

RESEARCH ARTICLE

Characterization of ADME genes variation in Roma and 20 populations worldwide

Tatjana Škarić-Jurić¹, Željka Tomas¹, Matea Zajc Petranović¹, Nada Božina², Nina Smolej Narančić¹, Branka Janičijević¹, Marijana Peričić Salihović^{1*}

1 Institute for Anthropological Research, Zagreb, Croatia, **2** Department for Pharmacogenomics and Therapy Individualization, University Hospital Center Zagreb, Department of Pharmacology, University of Zagreb School of Medicine, Zagreb, Croatia

* mpericic@inantro.hr



OPEN ACCESS

Citation: Škarić-Jurić T, Tomas Ž, Zajc Petranović M, Božina N, Smolej Narančić N, Janičijević B, et al. (2018) Characterization of ADME genes variation in Roma and 20 populations worldwide. PLoS ONE 13(11): e0207671. <https://doi.org/10.1371/journal.pone.0207671>

Editor: Tzen-Yuh Chiang, National Cheng Kung University, TAIWAN

Received: May 9, 2018

Accepted: November 5, 2018

Published: November 19, 2018

Copyright: © 2018 Škarić-Jurić et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This study was funded by the Croatian Science Foundation, <http://www.hrzz.hr/default.aspx?id=47>, IP-2014-09-4454 to MPS. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

The products of the polymorphic ADME genes are involved in Absorption, Distribution, Metabolism, and Excretion of drugs. The pharmacogenetic data have been studied extensively due to their clinical importance in the appropriate drug prescription, but such data from the isolated populations are rather scarce. We analyzed the distribution of 95 polymorphisms in 31 core ADME genes in 20 populations worldwide and in newly genotyped samples from the Roma (Gypsy) population living in Croatia. Global distribution of ADME core gene loci differentiated three major clusters; (1) African, (2) East Asian, and (3) joint European, South Asian and South American cluster. The *SLCO1B3* (rs4149117) and *CYP3A4* (rs2242480) genes differentiated at the highest level the African group of populations, while *NAT2* gene loci (rs1208, rs1801280, and rs1799929) and *VKORC1* (rs9923231) differentiated East Asian populations. The *VKORC1* rs9923231 was among the investigated loci the one with the largest global minor allele frequency (MAF) range; its MAF ranged from 0.027 in Nigeria to 0.924 in Han Chinese. The distribution of the investigated gene loci positions Roma population within the joined European and South Asian clusters, suggesting that their ADME gene pool is a combination of ancestral (Indian) and more recent (European) surrounding, as it was already implied by other genetic markers. However, when compared to the populations worldwide, the Croatian Roma have extreme MAF values in 10 out of the 95 investigated ADME core gene loci. Among loci which have extraordinary MAFs in Roma population two have strong proof of clinical importance: rs1799853 (*CYP2C9*) for warfarin dosage, and rs12248560 (*CYP2C19*) for clopidogrel dosage, efficacy and toxicity. This finding confirms the importance of taking the Roma as well as the other isolated populations' genetic profiles into account in pharmaco-therapeutic practice.

Introduction

Medication efficacy and adverse drug reactions are associated with specific genes' variants [1] that are related to the Absorption, Distribution, Metabolism and Elimination of drugs (ADME). Their polymorphic nature is the basis for the individual response to drug treatment,

together with a number of factors such as sex, age, weight, concomitant medication, health status, comorbidity level etc. [2]. The ADME genes' variation is markedly related to ethnicity and shows distinct geographic patterns [2, 3].

Numerous studies have been published on ADME-related genes' polymorphisms and their clinical importance. Some of them used large panels of genes and/or big samples from large population groups [1, 4–10], while others provided data for specific variants of interest and/or smaller samples from diverse geographic or clinical populations [11–14]. Currently, only sparse data are available on the prevalence of these gene variants in Roma (Gypsy) population [15–17]. Lacking large-scale pharmacogenomics information in this population presents a bottleneck for their healthcare improvement.

Although today's state-of-art methodology in personalized medicine is individual genotyping prior to medications, unfortunately this approach has not been routinely applied at points of care and usually follows after an adverse drug effect. Therefore, population ADME genes profiling is useful for clinical practice especially in cases when the population consists of different ancestry groups.

It has been estimated that about 15 million of Roma people live worldwide today from whom 10 million reside in Europe. About 40 thousand Roma live in Croatia [18]. However, their numbers are probably significantly underestimated due to ethnomimicry characteristically present in this population in addition to the avoidance of contacts with state officials, leaving many members of this population unregistered in census data.

Anthropologically, the Roma are the transnational minority population marked by common Indian ancestry. Various social and economic pressures caused gradual population fragmentation and formation of a complex network of numerous and often endogamous subgroups with specific languages (dialects), religions, and socio-cultural characteristics [19, 20]. Lasting isolation preserved their founding gene pool with different characteristics compared to the surrounding majority populations [21].

Their pronounced genetic differences from the majority Croatian population and traces of their ancestral origins have been detected in mitochondrial DNA [22], Y chromosome markers [23, 24] and various autosomal common [25] and rare disease loci [26]. The recent study of ADME gene's CYP2B6 polymorphisms proved the distinctive position of the Croatian Roma in the world's populational variability [27] and indicated the need for a systematic investigation of the most important pharmacogenes' variants in the Roma.

Since isolated populations usually have a unique genetic profile it is important to determine their ADME genes pattern in the context of broader global diversity [28, 29]. Therefore, in this study, we (1) present the allele frequencies for 95 polymorphisms in 31 core ADME-related genes for 20 worldwide populations as well as for the Roma population living in Croatia; (2) identify and describe the set of markers that mostly contribute to the separation of major population groups and, thus, give rise to specific geographic patterns of core ADME gene loci; and (3) elucidate the position of the Roma in the global ADME genetic landscape.

Results

The minor allele frequencies (MAFs) and sample sizes as well as the references in addition to the 1000 Genomes for all 95 SNPs used in the this study are presented in the [S1 Table](#). MAF always refers to the global minor allele as indicated in 1000 Genomes' database.

Genetic distance analyses were carried out to quantify genetic differentiation across 21 populations in this study and the dendrogram is reported in [Fig 1](#). As expected, there is a clear separation of clusters that correspond to the continental regions their member populations belong. The European and the South Asian populations cluster closely together. They are

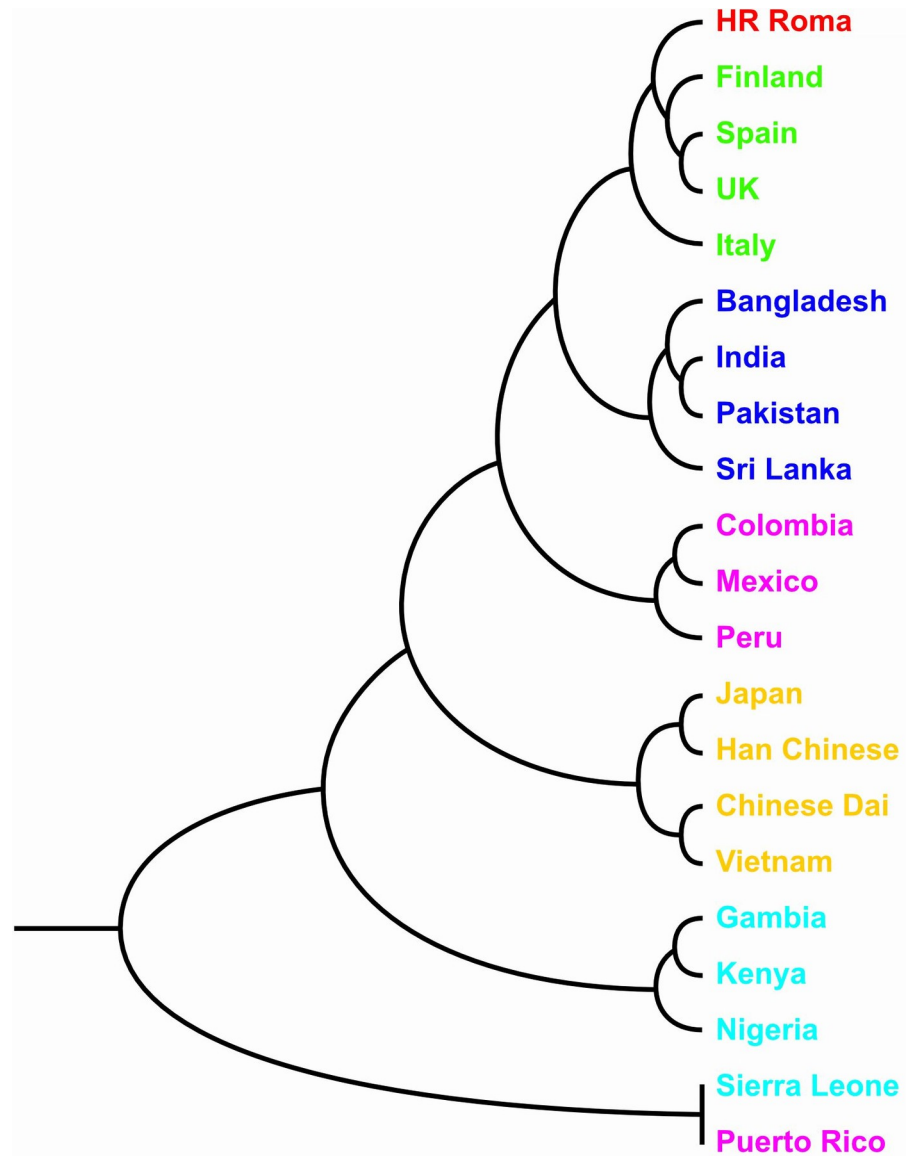


Fig 1. UPGMA dendrogram from Nei's genetic distance matrix based on the data on minor allele frequencies for 95 ADME core genes' loci in Croatian Roma and 20 populations worldwide.

<https://doi.org/10.1371/journal.pone.0207671.g001>

joined successively by the American, East Asian and African clusters. Finally, closely joined Sierra Leone and Puerto Rico populations cluster as a distinct subgroup further away from the rest of the populations.

The relationship between genetic variation and geographic distance was analyzed using correlation between matrices of genetic and geographic distances. The correlation is positive and significant (Pearson's $r = 0.300$, $p \leq 0.006$ after 1,000 permutations) indicating isolation by spatial distance at the global scale. Focusing on the Croatian Roma population, their genetic distances from the other 20 populations in this study were plotted against the geographic distances (Fig 2). Roma cluster well within the European populations, and are relatively close to the South Asian populations genetically despite their spatial distance of 5,000–7,500 km. Larger genetic distances exist between the Roma and spatially more distant East Asian as well

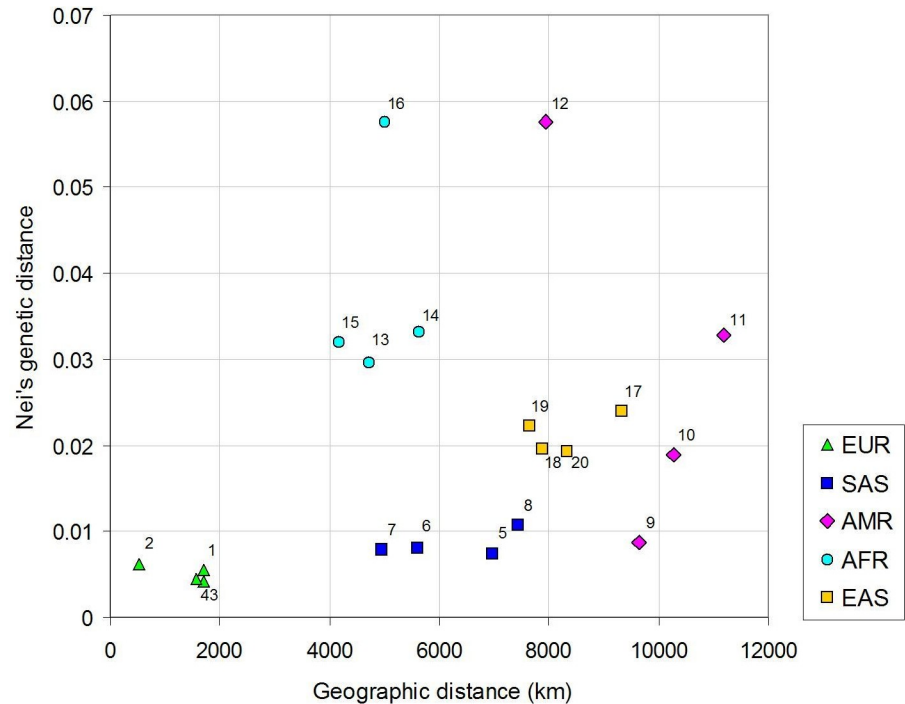


Fig 2. Genetic distances between the Croatian Roma and 20 populations worldwide in relation to their geographic distances. Legend: 1 = Finland; 2 = Italy; 3 = Spain; 4 = UK; 5 = Bangladesh; 6 = India; 7 = Pakistan; 8 = Sri Lanka; 9 = Colombia; 10 = Mexico; 11 = Peru; 12 = Puerto Rico; 13 = Gambia; 14 = Kenya; 15 = Nigeria; 16 = Sierra Leone; 17 = Japan; 18 = China-Dai; 19 = China-Han; 20 = Vietnam.

<https://doi.org/10.1371/journal.pone.0207671.g002>

as fairly dispersed American populations. Roma genetically differ the most from the African populations that are on the average closer to them geographically, confirming the genetic distinctiveness of the African region.

The Principal Component Analysis (PCA) was performed using MAF data for 95 ADME core genes' loci for 21 populations from various part of the world, including Croatian Roma. PCA revealed six PCs (76.4% of the total variance explained) reflecting the genetic relationships among the populations showing a pattern that is very similar to that obtained by genetic distance dendrogram. PC1 (28.4% of the total variance) separates four African countries from the rest of the world while PC2 (20.9% of the total variance) separates four East Asian countries. PC3 (accounting for 10.8% of the total genetic variance) separates South Asians from the remaining European-American cluster with the Croatian Roma being intermediate to Europeans and South Asians. The PC4 axis (accounting for 8.5% of the total variance) separates Americans from the European-Roma group and places the Roma population at the top of the positive pole of the axis (while negative pole is represented by Peru). PC5 (explains 4.3% of the total variance) differentiates European countries (placing at the opposite poles Finland and UK) while PC6 (accounting for 3.5% of the total variance) differentiates East Asian populations (opposing Han from Dai Chinese populations).

In order to elucidate the most characteristic continental single nucleotide polymorphisms (SNPs) among the core ADME genes' loci, we performed the gene-oriented Principal Component Analysis (gPCA). The gPCA revealed three significant components: gPC1 (explaining 77% of the total variance) was defined by the global range in MAF values (i.e. it contains not population-specific but locus-specific information). However, gPC2 (9.3%) and gPC3 (6.1% of the total variance) were population-specific.

Combining population- and gene-oriented approaches, Fig 3 shows scatterplot of the first two principal components from the PCA together with the presentation of the loci with the highest factor scores in the gPCA. The first two principal components (PCs) of the PCA, accounting for 49.3% of the total genetic variance, clearly separate three clusters that reflect the major genetic relationships among the populations: the African (AFR), the East Asian (EAS) and the joint South Asian, European and American (SAS, EUR and AMR, respectively). The AFR and EAS groups are related with the ADME genes that have the highest factor scores at gPC2 and gPC3, respectively.

As shown in Fig 3, the most distinctive African loci are rs4149117 (*SLCO1B3*) and rs2242480 (*CYP3A4*) that have the highest factor scores at the positive pole and (with smaller factor score values) rs4124874 (*UGT1A1*), rs9923231 (*VKORC1*), and rs1128503 (*ABCB1*) at the negative pole. In the combination with the results (factor loadings) from the population-oriented PCA (data not here presented), we consider rs4149117 (*SLCO1B3*) and rs2242480 (*CYP3A4*) as the most distinctive African loci.

The East Asian (EAS) gPC3 has 6 loci on the positive pole (factor scores higher than 2): rs9923231 (*VKORC1*), rs1143671, rs2257212, rs1143672, rs2293616 (all four placed within *SLC15A2*), and rs4149117 (*SLCO1B3*). At the negative pole of gPC3 there are three SNPs (rs1208, rs1801280, and rs1799929) all of them placed within *NAT2* gene. In the combination with the results of the population-based PCA (data not here presented), those results indicate that distinctive SNPs for the East Asians are the three *NAT2* gene SNPs and a *VKORC1* (rs9923231) SNP.

The largest global MAF differences (delta) between maximal and minimal MAF values for 95 SNPs are presented graphically in decreasing order (Fig 4) and the populations with extreme MAF values as well as the exact delta values are given in the S2 Table. The three SNPs characterized with the largest global diversity are: rs9923231 in *VKORC1* gene (range: 0.924 in China Han to the 0.027 in Nigeria, delta = 0.897), rs2242480 in *CYP3A4* gene (range: 0.909 in Kenya to the 0.071 in UK; delta = 0.838) and rs1048943 in *CYP1A1*2C* gene (range: 0.706 in Peru to the 0.000 in Gambia, Kenya, Nigeria, and Sierra Leone; delta = 0.706).

Focusing on the Croatian Roma, they have the extreme MAF values for 10 SNPs (Fig 4). Their highest MAFs among 21 world's populations have been found for rs1128503 (*ABCB1*), rs1902023 (*UGT2B15*), rs12248560 (*CYP2C19*17*), rs1799853 (*CYP2C9*), rs3758581 (*CYP2C19*), rs10509681 (*CYP2C8*3*), rs1138272 (*GSTP1*), rs8192709 (*CYP2B6*), and rs34059508 (*SLC22A1*) (and next to the highest MAF for rs28371725 (*CYP2D6*)). On the other side, Roma have the lowest MAF for rs4149117 (*SLCO1B3*) (and next to the lowest MAF value for the rs28399433 (*CYP2A6*)).

Discussion

The findings of numerous studies suggest that population differences in ADME genes show marked geographic and ethnic variation. However, most of the studies investigating these variations generally lack pharmacogenetic data on isolated populations. The same is true for the Roma, one of the world's largest transnational minority populations. Roma are an example of isolated population with specific migrational history whose gene pool is highly influenced by the genetic drift due to unique social and cultural features.

Therefore, in this paper we investigated the position that the Croatian Roma take within the world-wide variation in 95 ADME core genes' loci. The ADME genes' MAF information from twenty 1000 Genomes' populations is here enriched by the data found through systematic literature search.

The results of this study confirm the previous findings that distinguish three world's population clusters: African, East Asian and joint European, South Asian and Native American

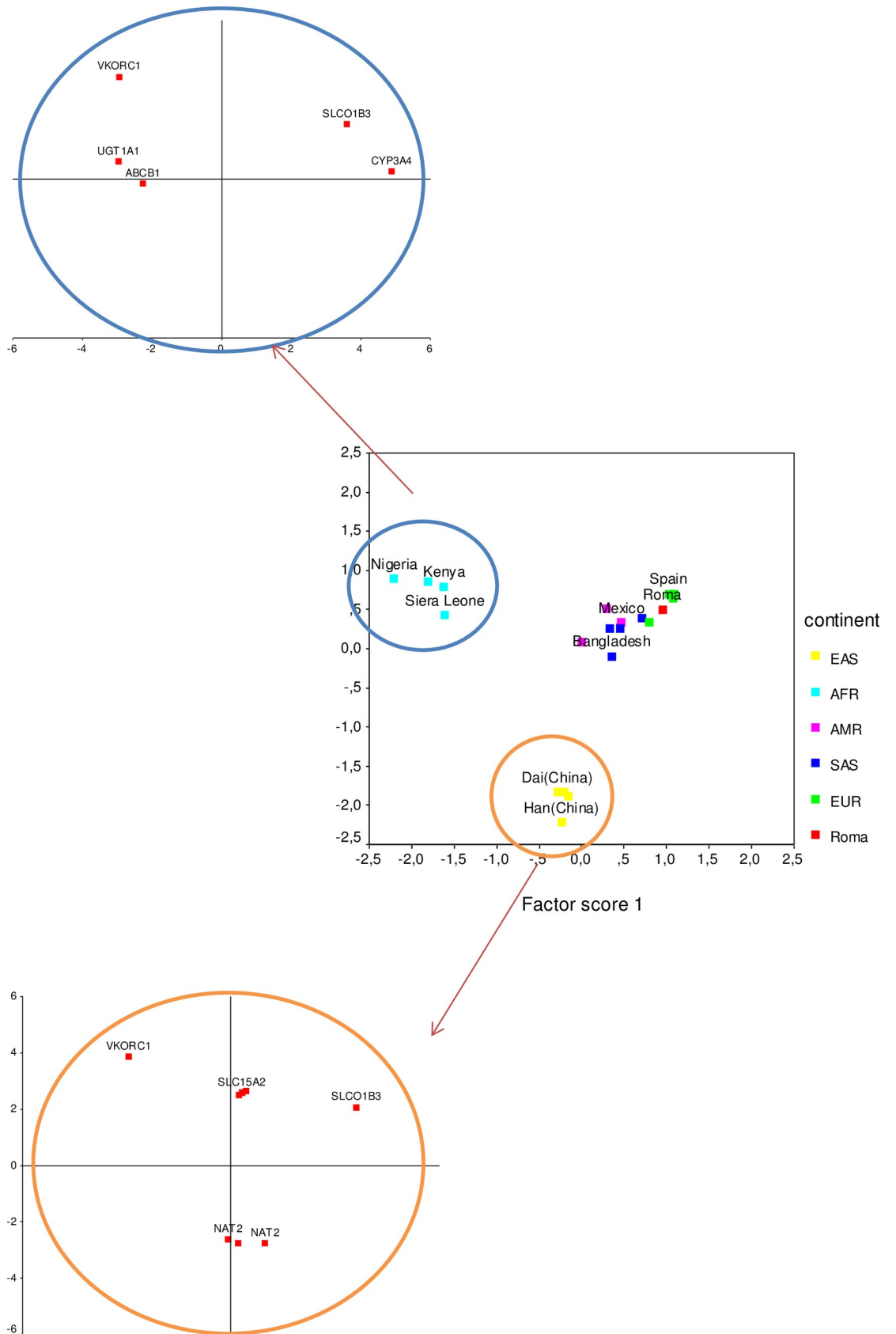
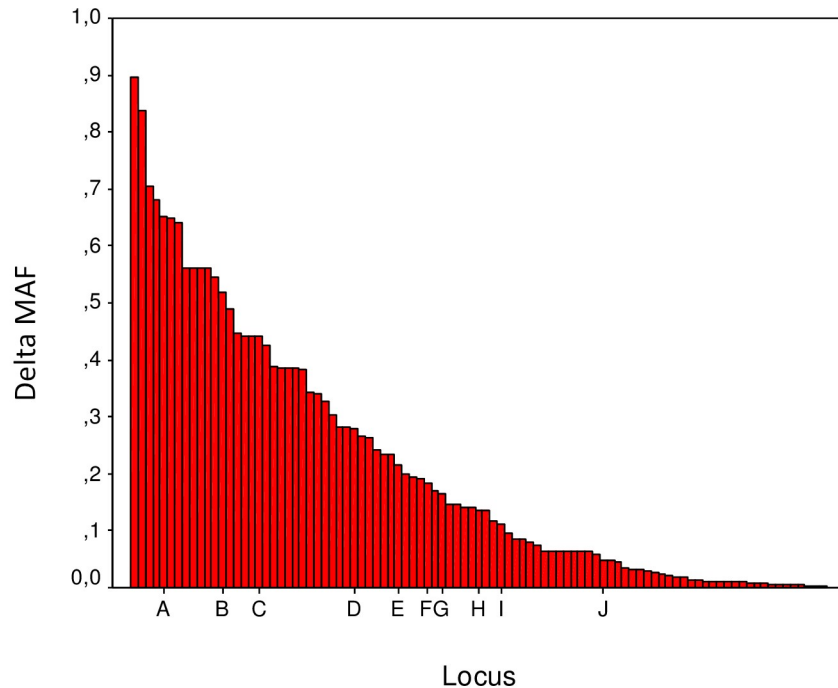


Fig 3. Principal component analyses (PCA) using the allele frequencies of the ADME core genes' loci. Scatterplot illustrates the grouping of 21 populations by the first two principal components of the population-based PCA and shows the loci with the largest factor score values as revealed by the gene-based PCA (gPC2 and gPC3).

<https://doi.org/10.1371/journal.pone.0207671.g003>

[30]. Additionally, this study identifies six ADME genes' SNP loci that most prominently distinguish continental groups which extends our knowledge about ADME global variation. They are: *VKORC1* (rs9923231), *SLCO1B3* (rs4149117), *CYP3A4* (rs2242480), and *NAT2* (rs1208, rs1801280, and rs1799929).



Roma are outliers (max. & min. frequencies) in 10 loci:

- | | | | |
|-----|----|------------|-----------------------|
| 1. | A# | rs4149117 | (<i>SLCO1B3</i>) |
| 2. | B | rs1128503 | (<i>ABCB1</i>) |
| 3. | C | rs1902023 | (<i>UGT2B15</i>) |
| 4. | D | rs12248560 | (<i>CYP2C19*17</i>) |
| 5. | E | rs1799853 | (<i>CYP2C9</i>) |
| 6. | F | rs3758581 | (<i>CYP2C19</i>) |
| 7. | G | rs10509681 | (<i>CYP2C8*3</i>) |
| 8. | H | rs1138272 | (<i>GSTP1</i>) |
| 9. | I | rs8192709 | (<i>CYP2B6</i>) |
| 10. | J | rs34059508 | (<i>SLC22A1</i>) |

Roma have the lowest MAF among 21 populations

Fig 4. The maximal global differences in minor allele frequencies (delta) for the selected 95 ADME core genes' loci in decreasing order.

<https://doi.org/10.1371/journal.pone.0207671.g004>

Among all the studied loci, locus rs9923231 in *VKORC1* gene takes a special position. It is the one characterized by the largest MAF differences among the populations worldwide ($\Delta = 0.897$) and is among few SNPs that define both African and East Asian continental clusters. Those findings are even more important having in mind that among the ADME genes *VKORC1* is the one with the most vital implications for medical decisions considering appropriate pharmacotherapy.

Striking differences in worldwide allele frequency distribution noticed in rs9923231 is mostly explained as a result of selection [31]. The minor allele frequency of 0.4614 in Roma is comparable to European populations and it is higher than in their ancestral south Asian populations. *VKORC1* gene encodes vitamin K 2,3-epoxide reductase complex 1, which is responsible for the conversion of vitamin K-epoxide to vitamin K [32]. Numerous studies suggest that *VKORC1* genotype seems to be the most important predictor of adequate warfarin dose [33, 34]. The rs9923231 has proved clinical importance concerning dosage and toxicity of warfarin, acenocoumarol and phenprocoumon.

Another locus, which shows large differences in allele frequencies among the investigated populations, is rs2242480, the intron variant in *CYP3A4* gene. This locus exhibits significant difference between Africans and non-Africans [35] which is evident in our gPC plot as well. Lakiotaki et al. [36] identified this variant to belong to the 10 most different ones in worldwide populations. Unlike the previously mentioned locus in *VKORC1*, the frequency of globally minor allele of rs2242480 in the investigated Roma population (0.3012) is higher than in European populations and corresponds to the frequency range of South Asian populations. *CYP3A4* is responsible for the metabolism of approximately 50–60% of clinical drugs used today, including acetaminophen, codeine, cyclosporine A, diazepam, and erythromycin. It is also important for the metabolism of steroid hormones [33, 34]. Although there is no clear evidence of an association, the rs2242480 is suspected to be connected with methadone toxicity [37], clopidogrel efficacy [38] and pharmacokinetics of tacrolimus [39] carbamazepine [40].

Locus rs4149117 in *SLCO1B3* gene also separates African cluster from the other populations. The G allele of the rs4149771 locus was found even two times more frequently in non-Africans (Europeans, Caucasian Americans, and Asians) than in Africans, in approximately 80% vs. 40% of subjects [41, 42]. Roma population from Croatia has the lowest minor allele frequency in Europe. *SLCO1B3* gene encodes for solute carrier organic anion transporter family member 1B3 normally expressed in the liver and involved in transporter functions to uptake large, non-polar drugs and hormones. The rs4149771, like the previous one, is suspected to be connected with carboplatin and paclitaxel toxicity [43] and with sunitinib efficacy [44].

NAT2, one of the most polymorphic ADME genes, encodes for a NAT2 protein, which is expressed mostly in the liver, small intestine and colon tissues as a typical xenobiotic metabolizing enzyme [45]. NAT2 gene variants differ among diverse populations and its genetic differentiation patterns are related to geography [46]. The minor alleles' frequencies of three NAT2 loci (rs1208, rs1801280, and rs1799929) in Roma sample are within the range of European populations. Isoniazid, a first line drug in tuberculosis (TB) treatment is metabolized by the NAT2 enzyme. Genetic variations in NAT2 affect the therapeutic response to isoniazid and other drugs detoxified by this enzyme.

Among loci with the most pronounced population differences as revealed by MAF Δ values is rs1048943 missense (Ile462Val) mutation within *CYP1A1* gene. *CYP1A1* is a member of the CYP1 family and participates in the metabolism of numerous xenobiotics, as well as endogenous substrates [47]. *CYP1A1* is a key enzyme in phase I metabolism of polycyclic aromatic hydrocarbons and in estrogen metabolism. This mutation defines haplotype *CYP1A1*2C*. The highest frequency of the mutated allele is noticed in the South American population from Peru while the absence of mutated allele is noticed in most of the African

populations. Such distribution implies its introduction after the Out of Africa migrational event and therefore its current distribution probably results from genetic drift. Roma population has the frequency of minor allele within the European range which is substantially lower than in the ancestral South Asian populations.

Although there is evidence of selection for some of the above mentioned genes, the overall large allele variations between populations more often result from genetic drift, migrations and other demographic events [48].

As it can be seen from our results, ADME core loci separate African and East Asian clusters from other Euro-Asian and American populations. This pattern is also confirmed by clustering of genetic distances. The Roma population is positioned within the European cluster and is close to the South Asian populations. Such results suggest that Roma ADME gene pool is a combination of two main layers: ancestral (Indian) and more recent (European). This is also evident from the analyses of the uniparental genetic markers [22, 23]. Similarly, Melegh et al. [49] found that the Roma are located on a PCA cline between Europeans and South Asians, but closer to Europeans by analyzing genome-wide SNP loci.

Although Roma population is found to be a member of the closely related European and South Asian clusters, it has the extreme MAF values in 10 out of 95 analyzed SNPs. Significant genetic differentiation from general Europeans in SNPs in the *CYP2C* and *CYP2D* subfamily regions was also found in previous research of isolated populations in Europe (Roma, Basques, and Orcadians) [50]. Among the former analyses of ADME polymorphisms in Roma populations, Tomas et al. [27] particularly studied 3 SNP loci in the *CYP2B6* gene, Spikey et al. [51] studied 4 SNP loci in the *MDR1* gene while Nagy et al. [15] analyzed 2 SNP loci in the *SLCO1B3* gene and all of them confirmed that the Roma differ considerably from geographically close majority populations, as well as from Indian populations in those particular loci. Such results are not surprising knowing Roma genetic history which is influenced by strict rules of group endogamy, reproductive isolation and specific mating practice and isolation over the past several centuries.

Our population genetics findings contribute to the knowledge of interpopulation differences in high-risk pharmacogenomics allele distribution. The Pharmacogenomics Knowledgebase (PharmGKB) is a source of clinically relevant information, including dosing guidelines, annotated drug labels, and potentially actionable gene–drug associations and genotype–phenotype relationships [52]. Several loci which have extraordinary MAFs in Roma population are listed in PharmGKB as loci with strong proof of clinical importance. Two of them have been clinically annotated as level 1A (strong evidence—included in the Clinical Pharmacogenetics Implementation Consortium–CPIC guidelines): rs1799853 (*CYP2C9*) for warfarin dosage, and rs12248560 (*CYP2C19*) for clopidogrel dosage, efficacy and toxicity, while the second locus has been also clinically annotated as level 2A (very important pharmacogene) for citalpram or escitalopram pharmacokinetics. Additionally, rs10509681 (*CYP2C8*) has been annotated as level 2A for rosiglitazone pharmacokinetics and rs1902023 (*UGT2B15*) as level 2B (moderate clinical evidence) for lorazepam or oxazepam (www.pharmagkb.org). The identification of high risk allele at loci whose genotypes have a direct influence on quality of drug intake in this population, shows the necessity of the assessment of unique genetic profile of Roma in order to achieve the most in the modulation of pharmacotherapy in this population.

Our data confirm that isolated populations take specific positions within the global ADME genetic landscape. This pinpoints that the pharmacogenetics guidelines of the well-defined majority populations cannot be used in pharmaco-therapeutic practice in population isolates, and confirms the necessity for defining their specific genetic profile.

Material and methods

Biological material used in this study was collected in multiple field studies, which were part of the on-going multidisciplinary anthropological, molecular-genetic and epidemiological research of Roma populations in Croatia. The fieldwork was carried out in several regions of Croatia with the highest number of Roma minority inhabitants according to the census data [18]. The participants were volunteers and were informed about the goals, methods and expectations of the study with the help of linguistically and culturally competent and trained Roma volunteers. The study protocol was approved by the Scientific Board and the Ethical Committee of the Institute for Anthropological Research in Zagreb, Croatia.

Genotyping of 439 DNA samples was done using KASP method. The KASP genotyping assay is a form of competitive allele-specific PCR combined with homogeneous fluorescent SNP genotyping system, which determines the alleles at a specific locus within genomic DNA [53]. This technology has been widely used on plant species, while recently it has been successfully applied to human samples too [54, 55]. From the list of evidence-proved genetic biomarkers associated with metabolism of drugs, which is available at www.pharmaadme.org, 137 single nucleotide polymorphisms (SNPs) were selected for genotyping using the KASP and 127 of them were genotyped successfully. Allele and genotype frequencies were calculated by direct counting method.

The present investigation of genetic diversity was based on SNPs from the ADME core list which were genotyped in both the Croatian Roma and in 20 populations with different genetic ancestry from the 1000 Genomes Project Phase 3 list. This limitation and the finding that four ADME SNPs genotyped were monomorphic, led to further reduction of the total number of SNPs so in the end a total of 95 SNPs located in 31 ADME genes were used for the analyses.

The 20 populations from the 1000 Genomes project belong to the five large continental regions, and each region is represented by four populations: (1) European (EUR): Finland, Italy, Spain, UK, (2) South Asian (SAS): Bangladesh, India, Pakistan, Sri Lanka, (3) African (AFR): Gambia, Kenya, Nigeria, Sierra Leone, (4) Central and South American (AMR): Colombia, Mexico, Peru, Puerto Rico, and (5) Eastern Asian (EAS): Dai Chinese, Han Chinese, Japan, Vietnam (only China is represented by two distinct populations—Han and Dai—since Han is a majority population while Dai represent here non-Han China populations).

In our analyses, we enriched the 1000 Genomes' data with those found in the publications citing any of the 95 investigated SNPs in the above mentioned populations. Selection criteria for using data from these publications, listed at the *e!Ensembl* browser for the each SNP, were: (1) clearly stated study geographical population and, where relevant, participants' ethnicity, (2) alleles frequencies and sample sizes, (3) samples come from the general population or control groups in case-control studies. These additional genotyping data enlarged the size of the 11 following 1000 Genomes populations: Italy, Spain, UK, India, Sri Lanka, Gambia, Kenya, Colombia, Mexico, Han Chinese and Japan. Allele frequencies for these populations were calculated by weighting samples for each population.

The genetic distance matrix, computed according to the method of Nei (1972), was subjected to hierarchic clustering routine using UPGMA (unweighted pair-group method using arithmetic averages) available in free software Phylip v3.697 (<http://evolution.gs.washington.edu/phylip.html>).

The Mantel test of correlation between genetic and geographic distances was performed using non-commercial software *IBD: Isolation by distance* v1.52 (available at <http://www.bio.sdsu.edu/pub/andy/IBD.html>). Geographic distances between the analyzed populations were calculated using two free online softwares: *iTouchMap* and *Movable Type Scripts*. *iTouchMap* calculates latitude and longitude of a point, and *Movable Type Scripts* calculated distance

between latitude/longitude points (available at <http://itouchmap.com/latlong.html> and <http://www.movable-type.co.uk/scripts/latlong.html>).

Principal component analysis (PCA) is a multivariate method that systematically identifies underlying variables, or principal components (PCs), that best differentiate a set of data [56]. Two analyses were run using the MAF data of the 95 ADME core SNPs in 21 populations. First, PCA was performed to investigate the grouping of 21 populations using the known genetic data. The second analysis, the gene-oriented PCA (gPCA), was run to investigate the clustering of SNPs using the a priori defined 21 populations in order to detect loci defining the population clusters obtained in PCA. The number of PCs considered in each analysis was determined from the scree plot. This statistics was performed using the SPSS software package 17.0.

Supporting information

S1 Table. The minor allele frequency (MAF) weighted values and sample sizes for the selected 95 ADME core genes' loci in Croatian Roma (present study) and in 20 populations worldwide. The references are provided in cases when data from the literature are used in addition to the 1000 Genomes data.

(XLSX)

S2 Table. Populations with maximal and minimal minor allele frequencies (MAF) values for the selected 95 ADME core genes' loci. The list is ordered by decreasing delta values (difference between maximal and minimal MAF).

(DOCX)

Acknowledgments

We are deeply grateful to the Roma people for their kindness and the interest for participation in this study.

Author Contributions

Conceptualization: Tatjana Škarić-Jurić, Nina Smolej Narančić, Branka Janićijević, Marijana Peričić Salihović.

Data curation: Željka Tomas, Matea Zajc Petranović.

Formal analysis: Tatjana Škarić-Jurić, Matea Zajc Petranović, Nina Smolej Narančić, Marijana Peričić Salihović.

Funding acquisition: Marijana Peričić Salihović.

Investigation: Tatjana Škarić-Jurić, Željka Tomas, Matea Zajc Petranović, Nina Smolej Narančić, Branka Janićijević, Marijana Peričić Salihović.

Methodology: Tatjana Škarić-Jurić, Nina Smolej Narančić, Branka Janićijević, Marijana Peričić Salihović.

Project administration: Marijana Peričić Salihović.

Resources: Branka Janićijević, Marijana Peričić Salihović.

Supervision: Nada Božina.

Validation: Nada Božina.

Visualization: Tatjana Škarić-Jurić, Nina Smolej Narančić.

Writing – original draft: Tatjana Škarić-Jurić, Nina Smolej Narančić, Branka Janičijević, Marijana Peričić Salihović.

Writing – review & editing: Tatjana Škarić-Jurić, Željka Tomas, Matea Zajc Petranović, Nada Božina, Nina Smolej Narančić, Branka Janičijević, Marijana Peričić Salihović.

References

1. Maisano Delser P, Fuselli S. Human loci involved in drug biotransformation: worldwide genetic variation, population structure, and pharmacogenetic implications. *Human genetics*. 2013; 132(5):563–77. Epub 2013/01/29. <https://doi.org/10.1007/s00439-013-1268-5> PMID: 23354977.
2. Ravindra Kumar M, Adithan C. Pharmacogenomics in the Indian population. In: Suarez-Kurtz G, editor. *Pharmacogenomics in Admixed Populations*: Landes Bioscience; 2007.
3. Li J, Lou H, Yang X, Lu D, Li S, Jin L, et al. Genetic architectures of ADME genes in five Eurasian admixed populations and implications for drug safety and efficacy. *Journal of medical genetics*. 2014; 51(9):614–22. Epub 2014/07/31. <https://doi.org/10.1136/jmedgenet-2014-102530> PMID: 25074363.
4. Pasanen MK, Neuvonen PJ, Niemi M. Global analysis of genetic variation in SLCO1B1. *Pharmacogenomics*. 2008; 9(1):19–33. Epub 2007/12/25. <https://doi.org/10.2217/14622416.9.1.19> PMID: 18154446.
5. Suarez-Kurtz G, Pena SDJ, Struchiner CJ, Hutz MH. Pharmacogenomic Diversity among Brazilians: Influence of Ancestry, Self-Reported Color, and Geographical Origin. *Frontiers in Pharmacology*. 2012; 3:191. <https://doi.org/10.3389/fphar.2012.00191> PubMed PMID: PMC3490152. PMID: 23133420
6. Kim JY, Cheong HS, Park T-J, Shin HJ, Seo DW, Na HS, et al. Screening for 392 polymorphisms in 141 pharmacogenes. *Biomedical Reports*. 2014; 2(4):463–76. <https://doi.org/10.3892/br.2014.272> PubMed PMID: PMC4051470. PMID: 24944790
7. Jittikoon J, Mahasirimongkol S, Charoenyingwattana A, Chaikledkaew U, Tragulpiankit P, Mangmool S, et al. Comparison of genetic variation in drug ADME-related genes in Thais with Caucasian, African and Asian HapMap populations. *Journal of human genetics*. 2016; 61(2):119–27. Epub 2015/10/02. <https://doi.org/10.1038/jhg.2015.115> PMID: 26423926.
8. Mizzi C, Dalabira E, Kumuthini J, Dzimiri N, Balogh I, Basak N, et al. A European Spectrum of Pharmacogenomic Biomarkers: Implications for Clinical Pharmacogenomics. *PloS one*. 2016; 11(9):e0162866. Epub 2016/09/17. <https://doi.org/10.1371/journal.pone.0162866> PMID: 27636550; PubMed Central PMCID: PMC45026342.
9. Mwinyi J, Kopke K, Schaefer M, Roots I, Gerloff T. Comparison of SLCO1B1 sequence variability among German, Turkish, and African populations. *European journal of clinical pharmacology*. 2008; 64(3):257–66. Epub 2008/01/11. <https://doi.org/10.1007/s00228-007-0409-y> PMID: 18185926.
10. Rajman I, Knapp L, Morgan T, Masimirembwa C. African Genetic Diversity: Implications for Cytochrome P450-mediated Drug Metabolism and Drug Development. *EBioMedicine*. 2017; 17:67–74. Epub 2017/02/27. <https://doi.org/10.1016/j.ebiom.2017.02.017> PMID: 28237373; PubMed Central PMCID: PMC45360579.
11. Phipps-Green AJ, Hollis-Moffatt JE, Dalbeth N, Merriman ME, Topless R, Gow PJ, et al. A strong role for the ABCG2 gene in susceptibility to gout in New Zealand Pacific Island and Caucasian, but not Maori, case and control sample sets. *Human molecular genetics*. 2010; 19(24):4813–9. Epub 2010/09/23. <https://doi.org/10.1093/hmg/ddq412> PMID: 20858603.
12. Brinar M, Cukovic-Cavka S, Bozina N, Ravic KG, Markos P, Ladic A, et al. MDR1 polymorphisms are associated with inflammatory bowel disease in a cohort of Croatian IBD patients. *BMC gastroenterology*. 2013; 13:57. Epub 2013/03/30. <https://doi.org/10.1186/1471-230X-13-57> PMID: 23537364; PubMed Central PMCID: PMC3616873.
13. Campa D, Sainz J, Pardini B, Vodickova L, Naccarati A, Rudolph A, et al. A comprehensive investigation on common polymorphisms in the MDR1/ABCB1 transporter gene and susceptibility to colorectal cancer. *PloS one*. 2012; 7(3):e32784. Epub 2012/03/08. <https://doi.org/10.1371/journal.pone.0032784> PMID: 22396794; PubMed Central PMCID: PMC3292569.
14. Zou JG, Ma YT, Xie X, Yang YN, Pan S, Adi D, et al. The association between CYP1A1 genetic polymorphisms and coronary artery disease in the Uygur and Han of China. *Lipids Health Dis*. 2014; 13:145. Epub 2014/09/06. <https://doi.org/10.1186/1476-511X-13-145> PMID: 25189712; PubMed Central PMCID: PMC4175619.
15. Nagy A, Sipeky C, Szalai R, Melegh BI, Matyas P, Ganczer A, et al. Marked differences in frequencies of statin therapy relevant SLCO1B1 variants and haplotypes between Roma and Hungarian populations. *BMC genetics*. 2015; 16:108. Epub 2015/09/04. <https://doi.org/10.1186/s12863-015-0262-4> PMID: 26334733; PubMed Central PMCID: PMC4559300.

16. Sipeky C, Csonge V, Jaromi L, Safrany E, Polgar N, Lakner L, et al. Vitamin K epoxide reductase complex 1 (VKORC1) haplotypes in healthy Hungarian and Roma population samples. *Pharmacogenomics*. 2009; 10(6):1025–32. Epub 2009/06/18. <https://doi.org/10.2217/pgs.09.46> PMID: 19530970.
17. Sipeky C, Weber A, Szabo M, Melegh BI, Janicsek I, Tarlos G, et al. High prevalence of CYP2C19*2 allele in Roma samples: study on Roma and Hungarian population samples with review of the literature. *Molecular biology reports*. 2013; 40(8):4727–35. Epub 2013/05/07. <https://doi.org/10.1007/s11033-013-2569-4> PMID: 23645039.
18. Statistics CBo. 2011.
19. Fraser A. *The Gypsies*. Oxford: Blackwell Publishers; 1992.
20. Hancock IF. *We Are the Romani People*. Hatfield: University of Herdforsshire Press; 2002.
21. Morar B, Gresham D, Angelicheva D, Tournev I, Gooding R, Guergueltcheva V, et al. Mutation History of the Roma/Gypsies. *American Journal of Human Genetics*. 2004; 75(4):596–609. PubMed PMID: PMC1182047. <https://doi.org/10.1086/424759> PMID: 15322984
22. Salihovic MP, Baresic A, Klaric IM, Cukrov S, Lauc LB, Janicijevic B. The role of the Vlach Roma in shaping the European Romani maternal genetic history. *American journal of physical anthropology*. 2011; 146(2):262–70. Epub 2011/09/15. <https://doi.org/10.1002/ajpa.21566> PMID: 21915846.
23. Klaric IM, Salihovic MP, Lauc LB, Zhivotovsky LA, Rootsi S, Janicijevic B. Dissecting the molecular architecture and origin of Bayash Romani patrilineages: genetic influences from South-Asia and the Balkans. *American journal of physical anthropology*. 2009; 138(3):333–42. Epub 2008/09/13. <https://doi.org/10.1002/ajpa.20933> PMID: 18785634.
24. Pokupcic K, Cukrov S, Klaric IM, Salihovic MP, Lauc LB, Blazanovic A, et al. Y-STR genetic diversity of Croatian (Bayash) Roma. *Forensic science international Genetics*. 2008; 2(2):e11–3. Epub 2008/12/17. <https://doi.org/10.1016/j.fsigen.2007.11.002> PMID: 19083796.
25. Zeljko HM, Škarić-Jurić T, Narančić NS, Tomas Ž, Barešić A, Salihović MP, et al. E2 allele of the Apolipoprotein E gene polymorphism is predictive for obesity status in Roma minority population of Croatia. *Lipids in Health and Disease*. 2011; 10(1):9. <https://doi.org/10.1186/1476-511x-10-9> PMID: 21244662
26. Baresic A, Peric Salihovic M. Carrier rates of four single-gene disorders in Croatian Bayash Roma. *Genetic testing and molecular biomarkers*. 2014; 18(2):83–7. Epub 2013/11/05. <https://doi.org/10.1089/gtmb.2013.0323> PMID: 24180318; PubMed Central PMCID: PMCPCMC3926160.
27. Tomas Ž, Kuhanec A, Škarić-Jurić T, Petranović MZ, Narančić NS, Janićijević B, et al. Distinctiveness of the Roma population within CYP2B6 worldwide variation. *Pharmacogenomics*. 2017; 18(17):1575–87. <https://doi.org/10.2217/pgs-2017-0105> PMID: 29095103.
28. Hatzikotoulas K, Gilly A, Zeggini E. Using population isolates in genetic association studies. *Briefings in functional genomics*. 2014; 13(5):371–7. Epub 2014/07/11. <https://doi.org/10.1093/bfpg/elu022> PMID: 25009120; PubMed Central PMCID: PMCPCMC4168662.
29. Panoutsopoulou K, Hatzikotoulas K, Xifara DK, Colonna V, Farmaki AE, Ritchie GR, et al. Genetic characterization of Greek population isolates reveals strong genetic drift at missense and trait-associated variants. *Nat Commun*. 2014; 5:5345. Epub 2014/11/07. <https://doi.org/10.1038/ncomms6345> PMID: 25373335; PubMed Central PMCID: PMCPCMC4242463.
30. Ramos E, Doumatey A, Elkahloun AG, Shriner D, Huang H, Chen G, et al. Pharmacogenomics, ancestry and clinical decision making for global populations. *The Pharmacogenomics Journal*. 2013; 14:217. <https://doi.org/10.1038/tpj.2013.24> <https://www.nature.com/articles/tpj201324#supplementary-information>. PMID: 23835662
31. Ross KA, Bigham AW, Edwards M, Gozdzik A, Suarez-Kurtz G, Parra EJ. Worldwide allele frequency distribution of four polymorphisms associated with warfarin dose requirements. *Journal of human genetics*. 2010; 55(9):582–9. Epub 2010/06/18. <https://doi.org/10.1038/jhg.2010.73> PMID: 20555338.
32. Wajih N, Hutson SM, Owen J, Wallin R. Increased production of functional recombinant human clotting factor IX by baby hamster kidney cells engineered to overexpress VKORC1, the vitamin K 2,3-epoxide-reducing enzyme of the vitamin K cycle. *The Journal of biological chemistry*. 2005; 280(36):31603–7. Epub 2005/07/21. <https://doi.org/10.1074/jbc.M505373200> PMID: 16030016.
33. Yin T, Miyata T. Warfarin dose and the pharmacogenomics of CYP2C9 and VKORC1—rationale and perspectives. *Thrombosis research*. 2007; 120(1):1–10. Epub 2006/12/13. <https://doi.org/10.1016/j.thromres.2006.10.021> PMID: 17161452.
34. Wadelius M, Chen LY, Eriksson N, Bumpstead S, Ghori J, Wadelius C, et al. Association of warfarin dose with genes involved in its action and metabolism. *Human genetics*. 2007; 121(1):23–34. Epub 2006/10/19. <https://doi.org/10.1007/s00439-006-0260-8> PMID: 17048007; PubMed Central PMCID: PMCPCMC1797064.
35. Chen X, Wang H, Zhou G, Zhang X, Dong X, Zhi L, et al. Molecular Population Genetics of Human CYP3A Locus: Signatures of Positive Selection and Implications for Evolutionary Environmental

- Medicine. *Environmental Health Perspectives*. 2009; 117(10):1541–8. <https://doi.org/10.1289/ehp.0800528> PubMed PMID: PMC2790508. PMID: 20019904
36. Lakiotaki K, Kanterakis A, Kartsaki E, Katsila T, Patrinos GP, Potamias G. Exploring public genomics data for population pharmacogenomics. *PloS one*. 2017; 12(8):e0182138. <https://doi.org/10.1371/journal.pone.0182138> PMID: 28771511
 37. Chen CH, Wang SC, Tsou HH, Ho IK, Tian JN, Yu CJ, et al. Genetic polymorphisms in CYP3A4 are associated with withdrawal symptoms and adverse reactions in methadone maintenance patients. *Pharmacogenomics*. 2011; 12(10):1397–406. Epub 2011/09/10. <https://doi.org/10.2217/pgs.11.103> PMID: 21902501.
 38. Angiolillo DJ, Fernandez-Ortiz A, Bernardo E, Ramirez C, Cavallari U, Trabetti E, et al. Contribution of gene sequence variations of the hepatic cytochrome P450 3A4 enzyme to variability in individual responsiveness to clopidogrel. *Arteriosclerosis, thrombosis, and vascular biology*. 2006; 26(8):1895–900. Epub 2006/04/29. <https://doi.org/10.1161/01.ATV.0000223867.25324.1a> PMID: 16645157.
 39. Liu F, Ou YM, Yu AR, Xiong L, Xin HW. Long-Term Influence of CYP3A5, CYP3A4, ABCB1, and NR1H2 Polymorphisms on Tacrolimus Concentration in Chinese Renal Transplant Recipients. *Genetic testing and molecular biomarkers*. 2017; 21(11):663–73. Epub 2017/09/26. <https://doi.org/10.1089/gtmb.2017.0088> PMID: 28945481.
 40. Zhu X, Yun W, Sun X, Qiu F, Zhao L, Guo Y. Effects of major transporter and metabolizing enzyme gene polymorphisms on carbamazepine metabolism in Chinese patients with epilepsy. *Pharmacogenomics*. 2014; 15(15):1867–79. Epub 2014/12/17. <https://doi.org/10.2217/pgs.14.142> PMID: 25495409.
 41. Smith NF, Marsh S, Scott-Horton TJ, Hamada A, Mielke S, Moss K, et al. Variants in the SLCO1B3 gene: interethnic distribution and association with paclitaxel pharmacokinetics. *Clinical pharmacology and therapeutics*. 2007; 81(1):76–82. Epub 2006/12/23. <https://doi.org/10.1038/sj.clpt.6100011> PMID: 17186002.
 42. Laitinen A, Niemi M. Frequencies of Single-Nucleotide Polymorphisms of SLCO1A2, SLCO1B3 and SLCO2B1 Genes in a Finnish Population. *Basic & Clinical Pharmacology & Toxicology*. 2010; 108(1):9–13. <https://doi.org/10.1111/j.1742-7843.2010.00605.x> PMID: 20560925
 43. Park HS, Lim SM, Shin HJ, Cho A, Shin JG, Lee MG, et al. Pharmacogenetic analysis of advanced non-small-cell lung cancer patients treated with first-line paclitaxel and carboplatin chemotherapy. *Pharmacogenetics and genomics*. 2016; 26(3):116–25. Epub 2015/12/08. <https://doi.org/10.1097/FPC.000000000000196> PMID: 26641474.
 44. Kloth JSL, Verboom MC, Swen JJ, van der Straaten T, Sleijfer S, Reyners AKL, et al. Genetic polymorphisms as predictive biomarker of survival in patients with gastrointestinal stromal tumors treated with sunitinib. *Pharmacogenomics J*. 2018; 18(1):49–55. Epub 2017/01/25. <https://doi.org/10.1038/tpj.2016.83> PMID: 28117434.
 45. Boukouvala S, Fakis G. Arylamine N-acetyltransferases: what we learn from genes and genomes. *Drug metabolism reviews*. 2005; 37(3):511–64. Epub 2005/11/01. <https://doi.org/10.1080/03602530500251204> PMID: 16257833.
 46. Sabbagh A, Langaney A, Darlu P, Gerard N, Krishnamoorthy R, Poloni ES. Worldwide distribution of NAT2 diversity: implications for NAT2 evolutionary history. *BMC genetics*. 2008; 9:21. Epub 2008/02/29. <https://doi.org/10.1186/1471-2156-9-21> PMID: 18304320; PubMed Central PMCID: PMCPMC2292740.
 47. McManus ME, Burgess WM, Veronese ME, Huggett A, Quattrochi LC, Tukey RH. Metabolism of 2Acetylaminofluorene and Benzo(a)pyrene and Activation of Food derived Heterocyclic Amine Mutagens by Human Cytochromes P-4501990. 3367–76 p.
 48. Hofer T, Foll M, Excoffier L. Evolutionary forces shaping genomic islands of population differentiation in humans. *BMC Genomics*. 2012; 13:107. <https://doi.org/10.1186/1471-2164-13-107> PMID: 22439654; PubMed Central PMCID: PMCPMC3317871.
 49. Melegh BI, Banfai Z, Hadzsiev K, Miseta A, Melegh B. Refining the South Asian Origin of the Romani people. *BMC genetics*. 2017; 18(1):82. Epub 2017/09/02. <https://doi.org/10.1186/s12863-017-0547-x> PMID: 28859608; PubMed Central PMCID: PMCPMC5580230.
 50. Pimenoff VN, Laval G, Comas D, Palo JU, Gut I, Cann H, et al. Similarity in recombination rate and linkage disequilibrium at CYP2C and CYP2D cytochrome P450 gene regions among Europeans indicates signs of selection and no advantage of using tagSNPs in population isolates. *Pharmacogenetics and genomics*. 2012; 22(12):846–57. Epub 2012/10/24. <https://doi.org/10.1097/FPC.0b013e32835a3a6d> PMID: 23089684.
 51. Sipeky C, Csonge V, Jaromi L, Safrany E, Maasz A, Takacs I, et al. Genetic variability and haplotype profile of MDR1 (ABCB1) in Roma and Hungarian population samples with a review of the literature. *Drug metabolism and pharmacokinetics*. 2011; 26(2):206–15. Epub 2010/12/24. PMID: 21178299.

52. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, et al. Pharmacogenomics knowledge for personalized medicine. *Clinical pharmacology and therapeutics*. 2012; 92(4):414–7. Epub 2012/09/21. <https://doi.org/10.1038/clpt.2012.96> PMID: 22992668; PubMed Central PMCID: PMC3660037.
53. Semagn K, Babu R, Hearne S, Olsen M. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Molecular Breeding*. 2014; 33(1):1–14. <https://doi.org/10.1007/s11032-013-9917-x>
54. Rothe H, Brandenburg V, Haun M, Kollerits B, Kronenberg F, Ketteler M, et al. Ecto-5'-Nucleotidase CD73 (NT5E), vitamin D receptor and FGF23 gene polymorphisms may play a role in the development of calcific uremic arteriopathy in dialysis patients—Data from the German Calciphylaxis Registry. *PloS one*. 2017; 12(2):e0172407. Epub 2017/02/18. <https://doi.org/10.1371/journal.pone.0172407> PMID: 28212442; PubMed Central PMCID: PMC5315275.
55. Landoulsi Z, Benromdhan S, Ben Djebara M, Damak M, Dallali H, Kefi R, et al. Using KASP technique to screen LRRK2 G2019S mutation in a large Tunisian cohort. *BMC Medical Genetics*. 2017; 18:70. <https://doi.org/10.1186/s12881-017-0432-5> PubMed PMID: PMC5501550. PMID: 28683740
56. Menozzi P, Piazza A, Cavalli-Sforza L. Synthetic maps of human gene frequencies in Europeans. *Science (New York, NY)*. 1978; 201(4358):786–92. Epub 1978/09/01. PMID: 356262.