Thrombosis Research 179 (2019) 37–44

Contents lists available at ScienceDirect

Thrombosis Research

journal homepage: www.elsevier.com/locate/thromres

Full Length Article

Genetic profiling revealed an increased risk of venous thrombosis in the Hungarian Roma population

Szilvia Fiatal[a,b,*], Péter Pikó[a,c], Zsigmond Kósa[d], János Sándor[a,b], Róza Ádány[a,b,c]

[a] Department of Preventive Medicine, Faculty of Public Health, University of Debrecen, Debrecen, Hungary
[b] WHO Collaborating Centre on Vulnerability and Health, Department of Preventive Medicine, Faculty of Public Health, University of Debrecen, Debrecen, Hungary
[c] MTA-DE Public Health Research Group of the Hungarian Academy of Sciences, Faculty of Public Health, University of Debrecen, Debrecen, Hungary
[d] Department of Health Visitor Methodology and Public Health, Faculty of Health, University of Debrecen, Nyíregyháza, Hungary

ARTICLE INFO

Keywords:
Genetic predisposition to disease
Roma
Single nucleotide polymorphisms
Venous thrombosis
Genetic risk scores

ABSTRACT

Background: Besides modifiable risk factors, genetic susceptibility may also explain the high cardiovascular disease burden of the Roma population.
Objectives: Aim of this study was to define the genetic susceptibility of Hungarian Roma to venous thrombosis (VT) and comparing it to that of the general population.
Methods: Fifty-two SNPs associated with VT (in *F2*, *F5*, *F9*, *F11*, *F15*, *FGA*, *FGB*, *FGG*, *CYP4V2*, *KLKB1* and *vWF*) were selected and analyzed in the group of Roma ($N = 962$) and general ($N = 1492$) subjects collected by cross-sectional studies. Allele frequencies and genetic risk scores (GRS, unweighted and weighted) were computed for the study groups and compared to estimate the joint effects of SNPs.
Results: The majority of the susceptible alleles were more prevalent in the Roma population, and both GRS and wGRS were found to be significantly higher in Roma than in the general population (GRS: 41.83 ± 5.78 vs. 41.04 ± 6.04; wGRS: 7.78 ± 1.28 vs. 7.46 ± 1.33, $p = .001$). Only 2.39% of subjects in the Roma population were in the bottom fifth of the wGRS (wGRS ≤ 0.19) compared with 3.62% of those in the general population ($p = .080$); 2.88% of the general subjects were in the top fifth of the wGRS (wGRS ≥ 10.02), while 4.26% of the Roma population were ($p = .066$).
Conclusion: In conclusion, the Roma population seems to have increased genetic susceptibility to VT. This might have important implications in the future in identifying possible new opportunities for targeted prevention and treatment for those subgroups in the populations who are at greater risk for VT development.

1. Introduction

Venous thrombosis (VT) is a complex multifactorial disease, and it evolves as a result of a complex interplay between genetic, environmental and behavioral factors [1]. The involvement of inheritable factors was confirmed by family and twin studies. During recent years, a large number of studies identified different single nucleotide polymorphisms (SNPs) that increase or decrease (mostly with modest extent) the relative risk for VT [2], but the majority of them focused on populations with European ancestry and other populations are rarely studied [3,4].

It became known from different studies that combinations of genetic variants improve the predictive ability of susceptibility models related to any cardiovascular trait or disease. The simultaneous use of the most common and strongest risk markers (with or without other non-genetic traditional risk factors) may have the desired discriminatory accuracy to distinguish between high-risk and low-risk subjects and may guide prophylactic and therapeutic decisions [5]. Consequently, genomic profiling has become very popular in studies of common diseases, particularly in studies comparing the genetic load of different ethnic groups [6–9].

According to a recent review, the incidence and prevalence of hereditary thrombophilia (HD) shows geographic and ethnic differences [10]. The Roma ethnic minority group, considered to have a South Asian origin [11] and frequently concentrated in severely deprived regions, is one of the main subjects of ethnicity-based studies in Europe. The main two reasons for this are 1) the Roma population, concentrated mainly in Bulgaria, Hungary, Slovakia and Romania, constitutes the largest ethnic minority in the continent with a population of 10–12 million [12] and 2) this minority represents a genetically isolated population, characterized by a high frequency of consanguinity and inbreeding [13,14]. Although research on the Roma population faces

* Corresponding author at: Department of Preventive Medicine, Faculty of Public Health, University of Debrecen, H-4028 Debrecen, Kassai út 26, Hungary.
E-mail address: fiatal.szilvia@sph.unideb.hu (S. Fiatal).

https://doi.org/10.1016/j.thromres.2019.04.031
Received 7 November 2018; Received in revised form 15 April 2019; Accepted 30 April 2019
Available online 02 May 2019
0049-3848/

many challenges in both data collection and methodology [15], the comparative studies on the health status for Roma adults unanimously showed that the Roma population has a significantly worse health status (including cardiovascular health) in comparison with the hosting population independently from the country where they live [16,17].

Until recently, only a few studies investigated the role of genetic factors in association with metabolic traits among the Roma population, and more detailed contributions of environmental and genetic factors to the increased risk of cardiovascular traits have been defined only recently. The majority of studies on selected genetic markers in Roma samples showed that the allele frequencies, the distribution of haplotype profiles, the genetic risk scores and consequently the genetic load differ significantly from the Hungarian general population [6,7,18,19].

To assess the prevalence of VT among the Roma population, as well as that of other diseases, is very difficult because of restrictions on collecting data on health by ethnic status in the countries where this minority lives [20]. Hereditary thrombophilia gained attention as a possible risk factor for pregnancy complications – however some controversy exists - such as intrauterine growth retardation, spontaneous abortions, still births or even preeclampsia [21–23]. Several studies noted the incidence of these complications are greater among the Roma compared to majority populations of the country where they live. It was concluded that, in addition to low-socioeconomic status, unhealthy diet, poor hygiene and low maternal age, genetic factors may also contribute to the disparity in the prevalence of pregnancy complications in the Roma population [24,25].

Genetic variation at VT risk loci has been profoundly characterized in European populations, and less comprehensively in other populations such as African, Hispanic or Asian [3] It is unclear whether differences exist in cumulative risk allele loads between the Roma and the majority populations, which may partly explain ethnic disparities in disease prevalence. The aim of our present study was to define and compare the VT associated genetic load between the Hungarian general and Roma populations using well-constructed genetic risk score models (unweighted and weighted). This information might have important implications in the future in identifying possible new opportunities for targeted prevention and treatment [26] for those subgroups in the populations who are at greater risk for VT development.

2. Materials and methods

2.1. Study design

Our study included samples of 1179 Hungarian Roma individuals (453 males and 726 females) living in segregated colonies (Roma) in North-East Hungary where the Roma are concentrated and 1542 individuals (732 males and 810 females) from the Hungarian general population (General) from recent cross sectional surveys on randomly selected representative cohorts of the two populations [16,27,28].

2.2. Study populations

2.2.1. Roma living in segregated colonies

Participants were enrolled from the regions of North Hungary and North Great Plain (Borsod-Abaúj-Zemplén, Heves and Szabolcs-Szatmár-Bereg, Hajdú-Bihar, Jász-Nagykun-Szolnok counties), where the majority of Roma colonies can be found, using a stratified multistage sampling method. The details of the sampling methodology and data collected are described elsewhere [16,28]. The sample is representative by age and sex to Roma population living in segregated colonies in North-East Hungary. As a part of the health examination surveys (HES), medical histories (lipid disorders, hypertension and type 2 diabetes mellitus) and socio-demographic characteristics (age, gender and level of education) were recorded, and physical examinations (body weight, height, waist circumference and blood pressure) were carried out for each participant. Blood samples were taken for laboratory (serum concentrations of triglyceride, HDL cholesterol and fasting blood glucose) and genotype investigations. In the present study, 962 samples were analyzed where complete clinical and genotype records of 20–64-year-old Roma adults were available.

2.2.2. Hungarian general population

A population-based disease monitoring system, the General Practitioners' Morbidity Sentinel Stations Programme (GPMSSP), provided the Hungarian reference sample [29]. Study subjects were recruited from the population of the GPMSSP. The methods of sampling applied and survey data collected are described in the Hungarian Metabolic Syndrome Survey (HMSS) [27]. As part of HES similar data were collected as in case of the Roma population. Blood samples were taken for laboratory tests (fasting blood samples were taken to determine serum triglyceride, high-density lipoprotein (HDL)-cholesterol and serum glucose levels) and for DNA isolation. The present study used DNA samples from 1492 20–64-year-old adults with complete records (clinical and genotype) to create the reference dataset. The sample is representative of the 24 year-old to 64-year-old Hungarian adult population in terms of geographic, age and sex distributions.

2.3. DNA isolation

DNA was isolated using the 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. SNP selection

First, 31 SNPs based on the study of de Haan et al. were selected. Recently, to more accurately predict a person's risk for developing VT, they constructed and tested two risk score genetic models: one model involved 31 SNPs (among them, 3 protective ones), and in the other model, the 5 most strongly associated risk increasing SNPs were put together from this SNP-panel. By using these models in combination with a non-genetic risk model, de Haan et al. demonstrated good discriminative (and diagnostic) accuracy in a large case-controlled study [30]. These SNPs were given top priority in the design of the microarray platform (Table S1. Step 1). In addition, 28 more risk increasing SNPs were selected in a structured PubMed search. Twenty-one of the 28 SNPs that had been repeatedly confirmed to be associated with VT in several studies received second-level priority: ten of the 28 SNPS were added from the work of Bezemer et al. who confirmed their association with VT on two study populations (LETS and MEGA-1) [31], nine from a study mapping the 4q35.2 locus and describing association with SNPs localized in the locus and VT, and two from a follow-up genome-wide association study [32,33]. Seven more SNPs were identified from a study on the association of these SNPs with non-fatal venous thrombosis in postmenopausal women and considered as a priority 3 group [34]. The selected SNPs, their genes and the related references, are summarized in Table S1 as Step 1.

During the assay design, two pools were created for genotyping by the service provider, the Mutation Analysis Core Facility (MAF) of the Karolinska University Hospital, Stockholm, Sweden. Based on data obtained in the genotyping process, 52 SNPs were selected for allele frequency comparison and GRS computation. Effect size estimates were available for 50 of the 52 SNPs. To avoid multicollinearity, only one SNP per LD block (the one with the strongest effect) was involved, thus finally 47 SNPs were included in the computation of wGRS (see details of SNPs' selection process in Table S1 Steps 1–4).

2.5. 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 MAF according to their protocols.

2.6. Statistical analyses

Statistical tests were conducted with Stata (version 12.0), IBM SPSS Statistics (version 23) and Haploview (version 4.2) software. A Shapiro-Wilk test for normality was performed. Variables with non-normal distribution (age, BMI, HDL−C, FG, TG, WC, SBP, DBP) were transformed using a two-step approach suggested by Templeton [35] and then compared with a Student's-*t*-test. Sex distribution was compared with a $\chi 2$ test. The existence of a Hardy-Weinberg equilibrium (HWE) and the differences of allele frequencies for all SNP variants between the two populations were evaluated with $\chi 2$ tests. The Bonferroni's adjustment was applied when several statistical tests were being performed simultaneously (the number of independent loci was 47, $p < .001$). Generally, the conventional p threshold of 0.05 was applied. Linkage disequilibrium between polymorphisms was analyzed using Haploview software (version 4.2) [36]. To reveal whether the association between genetic risk and ethnicity depends on the influence of other factors, multivariate linear regression analyses were conducted in which the genetic risk scores were the dependent variable, and the type of population (Roma or general), gender, and age, BMI and HDL−C, FG were considered as independent variables.

2.7. Computations of GRS and wGRS values

To reflect the combined effect of several selected SNPs, unweighted and weighted genetic scores were calculated (GRS and wGRS). Individuals with any missing SNP genotypes were excluded. In the GRS, each person was assigned a score based on the number of risk alleles they carried. Thus, risk allele homozygotes were coded as genotype “2”, heterozygotes as genotype “1”, while “0” indicated 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 [37]. By using these codes, a simple count score (unweighted) was calculated as it is described by Eq. (1) in which Gi is the number of the risk alleles for the ith SNP. This model sums up all risk alleles over all loci as a summary score assuming that all alleles have the same effect.

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

In the weighted approach, rather than giving equal weight to each SNP, SNPs with larger effects contributed more to the score. Weights were derived from the risk coefficient for each allele based on OR reported in previous association studies. Risk estimates for alleles were used for 47 SNPs (see in Table 3) The calculation of the weighted genetic risk score is described by Eq. (2). In this weighted score, weights (wβ_i) were derived from the OR; i.e., each allele was weighted using the natural log of the published OR.

These weights (wβ_i) were multiplied by 0, 1 or 2 according to the number of effect alleles carried by each person (Xi) [37–39]. Two-sided *t*-tests were used to compare the distribution of GRSs.

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

The 5-SNP risk model suggested for testing genetic susceptibility to VT by de Haan et al. [30] (since the discriminative accuracy of this 5-SNP model was almost as high as that of their 31-SNP model) was also adopted for computation of both unweighted (5-SNP GRS) and weighted (5-SNP wGRS) risk score calculations using the SNPs reported with the highest OR in the literature (rs6025/Leiden, *F5*; rs1799963, *F2*; rs8176719, *ABO*; rs2066865, *FGG*; rs2036914, *F11*). Study populations were stratified by quintiles of the weighted Genetic Risk Scores and the proportion of subjects belong to quintiles were compared.

2.8. Ethical approval

The study was approved by the Ethical Committee of the Medical Health Sciences Centre, University of Debrecen, Hungary (reference No. 2462-2006) and by the Ethical Committee of the Hungarian Scientific Council on Health Research (reference Nos. NKFP/1/0003/2005; 8907-O/2011-EKU) and all performed procedures were in accordance with the 1964 Helsinki declaration and its later amendments.

3. Results

3.1. Characteristics of the study sample

Only subjects for whom genotype data were available in addition to health and demographic data were considered. In total, 962 Roma and 1492 individuals representing the general Hungarian population were included in the baseline comparison. The proportion of females in Roma samples was significantly lower than in the General one (38.46% vs. 45.97%, $p < .001$). The age distribution of Roma samples was shifted towards the younger age groups and deviated from the General sample ($p < .001$). Significant differences ($p < .001$) were observed in the distribution of BMI values (kg/m^2), high-density-lipoprotein cholesterol levels (HDL−C, mmol/l), fasting-glucose levels (FG, mmol/l), waist circumference (cm), systolic and diastolic blood pressures (SBP, DBP, mmHg) but not in the case of triglyceride levels (TG, mmol/l, $p = .222$) (Table 1).

3.2. Comparison of allele frequencies

Genotype results were acquired for 2454 samples (Roma: $N = 962$; General: $N = 1492$). All SNPs were tested to determine whether the observed genotype frequencies were consistent with Hardy-Weinberg

Table 1
Characteristics of the study groups.

Characteristics	General ($N = 1492$)	Roma (N = 962)	*p*-Value
Female (%)	45.97%	38.46%	p < .001
Age (years; median, IQR)	45 (34–54)	40 (31–50)	p < .001
BMI (kg/m^2; median, IQR)	27.18 (23.87–30.86)	25.34 (21.30–31.10)	*p* < .001
HDL-C (mmol/L; median, IQR)	1.36 (1.12–1.67)	1.27 (1.05–1.54)	p < .001
FG (mmol/L; median, IQR)	4.50 (4.10–5.00)	4.80 (4.40–5.40)	p < .001
TG (mmol/L; median, IQR)	1.21 (0.81–1.90)	1.28 (0.90–1.96)	*p* = .222
WC (cm; median, IQR)	95 (84–104)	88 (77–102)	p < .001
SBP (mmHg; median, IQR)	125 (115–136)	120 (110–135)	p < .001
DBP (mmHg; median, IQR)	80 (75–85)	80 (70–81)	p < .001

IQR: interquartile range, BMI: body mass index, HDL−C: high density lipoprotein cholesterol, FG: fasting-blood glucose, TG: triglyceride, WC: waist circumference, SBP: systolic blood pressure, DBP: diastolic blood pressure.

Table 2
Comparison of protective and susceptibility allele frequencies (%) in the Hungarian General and Roma populations.

Number	Priority	Gene (short)	SNP	Effect allele	General population (n = 1492)	Roma population (n = 962)	p-Value
1	Priority 1	*ABO*	***rs8176719***	G	47.39%	**55.46%**	< 0.001
2		*CPB2*	**rs3742264**	A	**33.71%**	23.02%	< 0.001
3		*F11*	***rs2036914***	C	46.72%	48.44%	0.002
4			**rs2289252**	T	**41.12%**	33.06%	< 0.001
5			rs3822057	A	51.61%	53.79%	0.001
6		*F13A1*	rs5985	A[a]	25.50%	21.05%	0.02
7		*F13B*	**rs6003**	G	8.58%	**12.37%**	< 0.001
8		*F2*	***rs1799963***	A	2.04%	1.14%	0.054
9			**rs3136516**	G	52.04%	**61.02%**	< 0.001
10			rs3136520	T	2.61%	3.53%	0.087
11		*F5*	***rs6025***	T	4.29%	**11.12%**	< 0.001
12			**rs4524**	T[a]	73.19%	**80.46%**	< 0.001
13		*F9*	rs4149755	A	6.94%	4.26%	0.002
14			rs6048	A	76.46%	71.45%	0.003
15		*FGG*	***rs2066865***	T	24.36%	22.66%	0.042
16		*GP6*	**rs1613662**	C	13.34%	**20.22%**	< 0.001
17		*HIVEP1*	rs169713	C	22.86%	19.39%	0.014
18		*NAT8B*	rs2001490	C	35.86%	39.86%	0.021
19		*NR1I2*	**rs1523127**	C	39.14%	**47.66%**	< 0.001
20		*PROC*	rs1799809	G	41.09%	39.92%	0.678
21		*PROCR*	**rs2069951**	A	4.32%	**9.88%**	< 0.001
22			**rs2069952**	T	54.49%	**61.90%**	< 0.001
23		*SERPINC1*	**rs2227589**	A	11.73%	**21.41%**	< 0.001
24		*STXBP5*	**rs1039084**	A[a]	36.39%	**54.73%**	< 0.001
25		*TFPI*	rs8176592	T	68.00%	72.82%	0.001
26		*VWF*	**rs1063856**	C	**34.38%**	28.74%	< 0.001
27	Priority 2	*CYP4V2*	**rs3817184**	T	**35.99%**	29.78%	< 0.001
28			rs13146272	A	63.47%	59.77%	0.033
29			rs1053094	T	45.44%	42.31%	0.09
30			rs4253236	C	62.47%	60.34	0.176
31		*KLKB1*	rs3733402	A	49.36%	48.60%	0.058
32			**rs4253303**	A	**36.33%**	30.67%	< 0.001
33			**rs4253302**	A	81.94%	**86.33%**	< 0.001
34			**rs2292423**	A	**37.60%**	31.39%	< 0.001
35		*F11*	rs4253418	G	95.51%	96.73%	0.008
36		*PTPRF/KDM4A*	**rs11210892**	G	31.67%	**44.70%**	< 0.001
37		*HIVEP*	rs9380643	T	26.11%	26.35%	0.939
38		*APOH*	**rs1801690**	C	94.00%	**97.56%**	< 0.001
39		*MET*	**rs2237712**	G	4.96%	**7.54%**	< 0.001
40		*EPS8L2*	**rs3087546**	T	56.80%	**64.60%**	< 0.0001
41		*CASP8AP2*	rs369328	A	46.38%	43.87%	0.188
42		*SELP*	**rs6131**	T	**21.55%**	16.79%	< 0.001
43		*ZNF544*	rs6510130	G	2.04%	1.77%	0.628
44		*TACR1*	**rs881**	C	**81.70%**	69.33%	< 0.001
45		*F5*	**rs6016**	G	73.22%	**80.46%**	< 0.001
46		*CCDC181*	**rs3820059**	A	37.30%	**46.93%**	< 0.001
47		*TENM1*	**rs2266911**	C	82.14%	**91.16%**	< 0.001
48	Priority 3	*F5*	**rs3753305**	G	**43.77%**	42.83%	< 0.001
49			rs9332695	A	0.77%	0.16%	0.011
50		*FGA*	rs2070006	A	38.40%	38.57%	0.139
51		*FGB*	rs1800788	T	20.84%	21.21%	0.544
52		*FGG*	**rs2066954**	T	**37.89%**	32.07%	< 0.001

SNPs in bold characters showed highly significantly ($p < .001$) different frequencies in the two populations after multiple test correction. SNPs in bold and italics are involved in the 5-SNPs (unweighted and weighted) GRS models.

[a] Protective alleles.

equilibrium (HWE). No significant deviation from HWE was observed in the study populations. After correcting for multiple tests, a p value of < 0.001 was considered significant. The allele frequencies of fifty-two compared genetic variants are described in Table 2. Allele frequencies in the study populations were calculated on the basis of the obtained genotype distributions. Differences between the Roma and General populations remained significant for 29 SNPs after multiple test correction. Of 49 susceptibility SNPs, seventeen were more frequent in the Roma population ($p < .001$) and ten susceptibility SNPs were more common in the General population. Of the three protective alleles, two were more prevalent among Roma ($p < .001$).

3.3. Comparison of genetic risk scores

The GRS based on 52 SNPs was calculated for 962 Roma individuals and ranged from 24 to 57. The GRS for general subjects was calculated for 1492 individuals and ranged from 22 to 58. The GRS data were normally distributed in both the Roma and the General populations. The mean of the gene count scores was 41.83 ± 5.78 in the Roma and was 41.04 ± 6.04 in the General population. The distribution in the two study groups was found to be different using the two-sided t-test ($p = .001$) being right-shifted in the Roma population compared to the General (Fig. 1A). The GRS was found to be significantly higher among Roma in both sexes using the two-sided t-test (men: $GRS_{General} = 40.90 \pm 5.91$ vs. $GRS_{Roma} = 41.83 \pm 5.74$, $p = .008$; women: $GRS_{General} = 41.19 \pm 6.16$ vs. $GRS_{Roma} = 41.77 \pm 5.87$,

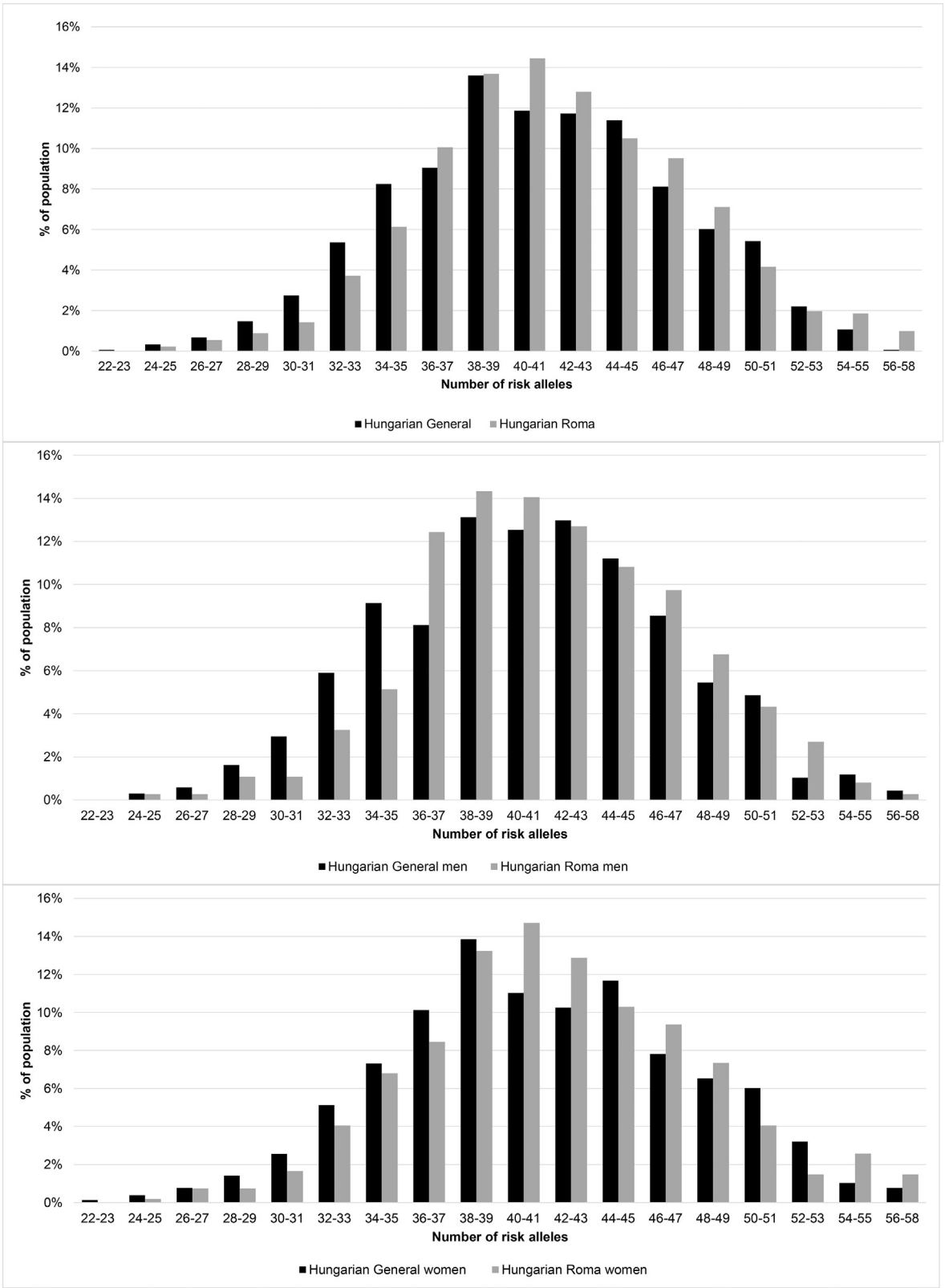

Fig. 1. Distribution of unweighted genetic risk scores based on 52 SNPs by study populations (A) and separately for men (B) and women (C) in both populations. Distribution of GRSs in the Hungarian Roma population (gray bars) is right-shifted compared to the Hungarian General population (black bars).

$p = .046$; see Fig. 1B and C)

The SNPs of the *KLKB1* (rs4253303, rs2292423) *CYP4V2* (rs3817184) and *F11* (rs2036914, rs3822057) were in linkage disequilibrium ($r^2 \geq 0.8$; see Fig. S1). Therefore, from among these variants, only rs2292423 of *KLKB1* and rs2036914 of the *F11* variants were taken into consideration in the computation of the weighted genetic risk score, resulting in a wGRS composed of 47 SNPs. Effect size estimates for these 47 SNPs are described in Table 3. The wGRS data

Table 3
SNPs used in the weighted genetic risk score computation, their genes, effect alleles, odd ratios and weighting numbers with original publications they were adopted from.

Number	Priority	Gene (short)	SNP	Effect allele[a]	OR	Weighting number (lnOR)	Reference
1	Priority 1	*ABO*	rs8176719	G	1.85	0.62	de Haan et al. [30]
2		*CPB2*	rs3742264	A	1.01	0.01	
3		*F11*	rs2036914	C	1.32	0.28	
4			rs2289252	T	1.26	0.23	
5		*F13A1*	rs5985	A[a]	0.93	−0.07	
6		*F13B*	rs6003	G	1.09	0.09	
7		*F2*	rs1799963	A	2.78	1.02	
8			rs3136516	G	1.19	0.17	
9			rs3136520	T	1.13	0.12	
10		*F5*	rs6025	T	3.79	1.33	
11			rs4524	T[a]	0.92	−0.08	
12		*F9*	rs4149755	A	1.24	0.21	
13			rs6048	A	1.08	0.08	
14		*FGG*	rs2066865	T	1.56	0.44	
15		*GP6*	rs1613662	C	1.15	0.14	
16		*HIVEP1*	rs169713	C	1.2	0.18	
17		*NAT8B*	rs2001490	C	1.1	0.10	
18		*NR1I2*	rs1523127	C	1.05	0.05	
19		*PROC*	rs1799809	G	1.17	0.16	
20		*PROCR*	rs2069951	A	1.3	0.26	
21			rs2069952	T	1.21	0.19	
22		*SERPINC1*	rs2227589	A	1.2	0.18	
23		*STXBP5*	rs1039084	A[a]	0.9	−0.11	
24		*TFPI*	rs8176592	T	1.06	0.06	
25		*VWF*	rs1063856	C	1.16	0.15	
26	Priority 2	*CYP4V2*	rs13146272	A	1.24	0.22	Li et al. [32]
27			rs1053094	T	1.26	0.23	
28			rs4253236	C	1.18	0.16	
29		*KLKB1*	rs3733402	A	1.23	0.21	
30			rs4253302	A	1.19	0.17	
31			rs2292423	A	1.25	0.22	
32		*F11*	rs4253418	G	1.31	0.27	
33		*APOH*	rs1801690	C	1.42	0.35	Bezemer et al. [31]
34		*MET*	rs2237712	G	1.38	0.32	
35		*EPS8L2*	rs3087546	T	1.12	0.11	
36		*CASP8AP2*	rs369328	A	1.13	0.12	
37		*SELP*	rs6131	T	1.21	0.19	
38		*ZNF544*	rs6510130	G	1.56	0.44	
39		*TACR1*	rs881	C	1.15	0.14	
40		*F5*	rs6016	G	1.27	0.24	
41		*CCDC181*	rs3820059	A	1.22	0.20	
42		*TENM1*	rs2266911	C	1.47 (men) 1.25 (women)	0.38 (men) 0.22 (women)	
43	Priority 3	*F5*	rs3753305	G	1.2	0.18	Smith et al. [34]
44			rs9332695	A	2.01	0.70	
45		*FGA*	rs2070006	A	1.25	0.22	
46		*FGB*	rs1800788	T	1.31	0.27	
47		*FGG*	rs2066954	T	1.23	0.21	

[a] Protective alleles.

were normally distributed in both the Roma and General populations. The average wGRS in the group of Roma was 7.78 ± 1.28, while it was 7.46 ± 1.33 for the General population ($p < .001$). The distribution curves of the wGRS values for the study populations are shown in Fig. 2.

Only 2.39% of subjects in the Roma population were in the bottom fifth (wGRS ≤ 5.19) of the weighted score compared with 3.62% of those in the General population ($p = .080$). In the Roma group, 4.26% of the individuals were in the top fifth (wGRS ≥ 10.02) of the wGRS compared with 2.88% of those in the General population ($p = .066$); i.e., the distribution of weighted genetic risk scores was found to be right-shifted in the Roma population compared to the General population (Table S2).

Comparing the results obtained in 5-SNP GRS and 5-SNP wGRS models did not change our inference about the greater genetic susceptibility to VT of the Roma population (5-SNP $GRS_{General} = 2.62 \pm 1.23$ vs.5-SNP $GRS_{Roma} = 2.77 \pm 1.23$, $p = .003$ and 5-SNP $wGRS_{General} = 1.25 \pm 0.71$ vs. 5-SNP $wGRS_{Roma} = 1.47 \pm 0.79$, $p < .001$).

Multivariate regression analyses (age, sex, BMI, HDL−C, FG, SBP and DBP were covariates) also suggested that the Roma population has higher unweighted and weighted genetic risk scores independent of covariates. In the multivariate model, Roma ethnicity was associated with increased genetic risk scores corresponding to an increase of 0.75 in GRS and 0.31 in wGRS. The results of multivariate regression analyses are shown in Table S3.

4. Discussion

It is generally accepted that most SNPs have a small effect size, so using multiple markers in combination can be more useful to translate results from genomic studies into tools for population health research. In the genomic era, identifying people who are at high risk of developing common CVDs may improve the benefits of personalized prevention programs by reducing the risk of onset or progression of disease. However, so far, only limited evidence is available on the application of genomic results in health and public health practice [2]. In the cases of rs8176719 (*ABO*) and rs6025 (*F5*/Leiden mutation), which are among the five polymorphisms most strongly associated with

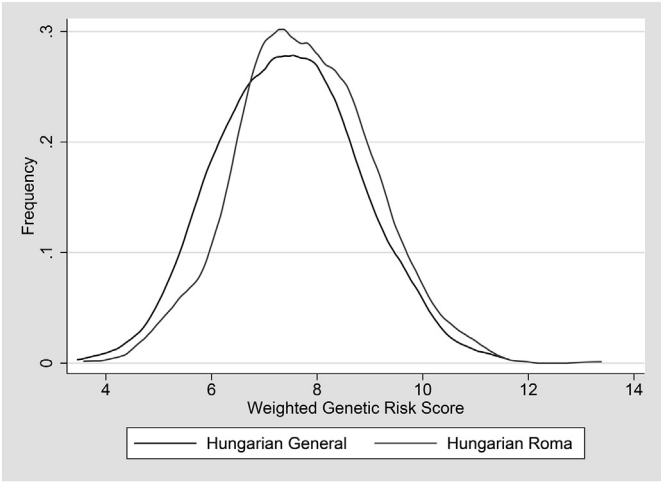

Fig. 2. Distribution curves of weighted genetic risk scores by study groups. The distribution curve of wGRS in Roma (gray line) is right shifted compared to the General population (black line).

VT according to de Haan et al. the allele frequency differences were robust between our study populations. These SNPs that have clear and important roles in the blood coagulation cascade occur much more frequently in the Roma population. In addition to SNPs related to the intrinsic (*F11*, *GP6*, *STXBP5* and *KLKB1*) and the common (*F13B*, *F2*, *F5*, *PROC*, and *PROCR*, *SERPINC1*) pathways of the blood coagulation cascade, several non-functional (*NR1I2*, *PTPRF/KDM4A*, *APOH*, *MET*, *EPS8L2*, *CCDC181* and *TENM1*) risk SNPs were also more prevalent among the Roma. We computed genetic risk scores based on SNPs consistently associated with VT and observed significant differences between our study populations. On average, the total number of risk alleles (GRS) were higher in the Roma, and the distribution of wGRS (in this case published weights to each SNPs were utilized) was shifted to the right, suggesting a higher genetic load among them.

Studies have shown a high prevalence of cardiovascular risk factors among the Roma independent of the countries where they live; this suggests that, in addition to the environmental/modifiable risk factors, genetic susceptibility may also contribute to the high cardiovascular disease morbidity and mortality burden among the Roma population [40,41]. Only sporadic single gene studies showed that the Roma population seems to be more susceptible to hereditary thrombophilia [42–44]. Two studies on small sized samples noted that the Roma in Hungary and Slovakia had a higher prevalence of *Factor V* Leiden mutations compared to the general populations (7.1% vs. 4.9% in Hungary [42]; 6.6% vs. 2.9% in Slovakia [44]), but in another study, the association of *factor V* Leiden and *MTHFR* C677T polymorphisms with pregnancy complications could be detected only in the Slovak majority patient group, not the Roma patients' group [43].

The *F5* rs6025 (*Factor 5* Leiden) and *F2* G20210A (*Prothrombin* 20210A variant) are the most common causes of hereditary thrombophilia in the general Caucasian population [45]. The prevalence of the risk alleles varies between 0 and 13.3% (*F5* Leiden mutation was absent from a sample of Inuit from Greenland but had a peek value of > 13% in Cyprus) and 1.7–3% (*F2*, G20210A) in Europeans [46,47], and both mutations are completely missing among south-east Asians and south Asians [48,49]. Our results revealed that, among the Roma, the prevalence of the A risk allele (Leiden) is almost three times higher than in the General population (11.12% vs. 4.29%, $p < .001$), but there was no significant differences in the case of *F2* G20210A in terms of allele frequency (2.04% vs. 1.14%, $p = .054$). The prevalence of *F5* Leiden showed a gradient in the continent, being high in southeast Europe and decreasing towards the north and east [47]. Balogh et al. detected the prevalence of 7.1% in the case of the Leiden mutation in a small sample of Roma ($N = 147$) and noted the allele frequency among the Roma population is among the highest possibly due to their migration route across the southeast Mediterranean region, co-existence with the Hungarian population and high inbreeding rate [42].

The Roma population in Hungary represents a genetically isolated population with a high consanguinity rate. Endogamy was proved by the gipsy origin of male partners in 90% of couples. The occurrence of first cousin couples was 16 times higher than that of the Hungarian population at large [14]. Based on these data, it can be assumed that a number of private founder mutations might have an impact on disease development among Roma. The founder mutations discovered in this minority were affiliated mainly to Mendelian diseases, however, it cannot be excluded that founder mutations associated with common complex diseases, such as VT, may also exist; however, have not yet been revealed.

A limitation of this study is that, although the majority of the Roma population is accumulated in a certain geographical region (north-east Hungary), our study sample cannot be considered as a representative sample for the Hungarian Roma population. Sampling excluded Roma who have assimilated with the general population. However, because many people are reluctant to self-define their ethnicity as Roma, it would be very difficult to overcome this constraint. It is also important to mention that the representative sample of the general Hungarian population included some people who are Roma, so it is possible that their inclusion resulted in a slight underestimation of the differences between the two populations.

The unweighted GRS models in risk assessment have remarkable limitations because they assume that all SNPs have an independent additive effect, and neither the different effect sizes of SNPs nor their interactions are considered. In addition, exposure to epigenetic factors, rare or structural variants, gene-environmental and gene-gene interactions were not taken into account; however, it is well-known that all of these factors can significantly modify genetic risk. It is important to highlight that more in-depth study is needed to clarify the role not only of the genetic factors but also of the interactions between genetic and environmental factors in the interpretation of ethnic differences, if they exist. Another drawback of our study is that data on VT disease status was not available in the case of the study populations, consequently genotype-phenotype associations and the discriminatory ability of the models could not be tested, but it is important to recall that the priority 1 group of SNPs applied in our study were reported to have a robust discriminatory effect.

In the weighted GRS models, effect size measures obtained from population of European ancestry were applied. According to *NHGRI-EBI* Catalogue of published genome-wide association studies, all the GWASs related to VT have been conducted in populations of European-descent and no research studies were published on other populations and no data are available for the Roma population at all. Additionally, in the case of candidate gene studies, effect size measures related to Roma subjects are not available. Although the question as to what extent the effect size estimates obtained on Caucasian populations are applicable for Roma in genetic risk assessment to VT, it is worth mentioning that in our recent research on HDL-C level related genetic loci, it was demonstrated that effect size measures obtained from GWAS can be used for risk estimation not only in the case of the Hungarian general population but also in the case of the Roma population [50].

The strength of our study is that a high number of SNPs – all the presently known ones with considerable size effects – were included in the analysis, among them the two risk models validated in case control study [30]. In addition, in our study, the numbers in the Roma population surveyed were the highest. This is the first genetic risk score-based study that was carried out to estimate genetic susceptibility to VT in the Roma population in comparison with the general population. We demonstrated that the Roma population is significantly more susceptible to VT using unweighted and weighted genetic risk scoring models. Interventions aiming to improve the Roma's cardiovascular health and to identify subgroups of high-risk individuals need to consider that this

ethnic population has an increased genetic propensity to VT.

Authorship contributions

R Á had the original idea for the comparative analysis, participated in the development of study design, interpreted the results, guided writing of the manuscript and was involved in finalizing the manuscript. Sz F contributed to doing the statistical analyses, interpreting the results, and writing the manuscript. P P performed the statistical analysis, interpreted the results. Zs K participated in the sampling design, performed the sampling and coordinated the Roma surveys. J S planned the sampling, performed the statistical analysis.

Disclosure of conflicts of interest

The authors declare that they have no conflict of interest.

Acknowledgements

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.thromres.2019.04.031.

References

[1] P.E. Morange, P. Suchon, D.A. Tregouet, Genetics of venous thrombosis: update in 2015, Thromb. Haemost. 114 (2015) 910–919.
[2] S. Fiatal, R. Adany, Application of single-nucleotide polymorphism-related risk estimates in identification of increased genetic susceptibility to cardiovascular diseases: a literature review, Front. Public Health 5 (2017) 358.
[3] M. Margaglione, E. Grandone, Population genetics of venous thromboembolism. A narrative review, Thromb. Haemost. 105 (2011) 221–231.
[4] N.A. Zakai, L.A. McClure, Racial differences in venous thromboembolism, J. Thromb. Haemost. 9 (2011) 1877–1882.
[5] N.T. Krarup, A. Borglykke, K.H. Allin, et al., A genetic risk score of 45 coronary artery disease risk variants associates with increased risk of myocardial infarction in 6041 Danish individuals, Atherosclerosis 240 (2015) 305–310.
[6] P. Piko, 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.
[7] J. Dioszegi, S. Fiatal, R. Toth, et al., Distribution characteristics and combined effect of polymorphisms affecting alcohol consumption behaviour in the Hungarian general and Roma populations, Alcohol Alcohol. 52 (2017) 104–111.
[8] Y.C. Klimentidis, N.E. Wineinger, A.I. Vazquez, G. de Los Campos, Multiple metabolic genetic risk scores and type 2 diabetes risk in three racial/ethnic groups, J. Clin. Endocrinol. Metab. 99 (2014) E1814–E1818.
[9] J.M. Keaton, J.N. Cooke Bailey, N.D. Palmer, et al., A comparison of type 2 diabetes risk allele load between African Americans and European Americans, Hum. Genet. 133 (2014) 1487–1495.
[10] G.E. Raskob, P. Angchaisuksiri, A.N. Blanco, et al., Thrombosis: a major contributor to global disease burden, Semin. Thromb. Hemost. 40 (2014) 724–735.
[11] B.I. Melegh, Z. Banfai, K. Hadzsiev, A. Miseta, B. Melegh, Refining the south Asian origin of the Romani people, BMC Genet. 18 (2017) 82.
[12] Commission E, Communication From the Comission to the European Parliament, the Council, the European Economic and Social Comitte and the Comitte of the Regions. An EU Framework for National Roma Integration Strategies Upto 2020, (2011).
[13] L. Kalaydjieva, D. Gresham, F. Calafell, Genetic studies of the Roma (Gypsies): a review, BMC Med. Genet. 2 (2001) 5.
[14] S. Assal, E. Susanszky, A. Czeizel, High consanguinity rate in Hungarian gipsy communities, Acta Paediatr. Hung. 31 (1991) 299–304.
[15] K. Kosa, R. Adany, Studying vulnerable populations: lessons from the Roma minority, Epidemiology 18 (2007) 290–299.
[16] Z. Kosa, A. Moravcsik-Kornyicki, J. Dioszegi, et al., Prevalence of metabolic syndrome among Roma: a comparative health examination survey in Hungary, Eur. J. Pub. Health 25 (2015) 299–304.
[17] B. Cook, G.F. Wayne, A. Valentine, A. Lessios, E. Yeh, Revisiting the evidence on health and health care disparities among the Roma: a systematic review 2003–2012, Int. J. Public Health 58 (2013) 885–911.
[18] V. Adamkova, J.A. Hubacek, D. Novakova, et al., Genetic and biochemical characteristics in the Roma minority in the South Bohemia Region, Neuro Endocrinol. Lett. 36 (Suppl. 2) (2015) 29–34.
[19] A. Poveda, M.E. Ibanez, E. Rebato, Common variants in BDNF, FAIM2, FTO, MC4R, NEGR1, and SH2B1 show association with obesity-related variables in Spanish Roma population, Am. J. Hum. Biol. 26 (2014) 660–669.
[20] R. Adany, Roma health is global ill health, Eur. J. Pub. Health 24 (2014) 702–703.
[21] W.J. Kist, N.G. Janssen, J.J. Kalk, et al., Thrombophilias and adverse pregnancy outcome - a confounded problem!, Thromb. Haemost. 99 (2008) 77–85.
[22] I.A. Greer, Thrombophilia: implications for pregnancy outcome, Thromb. Res. 109 (2003) 73–81.
[23] J.L. Kujovich, Thrombophilia and pregnancy complications, Am. J. Obstet. Gynecol. 191 (2004) 412–424.
[24] M. Bobak, J. Dejmek, I. Solansky, R.J. Sram, Unfavourable birth outcomes of the Roma women in the Czech Republic and the potential explanations: a population-based study, BMC Public Health 5 (2005) 106.
[25] P. Balazs, I. Rakoczi, A. Grenczer, K.L. Foley, Birth-weight differences of Roma and non-Roma neonates—public health implications from a population-based study in Hungary, Cent. Eur. J. Public Health 22 (2014) 24–28.
[26] M.W. Gillman, R.A. Hammond, Precision treatment and precision prevention: integrating “below and above the skin”, JAMA Pediatr. 170 (2016) 9–10.
[27] E. Szigethy, G. Szeles, A. Horvath, et al., Epidemiology of the metabolic syndrome in Hungary, Public Health 126 (2012) 143–149.
[28] R. Adany, K. Kosa, J. Sandor, M. Papp, G. Furjes, General practitioners' cluster: a model to reorient primary health care to public health services, Eur. J. Pub. Health 23 (2013) 529–530.
[29] G. Szeles, Z. Voko, T. Jenei, et al., A preliminary evaluation of a health monitoring programme in Hungary, Eur. J. Pub. Health 15 (2005) 26–32.
[30] H.G. de Haan, I.D. Bezemer, C.J. Doggen, et al., Multiple SNP testing improves risk prediction of first venous thrombosis, Blood 120 (2012) 656–663.
[31] I.D. Bezemer, L.A. Bare, C.J. Doggen, et al., Gene variants associated with deep vein thrombosis, JAMA 299 (2008) 1306–1314.
[32] Y. Li, I.D. Bezemer, C.M. Rowland, et al., Genetic variants associated with deep vein thrombosis: the F11 locus, J. Thromb. Haemost. 7 (2009) 1802–1808.
[33] P.E. Morange, I. Bezemer, N. Saut, et al., A follow-up study of a genome-wide association scan identifies a susceptibility locus for venous thrombosis on chromosome 6p24.1, Am. J. Hum. Genet. 86 (2010) 592–595.
[34] N.L. Smith, L.A. Hindorff, S.R. Heckbert, et al., Association of genetic variations with nonfatal venous thrombosis in postmenopausal women, JAMA 297 (2007) 489–498.
[35] G.F. Templeton, Article 4, Communications of the AIS, (2011).
[36] J.C. Barrett, B. Fry, J. Maller, M.J. Daly, Haploview: analysis and visualization of LD and haplotype maps, Bioinformatics 21 (2005) 263–265.
[37] P.J. Talmud, A.D. Hingorani, J.A. Cooper, et al., Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study, BMJ 340 (2010) b4838.
[38] P. Sebastiani, N. Solovieff, 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. 3 (2012) 26.
[39] T. Johnson, Efficient calculation for multi-SNP genetic risk score, American Society of Human Genetics Annual Meeting, 2012 (San Francisco, November 6–10).
[40] D. Bogdanovic, D. Nikic, B. Petrovic, et al., Mortality of Roma population in Serbia, 2002–2005, Croat. Med. J. 48 (2007) 720–726.
[41] B. Vozarova de Courten, M. de Courten, R.L. Hanson, et al., Higher prevalence of type 2 diabetes, metabolic syndrome and cardiovascular diseases in gypsies than in non-gypsies in Slovakia, Diabetes Res. Clin. Pract. 62 (2003) 95–103.
[42] I. Balogh, R. Poka, G. Losonczy, L. Muszbek, High frequency of factor V Leiden mutation and prothrombin 20210A variant in Romanies of Eastern Hungary, Thromb. Haemost. 82 (1999) 1555–1556.
[43] A. Bozikova, D. Gabrikova, J. Pitonak, et al., Ethnic differences in the association of thrombophilic polymorphisms with obstetric complications in Slovak and Roma (Gypsy) populations, Genet. Test. Mol. Biomarkers 19 (2015) 98–102.
[44] A. Bozikova, D. Gabrikova, A. Sovicova, et al., The frequency of factor V Leiden and prothrombin G20210A mutations in Slovak and Roma (Gypsy) ethnic group of Eastern Slovakia, J. Thromb. Thrombolysis 34 (2012) 406–409.
[45] V. Bafunno, M. Margaglione, Genetic basis of thrombosis, Clin. Chem. Lab. Med. 48 (Suppl. 1) (2010) S41–S51.
[46] J.S. Clark, G. Adler, N.N. Salkic, A. Ciechanowicz, Allele frequency distribution of 1691G > A F5 (which confers Factor V Leiden) across Europe, including Slavic populations, J. Appl. Genet. 54 (2013) 441–446.
[47] G. Lucotte, G. Mercier, Population genetics of factor V Leiden in Europe, Blood Cells Mol. Dis. 27 (2001) 362–367.
[48] L.N. Roberts, R.K. Patel, R. Arya, Venous thromboembolism and ethnicity, Br. J. Haematol. 146 (2009) 369–383.
[49] G. Himabindu, D. Rajasekhar, K. Latheef, et al., Factor V Leiden mutation is not a predisposing factor for acute coronary syndromes, Indian Heart J. 64 (2012) 570–575.
[50] P. Pikó, S. Fiatal, Z. Kósa, J. Sándor, R. Ádány, Generalisability of effect size estimates for SNPs related to HDL-C obtained on European populations: conference abstract, Eur. J. Pub. Health 27 (2017).