



OPEN

Genetic diversity of variants involved in drug response among Tunisian and Italian populations toward personalized medicine

Haifa Jmel^{1,2,3}, Stefania Sarno⁴, Cristina Giuliani⁴, Wided Boukhalfa^{1,2}, Sonia Abdelhak^{1,2}, Donata Luiselli⁵ & Rym Kefi^{1,2,3}✉

Adverse drug reactions (ADR) represent a significant contributor to morbidity and mortality, imposing a substantial financial burden. Genetic ancestry plays a crucial role in drug response. The aim of this study is to characterize the genetic variability of selected pharmacogenes involved with ADR in Tunisians and Italians, with a comparative analysis against global populations. A cohort of 135 healthy Tunisians and 737 Italians were genotyped using a SNP array. Variants located in 25 Very Important Pharmacogenes implicated in ADR were extracted from the genotyping data. Distribution analysis of common variants in Tunisian and Italian populations in comparison to 24 publicly available worldwide populations was performed using PLINK and R software. Results from Principle Component and ADMIXTURE analyses showed a high genetic similarity among Mediterranean populations, distinguishing them from Sub-Saharan African and Asian populations. The *Fst* comparative analysis identified 27 variants exhibiting significant differentiation between the studied populations. Among these variants, four SNPs rs622342, rs3846662, rs7294, rs5215 located in *SLC22A1*, *HMGCR*, *VKORC1* and *KCNJ11* genes respectively, are reported to be associated with ethnic variability in drug responses. In conclusion, correlating the frequencies of genotype risk variants with their associated ADRs would enhance drug outcomes and the implementation of personalized medicine in the studied populations.

The therapeutic failure and severe adverse drug reactions (ADR) remain one of the significant unmet medical challenges, representing major preventable causes of morbidity and mortality¹. ADRs are the most frequent cause of all hospital admissions and rank as the sixth leading cause of death. Furthermore, ADRs increase healthcare costs due to extended hospitalization stay and supplementary clinical testing for severe cases^{2,3}. ADR can be classified into two types:

Type A reactions represent dose-dependent and predictable reactions based on known pharmacological actions of the drug. Type A reactions are relatively common and include hypoglycemia induced by antidiabetic drugs and hemorrhagic reactions induced by anticoagulants, such as warfarin, acenocoumarol. These reactions represent 75% of the ADR, and are less likely to induce fatal consequences than type B. Despite their morbidity rate, ADR type A is not well recognized due to its progressive appearance⁴.

Type B reactions represent unpredictable pharmacological reactions and they are not necessarily dose-dependent. Type B skin reactions include skin disorders such as Stevens Johnson syndrome and toxic epidermal necrolysis. They represent serious reactions of skin hypersensitivity and liver damage^{4,5}.

¹Laboratory of Biomedical Genomics and Oncogenetics, Institut Pasteur de Tunis, Tunis, Tunisia. ²University of Tunis El Manar, Tunis, Tunisia. ³Genetic Typing DNA Service Pasteur Institute, Institut Pasteur de Tunis, Tunis, Tunisia. ⁴Laboratory of Molecular Anthropology & Centre for Genome Biology, Department of Biological, Geological and Environmental Sciences (BiGeA), University of Bologna, Bologna, Italy. ⁵Laboratory of Ancient DNA (aDNALab), Department of Cultural Heritage (DBC), University of Bologna, Ravenna, Italy. ✉email: rym.kefi@pasteur.utm.tn

Growing evidence suggests the involvement of complex and independent factors in the onset of ADRs and drug effectiveness. The genetic background plays a crucial role in pharmacological reactions^{6–11}. Over the last decades, pharmacogenetic studies identified several pharmacogenes and functional variants associated with drug response and helped in the identification of genes that can explain variation in drug response, known as pharmacogenes, as well as pharmacodynamics^{10–15}. Variant–drug response associations have been established for many of these genes and this information can be translated into the clinical setting^{16–18}. For example, approximately 18 percent of prescribed drugs carrying such pharmacogenomics labels in the USA¹⁹. Functional variants or risk alleles exhibit diverse distribution across ethnic groups, influencing the idiosyncratic response of drugs^{2,20,21}. Around a fifth of newly approved drugs showed differences associated with interethnic variability⁶. In this specific instance, ibuprofen employed in the treatment of rheumatoid arthritis, serves as an illustrative case. This drug was never approved in the U.S. and was withdrawn in the U.K. due to hepatotoxicity. However, ibuprofen continues to be available in Japan, where hepatotoxicity is not considered a significant concern^{22,23}. Pharmacogenetic stands out as an important determinant of inter-population variations in drug response²². This field focuses on identifying gene variants responsible for divergent drug responses, aiming to utilize them as predictive markers²⁴. Inter-population frequencies variability in the frequencies of functional variants is common in the genes encoding for phase I drug metabolizing cytochrome P450s enzymes (*CYP2D6*, *CYP2C9*, *CYP2C19*), phase II enzymes catabolism metabolites (*UGT1A1*, *VKORC1*), drug transporters (*SLCO1B1*, *ABCG2*, *ABCB1*) and drug receptor (*P2RY12*)²⁵.

Understanding ethnic contributions to drug efficacy and safety is crucial in the drug development process⁶. Establishing a registry containing pharmacogenetic data accessible to clinicians and the scientific community would be valuable for inferring population-level response rates and ADR risk for different medications. Such registers enable the identification of the populations requiring caution regarding the consumption of specific drugs^{23,26}. A major challenge of performing this type of population level analysis is to remedy the paucity of data on ADR rates in different populations and determine the inherent mechanisms causing these significant proportions of ADR. These studies are of great interest in detecting the association between host genotype, ethnicity, and drug response. This could help formulate specific country-level guidelines on drug response, which are enormously useful and cost-effective in clinical practice^{7,27}. While facing these challenges the past decade has witnessed a surge in pharmacogenetic studies due to the rapid increase in human genome data and the development of high-throughput genotyping technologies. Consequently, ADR occurrences have significantly decreased^{18,28}.

However, few previous studies have focused on exploring pharmacogenomic diversity of Mediterranean populations.

In this context, our aim is (1) to explore the diversity of variants located in pharmacogenes involved in drug biotransformation and ADR among individuals from the Tunisian and Italian populations and (2) to compare our findings with the variability observed among worldwide populations. We choose the Tunisian and Italian population because the Tunisian population is characterized by high genetic heterogeneity²⁹. This could be due to several historical events such as migration and invasion^{30,31}. Previous studies have revealed the genetic mosaicism of the Tunisian population, encompassing different ethnic groups, especially of European ancestry^{29,30,32}. Given the geographical proximity of Tunisia to Italy and the historical events between these populations, we suggest the presence of a common genetic background between them. This could contribute to a better understanding of the pharmacogenetic response to some drugs and thus to the development of tailored therapeutic strategies for these populations.

Results

Population structure at pharmacogenomics loci

Based on an extensive data mining search and PharmGKB interrogation, we selected 25 pharmacogenes implicated in drug response modulation and ADR (Table 1). A total of 148 variants located on the 25 pharmacogenes have been identified in the Tunisian population and kept after quality control steps for subsequent analysis. Only 138 variants have been identified as shared among Tunisian and Italian populations. After applying the threshold criteria of quality control, we have kept out 116 shared variants. To study the genetic diversity of these variants, we compared the genotypic data of the shared variants to those of all 1000 genomes project populations.

In order to assess the degree of similarity in genetic structure between the different populations, we calculated pairwise *F*_{st} values to evaluate the magnitude of differentiation among them. Pairwise *F*_{st} values between the Tunisian and the Italian populations ranged from 0.0014 to 0.004 (Supplementary Table 2). Comparing Tunisian and Italian to other populations, the lowest level of differentiation was observed between the Tunisian and European populations, whereas the highest divergence was observed with Sub Saharan African populations.

The MAF of highly selected actionable pharmacogene variants among TUN and ITA populations (rs622342, rs7294, rs5215 and rs3846662) were compared with EUR and AFR populations (Table 2). Our results revealed that rs622342 located in *SLC22A1* showed no significant differences between the TUN population and the EUR (N_ITA, C_ITA, S_ITA, TSI, and IBS) and AFR (YRI, LWK, ESN, GWD, MSL) populations ($p > 0.05/4 \times 14$). No significant differences were found between TUN and central, southern Italian, and AFR (YRI, ESN, and MSL) populations regarding the rs5215 variant located in *KCNJ11*. However, the MAF of rs5215 showed significant differences between the TUN population and northern Italian and SARD populations. In addition, The MAF of the rs7294 variant located in the *VKORC1* gene was significantly different between the TUN, EUR, and AFR subgroups ($p < 0.05/4 \times 14$). Regarding the rs3846662 variant located in the *HMGCR* gene, we found significant differences between TUN and AFR (YRI, LWK, ESN, GWD, MSL) populations.

PCA revealed that the Tunisian population is clustered with the Italian population (Central, North and the South), European (EUR) and American (AMR) populations. A greatest divergence was observed with the East Asian (EAS) and the African (AFR) populations (Fig. 1). The scatter plot shows that the PCA has the best percent

Drug class	Drug name	VIP	Chr	VIP variant	Toxicity level	Phenotype	Population	Refs.
Anticoagulant	Clopidogrel (Palvix)	<i>ABCB1</i>	7	rs1054564 rs2032582	2A 3	Cardiovascular complications Coronary artery disease	North African European	34,35
		<i>CYP2C19</i>	10	rs1057910 rs4244285	1A 1A	Acute coronary syndrome coronary artery disease/ Myocardial infarction	Mixed population	32,36
		<i>P2RY12</i>	3	rs6785930	4	Neurological events	European	37,38
	Warfarin Acenocoumarol	<i>CYP2C9</i>	10	rs1057910	1A	Hemorrhagic events	Asian	39
		<i>VKORC1</i>	16	rs9923231 rs7294	1B 1A	Hemorrhagic Venous thrombosis	African American _Asian	39,40
Antidiabetic	Biguanide/Metformin	<i>SLC22A1</i>	6	rs628031	3	Gastrointestinal toxicity	European	41
	Biguanide/Metformin	<i>SLC22A2</i>	6	rs12943590	3	Gastrointestinal toxicity and Hypoglycemia Nephrotoxicity		38
	Biguanide/Metformin	<i>SLC47A1</i>	17	rs2289669	3	Efficacy*	Chinese	42
	Biguanide/Metformin	<i>SLC47A2</i>	17	rs12943590	3	Efficacy**	Chinese	43
	Biguanide/Metformin	<i>AT M</i>	11	rs11212617	2B	Efficacy: Great reduction of Fasting Plasma Glucose	Caucasian Chinese	44 43
	TZD/Pioglitazone	<i>PPARG</i>	3	rs1801282	3	Oedema	European	45
	TZD/Rosiglitazone	<i>PGC-1alpha</i>	5	rs1801282	3	Efficacy	Mixed population	46
	TZD/Troglitazone	<i>RETN</i>	19	rs1862513		lower fasting blood glucose	Caucasian and Asian	47
		<i>Leptin LEPR (G)</i>	1	rs1137101	3	Hyperglycemia and T2D susceptibility	Chinese	48
	Sulfonylurea	<i>TNFA1pha</i>	2	rs1800629	3	Hyperglycemia	Egyptian	49
		<i>KCNJ11</i>	11	rs5219	2A	Severe hypoglycemia Microvascular Neuro- pathic complication		50,51
		<i>ABCC8</i>	11	rs757110	3	Anaphylactic shock Serum sickness-like syndrome Systemic allergic vasculitis	Espagnol German Chinese	51-53
		<i>KCNQ1</i>	11	rs2237892, rs2237895	3	Efficacy in term of post- prandial plasma glucose level	Chinese population	54
		<i>TCF7L2</i>	10	rs12255372	3	Efficacy	European	55
		<i>CYP2C9</i>	10	rs1057910 rs1799853	1A	Hypoglycemic attacks	American	56
Lipid-lowering	<i>Fluvastatin</i>	<i>SLCO1B1</i>	12	rs4149056	1	Myopathy	European	57
	<i>Lovastatin</i>	<i>ABCB1</i>	7	rs1128503	3	Hypercholesterolemia	European	6,58
	<i>Atorvastatin</i>	<i>ABCB1</i>	7	rs2032582	2A	Hypercholesterolemia	European	58
	<i>Pravastatin</i>	<i>ABCA1</i>	9	rs12003906		Myocardial infarction	European	60
		<i>CYP2C9</i>	10	rs1057910	1A	Epilepsy	Asian and European	61,62
		<i>PPARA</i>	22	rs4253728	3	Cardiovascular complica- tion	American	63
		<i>SLCO1B1</i>	12	rs4149056	1A	Myopathy	European	64
		<i>HMGCR</i>	5	rs17238540	4	Cardiovascular complica- tion	African and European	38
	<i>CYP3A5</i>	7	rs776746	3	Hypercholesterolemia and Myalgia	European	65	

Table 1. Basic Information about the selected pharmacogenes. The table describes the selected pharmacogenes and variants implicated in the ADR of several anticoagulants, antidiabetics, lipid lowering drugs, FINS: fasting insulin, glycated hemoglobin A1 (HbA1C), HOMA-IS: homeostasis model assessment-insulin resistance. *Higher reduction in Hba1c reduction, FINS and HOMA-IR. **Higher Hba1c reduction by elevating the circulating concentration of metformin. According to clinical guidelines, toxicity level of evidence is classified as follows: 1A, 1B: high toxicity, 2A, 2B: moderate toxicity, 3: low toxicity, 4: lack of research supporting the association of the variation with toxicity.

of explained variance for PC1, estimated at 8.4%. This is the highest percentage among the four prisms displaying all the orientation possibilities of the PCA (Fig. 2).

Gene	PGx-Variant	PGx Variants	SARD	N_ITA	C_ITA	S_ITA	TSI	IBS	CEU	GBR	FIN	YRI	LWK	ESN	GWD	MSL	Relevant Drug
<i>SLC22A1</i>	rs622342	C	0.031	0.116	0.116	0.116	0.116	0.002	0.031	0.062	0.031	0.376	0.731	0.284	0.731	0.284	Metformin
<i>VKORC1</i>	rs7294	C	1.42E-08	4.59E-10	4.59E-10	4.59E-10	4.59E-10	1.30E-06	1.42E-08	2.67E-09	1.42E-08	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	Warfarin, Acenocoumarol
<i>KCNJ11</i>	rs5215	C	0.0002	0.0007	0.0016	0.0015	0.0027	1.26E-05	0.0001	0.001	4.31E-05	1	5.79E-01	1	5.79E-01	1	Biguanide, Metformin
<i>HMGCR</i>	rs3846662	A	0.043	0.67	0.887	1	0.57	0.393	0.47	0.777	0.776	3.80E-15	4.02E-16	4.02E-16	2.92E-14	3.80E-15	Paravas-tatin

Table 2. Frequency comparison of actionable pharmacogene variants shared between Tunisian and Italian populations. The table shows the MAF comparison of highly selected actionable pharmacogene variants (rs622342, rs5215, rs7294, rs622342 and rs3846662) linked to ADR among TUN and ITA populations, obtained after Bonferroni correction. Variant with significant differences (p -values $< 0.05/4 \times 14$) are shown in boldface.

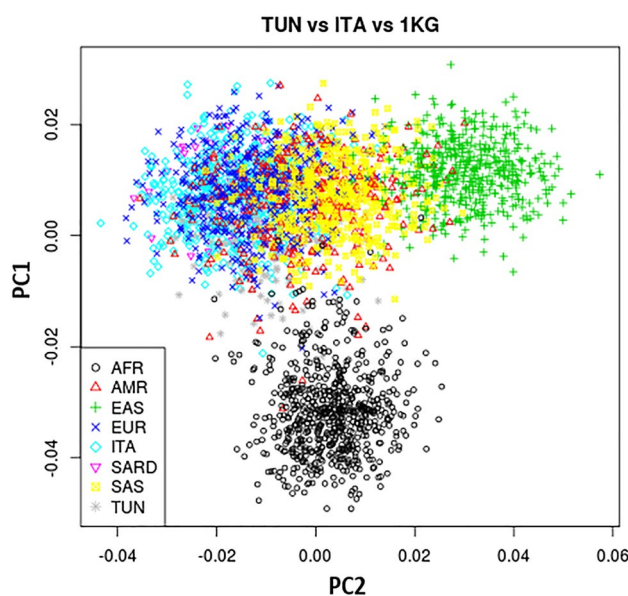


Figure 1. Principal Component Analysis Plot of the Tunisian “TUN” subpopulation, Italian (ITA) population and the 1000 genome project populations originating from Sub-Saharan Africa (AFR), America (AMR), East Asian (EAS), EUR (Europe), Sardinia (SARD), South Asian (SAS).

The comparison of the different studied populations clustered in four continents (African, European, American and Asian) shows a high level of similarity between Tunisian population and the Italian population. On the other hand, the African was distinguished from the European and the North American continents, projected on the opposed factorial axis (Fig. 3). The PCA regrouping Tunisian and Italian population sub-groups reveal the genetic similarity between these two populations (Fig. 4).

Next, ADMIXTURE analysis was employed to infer ancestral population groups. The $K = 3$ was identified as the most likely number of genetic clusters and corresponding to the three ancestral continental population groups; European (S_ITA, C_ITA, SARD, TSI, IBS, CEU, GBR, FIN), Asian (CHB, CHS, CDX) and Sub-Saharan African (LWK, YRI, MSL, GWD, TN_TC, TN_MC). The ADMIXTURE analysis graphic reveals the predominance of African and European components to this number of ancestor K ancestries (Fig. 5). The findings indicate that the Tunisian population, in terms of the studied pharmacogenes, were more similar to the European and particularly Mediterranean populations (average frequency of 45%). This result corroborated the patterns seen on PCA (Fig. 1). The evidence of F_{st} , PCA, and ADMIXTURE analyses shows that there is a high similarity among Mediterranean populations which are genetically divergent from the Sub-Saharan Africa populations. Then after the description of the general pattern of variation including all the SNPs we performed a F_{st} comparative analyses to identify the most divergent SNPs. We identified 26 variants with high levels of differentiation between Tunisian, Italian and other studied populations (Table 3). Among these variants, there are four SNPs rs622342, rs3846662, rs7294, rs5215 located respectively in *SLC22A1*, *HMGCR*, *VKORC1*, *KCNJ11*, and reported as associated with ethnic variability in drug response. The other 22 variants are not yet described to be associated with any pharmacogenetic variability (Table 3). Moreover, the F_{st} comparison analyses show that the variant rs5215 located on *KCNJ11* is the most differentiated variant between the Tunisian and Italian populations. This variant is

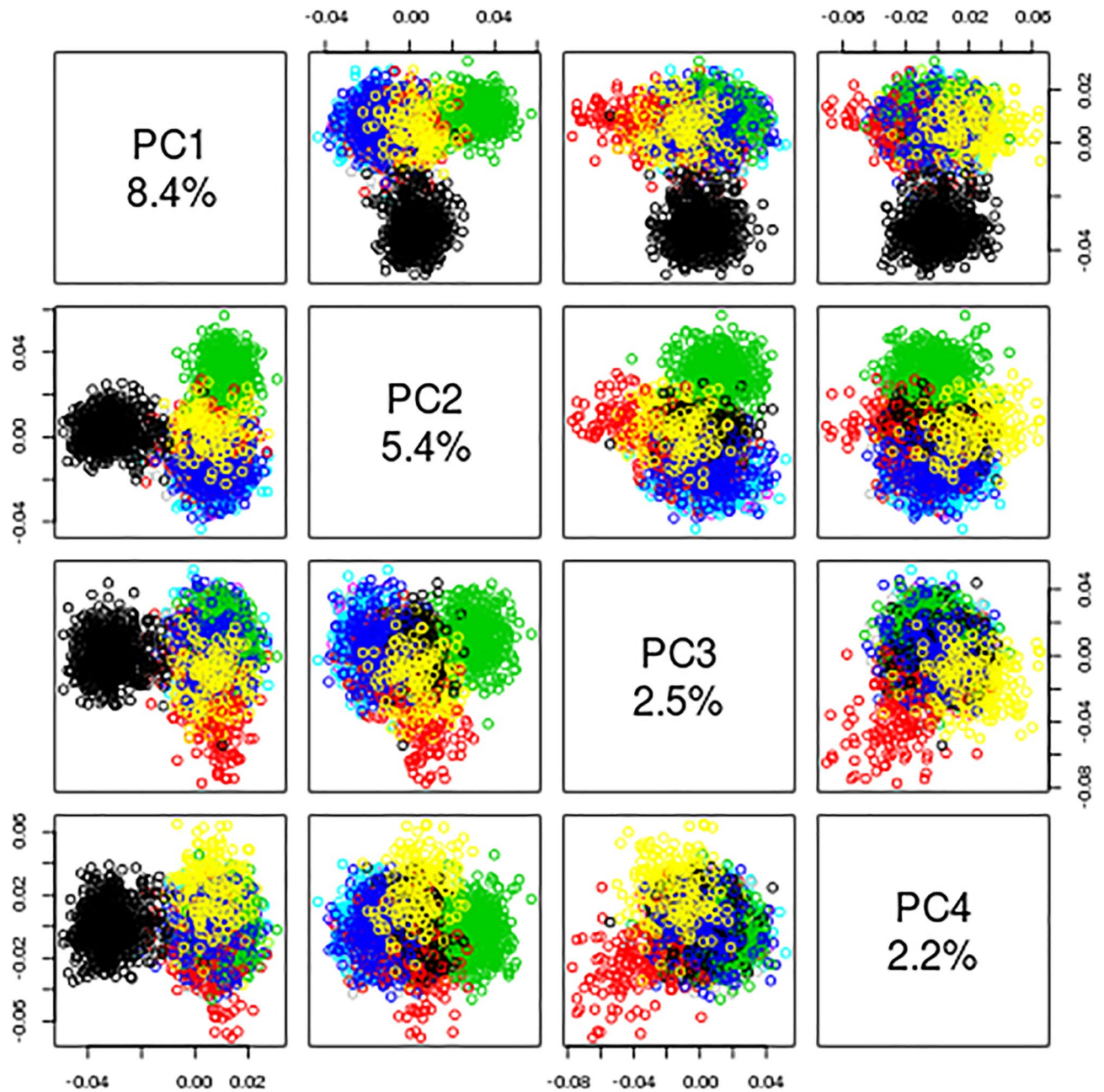


Figure 2. Scatter Plot analysis of the PCA representative of the Tunisian and Italian populations compared to the populations of the 1000 Genomes project. The result of the Scatter Plot shows that PCA (Fig. 1) has the best estimated orientation of 8.4%. This is the highest percentage between the four prisms showing all possibilities of the PCA orientations.

associated with the resistance to oral antidiabetics (Metformin, Glucamide, Sulfamides.) (Supplementary Fig. 1). The variant rs7294 located on *VKORC1* associated with warfarin dose requirement, shows a large differentiation between Tunisian and British populations from England and Scotland (GBR). However, this variant reveals a similarity among Tunisian, Italian and all the other studied populations (Table 3) (Supplementary Fig. 1). The variant rs622342 located on *SLC22A1* associated with the efficacy of metformin shows a significant differentiation between Tunisian and the Iberian (IBS) populations and notable similarity with Italian population (Table 3) (Supplementary Fig. 1). The rs3846662 variant located on *HMGCR*, associated with the reduced Pravastatin efficacy and smaller LDL cholesterol reduction, displays substantial genetic differences between the studied populations. Furthermore, these results were confirmed by genotype and allelic distributions of the four high selected VIP variants ($p < 0.05$), using the worldwide representative map (Figs. 6, 7). The genotype frequencies comparison shows that GG and AG genotypes of the variant rs3846662 located in *HMGCR* gene, associated with the reduced effect of pravastatin, represent high frequencies in the Sub-Saharan Africa population compared to other studied populations (Fig. 7). The frequency of GG, AG genotypes of rs3846662 variant in the Tunisian population are 24.81% vs 46.62%, these frequencies are very similar to those observed in the Italian populations which are

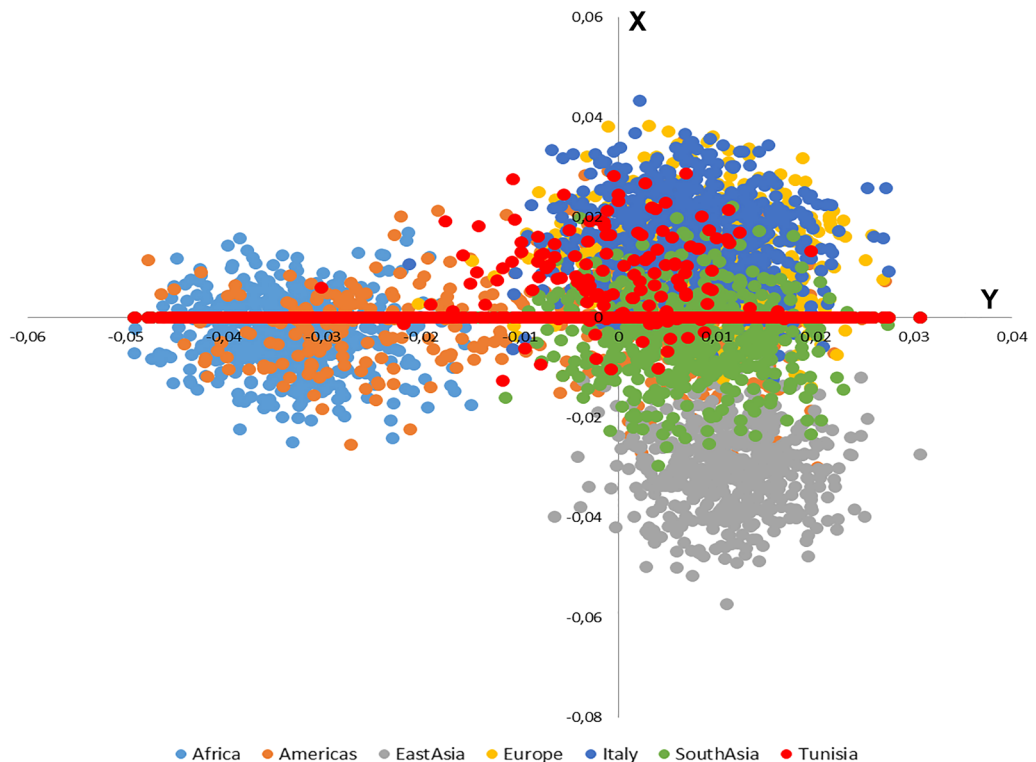


Figure 3. Comparison between the Tunisian and Italian populations with all the other European populations (combined in a single entity) and Sub-Saharan Africa populations of the 1000 Genomes project.

23.08% vs 45.66 in the Center, 19.85% vs 48.85% in the North, 21.79% vs 50.64% in the South and 12.77 vs 40.43 in Sardinia (Supplementary Table 4). These frequencies are moderately low compared to those of Sub-Saharan Africa which has a GG frequency equal to 95% (Supplementary Table 4) (Fig. 7).

The distribution of the genotypes for the variant rs7294 (C > T) located in *VKORC1* gene is as follows. The genotypes TT and CT associated with a high dose of warfarin are for the Tunisian population of 64.84% and 26.56%. These frequencies align closely with those observed in the Italian population with values of 44.03% vs 46.52% in the Center, 40.46% vs 45.8% in the North and 52.56% vs 35.90% in the South (Supplementary Table 4). The frequency of the recessive homozygotes TT is 8% in the Tunisian studied population, resembling that in the Italian population (Center 9.45%, North 13.7%, South 11.5%) (Supplementary Table 4). In the Tunisian population, the frequency of the rs7294-T allele is 21.88%. Similarly, in the Italian population, the frequency varies across regions, with values of 36.64% in the North, 32.71% in the Center, and 49% in the South (see Supplementary Table 3) (see Fig. 6). Comparatively, these frequencies are notably lower in both the Tunisian and Italian populations when contrasted with Sub-Saharan African populations, where the rs7294-T allele frequencies are 51.39% in YRI, 42.93% in LWK, and 45.13% in GWD (see Supplementary Table 3) (see Fig. 6).

Discussion

The distribution of genetic polymorphisms located in pharmacogenes involved in drug metabolism and transport are major determinants of treatment efficacy and adverse reactions. Biogeographical ancestry is an important factor leading to large diversity in drug effectiveness or adverse reactions. In the present study, we showed that some Very Important Pharmacogenes ‘VIP’ variants which involved in weight loss drugs, lipid-lowering, anti-hypertensive and oral antidiabetic agents, exhibited a great genetic variation among the studied populations. This variability impacts directly on the delivery of individualized medicine.

Genetic landscape of the selected pharmacogenes

The PCA, pairwise *F_{st}* values and ADMIXTURE analyses reveal a high similarity among Mediterranean populations concerning the studied pharmacogenes. In addition, these analyses exhibit a great genetic divergence between Mediterranean populations and Sub-Saharan Africa and Asian populations.

The different combinations of PCA show that Tunisian population is genetically related to the European and especially to the Italian population. Furthermore, the shift suggests slight genetic differences but the lack of distinct clusters suggests common ancestral links (Fig. 4). These observations were also reported for mitochondrial DNA, Y chromosomal and autosomal markers and interpreted as influences from different migration events^{30,31,33–36}. Obviously, differences in admixture history exert an important impact in the allelic and genotypic distribution of variants at the population level³⁶. In the present study, four polymorphisms characterized as clinically relevant VIP variants were selected, based on previous pharmacogenomics research. These four

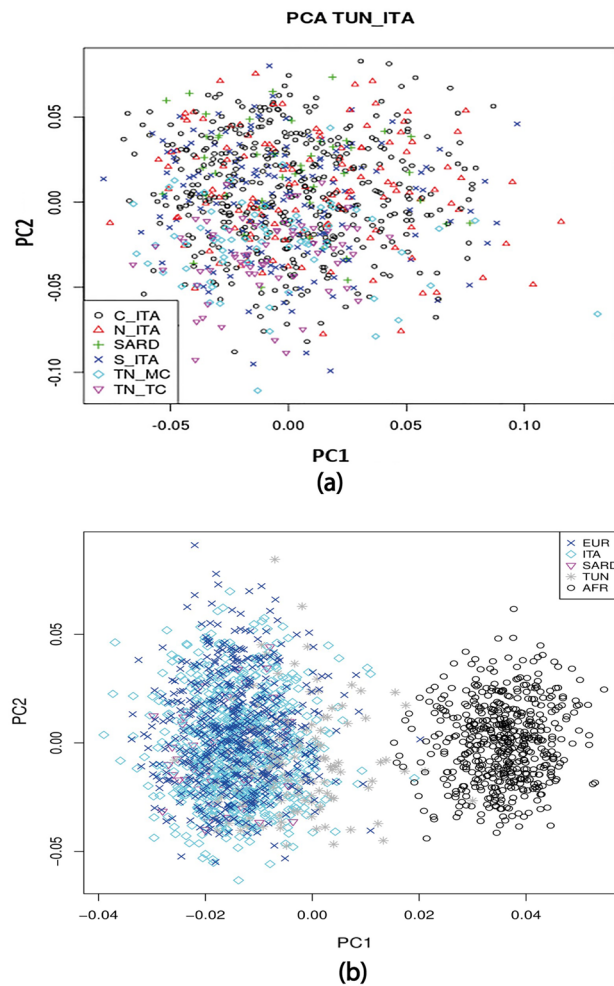


Figure 4. Principal Component Analysis plots of different population combinations. C_ITA: Italians from the Center, N_ITA: Italians from the North, Italian from Sardinia, S_ITA: Italians from the South, TN_MC: Tunisians originated from the coastal city of Monastir, TN_TC: Tunisians originated from the Capital Tunis, AFR: Sub-Saharan African populations.

polymorphisms (rs622342, rs3846662, rs7294 and rs5215) have been studied with a low level of clinical evidence in PharmGKB. As explained in this database, the low score may be based on a single study annotated in PharmGKB, or there may be several studies that failed to replicate the association. Therefore, further association studies should be performed to investigate these variants and to determine the population specificity in order to personalize treatments. Investigating the prevalence of pharmacogene risk variants in target groups, together with actual drug consumption data, will allow better therapeutic decisions³⁷.

Genetic variability of the selected pharmacogenes

Our genetic distribution study showed that TT and CT genotypes of rs7294 (C > T) located in the *VKORC1* gene, associated with a high dose of warfarin requirement, were more common in the GBR population than in other populations. This finding implies that the British population has poor warfarin metabolism, which raises the risk of a bleeding event. Our result was consistent with the study of Jones et al., 2005, who discovered that poor clinical outcomes were associated with suboptimal anticoagulation in British patients. New measures must be implemented to improve maintenance anticoagulation in British patients with nonvalvular atrial fibrillation³⁸.

Our result showed that the genotype frequency of GG and GA of the rs3846662 (A > G) variant located on *HMGCR* associated with the reduced effect of pravastatin was higher in Sub-Saharan Africa populations than in other studied populations. This data suggest that Sub-Saharan Africans were resistant to this treatment. This outcome was in agreement with those of Medina et al. 2008 which highlighted that GG and GA of the variant rs3846662 were associated with decreased induction of full-length transcripts of *HMGCR* and increased expression of the spliced *HMGCRv_1* transcript as compared with AA genotype. The author of this study explained that the variant rs3846662 genotype by itself was not significantly associated with statin response³⁹. The *HMGCRv_1* was suggested as a marker for statin therapy efficacy⁴⁰. However, it is important to note that rs3846662 is not the only determined of statin response⁴¹.

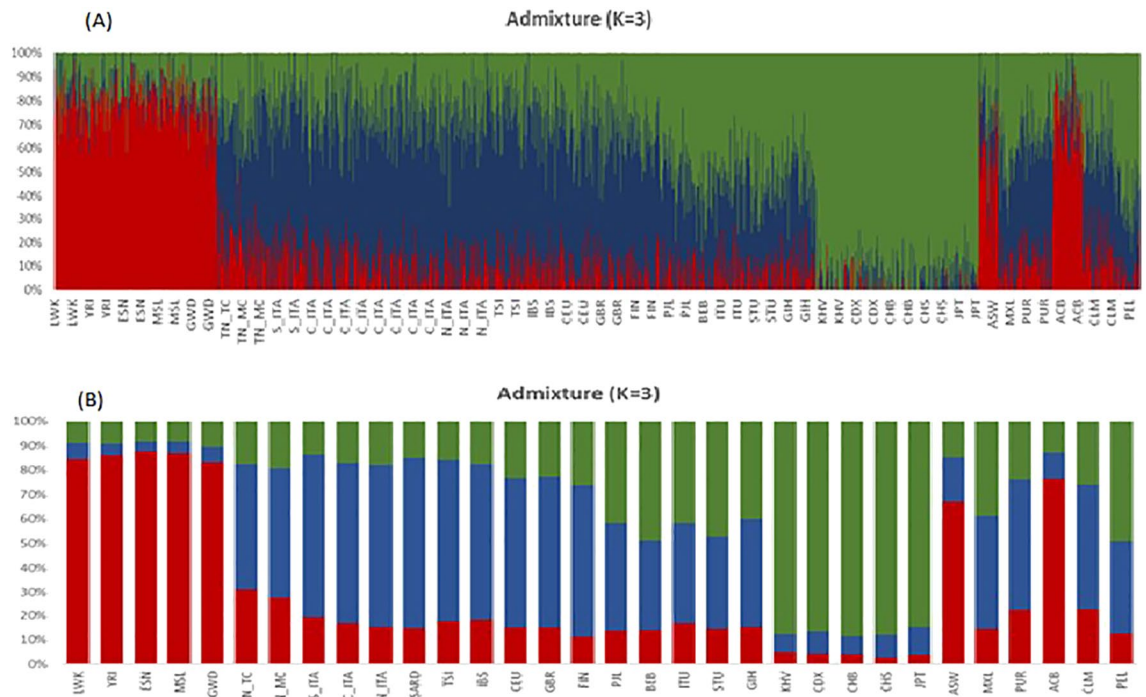


Figure 5. ADMIXTURE individual bar plots ($K=3$). (A) Admixture plot_27 populations, 116 variants studied—displaying results for runs with highest likelihood out of 27 runs in each cluster $K3$ to 10. Black vertical lines identify the population boundaries. The height extent of each color within an individual's color bar corresponds to the estimated membership of the individual in one of the clusters; each cluster is assigned a separate color. The bars with multiple colors can be interpreted as genetic admixture or as relative probabilities of belonging to the different clusters. The graphic of ADMIXTURE analysis demonstrates three components with the predominance of African, European components at the number of ancestor $K=3$. (B) Individual Admixture Proportions: individual are represented by a single vertical line broken into three colored segments, with lengths proportional to each three inferred clusters; red indicates African ancestry proportion, blue indicates European ancestry proportion and green indicates East Asian proportion. The coordinate indicates the proportion unit.

The comparison of the genotype and minor allelic distribution revealed that the VIP variants in Tunisian population were genetically similar to Italian and other European populations and divergent from Sub-Saharan Africa and Asian populations. Nonetheless, our study has shown some exceptions. The rs5215 (C > T) variant located in *KCJN11* gene showed the highest significant differentiation between Tunisian and Italian populations (Table 3). This variant was associated with the resistance to oral antidiabetic drugs such as (glimepiride, gliclazide, gliquidone, glipizide, glibenclamide). The reported variant was in perfect high LD with the variants rs5219 in both Tunisian and Italian populations. Multiple studies have revealed the association of rs5219 to the resistance of sulfamide-inducing hypoglycemia, hypoglycemic coma, hypersensitivity, hepatotoxicity, drug-induced erythema multiforme and photodermatitis⁴². Type 2 diabetes (T2D) patients with the CC genotype of rs5219 treated with metformin and sulfonamides showed a decreased likelihood of treatment failure and a higher HbA1c concentration than patients with TT genotype. We can speculate that the C allele of rs5215 is more represented in the Italian population and is also associated with the resistance to the two main widely administered oral antidiabetics, metformin and sulfamide⁴³. This result is in agreement with the study of Sesti et al. which reported the failure of the combination of sulfonylurea with metformin, rather than the failure of sulfonylurea treatment itself. The study included 525 Italian patients and showed that carriers of the C allele of rs5219 had a significantly higher probability of secondary sulfonylurea failure. The patients were treated with a combination of sulfonylurea as the first-choice drug and metformin as an add-on drug⁴⁴. The defective C allele is more prevalent in Italian and European populations than in the Tunisian and Sub-Saharan Africa populations. We hypothesize that Europeans have a higher risk of treatment failure when combining sulfonylurea and metformin than African individuals.

Furthermore, rs622342 (A > C) located in *SLC22A1* gene is associated with inter-individual metformin response variability. Our results showed that the risk allele "C" of rs622342 (*SLC22A1*) associated with Metformin intolerance showed significant differences between Tunisian, European and African populations. This variant has been reported by previous study as potentially associated with glycemic response to metformin in European and African populations. However, no metformin intolerance has been shown. This evidence may provide novel insight into gene-oriented personalized medicine for diabetes⁴⁵.

The carriers of CC and the CA genotype of rs622342 were associated with poorer response to metformin, as measured by a smaller reduction in HbA1c levels, compared with AA patients^{46,47}. Our finding revealed that carriers of the CC and CA genotypes were more frequent in Italian and European populations than in Tunisian populations. The Tunisian population appeared to be a better responder to oral antidiabetics. This information

	CEU	FIN	GBR	IBS	TSI	N_ITA	C_ITA	S_ITA	SARD	GWD	MSL	LWK	ESN	YRI
rs11023996		*												
rs11196212			*											
rs7294			*											
rs11196224			**											
rs622342				*										
rs10832785				*										
rs11603988				*										
rs10889503					**	*				**	**	**	*	*
rs887241						*								
rs316013								*						
rs3777406									*					
rs7931930									*					
rs234852								*	*					
rs12654264						**	**	**	**					
rs2228100						**	**	**	**					
rs1116714	*	*	*											
rs10064799	*	*			*		*							
rs7757997	*				*		*	*						
rs5215	**	**		**	*	*	*	*						
rs11656096	**	**	**	**	**	*	*		*					
rs1555485	*	*	*	*	*									
rs4916014										**	**	*	*	**
rs11208591										*	*	**	**	*
rs6007919										*	*	*	**	*
rs4915675										*	*	*	*	**
rs3846662										*	*	*	*	*

Table 3. Description of VIP variants through the comparison of their F_{st} values. The statistical test of Chi2 was used to compare the F_{st} values of the variant: Highly selected variants with statistical significance differentiation ($p < 0.05$) are with *. Variants with high statistical significance differentiation ($p < 0.01$) are in **. * for $p < 0.05$ and ** for $p < 0.01$.

could be clinically relevant predicting the glucose-lowering effects of metformin in different ethnicities before the start of therapy.

The rs3846662 variant was associated with the variation in the production of the HMGCR isoform, correlated with reduced sensitivity to statins. As a result, it may play a significant role in the interindividual variations in LDL, apolipoprotein B, and triglyceride concentrations during statin therapy. More precisely, the GG and AG rs3846662 variant decreased induction of full-length transcripts of HMGCR and increased expression of the spliced HMGCRv_1 transcript as compared with AA genotype³⁹. Our analysis revealed a high frequency of the GG and AG genotypes of rs3846662 in Sub-Saharan African populations compared with the other studied populations. These differences may affect the effectiveness of simvastatin and decrease the risk of myopathy⁴⁸. Taking these pieces of evidence into consideration, we hypothesize that Mediterranean and European populations are good responders to simvastatin as the treatment of choice for LDL decreases. In contrast, Sub-Saharan Africans are simvastatin-resistant, which explains the increased rate of cardiovascular disease in this group. This finding is consistent with the CAP study in which African-Americans had a weaker statin LDL-C response than Caucasians. However, our results contradict the finding of Akinyemi Oni-Orisan et al. who found that participants in East Asian, Hispanic, and African populations responded to statin medication with larger percentage reductions in LDL-C than participants in European populations⁵⁰. This outcome might be the result of the synergic effect of variants linked to the statin response variability.

The fourth SNP is rs7294, located on the *VKORC1* gene and is an essential anticoagulant cofactor in vitamin K metabolism. The rs7294 TT and CT carriers required a higher warfarin dose. Our analysis showed that Tunisian, Italian, and other European (except Spanish) populations presented similar frequencies of these genotypes. However, Sub-Saharan Africans have a lower frequency of these genotypes. These could explain the differences in warfarin dose administration among populations. This result is well illustrated in (Fig. 7). Our results are consistent with the study by Yang et al., who showed that Sub-Saharan African populations require a low dose of warfarin to achieve the therapeutic effect compared with Caucasian and Asian populations⁵¹. We mention that in Tunisia, since 2016, warfarin is no longer adopted in the therapy of thromboembolic complications. The anticoagulant acenocoumarol has become a clinician's therapeutic choice for the treatment of hypertension complications. Indeed, patient carrying TT and CT for the variant rs7294 may require higher doses of acenocoumarol than the CC wild-type genotype⁵². Previous studies have shown that the "T" recessive allele is the defective allele, associated with resistance to acenocoumarol. On the other hand, the "C" allele is associated with the efficacy of

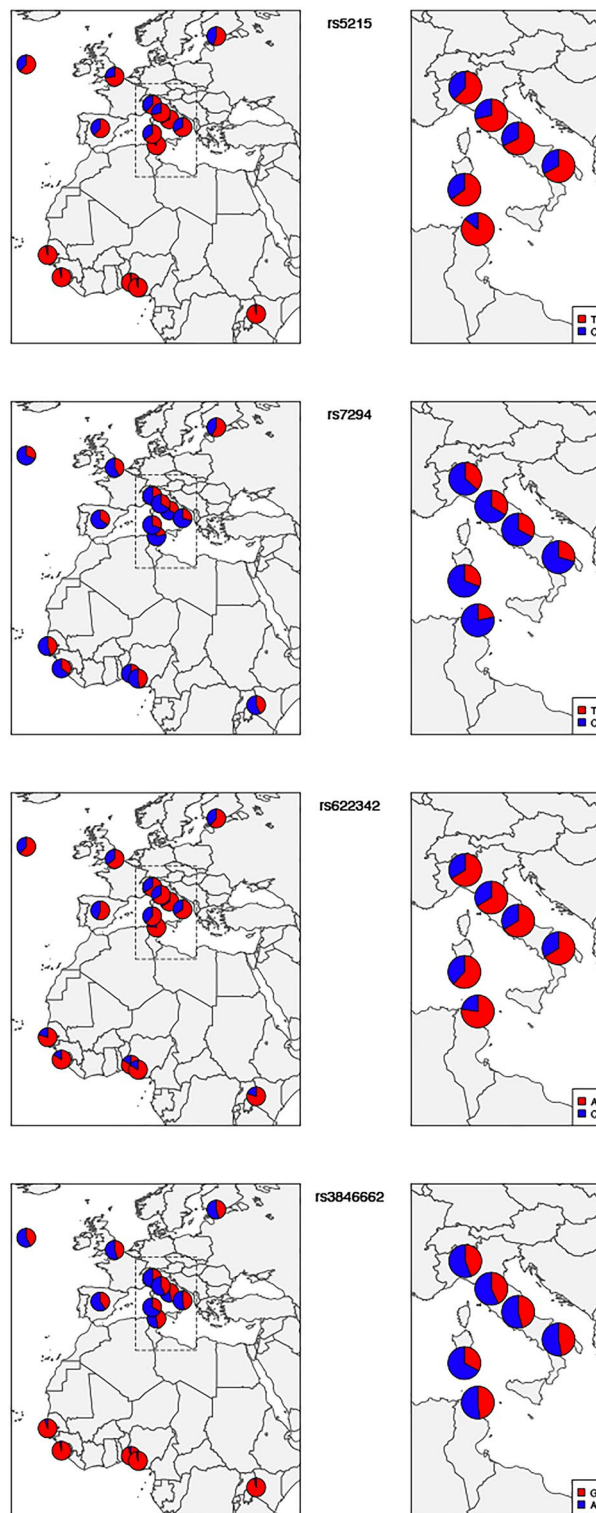


Figure 6. Distribution of allele frequencies of clinically relevant variants in Tunisian, Italian, European, African and South Asian populations. The graph shows an inter-population diversity of the frequency distribution of the dominant and recessive alleles of the variants rs622342, rs3846662, rs7294, rs5215 located respectively in the “pharmacogenes” *SLC22A1*, *HMGCR*, *VKORC1*, *KCNJ11*. Red and blue colors indicate allele frequencies. We generated geographical map using R software [v.3.2.4] (<https://www.rproject.org/>), by using the packages rworldmap (URL: <https://cran.r-project.org/web/packages/rworldmap/>) and mapplots (URL: <https://cran.r-project.org/web/packages/mapplots/index.html>).

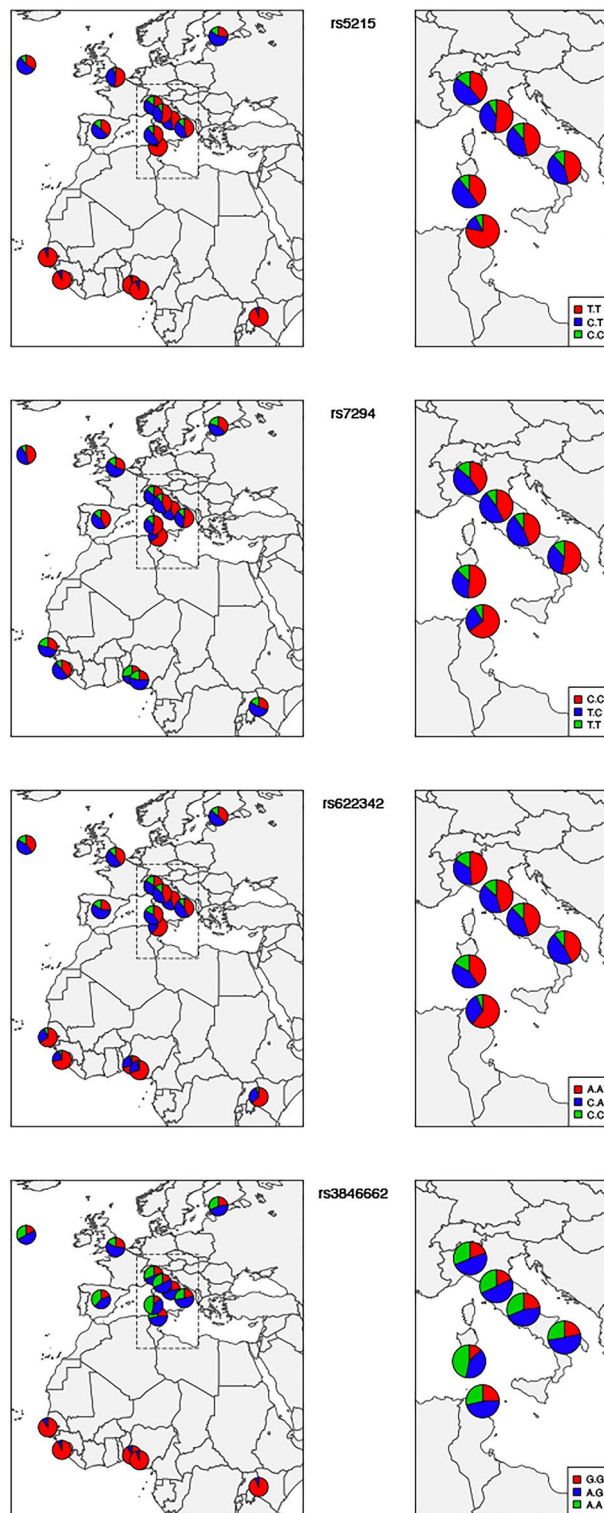


Figure 7. Distribution of genotypic frequencies of clinically relevant variants in Tunisian, Italian, European, African and South Asian populations. The graph shows an inter-population diversity of the genotype frequency distribution of the variants rs622342, rs3846662, rs7294, rs5215 located respectively in the “pharmacogenes” *SLC22A1*, *HMGCR*, *VKORC1*, *KCNJ11*. The red color indicates the frequency of wild homozygous, the blue color; the frequency of heterozygous and the color green; the frequency of recessive homozygous. We generated geographical map using R software [v.3.2.4] (<https://www.rproject.org/>), by using the packages rworldmap (URL: <https://cran.r-project.org/web/packages/rworldmap/>) and mapplots (URL: <https://cran.r-project.org/web/packages/mapplots/index.html>).

acenocoumarol. Our results showed a low prevalence of the risk "T" allele of the variant rs7294 in the Tunisian and Italian populations compared to Sub-Saharan Africa populations. Consequently, Tunisian patients carrying the ancestral allele "C" in double or single copy, may require a low dose of acenocoumarol to achieve the curative effect³². Based on these findings, we can hypothesize that the ancestral "C" allele is a protective allele for the Tunisian patients against hemorrhagic complications caused by an overdose of acenocoumarol. Our results oppose those of Ajmi et al. who declared the protective effect of rs7294 "T" allele against bleeding caused by acenocoumarol treatment⁵³. This study was conducted on 246 Tunisian patients who originated from the Sahel region. This discrepancy could be due to Tunisian high genetic heterogeneity^{30,54,55}. Thus, the T allele follows a south/north gradient from Africa to Europe, via Tunisia. This result highlights the need for genotyping Tunisian, Italian and other European patients for the variant rs7294 located on *VKORC1* gene in order to optimize the dose of acenocoumarol in to avoid the bleeding events in this susceptible populations.

The correlation of the genotype with the response to the drug and the knowledge of the frequencies of the risk allele associated with adverse drug effects in the different populations, improve the therapeutic results of the drugs. This has an important impact in the implementation of personalized medicine⁵⁶. Our results show that the Tunisian population has a complex demographic history of migrations and gene flow involving the pharmacogenes, within Africa, Europe, and the Middle East. This genetic heterogeneity could pave the way for these populations to benefit from precision medicine. These studies provide several opportunities for detecting the association of host genotypes and ancestry in order to formulate specific country-level guidelines on drug reactions which could be useful and cost effective in clinical practice.

The study of the genetic distribution of variants in genes known to influence the inter-population variability in drug response could be greatly advantageous and helpful to implement a personalized medicine. It limits the additional costs of hospitalization of patients following side effects of drugs. Furthermore, assessing ADRs is critical for determining risk factors and maximizing the benefits of drug therapy. More information about prescribed drugs and their side effects will aid in reducing ADRs and ensuring patient safety.

Populations and methods

Selection of the pharmacogenes

A subset of very important pharmacogenes (VIP) involved in the ADR of many drugs comprising weight loss drugs, lipid-lowering, antihypertensive and oral antidiabetics agents, were extracted from the PharmGKB database (<http://www.pharmgkb.org>), which provides an overview of significant genes involved in drug metabolism or response. VIP genes were chosen after a thorough review of several sources, including the FDA (Food and Drug Administration) biomarker list, FDA-approved drug labels with pharmacogenetic information, and Clinical Pharmacogenetic Implementation Consortium (CPIC) nominations (<https://cpicpgx.org>). A gene was also considered as VIP, if it was associated with a large number of variant annotations and a high level of clinical annotations. In addition, an extensive bibliographic search was used to select VIP variant comprising genetic polymorphisms located on VIP and having clinical evidence level association with ADR according to the PharmGKB database (Table 1).

Genotyping data and quality control

Genotyping data was generated in a previously published Affymetrix Chip 6.0 genotyping array⁵⁷ that investigated 135 healthy Tunisian individuals (participants were from the Capital Tunis and the coastal city of Monastir). The present study was designed and performed in accordance with relevant guidelines and regulations and according to ethical principles for medical research involving human subjects stated by the WMA Declaration of Helsinki. It was approved by the Ethics Committee of the Institut Pasteur (Tunis, Tunisia-Registration numbers IRB00005445, FWA00010074, and PV09/06, IRB# 0000000044). All participants provided written informed consent. Similarly, we used Illumina Chip genotyping data from 737 healthy Italian individuals originating from Northern, Central, Southern Italy and Sardinia. Genomics data was obtained from all participants within the framework of the study by Boattini et al.⁵⁸ (protocol n. 85/2009/U/Tess approved by the Bologna S.Orsola-Malpighi University Hospital ethics committee). In addition, we used public genotyping data to perform comparisons. Indeed, the 1000 Genomes project populations were investigated including 2504 individuals from East Asia; [Han Chinese in Beijing, China (CHB), the Chinese population of metropolitan Denver, Colorado, USA (CHD), the Chinese Dai in Xishuangbanna, China (CDX) and Japanese in Tokyo, Japan; (JPT)], South Asia; [Gujarati INS from Houston, Texas (GIH), Punjabi from Lahore, Pakistan (PJL)], Bengali from Bangladesh (BEB), SL Tamils from the UK (STU), Indian Telugu from the UK (ITU)], Africa; [Gambian in Western Divisions in the Gambia (GWD), African Caribbean (ACB), African Ancestry in SW USA (ASW), Esan in Nigeria (ESN), Mende in Sierra Leone [MSL], Luhya in Webuye, Kenya (LWK)] and Europe; [Utah residents with Northern and Western European ancestry (CEU)], Tuscany (TSI) in Italy, Finnish in Finland (FIN), British in England and Scotland (GBR), Iberian Population in Spain (IBS)] (Supplementary Table 1).

Data were extracted from 1000 Genomes phase 3 release (<http://www.1000genomes.org>).

From the genotyping data, we extracted all variants located in the chromosomal region of the selected pharmacogenes. The genotyping data of the 135 Tunisians was merged with the genotyping data of the 737 healthy Italians, to identify the shared common variants between the two populations, using the software PLINK v2.

Variants were excluded if they were deviating from the Hardy–Weinberg equilibrium (HWE) (p -value $< 10^{-2}$) and minor allele frequency (MAF) $< 5 \times 10^{-2}$. Individuals with missing genotype and genotyping quality less than 95%, were not included in the study.

Statistical analysis

To assess the variance in pharmacogenes variants across different populations, we calculated Wright's fixation index (F_{st}). The pairwise comparison of F_{st} values among Tunisian, Italian, and 1000 Genome project populations was carried out using the *Hierfstat* package in R. The χ^2 statistical analysis test was used to compare the prevalence of risk alleles between the Tunisian, Italian and worldwide studied populations. The Bonferroni adjustment was applied to the significance level set at the 5% p-value threshold by the number of loci studied and the number of populations.

Principal component analysis

The merge of genotyping data of all investigated populations was thinned with the PLINK software to perform the data pruning. We excluded variants in strong LD within a sliding window of 50 variants advanced by 10 SNPs at the same time, to infer cryptic population structure from genomic data. Cryptic population structure defines a population structure that is difficult to detect using visible characters but may be significant in genetic terms⁵⁹. It is important to shape the false matches due to the probabilistic assignment of population. The pruned genotypic data was used to perform Principal Component Analysis "PCA", based on Identity-By-Descent measures using the *SNPrelate* R package⁶⁰. The *Scatter Plot* analysis was designed to study the correlation between the important variants. Allele frequency variation and conventional metrics such as fixation index (F_{st}) were conducted to have a robust population differentiation parameter⁶¹. The fixation index represents the percentage of total genetic variation at a given locus that differs between populations. This index is influenced by the frequency of minor alleles (MAF) and the size of the population sample⁶¹. The Manhattan plot analysis was conducted to depict a pairwise comparison of F_{st} values to evaluate the differentiation of variant frequencies between studied populations.

Genetic structure

We estimated the admixture proportions by applying the unsupervised clustering algorithm implemented in the software ADMIXTURE on the pruned dataset⁶². We performed a series of admixture runs from $K=2$ through 10 and we used the cross-validation (CV) error to identify the best predictive model. Next, we examined the variants with the highest level of differentiation between Tunisian, Italian and the other populations and drew the map representing the genetic distribution of the selected variants with R script.

Conclusion

Finally, the present findings represent the first set of shared pharmacogenetic data for the Tunisian and Italian populations. These findings provide a valuable basis for further functional pharmacogenetic research, which is of great use in patient treatment in these populations. In addition, this study highlights the importance of the Tunisian-Italian clinical expertise exchange in disease management and personalized medicine implementation in the Mediterranean region. Our study revealed several shared VIP variants implicated in ADR between Tunisian and Italian populations. Thus, the identification of novel associated pharmacogenes in one population has a high probability of being replicated in other populations. The exchange of genetic and clinical expertise between these populations will greatly help to decrease treatment failure and minimize medication and hospitalization costs. Correlation between genotype and allelic frequencies of risk variants and their associated ADR would therefore improve drug outcomes and have a significant impact on the adoption of personalized medicine in people around the world. The establishment of a data bank or a registry including the pharmacogenetic data of specific populations, available to clinicians and the scientific community, is very useful. These registers allow the identification of populations which must be prudent concerning the consumption of certain drugs.

Data availability

The data generated and analyzed during the current study are available in the European Variation Archive (EVA) repository at EMBL-EBI under accession number PRJEB61743 (<https://www.ebi.ac.uk/eva/?evastudy=PRJEB61743>).

Received: 1 March 2023; Accepted: 21 February 2024

Published online: 10 March 2024

References

- Chan, S. L., Jin, S., Loh, M. & Brunham, L. R. Progress in understanding the genomic basis for adverse drug reactions: A comprehensive review and focus on the role of ethnicity. *Pharmacogenomics* **16**, 1161–1178. <https://doi.org/10.2217/pgs.15.54> (2015).
- Montastruc, J. L. *et al.* Fatal adverse drug reactions: A worldwide perspective in the World Health Organization pharmacovigilance database. *Br. J. Clin. Pharmacol.* **87**, 4334–4340 (2021).
- Sultana, J., Cutroneo, P. & Trifirò, G. Clinical and economic burden of adverse drug reactions. *J. Pharmacol. Pharmacother.* **4**, S73–S77. <https://doi.org/10.4103/0976-500X.120957> (2013).
- Coleman, J. J. & Pontefract, S. K. Adverse drug reactions. *Clin. Med.* **16**, 481 (2016).
- Rawlins, M., Thompson, J. & Davies, D. *Textbook of Adverse Drug Reactions* (Oxford University Press, 1977).
- Lakiotaki, K., Kanterakis, A., Kartsaki, E., Katsila, T. & Patrinos, G. P. Exploring public genomics data for population pharmacogenomics. *PLoS ONE* **12**, e0182138. <https://doi.org/10.1371/journal.pone.0182138> (2017).
- Giri, A. K. *et al.* Pharmacogenetic landscape of clopidogrel in north Indians suggest distinct interpopulation differences in allele frequencies. *Pharmacogenomics* **15**, 643–653. <https://doi.org/10.2217/pgs.13.241> (2014).
- Umamaheswaran, G., Kumar, D. K. & Adithan, C. Distribution of genetic polymorphisms of genes encoding drug metabolizing enzymes & drug transporters - a review with Indian perspective. *Indian J. Med. Res.* **139**, 27–65 (2014).
- Wood, A. J. & Zhou, H. H. Ethnic differences in drug disposition and responsiveness. *Clin. Pharmacokinet.* **20**, 350–373. <https://doi.org/10.2165/00003088-199120050-00002> (1991).

10. Cheng, Y. *et al.* Analysis of very important pharmacogenomics variants in the Chinese Lahu population. *Pharmacogenom. Pers. Med.* **14**, 1275–1289 (2021).
11. Badary, O. A. Pharmacogenomics and COVID-19: Clinical implications of human genome interactions with repurposed drugs. *Pharmacogenom. J.* **21**, 275–284 (2021).
12. Wang, L. *et al.* Genetic polymorphisms of pharmacogenomic VIP variants in the Uyur population from northwestern China. *BMC Genet.* **16**, 66. <https://doi.org/10.1186/s12863-015-0232-x> (2015).
13. He, Y. Y. *et al.* Association of KCNJ11 and ABCC8 genetic polymorphisms with response to repaglinide in Chinese diabetic patients 1. *Acta Pharmacol. Sin.* **29**, 983–989 (2008).
14. Weinshilboum, R. Inheritance and drug response. *N. Engl. J. Med.* **348**, 529–537. <https://doi.org/10.1056/NEJMra020021> (2003).
15. Dalleau, K. *et al.* Learning from biomedical linked data to suggest valid pharmacogenes. *J. Biomed. Semant.* **8**, 16. <https://doi.org/10.1186/s13326-017-0125-1> (2017).
16. Bank, P. C., Swen, J. J. & Guchelaar, H.-J. Implementation of pharmacogenomics in everyday clinical settings. *Adv. Pharmacol.* **83**, 219–246 (2018).
17. Zhou, Z.-W. *et al.* Clinical association between pharmacogenomics and adverse drug reactions. *Drugs* **75**, 589–631 (2015).
18. Cacabelos, R., Cacabelos, N. & Carril, J. C. The role of pharmacogenomics in adverse drug reactions. *Expert Rev. Clin. Pharmacol.* **12**, 407–442 (2019).
19. Lauschke, V. M. & Ingelman-Sundberg, M. Prediction of drug response and adverse drug reactions: From twin studies to next generation sequencing. *Eur. J. Pharm. Sci.* **130**, 65–77 (2019).
20. Taylor, C. *et al.* A review of the important role of CYP2D6 in pharmacogenomics. *Genes* **11**, 1295 (2020).
21. Zhang, F. & Finkelstein, J. Inconsistency in race and ethnic classification in pharmacogenetics studies and its potential clinical implications. *Pharmacogenom. Pers. Med.* **12**, 107–123 (2019).
22. Ingelman-Sundberg, M. Pharmacogenetics: An opportunity for a safer and more efficient pharmacotherapy. *J. Intern. Med.* **250**, 186–200 (2001).
23. Ramamoorthy, A., Pacanowski, M., Bull, J. & Zhang, L. Racial/ethnic differences in drug disposition and response: Review of recently approved drugs. *Clin. Pharmacol. Ther.* **97**, 263–273 (2015).
24. Miteva-Marcheva, N. N., Ivanov, H. Y., Dimitrov, D. K. & Stoyanova, V. K. Application of pharmacogenetics in oncology. *Biomark. Res.* **8**, 1–10 (2020).
25. Bonifaz-Pena, V. *et al.* Exploring the distribution of genetic markers of pharmacogenomics relevance in Brazilian and Mexican populations. *PLoS ONE* **9**, e112640. <https://doi.org/10.1371/journal.pone.0112640> (2014).
26. Chan, S. L. *et al.* Genetic diversity of variants involved in drug response and metabolism in Sri Lankan populations: Implications for clinical implementation of pharmacogenomics. *Pharmacogenet. Genom.* **26**, 28–39. <https://doi.org/10.1097/fpc.0000000000000182> (2016).
27. Ingelman-Sundberg, M., Sim, S. C., Gomez, A. & Rodriguez-Antona, C. Influence of cytochrome P450 polymorphisms on drug therapies: Pharmacogenetic, pharmacoeconomic and clinical aspects. *Pharmacol. Ther.* **116**, 496–526. <https://doi.org/10.1016/j.pharmthera.2007.09.004> (2007).
28. Bonifaz-Pena, V., Struchiner, C., Roela, R., Furuya-Mazzotti, T. & Chammas, R. Exploring the distribution of genetic markers of pharmacogenomics relevance in Brazilian and Mexican populations (vol 9, e112640, 2014). *PLoS ONE* **10** (2015).
29. Frigi, S. *et al.* Mitochondrial DNA analysis of Tunisians reveals a mosaic genetic structure with recent population expansion. *Homo* **68**, 298–315. <https://doi.org/10.1016/j.jchb.2017.05.003> (2017).
30. Kefi, R. *et al.* Phylogeny and genetic structure of Tunisians and their position within Mediterranean populations. *Mitochondr. DNA* **26**, 593–604. <https://doi.org/10.3109/19401736.2013.879649> (2015).
31. Fadhlaoui-Zid, K. *et al.* Genetic structure of Tunisian ethnic groups revealed by paternal lineages. *Am. J. Phys. Anthropol.* **146**, 271–280. <https://doi.org/10.1002/ajpa.21581> (2011).
32. Jmel, H. *et al.* Pharmacogenetic landscape of Metabolic Syndrome components drug response in Tunisia and comparison with worldwide populations. *PLoS ONE* **13**, e0194842. <https://doi.org/10.1371/journal.pone.0194842> (2018).
33. Henn, B. M. *et al.* Genomic ancestry of North Africans supports back-to-Africa migrations. *PLoS Genet.* **8**, e1002397. <https://doi.org/10.1371/journal.pgen.1002397> (2012).
34. Cherni, L. *et al.* Genetic variation in Tunisia in the context of human diversity worldwide. *Am. J. Phys. Anthropol.* **161**, 62–71. <https://doi.org/10.1002/ajpa.23008> (2016).
35. Ben Halima, T. *et al.* Palladium-catalyzed Suzuki–Miyaura coupling of aryl esters. *J. Am. Chem. Soc.* **139**, 1311–1318 (2017).
36. Bonifaz-Pena, V., Struchiner, C., Roela, R., Furuya-Mazzotti, T. & Chammas, R. Exploring the distribution of genetic markers of pharmacogenomics relevance in Brazilian and Mexican populations. *PLoS ONE* **9**, e112640 (2015).
37. Oslin, D. W. *et al.* Effect of pharmacogenomic testing for drug-gene interactions on medication selection and remission of symptoms in major depressive disorder: The PRIME care randomized clinical trial. *Jama* **328**, 151–161. <https://doi.org/10.1001/jama.2022.9805> (2022).
38. Jones, M. *et al.* Evaluation of the pattern of treatment, level of anticoagulation control, and outcome of treatment with warfarin in patients with non-valvular atrial fibrillation: A record linkage study in a large British population. *Heart* **91**, 472–477 (2005).
39. Medina, M. W., Gao, F., Ruan, W., Rotter, J. I. & Krauss, R. M. Alternative splicing of 3-hydroxy-3-methylglutaryl coenzyme A reductase is associated with plasma low-density lipoprotein cholesterol response to simvastatin. *Circulation* **118**, 355–362. <https://doi.org/10.1161/circulationaha.108.773267> (2008).
40. Medina, M. W. & Krauss, R. M. The role of HMGCR alternative splicing in statin efficacy. *Trends Cardiovasc. Med.* **19**, 173–177. <https://doi.org/10.1016/j.tcm.2009.10.003> (2009).
41. Medina, M. W., Sangkuhl, K., Klein, T. E. & Altman, R. B. PharmGKB: Very important pharmacogene–HMGCR. *Pharmacogenet. Genom.* **21**, 98–101. <https://doi.org/10.1097/FPC.0b013e328336c81b> (2011).
42. Daniels, M. *et al.* Pharmacogenomics in type 2 diabetes: Oral antidiabetic drugs. *Pharmacogenom. J.* **16**, 399–410 (2016).
43. Ordelheide, A.-M., Hrabě de Angelis, M., Häring, H.-U. & Staiger, H. Pharmacogenetics of oral antidiabetic therapy. *Pharmacogenomics* **19**, 577–587 (2018).
44. Sesti, G. *et al.* The E23K variant of KCNJ11 encoding the pancreatic β -cell adenosine 5'-triphosphate-sensitive potassium channel subunit Kir6. 2 is associated with an increased risk of secondary failure to sulfonylurea in patients with type 2 diabetes. *J. Clin. Endocrinol. Metab.* **91**, 2334–2339 (2006).
45. Peng, A. *et al.* Association between organic cation transporter genetic polymorphisms and metformin response and intolerance in T2DM individuals: A systematic review and meta-analysis. *Front. Public Health* **11**, 1183879. <https://doi.org/10.3389/fpubh.2023.1183879> (2023).
46. Tarasova, L. *et al.* Association of genetic variation in the organic cation transporters OCT1, OCT2 and multidrug and toxin extrusion 1 transporter protein genes with the gastrointestinal side effects and lower BMI in metformin-treated type 2 diabetes patients. *Pharmacogenet. Genom.* **22**, 659–666 (2012).
47. Dawed, A. Y. *et al.* Variation in the plasma membrane monoamine transporter (PMAT)(encoded by SLC29A4) and organic cation transporter 1 (OCT1)(encoded by SLC22A1) and gastrointestinal intolerance to metformin in type 2 diabetes: An IMI DIRECT study. *Diabetes Care* **42**, 1027–1033 (2019).
48. Medina, M. W. & Krauss, R. M. The role of HMGCR alternative splicing in statin efficacy. *Trends Cardiovasc. Med.* **19**, 173–177 (2009).

49. Krauss, R. M. *et al.* Variation in the 3-hydroxyl-3-methylglutaryl coenzyme a reductase gene is associated with racial differences in low-density lipoprotein cholesterol response to simvastatin treatment. *Circulation* **117**, 1537–1544. <https://doi.org/10.1161/circulationaha.107.708388> (2008).
50. Oni-Orisan, A. *et al.* Characterization of statin low-density lipoprotein cholesterol dose-response using electronic health records in a large population-based cohort. *Circulation* **111**, e002043 (2018).
51. Yang, L., Ge, W., Yu, F. & Zhu, H. Impact of VKORC1 gene polymorphism on interindividual and interethnic warfarin dosage requirement: A systematic review and meta analysis. *Thromb. Res.* **125**, e159–e166 (2010).
52. Krishna Kumar, D. *et al.* An acenocoumarol dosing algorithm exploiting clinical and genetic factors in South Indian (Dravidian) population. *Eur. J. Clin. Pharmacol.* **71**, 173–181. <https://doi.org/10.1007/s00228-014-1791-x> (2015).
53. Ajmi, M. *et al.* Influence of genetic and non-genetic factors on acenocoumarol maintenance dose requirement in a Tunisian population. *Eur. J. Clin. Pharmacol.* **74**, 711–722. <https://doi.org/10.1007/s00228-018-2423-7> (2018).
54. Fadhlaoui-Zid, K. *et al.* Sousse: Extreme genetic heterogeneity in North Africa. *J. Hum. Genet.* **60**, 41–49. <https://doi.org/10.1038/jhg.2014.99> (2015).
55. Diatchenko, L. *et al.* Genetic basis for individual variations in pain perception and the development of a chronic pain condition. *Hum. Mol. Genet.* **14**, 135–143. <https://doi.org/10.1093/hmg/ddi013> (2005).
56. Roden, D. M. & Tyndale, R. F. Genomic medicine, precision medicine, personalized medicine: What's in a name?. *Clin. Pharmacol. Ther.* **94**, 169–172. <https://doi.org/10.1038/clpt.2013.101> (2013).
57. Shan, J. *et al.* Genome scan study of prostate cancer in Arabs: Identification of three genomic regions with multiple prostate cancer susceptibility loci in Tunisians. *J. Transl. Med.* **11**, 1–8 (2013).
58. Boattini, A. *et al.* Uniparental markers in Italy reveal a sex-biased genetic structure and different historical strata. *PLoS ONE* **8**, e65441 (2013).
59. Falush, D., Stephens, M. & Pritchard, J. K. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587. <https://doi.org/10.1093/genetics/164.4.1567> (2003).
60. Zheng, X. *et al.* A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
61. Berner, C. *et al.* Dota 2 with large scale deep reinforcement learning. *arXiv Preprint arXiv:1912.06680* (2019).
62. Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinform.* **12**, 1–6 (2011).
63. de Keyser, C. E. *et al.* Genetic variation in the PPARA gene is associated with simvastatin-mediated cholesterol reduction in the Rotterdam Study. *Pharmacogenomics* **14**, 1295–1304. <https://doi.org/10.2217/pgs.13.112> (2013).
64. Link, E. *et al.* SLCO1B1 variants and statin-induced myopathy—A genomewide study. *N Engl. J. Med.* **359**, 789–799. <https://doi.org/10.1056/NEJMoa0801936> (2008).
65. Kivistö, K. T. *et al.* Lipid-lowering response to statins is affected by CYP3A5 polymorphism. *Pharmacogenetics* **14**, 523–525. <https://doi.org/10.1097/01.fpc.0000114762.78957.a5> (2004).

Acknowledgements

This work was funded by the European Commission Grant agreement no. 279171–1 for FP7 project MEDIGENE and supported by the Tunisian Ministry of Public Health and the Ministry of Higher Education and Scientific Research (LR16IPT05). H.J. is a recipient of a MOBIDOC fellowship, funded by The Ministry of Higher Education and Scientific Research through the PromEssE project and managed by the National Agency for the Promotion of Scientific Research. Moreover, this work was funded by Pfizer International (Pfizer2011) and the Italian Ministry of Education, University and Research (PRIN2010EL8TXP_006 to D.L.).

Author contributions

H.J., S.S. S.A., D.L. and R.K. conceived the study; H.J., S.S performed bioinformatic and statistical analyses; H.J. wrote the manuscript. H.J., W.B., C.G., S.A., D.L. and R.K. reviewed and edited the manuscript. R.K. supervised the work.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-55239-7>.

Correspondence and requests for materials should be addressed to R.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024