The genetic risk for hypertension is lower among the Hungarian Roma population compared to the general population

Beáta Soltezs1, Péter Pikó2, János Sándor3,4, Zsigmond Kósa5, Róza Ádány2,3,4, Szilvia Fiatal3,4,*

1 Doctoral School of Health Sciences, Department of Preventive Medicine, Faculty of Public Health, University of Debrecen, Debrecen, Hungary, 2 MTA-DE Public Health Research Group of the Hungarian Academy of Sciences, Faculty of Public Health, University of Debrecen, Debrecen, Hungary, 3 Department of Preventive Medicine, Faculty of Public Health, University of Debrecen, Debrecen, Hungary, 4 WHO Collaborating Centre on Vulnerability and Health, Department of Preventive Medicine, Faculty of Public Health, University of Debrecen, Debrecen, Hungary, 5 Department of Health Visitor Methodology and Public Health, Faculty of Health, University of Debrecen, Nyíregyháza, Hungary

* fiatal.szilvia@sph.unideb.hu

Abstract

Estimating the prevalence of cardiovascular diseases (CVDs) and risk factors among the Roma population, the largest minority in Europe, and investigating the role of genetic or environmental/behavioral risk factors in CVD development are important issues in countries where they are significant minority. This study was designed to estimate the genetic susceptibility of the Hungarian Roma (HR) population to essential hypertension (EH) and compare it to that of the general (HG) population. Twenty EH associated SNPs (in AGT, FMO3, MTHFR-NPPB, NPPA, NPPA-AS1, AGTR1, ADD1, NPR3-C5orf23, NOS3, CACNB2, PLCE1, ATP2B1, GNB3, CYP1A1-ULK3, UMOD and GNAS-EDN3) were genotyped using DNA samples obtained from HR (N = 1176) and HG population (N = 1178) subjects assembled by cross-sectional studies. Allele frequencies and genetic risk scores (unweighted and weighted genetic risk scores (GRS and wGRS, respectively) were calculated for the study groups and compared to examine the joint effects of the SNPs. The susceptibility alleles were more frequent in the HG population, and both GRS and wGRS were found to be higher in the HG population than in the HR population (GRS: 18.98 ± 3.05 vs. 18.25 ± 2.97, p<0.001; wGRS: 1.4 [IQR: 0.93–1.89] vs. 1.52 [IQR: 0.99–2.00], p<0.01). Twenty-seven percent of subjects in the HR population were in the bottom fifth (GRS ≤ 16) compared with 21% of those in the HG population. Thirteen percent of people in the HR group were in the top fifth (GRS ≥ 22) of the GRS compared with 21% of those in the HG population (p<0.001), i.e., the distribution of GRS was found to be left-shifted in the HR population compared to the HG population. The Roma population seems to be genetically less susceptible to EH than the general one. These results support preventive efforts to lower the risk of developing hypertension by encouraging a healthy lifestyle.
Introduction

Under the leadership of the WHO, all Member States agreed in 2013 on global mechanisms to reduce the avoidable burden of noncommunicable diseases (NCDs). This plan aims to reduce the number of premature deaths from NCDs by 25% through nine voluntary global targets by 2025. Two of the global targets directly focus on preventing and controlling cardiovascular diseases (CVDs), which are the number one cause of death globally [1]. Elevated blood pressure is one of the most important independent modifiable risk factors for CVDs. Essential hypertension (EH) is the common form of hypertension and constitutes 95% of all hypertension cases [2]. Hungary still has one of the worst CVD mortality profiles when compared to other European Union countries [3]. According to recent WHO data, the early death caused by CVDs is three times more in Hungary than the European Union [4].

In our recent health examination survey [5], in which the prevalence of metabolic syndrome and its components were defined, the prevalence of hypertension was significantly lower among the Hungarian Roma (HR) population compared to the general Hungarian (HG) population (40.40% vs. 48.44%, p < 0.001, respectively). Furthermore, a significant difference between the Hungarian Roma population and the general Hungarian population among males (43.08% vs. 53.01%, p = 0.007, respectively) but only a distinct trend toward significance among females (38.68% vs. 44.32%, p = 0.063, respectively) was found.

Roma is the largest ethnic group in Europe, with an estimated population of 10–12 million [6]. Hungary is one of the countries with the highest share of the Roma population. It was estimated by the latest national census in 2011 that approximately 3.2% of the total population of the country is Roma; notwithstanding, their estimated representation is much higher, up to 8.7% of the total population [7].

Several studies estimated the prevalence of hypertension among the Roma population compared to the majority population; however, the results were not fully consistent. Studies pointed out the lower hypertension prevalence in the Roma population compared to the general population of Romania and Croatia [8] [9] [10] [11]. Other studies, carried out in Slovakia, Spain and Italy, did not find any differences in the prevalence of hypertension among the Roma population compared to the general population [12] [13] [14] [15] [16].

EH is considered to be a polygenic and multifactorial disease influenced by environmental determinants (salt and fat intake; physical inactivity) [17]. The genetic contribution to blood pressure variation ranges from approximately 30 to 50% based on the heritability estimates of family and twin studies [18] [19]. Blood pressure (BP) is controlled by a complex network of interacting biochemical and physiological pathways (involving cardiac contractility, extracellular fluid volume homeostasis, and vascular tone through neural, renal or endocrine systems). The knowledge of these systems offered the opportunity to investigate the possible role of genes encoding proteins using a candidate gene approach [20] [21] [22]. Numerous GWASs have discovered variants in different loci within the genome that are significantly associated with blood pressure and hypertension [23]. Individual risk alleles found in GWASs and candidate gene studies explain only a very small proportion of the variation in systolic and diastolic blood pressure; consequently, the predictive value of the individual alleles for the risk of the trait are limited. Summarizing the multiple susceptible and protective alleles of single nucleotide polymorphisms (SNPs) by polygenic/genetic risk score (GRS) computation offers the opportunity to estimate the genetic load on different diseases on both the individual and population levels [24] [25] [26].

Our intention was to investigate whether disparities exist in the cumulative risk allele loads between the Hungarian Roma and the general populations, which may at least to a certain extent elucidate ethnic differences in hypertension prevalence. According to findings of this
study preventive interventions such as population-wide life-style modification and the use of genetic information for risk stratification should be specifically tailored in case of the population where genetic susceptibility to essential hypertension is found to be more defined.

**Materials and methods**

**Study design**

The subjects in this study included 1176 Hungarian Roma individuals living in segregated settlements in Northeast Hungary and 1178 individuals from the Hungarian general population. Roma subjects were derived from cross-sectional studies. Details of the sample collection can be found elsewhere [5] [27]. General individuals were collected through the country’s population-based disease registry called the General Practitioners’ Morbidity Sentinel Stations Programme (GPMSSP) using a stratified multistage sampling method.

**Characterization of the study populations**

- **Hungarian general population.** The GPMSSP, established in 1998 to monitor the occurrence of chronic noncommunicable diseases of great public health importance, provided an adequate method for generating the Hungarian reference sample in a cost-effective way [28]. In the initial phase of the Programme, only four counties were involved (Hajdú-Bihar, Győr-Moson-Sopron, Szabolcs-Szatmár-Bereg and Zala Counties). However, later, the Programme was extended to additional regions of the country (two counties from Central Hungary, Komárom-Esztergom and Bács-Kiskun, Baranya from Southern Transdanubia, and Heves from Northern Hungary) [29].

  Participants in this present study were collected at the initial four-county phase of the Programme. The source population of the study involved males and females older than 20 years of age and represented the Hungarian adult population on the basis of geographic, age and gender distributions. As a part of the initial survey (demographic and anthropometric data were collected); blood samples were also taken for routine laboratory tests and DNA isolation. On the whole, 1196 blood samples were collected for DNA preparation, out of which 1178 samples yielded good quality DNA for our present study.

- **Hungarian Roma population.** The DNA samples of the Roma population were obtained from two cross-sectional surveys. First, a comparative health examination survey, which enrolled Roma from North-East Hungary (Hajdú-Bihar and Szabolcs-Szatmár-Bereg counties) where the majority of Roma colonies can be found, was utilized. Details of the sampling methodology and the data collection are described elsewhere [5]. As a part of the survey, medical histories and sociodemographic characteristics were recorded, and a physical examination was performed for all participants. Blood samples were taken for laboratory investigations and genotype assessments.

  Second, an additional source of the Roma samples was the recently launched ‘Public Health Focused Model Programme for Organising Primary Care Services Backed by a Virtual Care Service Centre’, which was developed in the framework of the Swiss-Hungarian Cooperation Programme [27]. The intervention area of the Programme is found in the two most disadvantaged regions of Hungary (North Hungary and the North Great Plain). In these regions (Hajdú-Bihar, Borsod-Abaúj-Zemplén, Jász-Nagykun-Szolnok and Heves counties), one of the services that is delivered by the GPs’ cluster is a health status assessment, which provided the opportunity to further increase the collection of DNA samples from the Roma population. Altogether, 1292 samples representative of the Roma population living in Northern-East Hungary by age and gender are available in our repository. The DNA pool in this study consists of 1176 randomly selected Roma individuals.
DNA isolation

DNA extraction was performed from EDTA-anticoagulated whole blood samples using a MagNA Pure LC system (Roche Diagnostics GmbH, Mannheim, Germany) with a MagNA Pure LC DNA Isolation Kit–Large Volume (Cat. No. 03310515001, Roche Diagnostics GmbH, Mannheim, Germany) according the manufacturer’s instructions. Extracted DNA was eluted in 200 μl MagNA Pure LC DNA Isolation Kit–Large Volume Elution Buffer.

SNP literature search and selection criteria

Systematic literature search on the PubMed and HuGE Navigator [30] databases and on the NHGRI-EBI GWAS Catalogue [31] was conducted to identify the single-nucleotide polymorphisms (SNPs) most strongly associated with EH. Details of the systematic review search can be found in the supplementary material (Fig 1 in S1 File). The following keywords and all possible combinations were used for searching: essential hypertension, blood pressure, molecular genetics, genomics, genes, single-nucleotide polymorphism, genetic variants, gene polymorphism, common gene variants, genome-wide association study (GWAS), candidate gene study, case-control study, meta-analysis, review, association.

Studies were selected if they met the following criteria: (1) investigated the association between SNPs and EH using a statistically acceptable sample size, (2) provided information about susceptibility/protective alleles of the SNPs, (3) evaluated EH as an outcome and excluded secondary forms of hypertension or other types of monogenic hypertension, (4) defined EH as systolic blood pressure higher than 140 mmHg and/or diastolic blood pressure higher than 90 mmHg, or study individuals obtained antihypertensive medication, (5) full texts were available and written in the English language and (6) conducted in humans. Additional studies were also examined by reviewing references of the selected articles.

The adequate sample size for the study groups was computed using the online calculator OSSE (http://osse.bii.a-star.edu.sg/calculation1.php), assuming a power of 80% and an alpha level of 0.05 for a 1:1 case to control ratio. The allele frequencies for CEU (Utah Residents (CEPH) with Northern and Western Ancestry) and for GIH (Gujarati Indian from Houston, Texas) populations from the 1000 Genomes Project, Phase 3, were applied in the sample size estimation considering the fact that the Roma population arrived at the Balkans from North India and then migrated to Europe [32].

As a result of the systematic literature search, 30 SNPs were identified (Table 1, Step 1 in S1 File). During the assay design, a pool of 23 SNPs was created for genotyping by the Mutation Analysis Core Facility of the Karolinska University Hospital (Sweden). Based on data obtained in the genotyping process, 20 SNPs were chosen for allele frequency comparison and GRS computation. Effect size estimates from GWASs were available for 19 SNPs. Finally, 19 SNPs were included in the computation of the wGRS (see details of SNPs’ selection process in Table 1, Steps 1–5 in S1 File).

Genotyping

The genotyping was performed on a Mass-ARRAY platform (Sequenom Inc., San Diego, CA, USA) with iPLEX Gold chemistry [47]. The validation of assays, concordance analysis and quality control analysis were conducted by the facility according to their standard protocols. The total genotyping success rate was 97.8%, resulting in genotype information for 2343 individuals (1167 general and 1176 Roma).
Table 1. List of SNPs and their loci with the effect alleles and effect size estimates included in the study.

<table>
<thead>
<tr>
<th>SNP ID</th>
<th>Locus*</th>
<th>Chromosome</th>
<th>Functional consequence^b</th>
<th>IUPAC code followed by nucleic acid changec</th>
<th>Effect allele</th>
<th>Genetic model applied in the publication</th>
<th>Published effect for weighting</th>
<th>Type of SNP</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4762</td>
<td>AGT</td>
<td>1</td>
<td>missense variant</td>
<td>r = G&gt;A</td>
<td>T per allele</td>
<td>1.19 (1.07–1.33; 0.002)</td>
<td>NA candidate</td>
<td></td>
<td>[33] [34]</td>
</tr>
<tr>
<td>rs5049</td>
<td>AGT</td>
<td>1</td>
<td>2kb upstream variant</td>
<td>Y = C&gt;T</td>
<td>A per allele</td>
<td>1.37 (1.17–1.59; 0.000006)</td>
<td>NA candidate</td>
<td></td>
<td>[33]</td>
</tr>
<tr>
<td>rs6999</td>
<td>AGT</td>
<td>1</td>
<td>missense variant</td>
<td>R = A&gt;G</td>
<td>C per allele</td>
<td>1.20 (1.11–1.29; &lt;0.0001)</td>
<td>NA candidate</td>
<td></td>
<td>[35] [34]</td>
</tr>
<tr>
<td>rs2266782</td>
<td>FMO3</td>
<td>1</td>
<td>missense variant</td>
<td>r = G&gt;A</td>
<td>A NA NA NA NA candidate</td>
<td>[33]</td>
<td></td>
<td>[36]</td>
<td></td>
</tr>
<tr>
<td>rs17367504</td>
<td>MTHFR-NPPB</td>
<td>1</td>
<td>intron variant</td>
<td>R = A&gt;G</td>
<td>G per allele</td>
<td>NA -0.103 (-; 2.3x10^-16) GWAS [37]</td>
<td>[33]</td>
<td></td>
<td>[38]</td>
</tr>
<tr>
<td>rs5068</td>
<td>NPPA</td>
<td>1</td>
<td>3 prime UTR variant</td>
<td>D = A&gt;G,T</td>
<td>G per allele</td>
<td>0.85 (0.79–0.92; 4x10^-7)</td>
<td>NA candidate</td>
<td>[33]</td>
<td></td>
</tr>
<tr>
<td>rs198358</td>
<td>NPPA-AS1</td>
<td>1</td>
<td>non coding transcript variant</td>
<td>y = T&gt;C</td>
<td>C per allele</td>
<td>0.90 (0.85–0.95; 2x10^-7)</td>
<td>NA candidate</td>
<td>[33]</td>
<td></td>
</tr>
<tr>
<td>rs5186</td>
<td>AGTR1</td>
<td>3</td>
<td>3 prime UTR variant</td>
<td>M = A&gt;C</td>
<td>C recessive model (CC vs. AC + AA)</td>
<td>7.3 (1.9–31.9; 0.0005)</td>
<td>NA candidate</td>
<td>[33]</td>
<td></td>
</tr>
<tr>
<td>rs4961</td>
<td>ADD1</td>
<td>4</td>
<td>missense variant</td>
<td>E = G&gt;A,T</td>
<td>T dominant model (TT + GT vs. GG)</td>
<td>1.60 (1.32–1.92; 1.09x10^-6)</td>
<td>NA candidate</td>
<td>[40] [41]</td>
<td></td>
</tr>
<tr>
<td>rs1173771</td>
<td>NPR3-C5orf23</td>
<td>5</td>
<td>~20kb of both C5orf23 and NPR3</td>
<td>R = A&gt;G</td>
<td>G per allele</td>
<td>NA 0.062 (-; 5.2x10^-16) GWAS [37]</td>
<td>[33]</td>
<td></td>
<td>[38]</td>
</tr>
<tr>
<td>rs1799983</td>
<td>NOS3</td>
<td>7</td>
<td>missense variant</td>
<td>F = T&gt;A,G</td>
<td>T per allele</td>
<td>1.038 (1.034–1.043; 2.63x10^-16)</td>
<td>NA GWAS</td>
<td>[33]</td>
<td></td>
</tr>
<tr>
<td>rs2070744</td>
<td>NOS3</td>
<td>7</td>
<td>intron variant</td>
<td>Y = C&gt;T</td>
<td>C per allele</td>
<td>1.04 (1.038–1.041; 6.42x10^-6)</td>
<td>NA GWAS</td>
<td>[33]</td>
<td></td>
</tr>
<tr>
<td>rs1813353</td>
<td>CACNB2(3')</td>
<td>10</td>
<td>intron variant</td>
<td>y = T&gt;C</td>
<td>T per allele</td>
<td>NA 0.078 (-; 6.2x10^-16) GWAS [37]</td>
<td>[33]</td>
<td></td>
<td>[38]</td>
</tr>
<tr>
<td>rs4373814</td>
<td>CACNB2(5')</td>
<td>10</td>
<td>~10kb 5' of CACNB2</td>
<td>L = G&gt;C,T</td>
<td>G per allele</td>
<td>NA -0.046 (-; 8.5x10^-5) GWAS [37]</td>
<td>[33]</td>
<td></td>
<td>[38]</td>
</tr>
<tr>
<td>rs932764</td>
<td>PLCE1</td>
<td>10</td>
<td>intron variant</td>
<td>R = A&gt;G</td>
<td>G per allele</td>
<td>NA 0.055 (-; 9.4x10^-8) GWAS [37]</td>
<td>[33]</td>
<td></td>
<td>[38]</td>
</tr>
<tr>
<td>rs2681472</td>
<td>ATP2B1</td>
<td>12</td>
<td>intron variant</td>
<td>R = A&gt;G</td>
<td>A per allele</td>
<td>NA 0.15 (0.02; 1.75x10^-11) GWAS [24]</td>
<td>[33]</td>
<td></td>
<td>[38]</td>
</tr>
<tr>
<td>rs5443</td>
<td>GNB3</td>
<td>12</td>
<td>synonymous variant</td>
<td>Y = C&gt;T</td>
<td>T per allele</td>
<td>2.3 (1.7–3.3; 0.00002)</td>
<td>NA candidate</td>
<td>[33]</td>
<td></td>
</tr>
<tr>
<td>rs1378942</td>
<td>CYP1A1-ULK3</td>
<td>15</td>
<td>intron variant</td>
<td>I = C&gt;A,T</td>
<td>C per allele</td>
<td>NA 0.073 (-; 1.0x10^-8) GWAS [37]</td>
<td>[33]</td>
<td></td>
<td>[38]</td>
</tr>
<tr>
<td>rs13333226</td>
<td>UMOD</td>
<td>16</td>
<td>intron variant</td>
<td>R = A&gt;G</td>
<td>G per allele</td>
<td>0.87 (0.84–0.91; 3.6x10^-11)</td>
<td>NA GWAS</td>
<td>[33]</td>
<td></td>
</tr>
<tr>
<td>rs6015450</td>
<td>GNAS-EDN3</td>
<td>20</td>
<td>intron variant</td>
<td>R = A&gt;G</td>
<td>G per allele</td>
<td>NA 0.11 (-; 4.2x10^-14) GWAS [37]</td>
<td>[33]</td>
<td></td>
<td>[38]</td>
</tr>
</tbody>
</table>

^Locus data for each of the SNPs listed in the table are derived from the abovementioned references, according to gene(s) reported by the authors.

^Data on the functional consequences of SNPs were derived from dbSNP Build 153 database [45]. However in case of two SNPs the following data were used: in case of rs1173771 and rs4373814 SNPs the data defined by the references were used.

^The alleles for each of the of SNPs were extracted from the dbSNP Build 153 database [45], then the IUPAC codes of SNPs were defined manually based on Johnson, 2010 [46].

NA—not applicable

[33] [34] [35] [36] [37] [38] [39] [40] [41] [42] [43] [44] [45] [46]

References:

https://doi.org/10.1371/journal.pone.0234547.t001
Statistical analysis

Statistical analyses were performed using PLINK (version 1.07), STATA (version 12), Haploview (version 4.1) and MS Excel (version 2016) software. Age, gender, and body mass index (BMI) were assigned as confounding variables. A Shapiro-Wilk test for normality was performed. The nonparametric Kolmogorov-Smirnov test was used to compare the age and BMI values of the study groups. Allele frequencies were calculated on the basis of the obtained genotype distributions. The existence of Hardy-Weinberg equilibrium, the differences in allele distribution and the gender distribution were investigated with $\chi^2$ tests. To correct for multiple testing, the Meff (effective number of independent marker loci) method was applied, which takes into account the correlation between the SNPs. For this calculation, the online software SNPSpD was used (available at http://neurogenetics.qimrberghofer.edu.au/SNPSpD) [48]. A corrected p-value of 0.0025 was obtained. Generally, a conventional p-threshold of 0.05 was applied. Linkage disequilibrium (LD) between polymorphisms was defined using the Haploview software [49]. The pairwise LD between the markers was measured and visualized in the format of $r^2$ (squared coefficient of correlation), with a threshold of ≥0.8 for the study populations (see Fig 1). In order to investigate whether the association between genetic risk and ethnicity depends on the influence of other factors (age, gender, BMI were available), multivariate linear regression analyses were conducted (see details in Computations using GRS and wGRS values section). Furthermore, using blood pressure phenotype data—which was available only in case of the Roma sample—multivariate logistic and linear regression models were developed to analyze the association of GRSs with hypertension (as binary outcome) and with systolic or diastolic blood pressure (as continuous outcome), in unadjusted and adjusted models (see details in Computations using GRS and wGRS values section).

![Fig 1. LD pattern of SNPs associated with hypertension for the HG (upper) and HR (lower) populations.](https://doi.org/10.1371/journal.pone.0234547.g001)
Linkage analysis were conducted separately in the study populations. According to the LD map which generated by Haploview software (version 4.1), there were not observed multicollinearity between the polymorphisms, based on the LD pattern none of the pairwise LD of the studied SNPs reached the $r^2$ threshold of ≥0.8, thus it was not necessary to prune SNP from the analysis. The numbers above the LD plot show the rs numbers of SNPs. Numbers in squares are $r^2$ values. The colour scheme is the $r^2$ colour scheme (white $r^2 = 0$, shades of grey $0 < r^2 < 1$).

**Computations using GRS and wGRS values**

To obtain the combined effect of the selected SNPs, unweighted (GRSs) and weighted genetic risk scores (wGRSs) were calculated. For subjects whose SNP genotype data were incomplete, an expected value was imputed using the effect allele frequency observed in study populations (according to the PLINK software manual SNP scoring routine chapter, available at [http://zzz.bwh.harvard.edu/plink/dist/plink-doc-1.07.pdf](http://zzz.bwh.harvard.edu/plink/dist/plink-doc-1.07.pdf)). Variants, their effect alleles and weights included in the GRS and wGRS computation are depicted in Table 1.

In the GRS computation, each person was assigned a score based on the number of risk alleles carried. Thus, risk allele homozygotes were coded as genotype “2”, heterozygotes were coded as genotype “1”, and “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 other allele homozygotes [50]. By using these codes, a simple count score (unweighted GRS) was computed as described by Eq (1), in which $G_i$ is the number of risk alleles for the $i$th SNP. This model sums all risk alleles over all loci as a summary score assuming that all alleles have the same effect.

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

In the weighted approach (wGRS), rather than giving equal weight to each SNP, SNPs with larger effects contributed more to the score. The effect size measures were utilized from studies that published significant effect size measure results on population samples of European subjects. Risk estimates for alleles were available for 19 SNPs in 16 genes or loci (Table 1). Formerly, we showed that the effect of the vast majority of the SNPs on HDL-C levels could be replicated in the HG and Roma populations, which indicates that the effect size measurements obtained from the literature on European populations can be used for risk estimation for the Roma population [51]. In the case of the rs4762 variant, the published “pooled OR” (OR of European + Asian + Mixed populations) was used because the data for Europeans were not significant. In addition, in the same meta-analysis, the rs5049 variant was described mainly among Asians, and no evidence for heterogeneity of the effect size measure across studies was reported [33]. The computation of the wGRS is described by Eq (2). In this weighted score, weights ($w_{β_i}$) were derived from the risk coefficient for each allele based on beta values (or odds ratio if applicable, as the natural logarithm of the published OR for hypertension). These weights ($w_{β_i}$) were multiplied by 0, 1 or 2 according to the number of effect alleles carried by each person ($X_i$) [50] [52] [53]. In the case of two SNPs (rs4961 and rs5186), different allele coding was needed because the risk coefficients were reported for one or two copies of each effect allele as a single group (in the case of rs4961, the T allele was coded in a dominant model TT + GT vs. GG; in the case of rs5186, the C allele was coded in a recessive model CC vs. AC + CC). Consequently, the risk coefficients were multiplied by a score of 1 or 0, respectively [54].

$$wGRS = \sum_{i=1}^{I} w_{β_i}X_i$$  \hspace{1cm} (2)
The nonparametric two-sample Mann-Whitney U test and two-sample Kolmogorov-Smirnov
test were used to compare the distribution of GRS and wGRS, respectively. The $\chi^2$
test was used to compare GRSs divided into quintiles by population. To discover whether the association
between genetic risk and ethnicity depends on the influence of other factors (age, gender, BMI
were available), multivariate linear regression analyses were conducted in which GRSs were the
dependent variable, while ethnicity, gender, age and BMI were considered as independent vari-
ables. For the wGRSs, age and BMI values were non-normally distributed in the study populations
and were therefore transformed based on the two-step method described by Templeton [55].

To investigate whether the SNPs are suitable for GRS computation, regression models were
applied. Using blood pressure phenotype data—only in case of the Roma sample were available—
multivariate logistic and linear regression models were developed to analyze the association of
GRSs with hypertension (as binary outcome) and with systolic or diastolic blood pressure (as con-
tinuous outcome), in which Model 1 was unadjusted and Model 2 was adjusted for gender, age
and BMI. Hypertension status was considered as a dependent variable, while GRS, gender, age,
and BMI were considered as independent variables. A binary variable for hypertension status was
defined according to the consensus definition of the International Diabetes Federation (IDF) by a
systolic blood pressure $\geq$130 mm Hg, a diastolic blood pressure $\geq$85 mm Hg or the use of antihy-
pertensive medication. In the multivariate linear regression model for those subjects who were
under antihypertensive treatment control, we applied an imputation method by adding a constant
10 mmHg to the measured systolic blood pressure values and 5 mmHg to diastolic values [56].
The wGRS, GRS, age, BMI, systolic and diastolic blood pressure values were non-normally dis-
tributed in the study populations and were therefore transformed [55].

Ethics statement
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. All subjects
gave their written informed consent for the study. This study was approved by the Ethical
Committee of the Hungarian Scientific Council on Health (reference Nos. ETT-TUKEB 8907-
0/2011-EKU (285/PI/11.) TUKEB 2213-5/2013/EKU (233/2013). This article does not contain
any studies with animals performed by any of the authors.

Results
Characteristics of the study sample
In total, 1176 Roma and 1167 general individuals were included in the analyses. The mean age
was 40.98 ± 12.84 years in the Roma population and 47.31 ± 17.02 years in the HG population.
The mean age of the two study populations was different ($p<0.001$). The proportion of male
individuals was lower in the Roma sample (HR: 41% vs. HG: 46%, $p = 0.01$). The mean body
mass index (BMI) was higher in the Roma sample compared with the HG sample (26.55 ± SD
6.54 kg/m$^2$ vs. 26.10 ± SD 4.88 kg/m$^2$, $p<0.001$).

In the case of rs1799998 and rs3918226, polymorphisms deviated from HWE in the Hun-
garian general population ($p<0.05$) and were thus excluded from further analyses.

Comparison of allele frequencies
Allele frequency differences between the Roma and general populations are significant for 7
SNPs after multiple test correction ($p<0.0025$, Table 2). Of the 5 protective SNPs examined,
only one (the rs13333226 variant of the UMOD gene) was significantly more frequent in the
HG population. The majority of susceptibility alleles were more prevalent in the HG population; the difference in the prevalence of five alleles (the rs5186 allele of the AGTR1 gene, the rs2681472 allele of the ATP2B1 gene, the rs6015450 allele at the GNAS-EDN3 locus, the rs2070744 allele of the NOS3 gene, and the rs932764 allele of the PLCE1 gene) reached a high level of significance ($p < 0.0025$). In the case of five other alleles (the rs4762 and rs5049 alleles of the AGT gene, the rs2266782 allele of the FMO3 gene, the rs1799983 allele of the NOS3 gene, and the rs1173771 allele at the NPR3-C5orf23 locus), the significance was found to be nominal ($p < 0.05$), and the other two SNPs (the rs4961 allele of the ADD1 gene and the rs1813353 allele of the CACNB2(3') gene) did not differ significantly. Three susceptibility variants were more common among the Roma population; the difference reached a high level of significance ($p < 0.0025$) only in the case of the rs1378942 variant at the CYP1A1-ULK3 locus, while only nominal significance ($p < 0.05$) was observed in the case of the rs699 variant of the AGT gene, and the frequency of the rs5443 variant of the GNB3 gene did not differ significantly.

### Linkage analysis

Linkage analysis was conducted separately in the study populations (Fig 1). Based on the LD pattern, none of the pairwise LDs of the studied SNPs reached the $r^2$ threshold of $\geq 0.8$.

### Comparison of genetic risk scores

**GRS results.** The GRS (based on 20 SNPs) for the HR population ranged from 9 to 26. The GRS for HG subjects ranged from 11 to 29. GRS data were normally distributed for the

<table>
<thead>
<tr>
<th>Locus</th>
<th>SNPs</th>
<th>Allele</th>
<th>Hungarian general population (N = 1167)</th>
<th>Hungarian Roma population (N = 1176)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CACNB2(5')</td>
<td>rs4373814</td>
<td>G</td>
<td>52.29</td>
<td>55.01</td>
<td>0.069</td>
</tr>
<tr>
<td>MTHFR-NPPB</td>
<td>rs17367504</td>
<td>G</td>
<td>13.62</td>
<td>11.81</td>
<td>0.070</td>
</tr>
<tr>
<td>NPPA</td>
<td>rs5068</td>
<td>G</td>
<td>4.80</td>
<td>5.56</td>
<td>0.253</td>
</tr>
<tr>
<td>NPPA-AS1</td>
<td>rs198358</td>
<td>C</td>
<td>23.58</td>
<td>25.07</td>
<td>0.249</td>
</tr>
<tr>
<td>UMOD</td>
<td>rs13333226</td>
<td>G</td>
<td>17.26</td>
<td>11.41</td>
<td>$&lt;0.001$</td>
</tr>
</tbody>
</table>

SNPs in bold showed highly significantly ($p < 0.0025$) different frequencies in the two populations after multiple test correction.

https://doi.org/10.1371/journal.pone.0234547.t002
HG population, while non-normally distributed for the HR population. The mean of the gene count score was 18.25 ± 2.97 in the HR group and 18.98 ± 3.05 in the HG group. The study groups were different by GRS distribution according to the Mann-Whitney U test (p < 0.001, Fig 2).

The multivariate linear regression analysis performed on the GRSs confirmed the association between genetic risk and ethnicity independently of the effect of age, gender and BMI (p < 0.001, Table 3, part A).

![Fig 2. The distributions of GRSs in the HG (black) and HR (white) populations were significantly different (p < 0.001).](https://doi.org/10.1371/journal.pone.0234547.g002)

Table 3. The multivariate linear regression analysis performed on GRSs to confirm the association between genetic risk and ethnicity.

<table>
<thead>
<tr>
<th>A) dependent variable: GRS</th>
<th>R Square = 0.0167</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent variables</td>
<td>Coefficient</td>
</tr>
<tr>
<td>ethnicity (General vs. Roma)</td>
<td>-0.742</td>
</tr>
<tr>
<td>gender (men vs. women)</td>
<td>-0.201</td>
</tr>
<tr>
<td>age</td>
<td>-0.001</td>
</tr>
<tr>
<td>BMI</td>
<td>-0.010</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B) dependent variable: wGRS</th>
<th>R Square = 0.0085</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent variables</td>
<td>Coefficient</td>
</tr>
<tr>
<td>ethnicity (General vs. Roma)</td>
<td>-0.125</td>
</tr>
<tr>
<td>gender (men vs. women)</td>
<td>-0.050</td>
</tr>
<tr>
<td>age</td>
<td>-0.0001</td>
</tr>
<tr>
<td>BMI</td>
<td>-0.004</td>
</tr>
</tbody>
</table>

β: relative strength of predictors

The wGRS, age and BMI values were non-normally distributed and were transformed using a two-step approach suggested by Templeton [55]. Multivariate regression analysis using age, gender and BMI as covariates did not change the inference neither for the GRS nor for wGRS.

https://doi.org/10.1371/journal.pone.0234547.t003
The multivariate logistic regression analysis showed that carrying 1 additional risk allele in the Roma study subjects was associated with a 7% increase in the odds of hypertension, independently of the effect of age, gender and BMI (OR = 1.07; 95% CI: 1.02–1.12; p = 0.008, Table 4, part A).

The GRSs were significantly associated with systolic (β = 0.401 mmHg, 95% CI 0.052–0.750, p = 0.024) but not with diastolic blood pressure (β = 0.149 mmHg, 95% CI—0.045–0.344, p = 0.132) in the multivariate linear regression analyses (Table 5, part A).

Twenty-seven percent of subjects in the HR population were in the bottom fifth (GRS < 16) of the gene count score compared with 21% of those in the HG population. Thirteen percent of people in the HR group were in the top fifth (GRS ≥ 22) of the GRSs compared with 21% of those in the HG population (p < 0.001), i.e., the distribution of GRSs was found to be shifted to the left in the HR population compared to the HG population.

The weighted genetic risk score was significantly associated with both systolic (β = 1.906, 95% CI 0.572–3.240, p = 0.005) and diastolic blood pressure (β = 1.035, 95% CI: 0.293–1.778, p = 0.006) in the multivariate linear regression analyses (Table 5, part B).

Table 4. Association between GRSs and hypertension risk in the Hungarian Roma population.

<table>
<thead>
<tr>
<th>A) dependent variable: Hypertension status</th>
<th>model No.</th>
<th>independent variables</th>
<th>OR</th>
<th>95% CI</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1a</td>
<td>GRS</td>
<td>1.05</td>
<td></td>
<td>1.01–1.09</td>
<td>0.027</td>
</tr>
<tr>
<td>Model 2b</td>
<td>GRS</td>
<td>1.07</td>
<td></td>
<td>1.02–1.12</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>gender (men vs. women)</td>
<td>0.91</td>
<td>0.68–1.22</td>
<td>0.530</td>
<td></td>
</tr>
<tr>
<td></td>
<td>age</td>
<td>1.09</td>
<td></td>
<td>1.08–1.11</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>BMI</td>
<td>1.14</td>
<td></td>
<td>1.11–1.17</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B) dependent variable: Hypertension status</th>
<th>model No.</th>
<th>independent variables</th>
<th>OR</th>
<th>95% CI</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1a</td>
<td>wGRS</td>
<td>1.06</td>
<td></td>
<td>0.91–1.25</td>
<td>0.457</td>
</tr>
<tr>
<td>Model 2b</td>
<td>wGRS</td>
<td>1.12</td>
<td></td>
<td>0.92–1.37</td>
<td>0.248</td>
</tr>
<tr>
<td></td>
<td>gender (men vs. women)</td>
<td>0.90</td>
<td>0.67–1.21</td>
<td>0.504</td>
<td></td>
</tr>
<tr>
<td></td>
<td>age</td>
<td>1.09</td>
<td></td>
<td>1.08–1.11</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>BMI</td>
<td>1.14</td>
<td></td>
<td>1.11–1.16</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

*Model 1 is unadjusted
bModel 2 is adjusted for gender, age and BMI

https://doi.org/10.1371/journal.pone.0234547.t004

The multivariate logistic regression analysis showed that carrying 1 additional risk allele in the Roma study subjects was associated with a 7% increase in the odds of hypertension, independently of the effect of age, gender and BMI (OR = 1.07; 95% CI: 1.02–1.12; p = 0.008, Table 4, part A).

The GRSs were significantly associated with systolic (β = 0.401 mmHg, 95% CI 0.052–0.750, p = 0.024) but not with diastolic blood pressure (β = 0.149 mmHg, 95% CI—0.045–0.344, p = 0.132) in the multivariate linear regression analyses (Table 5, part A).

Twenty-seven percent of subjects in the HR population were in the bottom fifth (GRS ≤ 16) of the gene count score compared with 21% of those in the HG population. Thirteen percent of people in the HR group were in the top fifth (GRS ≥ 22) of the GRSs compared with 21% of those in the HG population (p < 0.001), i.e., the distribution of GRSs was found to be shifted to the left in the HR population compared to the HG population.

**wGRS results.** The applicable effect size estimate for the rs2266782 allele in the FMO3 gene was not publicly available; therefore, 19 SNPs were used for the weighted genetic risk score computation (Table 1). The HR study group had a lower weighted genetic risk than the HG group, as the median wGRS in the HR group was 1.40 (IQR: 0.93–1.89), while for the HG individuals, the median wGRS was 1.52 (IQR: 0.99–2.00). The lower HR median suggests that the HR population has a lower weighted genetic risk score on average (p < 0.01). The boxplot of the wGRSs estimated from the study populations (HG and HR) is shown in Fig 3. wGRS data were non-normally distributed with skewness of 0.56 (SE = 0.07) and kurtosis of 0.59 (SE = 0.14) in the HR population and with skewness of 0.79 (SE = 0.07) and kurtosis of 0.91 (SE = 0.14) in the HG population.

The multivariate linear regression model was able to confirm the association between genetic risk and ethnicity independently of the effect of age, gender and BMI (p < 0.01, Table 3, part B).

The multivariate logistic regression analysis was not able to confirm the association of wGRS with hypertension risk in the Hungarian Roma population (OR = 1.12; 95% CI 0.92–1.37; p = 0.248, Table 4, part B).

The weighted genetic risk score was significantly associated with both systolic (β = 1.906, 95% CI 0.572–3.240, p = 0.005) and diastolic blood pressure (β = 1.035, 95% CI: 0.293–1.778, p = 0.006) in the multivariate linear regression analyses (Table 5, part B).
A total 21.51% of Roma subjects were in the bottom quintile (wGRS ≤ -0.54) of the wGRSs compared with 18.51% of those in the general population. In the Roma group, 18.37% of the individuals were in the highest quintile (wGRS ≥ 4.92) of the wGRSs compared with 21.59% of those in the general population (p = 0.029) (Table 2 in S1 File).

**Discussion**

EH is a well-established risk factor for adverse cardiovascular outcomes. Estimating the prevalence of cardiovascular risk factors among the Roma population and investigating the role of genetic or environmental/behavioral/cultural risk factors in the development of CVDs is an important issue in countries where they are significant minorities. Our study is the first to examine the genetic susceptibility of the Roma population to EH by comparing allele frequencies and combining the effect of multiple hypertension-associated alleles into GRSs. In addition, the number of Roma subjects surveyed is relatively high.

Analysis of the biological, environmental, social and physiological domains related to ethnicity is an essential component of the integrative research process that aims to prevent the development of disease, including those that differ in prevalence among ethnic groups. Some
studies have already reported that the cardiovascular risk load in the Roma population differs from the majority population of the country where they live [8] [11]. Only a few studies have investigated the possible role of genetic factors associated with cardiovascular traits [57] [58], and the genetic load of the Roma population related to hypertension has not been investigated until now.

In summary, the susceptibility alleles showing significant differences in their frequency between the two populations were more prevalent in the HG population. In addition to the simple unweighted scores, effect size estimates from GWASs were utilized to model and compare the genetic risk related to hypertension in the two study populations. It was shown that the average risk scores (both GRs and wGRs) were significantly lower among the Roma population compared to the HG population even after adjusting for the effects of possible confounders. The multivariate linear regression analysis performed on the GRs was able to confirm the association between genetic risk and ethnicity independently of the effect of age, gender and BMI. In conclusion, decreased genetic susceptibility to hypertension in the Roma was observed compared to the HG population.

Despite its uniqueness, there are obvious limitations of this study. First, the list of data collected for the representative sample of the general HG was limited because data on blood pressure were not included in the analyses. Consequently, it was not possible to investigate the strengths of the association of different blood pressure phenotypes (systolic and diastolic blood pressure) with the GRs by regression models. Another disadvantage is that the Roma study population was not representative of the overall Roma population of Hungary because those Roma who have assimilated with the general population were excluded by the sampling.
method. However, because many people unwilling to self-report their ethnicity as Roma this pressure would be very difficult to manage. Third, unweighted GRS models have limitations. It assumes that alleles have a separate additive effect, and neither the effect sizes nor their possible interactions are noted. Furthermore, effect of epigenetic factors, or other structural/rare variants, and interactions (gene-environmental and gene-gene) were not taken into consideration in this study; however, it is accepted that these factors can alter genetic risk. It is important to point out that studies are needed to ascertain the impact not only of the genetic components but also of the interactions between inheritable and additionally environmental factors in the interpretation of ethnic disparities.

For the wGRS effect, size measures obtained from the population of European ancestry were applied. According to the catalogue of published genome-wide association studies (NHGRI-EBI), the GWASs related to EH have been conducted mostly on populations of European and African American descent. Additionally, effect size estimates relevant to Roma population from candidate gene studies are unavailable. Although the question as to what extent the effect size measures estimated among Caucasians are relevant for the Roma population in genetic risk assessment for essential hypertension, it is important to refer to our recent study on HDL-C levels associated genetic loci. We demonstrated that effect size estimates obtained from genome wide association studies can be utilized for risk estimation not only for the general population but also in the case of the Roma.

Applying multiple markers in combination instead of using single SNPs with small effect size can be more advantageous if we would like to translate results from genomic studies to population health research. In the genomic era, identifying people or populations who are at high risk of developing common CVDs may augment the advantage of prevention programmes by decreasing the risk of the development of CVDs. It is important to highlight that more research is needed because, so far, only limited evidence is available on the application of genetic/genomic results in public health practice [59]. Our present finding suggests that prevention of EH in the Roma population should focus on harmful environmental or behavioral factors rather than their genetic propensity.

Supporting information
S1 File.
(PDF)

Author Contributions
Conceptualization: Zsигmόnd Kόsα, Rόза Ádάnу, Szόlvıa Fıatal.
Formal analysis: Beáta Soldtész, Péter Pikó.
Investigation: Zsигmόnd Kόsα.
Methodology: Péter Pikó, János Sándor, Zsигmόnd Kόsα.
Supervision: Rόза Ádάnу, Szόlvıa Fıatal.
Validation: Szόlvıa Fıatal.
Visualization: Beáta Soldtész, Péter Pikó, Rόza Ádάnу.
Writing – original draft: Beáta Soldtész, Szόlvıa Fıatal.
Writing – review & editing: Rόza Ádάnу, Szόlvıa Fıatal.
References


