. Jermendy, G. Paragh, Z. Voko, Epidemiology of the metabolic syndrome in Hungary, *Public Health* 126 (2012) 143–149.
- [3] Z. Kosa, A. Moravcsik-Kornyicki, J. Dioszegi, B. Roberts, Z. Szabo, J. Sandor, R. Adany, Prevalence of metabolic syndrome among Roma: a comparative health examination survey in Hungary, *Eur. J. Public Health* 25 (2015) 299–304.
- [4] G.F. Templeton, A two-step approach for transforming continuous variables to normal: implications and recommendations for is research, *Commun. Assoc. Inf. Syst.* (2011) 28 (Article 4).