

*Article***Genetic variation among pharmacogenes in the Sardinian population**

Maria Laura Idda^{1†}, Magdalena Zoledziwska^{2†}, Silvana Anna Maria Urru^{3,4}, Gregory McInnes⁵, Alice Bilotta⁶, Viola Nuvoli⁶, Valeria Lodde⁷, Sandro Orrù⁸, David Schlessinger⁹, Francesco Cucca^{2,7†}, Matteo Floris^{1,7†*}

¹Institute for Genetic and Biomedical Research (IRGB-CNR), 07100 Sassari, Italy; laura.idda@irgb.cnr.it (M.L.I); matteo.floris@gmail.com (M.F.)

²Institute for Genetic and Biomedical Research, National Research Council, 09042 Monserrato Italy; magdalena.zoledziwska@irgb.cnr.it (M.Z.); fcucca@uniss.it (F.C.)

³Hospital Pharmacy Unit, Trento General Hospital, Autonomous Province of Trento, 38122 Trento, Italy; silvanaurru@gmail.com (S.A.M.U.)

⁴Department of Chemistry and Pharmacy, School of Hospital Pharmacy, University of Sassari, 07100 Sassari, Italy;

⁵Biomedical Informatics Training Program, Stanford University, 94305 Stanford, California, USA; gregorymcinnes@gmail.com (G.MI.)

⁶School of Medicine, Department of Biomedical Sciences, University of Sassari, 07100 Sassari, Italy; alicebilotta@gmail.com (A.B.); violanuvoli.flor@gmail.com (V.N.)

⁷Department of Biomedical Sciences, University of Sassari, 07100 Sassari, Italy; valerialodde@gmail.com (V.L.)

⁸Medical Genetics, Department of Medical Sciences and Public Health, University of Cagliari, 09126 Cagliari, Italy; s.orrù@unica.it (S.O.)

⁹Laboratory of Genetics, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA; SchlessingerD@grc.nia.nih.gov (D.S.)

*Correspondence: Matteo Floris: matteo.floris@gmail.com; Tel.: +39-(0)-3334857679

† These authors contributed equally

Abstract: Pharmacogenetics (PGx) aims to identify the genetic factors that determine inter-individual differences in response to drug treatment maximizing efficacy while decreasing the risk of adverse events. Estimating the prevalence of PGx variants involved in drug response, is a critical preparatory step for large-scale implementation of a personalized medicine program in a target population. Here, we profiled pharmacogenetic variation in fourteen clinically

relevant genes in a representative sample set of 1,577 unrelated sequenced Sardinians -- an ancient island population that well accounts for genetic variation in Europe as a whole and, at the same time are enriched in genetic variants that are very rare elsewhere. To this end, we used PGxPOP, a PGx allele caller based on the guidelines created by the Clinical Pharmacogenetics Implementation Consortium (CPIC), to identify the main phenotypes associated with the PGx alleles most represented in Sardinians. We estimate that 99.43% of Sardinian individuals may potentially respond atypically to at least one drug, that on average each individual is expected to have an abnormal response to about 17 drugs, and that for 27 drugs the fraction of the population at risk of atypical responses to therapy is more than 40%. Finally, we highlighted 174 pharmacogenetic variants for which the minor allele frequency is at least 10% higher among Sardinians as compared to other European populations, a fact that may contribute to substantial interpopulation variability in drug response phenotypes. This study provides baseline information for further large-scale pharmacogenomic investigations in the Sardinian population and underlines the importance of the PGx characterization of diverse than European populations as Sardinians.

Keywords: Pharmacogenetics, pharmacogenomics, drug response

1. Introduction

Drug treatments are characterized by adverse reactions and substantial difference in efficacy in different patients. Surprisingly, at least 5% of hospital admissions are directly related to an adverse drug reaction (ADR, including allergic, pseudoallergic, and exaggerated pharmacological reactions to medications), with an overall fatality of 0.15% and an annual cost of > 500 M\$ only for the UK National Health Service [1,2]. These data highlight the social and economic costs of ADRs and the urgent need to find effective strategies to ameliorate drug efficacy and reduce ADR.

The same drug, once other parameters are fixed, can have different therapeutic effects in different people and/or ethnic groups [3]. The analysis of the genetic variability modulating the individual's drug response (*pharmacogenetics*, *PGx*) has thus received great attention for its capacity to provide a new way to optimize drug therapies in terms of optimal dosing to improve drug efficacy and reduce toxicity risk [4]. As a result, a patients may receive the right drug at the right dose the first time they consult their doctors such that efficacy is guaranteed reducing

the risk of ADR. From a pharmaceutical point of view, PGx variants can influence pharmacokinetics and pharmacodynamics drugs thus influencing dosing, formulation sensitivity and drug-hypersensitivity reactions.

An individual's drug response can be assessed by the identification by genotyping arrays of well-characterized genetic variants in key genes implicated in drug processing or by the identification of specific haplotypes. For example, the gene *CYP2D6* is characterized by the presence of over 100 haplotypes, which share SNPs and include gene duplications and deletions, strongly influencing the metabolism and/or bioactivation of many clinically used drugs. In this example, phenotypes are assigned to haplotypes that contains more than one relevant SNP rather than to a specific SNPs [5].

The interest in ameliorating drug efficacy while reducing ADR promote the development of tools to properly analyze the correlation between variability in the genome and individual's drug response. For example, the *Pharmacogenetics and pharmacogenomics knowledgebase* (PharmGKB: <http://www.pharmgkb.org>) covers much information about pharmacogenomics and provides a convenient approach for researchers. The *Pharmacogenetics of Membrane Transporters (PMT) database* is another tool used for the exchange of chemicals and signals in and around biofilms [6,7].

Furthermore, the increasing availability of accurate classifications of genetic variants and haplotypes of pharmacogenetic interest, together with guidelines for their clinical translatability, allow analyze the potential impact of pharmacogenetics programs in many populations for which large-scale genomic resources exist [8,9].

Sardinians, a population for which large-scale genomic data are available, is particularly well suited for genetic studies. Sardinians are the contemporary human population that has retained the highest degree of inheritance from early European farmers who lived in the Neolithic period along with significant ancestry from western hunter gatherers who lived in the late Paleolithic period [10,11]. This is due to founder effects during the initial settlement of the island and the scarcity of gene flow from other populations during later periods [10,11]. As a result of its past evolutionary history the Sardinians are now a reservoir of ancient European genetic variants that are currently very rare elsewhere and may have relevant clinical consequences [12][13][14][15]. Genetic factors and the distinct genetic structure of the Sardinians thus present an excellent opportunity also to look for new pharmacogenetic information.

Here, we profiled pharmacogenetic variation in fourteen clinically relevant genes in 1,577 unrelated sequenced Sardinians. We used PGxPOP [16], a PGx allele caller based on the guidelines created by the Clinical Pharmacogenetics Implementation Consortium (CPIC), to identify the main phenotypes associated with the PGx alleles most represented in Sardinians. We estimate that 99.43% of Sardinian individuals may potentially respond atypically to at least one drug, and that on average each individual is expected to have an abnormal response to about 17 drugs. Furthermore, we highlighted differences in haplotype and diplotype frequencies of star alleles as compared to other populations; we estimated that for 27 drugs the fraction of the population at risk of atypical responses to therapy is more than 40%. These findings represent the foundation for further large-scale and more detailed pharmacogenomic investigations in Sardinia and, at the same time, underline the importance of the pharmacogenomics characterization of ethnically diverse than European populations as exemplified by Sardinians.

2. Results

2.1. Haplotype and phenotype calling

To identify clinically relevant pharmacogenetic variation in 14 important genes for which Clinical Pharmacogenetics Implementation Consortium (CPIC) has created detailed gene/drug clinical practice guidelines, we processed the genomic sequence data from our SardiNIA cohort (1,577 individuals) with PGxPOP [16]. PGxPOP is a PGx matching engine that is based on PharmCAT and uses its PGx allele definitions to characterize PGx haplotype, diplotype and phenotype frequencies; it extends the capabilities of PharmCAT by generating diplotypes from population scale datasets [17]. In the analyzed Sardinia cohort, 99.43% of the 1,577 participants carry at least one diplotype associated to predicted non-typical response phenotype across the 14 pharmacogenes analyzed (**Table 1**); furthermore, for each participant we were able to predict an average of about 4 phenotypes of non-typical drug response (Mean=3.44, Min=1, Max=8, **Figure 1**). These numbers are in line with what has been observed in the UK Biobank cohort, where participants were previously reported carry on average 3.7 nontypical response diplotypes for the 14 pharmacogenes, with 99.5% of participants carrying at least 1 nontypical drug response diplotype [16].

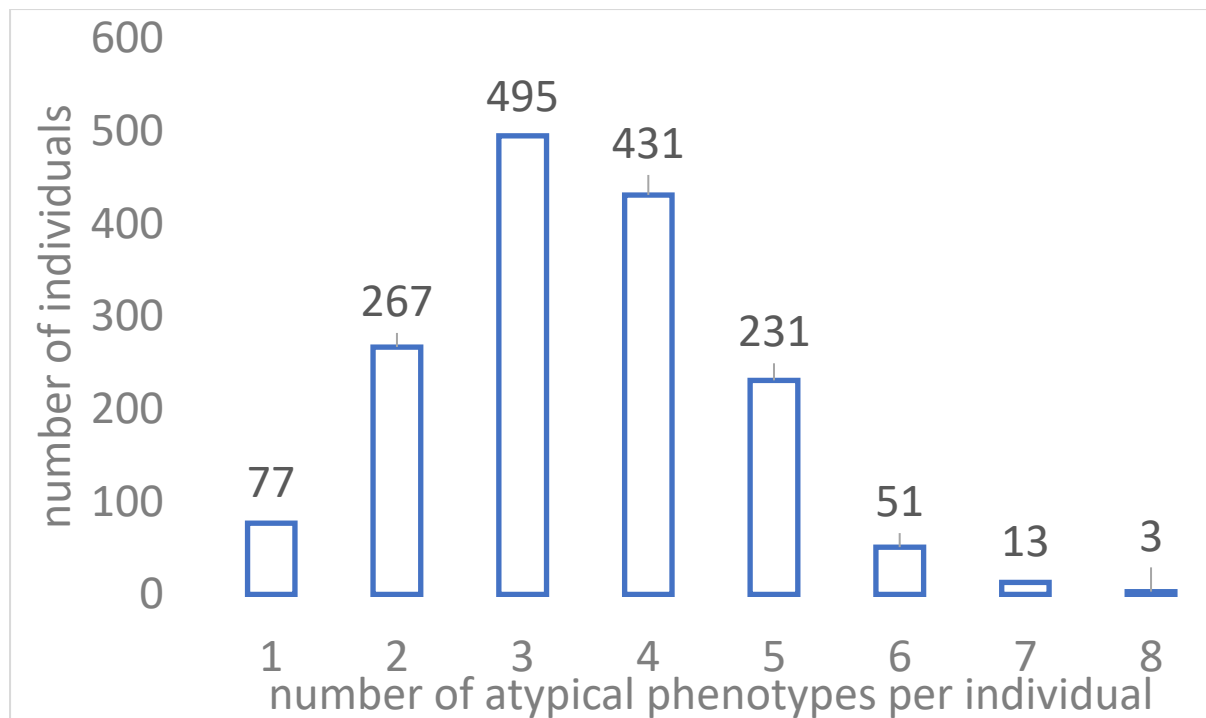


Figure 1. The figure shows the distribution of the number of abnormal drug response phenotypes detected in the 1,577 individuals in the Sardinia cohort by PGxPop.

Although in general we observed some agreement with the observations in UKBB, we noticed some differences in terms of haplotype and diplotype frequencies. Largest absolute discordance in star allele frequencies ($\Delta \text{MAF} \geq 10\%$) between Sardinians and individuals with European ancestry of the UK Biobank cohort were observed for *CYP2D6**1 and *119 (results that can be affected by the fact that structural variants were not called in this analysis), *CYP4F2* alleles *1, *2 (this being extremely rare in Sardinians) and *2+*3, *VKORC1* alleles -1639A and -1639G, and for *SLCO1B1* alleles *1A and *14 (**Suppl Table ST3**).

In addition, out of 133 diplotypes called by PGxPOP for the 14 genes analysed, only 13 diplotypes (10%) have an absolute difference in frequency ranging from 10% to 25%, as compared to UKBB European individuals (**Suppl Table ST4**).

We detected a total of 22 different phenotypes for the 14 pharmacogenes, but for *CYP2D6* and *UGT1A1* our analysis did not detect any non-typical drug response diplotype (**Table 2**). Based on these analyses, we can predict that, for at least 27 drugs, about one third of the population is at risk of an atypical response [**Figure 2**] and that on average each individual is expected to have an abnormal response to about 17 drugs [**Table 1**].

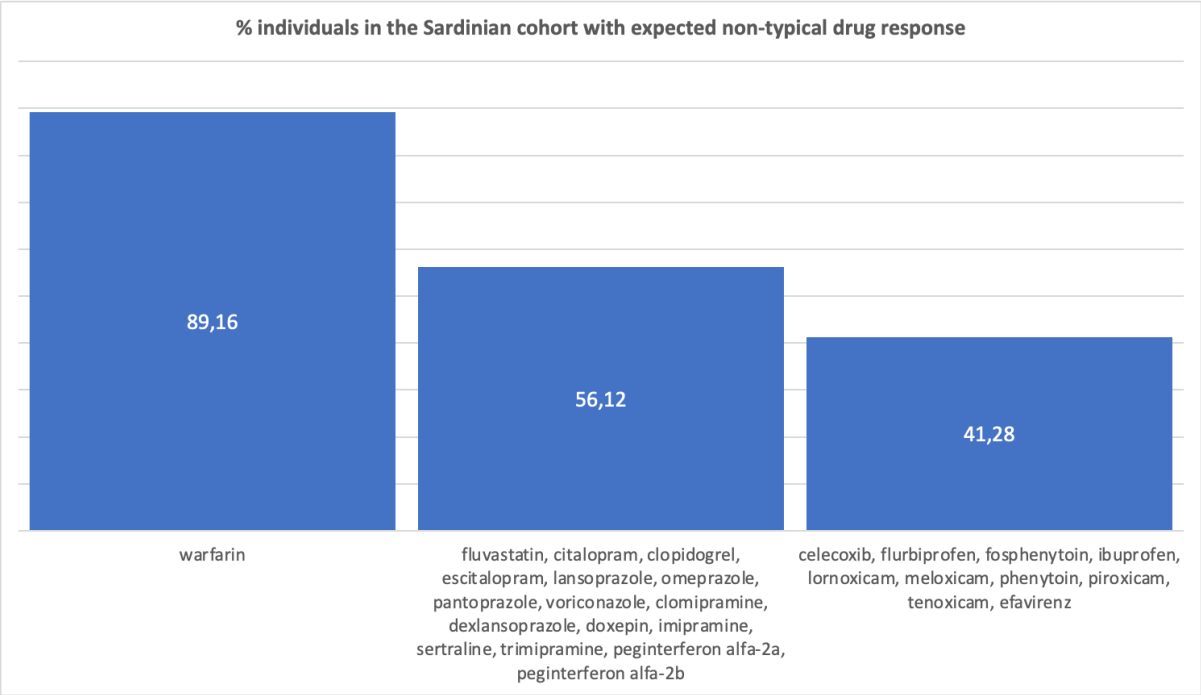


Figure 2. The figure shows for each drug - whose response is affected by at least one of the abnormal drug response phenotypes identified by PGxPop - the percentage of individuals in the Sardinia cohort who are at risk of atypical response.

Table 1. The table shows how many individuals are at risk of an atypical drug response, in descending order of number of drugs.

NUMBER OF INDIVIDUALS	NUMBER OF DRUGS FOR WHICH AN ATYPICAL RESPONSE IS EXPECTED
1	39
3	38
6	37
12	36
16	35
14	34
5	33
7	32
5	31
19	30
38	29
57	28
47	27
49	26
40	25
45	24
43	23
51	22

36	21
34	20
42	19
92	18
132	17
125	16
127	15
89	14
58	13
58	12
22	11
22	10
12	9
11	8
2	7
9	6
27	5
49	4
68	3
55	2
40	1
TOT = 1568 (99.43%)	

Table 2. Predicted non-typical response phenotype across the 14 pharmacogenes analyzed in the Sardinia cohort; each cell contains the number of individuals in the cohort to whom atypical response phenotypes (highlighted in **bold**) are assigned by PGxPop, divided by pharmacogene.

	Pharmacogenes													
Phenotypes *	CFTR	CYP2B6	CYP2C19	CYP2C9	CYP2D6	CYP3A5	CYP4F2	DPYD	IFNL3	NUDT15	SLCO1B1	TPMT	UGT1A1	VKORC1
Decreased Function											324			
Decreased warfarin dose														414
Favorable response genotype									721					
Increased dose phenotype							268							
Indeterminate		50			207		11				133	2		
Intermediate Function												101		
Intermediate Metabolizer		552	450	606		304		117		2				
Ivacaftor irrelevant	1555													
Ivacaftor non-responsive	14													
Ivacaftor responsive	8													
Normal dose phenotype							534							
Normal Function											533	1472		
Normal Metabolizer		862	715	925	622	16		1460		1575			815	
Normal warfarin dose														385
Not available		8		2	748	14	764				201		762	
Poor Function											51	2		
Poor Metabolizer		85	32	44		1243								
Possible increased function											335			
Possibly decreased warfarin dose														778
Rapid Metabolizer		20	343											
Ultrarapid Metabolizer			37											
Unfavorable response genotype									856					
Grand Total	1577	1577	1577	1577	1577	1577	1577	1577	1577	1577	1577	1577	1577	1577
Total number of non-typical drug response phenotypes	14	552	519	650	0	1547	268	117	856	0	375	103	0	1192

The genes with the highest number of non-typical phenotypes are *VKORC1* (N= 1,192; 76%), *CYP2C19* (N=862; 55%) and *IFNL3* (N= 856; 54%). Genes with the most 'not available' phenotypes are *CYP2D6* (47.43% of subjects), *UGT1A1* (48.31%), *CYP4F2* (48.45%) and *SLCO1B1* (12.75%). Alleles with unknown or uncertain function, leading to an "indeterminate" phenotype, were found in 6 genes (*CYP2B6*, *CYP2D6*, *CYP4F2*, *SLCO1B1*, *TPMT* and *UGT1A1*).

If we focus in particular the frequency of non-typical drug response phenotypes in the Sardinian cohort compared to the European sub-population of UK Biobank, those more frequent in Sardinians are "Decreased warfarin dose" for the *VKORC1* gene (26% vs 14%) and "Intermediate metabolism" for the *CYP3A5* gene (19% vs 13%); for 5 other non-typical drug response phenotypes, the difference in frequency is between 1 and 5%, while for all other non-typical phenotypes the frequency in the Sardinian population is always lower than in the European UK Biobank cohort (full details are available in **Suppl Table ST5**).

2.2. Actionable pharmacogenomic variants in Sardinian genomes

We extended the above analysis by considering the carrier status of clinically actionable variants (PharmGKB level 1A/1B) in each of the 1,577 Sardinian unrelated individuals.

Among the 3,073 clinical annotations in the PharmGKB database (accessed March 2022), 141 single-nucleotide variants (SNVs) and 132 haplotype variants have the highest level of evidence (1A/1B or 2A/2B), while 2,652 SNVs and 247 haplotype variants have a lower level (3 or 4). The SNV variants were then overlapped with the variants in Sardinian population and their prevalence was evaluated based on their allele frequencies in both a dataset of 1,577 unrelated Sardinians and in the gnomAD dataset (non-Finnish Europeans), which established extensive interpopulation differences. In this analysis, we considered only variants in non-cytochrome genes, yielding results as follows.

Among the variants associated with at least one of the highest levels of evidence, we identified in our analysis 13 variants for which the absolute difference in allele frequency between Sardinians and Europeans is at least 5%, of which 5 have an MAF in Sardinia that is at least 10% higher than the Europeans (**Table 3 for a summary, Table 4 for clinical annotations**). These 13 high-evidence pharmacogenetic variants are involved – according to 26 PharmGKB clinical annotations -- in the response to 9 different drugs: the anticoagulants Acenocumarol, Penprocoumon, and Warfarin; the cholesterol-lowering agent Pravastatin; the anticancer agents Capecitabine, Fluorouracil; and the immunosuppressants Etanercept, Methotrexate, and

Rituximab. Many of the variant-drug pairs (2 out of 3) are relevant for dosage and efficacy of the analyzed drug, while the remaining are relevant for toxicity (**Table 4**). Interestingly, about half of these differentiated pharmacogenetic variants with high priority level are relevant for drugs directed to pediatric populations. In more detail, 4 variants in 4 genes (*VKORC1*, *CYP4F2*, *DPYD* and *MTHFR*) are marked as important for pediatric prescription of 4 drugs (Phenprocoumon, Warfarin, Fluorouracil and Methotrexate). For 4 of these variants, we have at least one alert with Evidence level 1A or 1B (rs2108622, rs9923231 and rs9934438, whose minor allele is about 10% more frequent in Sardinians than in Europeans, whereas the rs1801265 minor allele is about 6% less frequent in Sardinians) (**Table 4**).

Among the other pharmacogenetic variants with lower levels of evidence (3 and 4), we identified 169 variants for which the minor allele frequency is at least 10% higher in Sardinians compared to Europeans, and a further 72 variants with an opposite trend (**Table 5 for a summary, Suppl Table ST6 for clinical annotations**). The 169 low-evidence variants are involved in the response to 201 different drugs (a total of 405 variant-drug pairs). Most variant-drug pairs are relevant for drug efficacy (39% of annotations) and for toxicity (37% of annotations) (**Suppl Table ST6**). Overall, 41 of the variant-drug pairs (about 10%) are relevant for drugs directed to the pediatric population.

For another 110 variant-drug pairs with lower levels of evidence (5 of which are relevant for the pediatric population), 72 unique variants -- for which the minor allele frequency is at least 10% lower in Sardinians compared to Europeans -- are involved in the response to 91 different drugs.

Table 3. Summary of the variants associated with levels of evidence (1A, 1B, 2A or 2B) and with absolute difference in allele frequency between Sardinians and Europeans >= 5%. Abbreviations: BP= Position; RSID= Reference SNP cluster identifier; CHR= Chromosome; A2=ALT allele

CHR	BP	RSID	A2	SARD_A2_FRQ	ALFA_A2_FRQ	DELTA
1	161514542	rs396991	C	0,528	0,317	0,211
16	31104509	rs8050894	G	0,523	0,374	0,148
19	15990431	rs2108622	T	0,415	0,294	0,120
16	31107689	rs9923231	T	0,510	0,392	0,117
16	31104878	rs9934438	A	0,510	0,395	0,114
6	39325078	rs20455	G	0,463	0,364	0,099
1	11856378	rs1801133	A	0,424	0,349	0,075
16	31105554	rs2884737	C	0,353	0,278	0,074
1	97981395	rs1801159	C	0,252	0,199	0,053
CHR	BP	RSID	A2_SARD	A2	ALFA_EUR_FRQ	DELTA
1	98348885	rs1801265	G	0,159	0,215	-0,056
21	46957794	rs1051266	C	0,511	0,568	-0,057
6	31543031	rs1800629	A	0,050	0,159	-0,109
16	31103796	rs2359612	G	0,490	0,604	-0,114

Table 4. Clinical annotations of the variants associated with levels of evidence (1A, 1B, 2A or 2B) and with absolute difference in allele frequency between Sardinians and Europeans >= 5%. Pediatric population column: 1=yes.

Clinical Annotation ID	Variant/Hapl otypes	is_top	Gene	Level of Evidence	Phenotype Category	Drug(s)	Phenotype(s)	PEDIATRIC Population
1184661194	rs2108622	rs2108622	CYP4F2	2A	Dosage	acenocoumarol	Atrial Fibrillation	0
981204044	rs9923231	rs9923231	VKORC1	1A	Dosage	acenocoumarol		0
1183704228	rs9934438	rs9934438	VKORC1	2A	Dosage	acenocoumarol		0
1451237940	rs9923231	rs9923231	VKORC1	1A	Dosage	phenprocoumon		1

1451244040	rs9934438	rs9934438	VKORC1	2A	Dosage	phenprocoumon		1
655385400	rs2108622	rs2108622	CYP4F2	1A	Dosage	warfarin		1
982035703	rs2884737	rs2884737	VKORC1	2A	Dosage	warfarin		0
655385392	rs9934438	rs9934438	VKORC1	1B	Dosage	warfarin		1
655385028	rs8050894	rs8050894	VKORC1	1B	Dosage	warfarin		0
655385012	rs9923231	rs9923231	VKORC1	1A	Dosage	warfarin		1
655385024	rs2359612	rs2359612	VKORC1	1B	Dosage	warfarin		0
655384799	rs1800629	rs1800629	TNF	2B	Efficacy	etanercept	Arthritis, Psoriatic; Arthritis, Rheumatoid; Crohn Disease; Inflammation; Psoriasis; Spondylitis, Ankylosing	0
1451245360	rs1051266	rs1051266	SLC19A1	2A	Efficacy	methotrexate	Arthritis, Rheumatoid	0
655384621	rs20455	rs20455	KIF6	2B	Efficacy	pravastatin	Coronary Disease; Myocardial Infarction	0
1444608384	rs396991	rs396991	FCGR3A	2B	Efficacy	rituximab	Arthritis, Rheumatoid; Neuromyelitis Optica	0
1447672998	rs9923231	rs9923231	VKORC1	2A	Efficacy	warfarin	time to therapeutic inr	1
1447673015	rs9923231	rs9923231	VKORC1	2A	Efficacy	warfarin	time in therapeutic range	1
1451286320	rs1801159	rs1801159	DPYD	1A	Toxicity	capecitabine	Neoplasms	0
1451287240	rs1801265	rs1801265	DPYD	1A	Toxicity	capecitabine	Neoplasms	0
981201981	rs1801265	rs1801265	DPYD	1A	Toxicity	fluorouracil	Neoplasms	1
981201962	rs1801159	rs1801159	DPYD	1A	Toxicity	fluorouracil	Neoplasms	0
827848365	rs1801133	rs1801133	MTHFR	2A	Toxicity	methotrexate	Drug Toxicity;hematotoxicity; Leukopenia; Lymphoma; mucositis; Neoplasms; Neutropenia; Osteosarcoma; Precursor Cell Lymphoblastic Leukemia-Lymphoma; primary central nervous system lymphoma; Thrombocytopenia; Toxic liver disease	1
655385307	rs1801133	rs1801133	MTHFR	2A	Toxicity	methotrexate	Arthritis, Juvenile Rheumatoid; Arthritis, Psoriatic; Arthritis, Rheumatoid; Drug Toxicity	1
1451243676	rs9923231	rs9923231	VKORC1	2A	Toxicity	phenprocoumon	Hemorrhage;over-anticoagulation; time above therapeutic range	0
1449269910	rs9923231	rs9923231	VKORC1	2A	Toxicity	warfarin	Hemorrhage	1
1447673005	rs9923231	rs9923231	VKORC1	1B	Toxicity	warfarin	over-anticoagulation	1

Table 5. Summary of the variants associated with lower levels of evidence (3 or 4) and with difference in allele frequency between Sardinians and Europeans $\geq 20\%$. Abbreviations: POS= position; RSID= Reference SNP cluster identifier; CHR= Chromosome; A2=ALT allele

CHR	POS	RSID	A2	SARD_A2_FRQ	ALFA_A2_FRQ	DELTA
1	1:207753621	rs2274567	G	0,608	0,195	0,413
6	6:31093482	rs3131003	A	0,739	0,432	0,308
6	6:31093587	rs3815087	A	0,507	0,217	0,290
13	12:13953118	rs2058878	A	0,602	0,338	0,265
16	12:85243681	rs6539870	G	0,486	0,221	0,264
4	3:45732515	rs2742421	G	0,687	0,427	0,260
6	6:31603770	rs11229	G	0,435	0,176	0,259
15	15:30193316	rs813676	C	0,761	0,504	0,257
6	6:31107361	rs2233945	A	0,417	0,163	0,254
6	6:31604591	rs10885	T	0,436	0,182	0,254
6	6:31018546	rs2523864	T	0,679	0,433	0,245
19	19:15959200	rs2189784	A	0,667	0,423	0,245
6	6:31022113	rs3873352	G	0,336	0,091	0,245
6	6:43737794	rs13207351	G	0,611	0,367	0,244
3	3:151090996	rs9859552	T	0,338	0,094	0,244
6	6:31543101	rs361525	A	0,292	0,054	0,238
6	6:31542308	rs1799964	C	0,447	0,214	0,233
7	6:78173281	rs130058	A	0,414	0,183	0,231
5	5:158750769	rs3213094	T	0,427	0,206	0,221
7	7:86331756	rs2189814	C	0,375	0,159	0,216
11	10:32202069	rs2799018	T	0,611	0,396	0,215
1	1:161514542	rs396991	C	0,528	0,317	0,211
11	10:92619161	rs7905446	G	0,480	0,276	0,205
6	6:33047612	rs3097671	C	0,366	0,163	0,203
10	1:29161999	rs2236855	A	0,386	0,185	0,201

3. Material and methods

3.1. Study population and data sets

In this study we focused on a subset of 1,577 unrelated sequenced Sardinian samples from the Sardinian sequence-based reference panel: these were mainly the unrelated parents of a larger sample set of 3,514 individuals (mainly trios) sequenced at low-coverage (average coverage 4.2x) which also included their children. The sample set included 2,090 individuals belonging to the SardiNIA cohort study in the subregion of Ogliastra and 1,424 individuals deriving from a case-control study on autoimmunity collected across the Island [11,14]. All participants signed informed consent to study protocols approved by the Sardinian Regional Ethics Committee (protocol no. 2171/CE).

3.1.1. Genotyping and imputation

All the genetic analyses were performed using a genetic map based on 6,602 samples genotyped with 4 Illumina arrays (OmniExpress, ImmunoChip, Cardio-MetaboChip and ExomeChip) as previously described [14]. Imputation was performed on a genome-wide scale using a Sardinian sequence-based reference panel of 3,514 individuals and the software Minimac (<https://genome.sph.umich.edu/wiki/Minimac3>) on pre-phased genotypes. After imputation, only markers with imputation quality (RSQR) > 0.3 for estimated minor allele frequency (MAF) $\geq 1\%$ or > 0.6 for MAF < 1% were retained for further analyses, yielding ~22 million variants (20,143,392 SNPs and 1,688,858 indels).

3.2. Variant calling and relatedness

Variant calling and phase assignment procedures leveraged the family structure of the extended sample of 3,514 individuals. Allele frequencies were estimated from the subset of 1,577 unrelated individuals and compared to the corresponding frequencies available in gnomAD project v2 (European non-Finnish subset) [18].

Relatedness was estimated by computing the genome-wide proportion of pairwise IBD (π) on a random set of 1 million SNPs with an MAF > 0.05 in 1,000 Genomes populations (Phase 3 v5) [<https://www.internationalgenome.org/home>]. Starting from the dataset of 3,514

individuals, for each pair of individuals with $\pi > 0.05$, we preferentially removed the offspring if it occurred in a trio; otherwise, we removed the individual with the larger summed value of π across all other relationships with $\pi > 0.05$, leading a total of 1,577 samples for the allele frequency analyses.

3.3. Pharmacogenetics resources

3.3.1. Haplotype and phenotype calling

We used PGxPOP [<https://github.com/PharmGKB/PGxPOP>] [16], to calculate specific haplotype, diplotype, and phenotype frequencies for fourteen genes in our multisample VCF file. PGxPOP uses the PharmCAT allele definition files [<https://github.com/PharmGKB/PharmCAT>][17], and reports exact matches to the allele definitions based on the provided phased genotype data.

Gene phenotypes were determined to have a non-typical response if any CPIC guidance recommended an alternate dosage or drug for that phenotype, as defined by McInnes et al [16] (**Suppl. Table ST1**).

Currently, the CPIC guidelines (<https://cpicpgx.org/genes-drugs/>, accessed on 14 July 2022) indicate a set of 64 recommendations for the 14 genes profiled by PGxPOP, for a total of 53 drugs. In this work, we only considered gene/drug pairs whose CPIC evidence levels are class A and B (for which there is a sufficient level of evidence to suggest a recommendation) (**Suppl. Table ST2**).

Frequencies in Sardinia of star allele haplotypes and diplotypes, and estimated prevalence of non-typical response phenotypes were compared with the corresponding values in UKBiobank populations reported by McInnes et al [16] (**Suppl. Tables ST3, ST4 and ST5**), where the burden of nontypical response phenotypes for each individual is estimated by counting the number of diplotypes with predicted nontypical response phenotypes across 14 genes with phenotypes. Gene phenotypes were determined to have a nontypical response if any CPIC guidance recommended an alternate dosage or drug for that phenotype.

Structural variants (SVs) were not called for any of the considered genes.

3.3.2. Allele frequency analysis of PGx actionable variants from PharmGKB

The full list of PharmGKB [19] clinically relevant variants assigned to evidence classes 1A, 1B, 2A and 2B, downloaded from the PharmGKB website ([<https://www.pharmgkb.org/>], accessed March 2022). Level 1A clinical annotations are typical of those variant-drug combinations characterized by the existence of precise variant-specific prescriptive guidance, such as in an FDA-approved drug label.

Level 1B clinical annotations describe variant-drug combinations supported by at least two independent publications but no specific prescriptive guidance, as was required for the previous annotations. Level 2A variants are in PharmGKB's Tier 1 Very Important Pharmacogenes (VIPs), i.e., in known pharmacogens, and thus with high probability of phenotype causation. Variants in Tier 2B clinical annotations are not present in PharmGKB Tier 1 VIPs and describe variant-drug combinations with a moderate level of evidence supporting association.

For more details about the PharmGKB scoring system, readers can refer to the documentation available at <https://www.pharmgkb.org/page/clinAnnLevels>.

4. Discussion and Conclusions

We estimated the potential impact of the large-scale introduction of pharmacogenetic practices in the Sardinian population by evaluating the prevalence of clinically relevant pharmacogenetic variants in a core set of 1,577 unrelated sequenced individuals, representative of the entire population; Sardinia consist of 1.5M residents in the island and a similar number of individuals of Sardinian descent spread across the world. To this end, we used PGxPOP [16], a PGx matching engine that is based on PharmCAT and uses its PGx allele definitions to characterize PGx allele and phenotype frequencies. Using this analysis, it was possible to estimate the theoretical number of Sardinian individuals exposed to adverse reactions to a range of drugs. In more detail, the frequencies of two phenotypes ("Decreased warfarin dose" and "Possibly decreased warfarin dose") involving warfarin, a widely used anticoagulant drug, are among the most interesting findings from this analysis. The two atypical phenotypes are determined by diplotypes of the *VKORC1* gene [20] and affect a total of 1,192 individuals in our cohort (i.e., about 3 of 4 individuals). Overall common genetic variants in this gene, but also in *CYP2C9*, *CYP4F2*, and the CYP2C cluster (e.g., rs12777823), plus known nongenetic factors, account for 50% of warfarin dose variability [20].

Other phenotypes potentially affecting a large proportion of the population are the "Intermediate Metabolizer" and "Poor Metabolizer" phenotypes, which are determined by cytochrome *CYP2C9* diplotypes and affect a total of 650 individuals in our cohort (about 41%

of the cohort analyzed). These phenotypes have important effects on the ADME-Tox of Nonsteroidal Anti-Inflammatory Drugs such as celecoxib, flurbiprofen, lornoxicam, and ibuprofen. According to CPIC guidelines [21], the diplotypes involved may result in a higher-than-normal risk of adverse events especially in individuals with other factors affecting clearance of these drugs, such as hepatic impairment or advanced age. The same guidelines suggest a reduced dosage of these drugs and monitoring of adverse effects. The same cautions can be extended to other drugs such as meloxicam, piroxicam and tenoxicam.

An important finding concerns two atypical phenotypes related to the *SLCO1B1* gene ("Decreased Function", N=324 individuals, and "Poor function", N=51), which globally affect almost 1 in 4 individuals, and are important for the metabolism of important drugs such as Atorvastatin (second among the top thirty active drugs both for consumption and expenditure in Italy) [22], and Fluvastatin, Lovastatin, Pitavastatin, Pravastatin, Rosuvastatin and Simvastatin. According to CPIC guidelines [23], these phenotypes can impact the starting dose and suggest an adjustment of doses based on disease-specific guidelines. According to suggestions in the same guidelines, prescribers should be aware of possible increased risk for myopathy.

We can then hypothesize that an important impact in the frequency of adverse effects could be caused by the high diffusion of atypical phenotypes ("Intermediate metabolizer", N=450, "Poor metabolizer", N=32, "Rapid metabolizer", N=343 and "Ultrarapid Metabolizer", N=37) attributable to diplotypes of the gene *CYP2C19*, involved in the metabolism of some of the most widely used antidepressants in Italy, including escitalopram and sertraline. According to the guidelines [24], among the problems caused by an incorrect dosage are increased risk for adverse cardiac and cerebrovascular events.

Special attention should be paid to the 103 individuals (approximately 6.5%) who are at high risk of severe toxicity due to antineoplastic drugs such as azathioprine, mercaptopurine, and thioguanine because of atypical phenotypes determined by diplotypes of the *TPMT* gene.

In a second phase of analysis, we aimed to identify the variants of pharmacogenetic interest that are more differentiated in Sardinia than in the general European population (taking as reference the genetic data of gnomAD version 2.1). In this analysis, we distinguished highly relevant PGx variants (levels of evidence 1A, 1B, 2A and 2B) from those of lower relevance (levels 3 and 4).

The strongest difference in terms of allele frequency was seen for the rs396991 variant located in *FCGR3A* gene and could be relevant for patients treated with Rituximab, according to a 2B

level of evidence documented by PharmGKB [25]. In fact, the C allele of the rs396991 variant has a frequency 1.5 times higher in Sardinia (AF=0.528) than in the rest of Europe (AF=0.344). This difference may significantly affect the efficacy of Rituximab -- used in the treatment of certain type of cancer and autoimmune disorders including Rheumatoid Arthritis and Neuromyelitis Optica: indeed, patients with a CC genotype may have an increased response to the drug compared to patients with AA and AC genotype.

Another variant of special interest is rs8050894, for which the frequency of the G allele is 1.34 times higher in Sardinia (AF=0.523) than in the general European population (AF=0.389). This variant has a role -- supported on a type 1B level of evidence -- in influencing warfarin dosage: according to the guidelines, patients with the GG genotype may require a lower dose of warfarin as compared to patients with the CC genotype. The variant is part of a haplotype of variants in the *VKORC1* gene, all of which are associated with warfarin dosing. Among them, the one with the strongest level of evidence is rs9923231, whose T allele is 1.309 times more frequent in Sardinians (AF=0.509 vs 0.389). This last variant is also relevant for the pediatric population and has relevance not only for the dosage of Warfarin, Acenocoumarol and Phenprocoumon, but also for the resultant efficacy and toxicity of these drugs. Of note, the genotypes of *VKORC1*-1639G>A (rs9923231) are mentioned in the FDA Label of Warfarin. Warfarin inhibits *VKORC1* to prevent regeneration of a reduced form of vitamin K necessary for clotting factor activation [26]. The common variants, noted in our analysis, are located in the 5'UTR and introns of the *VKORC1* gene and are associated with reduced gene expression and related effects on warfarin dosage. Warfarin and Acenocoumarol are common oral anticoagulant prescribed for the treatment and prevention of thromboembolic events for which genetic variants in several genes (*CALU*, calumenin; *CYP*, cytochrome P450 family members; *GGCX*, gamma-glutamyl carboxylase; *NQO1*, NAD(P)H quinone dehydrogenase 1; *VKORC1*, vitamin K epoxide reductase) have been associated with the need for carefully calibrated dosage to prevent bleeding episodes.

The influence of *VKORC1* polymorphisms on vitamin K antagonist dose requirements provides a remarkable example of pharmacogenomic diversity worldwide; this is documented by the International Warfarin Pharmacogenomic Consortium (IWPC) datasets, comprising 5700–6200 patients recruited from four continents and ascribed to three 'racial' groups, namely Asians, Blacks (mainly African Americans) and Whites [27].

Furthermore, considering the 19 HLA alleles associated with adverse events to the therapy, with highest level of evidence, mention should be made of HLA-B*58:01, which has been shown to have a strong effect on the development of severe cutaneous adverse reactions

(SCARs) including Stevens - Johnson syndrome and toxic epidermal necrolysis after treatment with allopurinol, the common treatment for hyperuricemia and gout. However, the frequency of HLA-B*58:01 significantly differs between different ethnic groups. The frequency of HLA-B*58:01 is the highest in Han-Chinese (20%), Korean (12%), and Thai (13%), but is much less frequent in Japanese (0.1%) The same allele, however, is also much more frequent in Sardinians (11%) than in other European populations (France 1.5%) [28]. We believe that this evidence is of particular relevance, given that Sardinia has the highest percentage of reports of adverse events [22] following allopurinol administration of the total number of adverse reports registered in Italy (1.9% compared to an average of 0.41% in the other Italian regions).

Among the most differentiated variants with lower levels of evidence, two independent variants rs3815087 (allele A) and rs3131003 (allele A) located in *PSORS1C1* region, are of particular interest: both variants, highly frequent in Sardinians compared to European populations (delta frequency > 29%), have been associated with epidermal necrolysis and Stevens-Johnson syndrome [29,30] after allopurinol therapy (evidence levels 3 and 4, respectively) and show coincident, strong association with psoriasis ($p=1.2E-294$, $OR=2.93$; $p=1.4E-105$, $OR=1.64$) [<https://genetics.opentargets.org>; rs3815087 and rs3131003 variants respectively]. They are very common in Sardinia (AR 50% and 74.6%), and thus screening for these variants before therapy is important. It is thus not surprising that the variant rs2233945, localized in the same *PSORS1C1* gene modulates the response to etanercept, a TNF inhibitor used for psoriasis and other autoimmune disorders including rheumatoid arthritis. Allele A in that locus has been associated with increased etanercept efficacy in comparison to allele C: and at the same time allele A has been associated with protection from psoriasis. This variant has been in linkage disequilibrium with one canonically described for allopurinol adverse events, rs9263726, the variant tag for HLA-B*58:01.

The impact of pharmacogenetic variation on health is thus patent, and it may be useful to consider that the impact falls disproportionally on the elderly. The effects of aging result first from 1) the relatively high usage of pharmaceutical drugs in the elderly (<https://www.kff.org/health-reform/issue-brief/data-note-prescription-drugs-and-older-adults/>; <https://hpi.georgetown.edu/rxdrugs/>) and 2) the relatively greater sensitivity to drugs in the elderly, so that the dosages, which are determined on younger adults, are often excessive for them, increasing the risk of side effects.

But variants in pharmacogenes can then further exacerbate the problems. Concerning germline mutations, we have dealt here with the effect of single variants on single drugs, but we know little about the effect of genetics on combinations of multiple drugs. This is again especially important for the elderly, who - compared to the middle-aged individuals who normally participate in clinical trials - are more frequently simultaneously prescribed multiple drugs (often in a 'therapeutic cascade', in which the side effects of one drug are treated with another drug). Correspondingly, additive or synergistic incidence of side effects and intensification of genetic variant effects can be expected. Furthermore, with age each individual accumulates new somatic mutations, and the liver -- in which the majority of the genes relevant to ADME-Tox are expressed -- is one of the tissues most exposed to environmental mutagens. It therefore tends to accumulate somatic mutations that can potentially further alter the function of pharmacogenes [31,32].

although the extent to which this occurs has not been quantified. Further analyses should define more precisely the relative load of genetic risk as a function of age and the measures -- like lower doses and substitution of drugs by others with a different genetic risk profile - that may mitigate it.

There are some limitations to this study that could be met in the future. First, only by using high coverage sequencing data on the exome or genome it will be possible to define with certainty the existence and prevalence of rarer variants with important effects specific to the Sardinian population. Second, we did not assess structural variants, but for example, *CYP2D6* is well-known to have structural variants including copy number variability and gene rearrangements between *CYP2D7*-*CYP2D6* known as *hybrid tandems*. Two other limitations currently preclude estimation of the potential economic cost of non-stratification of patients based on genetic characteristics: the absence of personal data on drug prescriptions (which would allow us to understand how many and which individuals are really at risk of adverse reactions to the drugs they use) and global data on consumption in Sardinia (which would allow us to make pharmaco-economic estimates). Nevertheless, this is to date the largest study of pharmacogenetic allele and phenotype frequencies in Sardinia - and in fact in Italy as a whole. The findings demonstrate the value of characterizing allele frequencies in diverse populations and highlights the need for more PGx research on understudied populations, an important step in the corresponding refined implementation of modern personalized medicine.

Supplementary Materials: Supplementary materials can be found at To be added

Author Contributions: Conceptualization: M.F.; Methodology: F.C., M.F.; Software: M.F.; Validation: M.L.I., M.Z., S.A.M.U., A.B.; Data curation: M.L.I., M.Z., A.B., V.N., V.L.; Resources: S.O., S.A.M.U.; Writing - Original Draft: M.L.I., M.Z., M.F.; Writing - Review & Editing: M.F., M.L.I., M.Z., G.M., D.S., F.C., V.L.

Funding: The Authors acknowledge funding from *University of Sassari* and *Fondazione di Sardegna* (grants “fondo di Ateneo per la ricerca 2019”, “fondo di Ateneo per la ricerca 2020”, and “Bando competitivo Fondazione di Sardegna - 2016 per progetti di ricerca con revisione tra pari”).

Acknowledgments: We thank Dr. Maristella Steri (IRGB-CNR) and Dr. Michele Marongiu (IRGB-CNR) for technical help and advice.

Conflict of Interest: The authors declared no competing interests for this work.

References

- [1] Bouvy JC, De Bruin ML, Koopmanschap MA. Epidemiology of adverse drug reactions in Europe: a review of recent observational studies. *Drug Saf* 2015;38:437–53. <https://doi.org/10.1007/s40264-015-0281-0>.
- [2] Pirmohamed M, James S, Meakin S, Green C, Scott AK, Walley TJ, et al. Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. *BMJ* 2004;329:15–9. <https://doi.org/10.1136/bmj.329.7456.15>.
- [3] Santos RV, da Silva GO, Gibbon S. Pharmacogenomics, human genetic diversity and the incorporation and rejection of color/race in Brazil. *Biosocieties* 2015;10:48–69. <https://doi.org/10.1057/biosoc.2014.21>.
- [4] Auwerx C, Sadler MC, Reymond A, Kotalik Z. From pharmacogenetics to pharmacogenomics: Milestones and future directions. *HGG Adv* 2022;3:100100. <https://doi.org/10.1016/j.xhgg.2022.100100>.
- [5] Gaedigk A, Riffel AK, Leeder JS. CYP2D6 Haplotype Determination Using Long Range Allele-Specific Amplification: Resolution of a Complex Genotype and a Discordant Genotype Involving the CYP2D6*59 Allele. *J Mol Diagn* 2015;17:740–8. <https://doi.org/10.1016/j.jmoldx.2015.06.007>.
- [6] Sim SC, Altman RB, Ingelman-Sundberg M. Databases in the area of pharmacogenetics. *Hum Mutat* 2011;32:526–31. <https://doi.org/10.1002/humu.21454>.
- [7] Badagnani I, Sorani M, Edwards RH, Brown C, Castro RA, Huang CC, et al. PharmGKB submission update: VI. PMT submissions of genetic variations in neurotransmitter transporters (SLC6, SLC17, and SLC18) to the PharmGKB network. *Pharmacol Rev* 2006;58:5–6. <https://doi.org/10.1124/pr.58.1.3>.
- [8] Sahana S, Bhoyar RC, Sivadas A, Jain A, Imran M, Rophina M, et al. Pharmacogenomic landscape of Indian population using whole genomes. *Clin Transl Sci* 2022;15:866–77. <https://doi.org/10.1111/cts.13153>.

- [9] Sivadas A, Scaria V. Pharmacogenomic survey of Qatari populations using whole-genome and exome sequences. *Pharmacogenomics J* 2018;18:590–600. <https://doi.org/10.1038/s41397-018-0022-8>.
- [10] Marcus JH, Posth C, Ringbauer H, Lai L, Skeates R, Sidore C, et al. Genetic history from the Middle Neolithic to present on the Mediterranean island of Sardinia. *Nat Commun* 2020;11:939. <https://doi.org/10.1038/s41467-020-14523-6>.
- [11] Chiang CWK, Marcus JH, Sidore C, Biddanda A, Al-Asadi H, Zoledziwska M, et al. Genomic history of the Sardinian population. *Nat Genet* 2018;50:1426–34. <https://doi.org/10.1038/s41588-018-0215-8>.
- [12] Zoledziwska M, Sidore C, Chiang CWK, Sanna S, Mulas A, Steri M, et al. Height-reducing variants and selection for short stature in Sardinia. *Nat Genet* 2015;47:1352–6. <https://doi.org/10.1038/ng.3403>.
- [13] Danjou F, Zoledziwska M, Sidore C, Steri M, Busonero F, Maschio A, et al. Genome-wide association analyses based on whole-genome sequencing in Sardinia provide insights into regulation of hemoglobin levels. *Nat Genet* 2015;47:1264–71. <https://doi.org/10.1038/ng.3307>.
- [14] Sidore C, Busonero F, Maschio A, Porcu E, Naitza S, Zoledziwska M, et al. Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. *Nat Genet* 2015;47:1272–81. <https://doi.org/10.1038/ng.3368>.
- [15] Steri M, Orrù V, Idda ML, Pitzalis M, Pala M, Zara I, et al. Overexpression of the Cytokine BAFF and Autoimmunity Risk. *N Engl J Med* 2017;376:1615–26. <https://doi.org/10.1056/NEJMoa1610528>.
- [16] McInnes G, Lavertu A, Sangkuhl K, Klein TE, Whirl-Carrillo M, Altman RB. Pharmacogenetics at Scale: An Analysis of the UK Biobank. *Clin Pharmacol Ther* 2021;109:1528–37. <https://doi.org/10.1002/cpt.2122>.
- [17] Sangkuhl K, Whirl-Carrillo M, Whaley RM, Woon M, Lavertu A, Altman RB, et al. Pharmacogenomics Clinical Annotation Tool (PharmCAT). *Clin Pharmacol Ther* 2020;107:203–10. <https://doi.org/10.1002/cpt.1568>.
- [18] Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581:434–43. <https://doi.org/10.1038/s41586-020-2308-7>.
- [19] Whirl-Carrillo M, Huddart R, Gong L, Sangkuhl K, Thorn CF, Whaley R, et al. An Evidence-Based Framework for Evaluating Pharmacogenomics Knowledge for Personalized Medicine. *Clin Pharmacol Ther* 2021;110:563–72. <https://doi.org/10.1002/cpt.2350>.
- [20] Johnson JA, Caudle KE, Gong L, Whirl-Carrillo M, Stein CM, Scott SA, et al. Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Pharmacogenetics-Guided Warfarin Dosing: 2017 Update. *Clin Pharmacol Ther* 2017;102:397–404. <https://doi.org/10.1002/cpt.668>.
- [21] Theken KN, Lee CR, Gong L, Caudle KE, Formea CM, Gaedigk A, et al. Clinical Pharmacogenetics Implementation Consortium Guideline (CPIC) for CYP2C9 and Nonsteroidal Anti-Inflammatory Drugs. *Clin Pharmacol Ther* 2020;108:191–200. <https://doi.org/10.1002/cpt.1830>.
- [22] Zito S, Fortinguerra F, Pierantozzi A, Ambrosino F, Traversa G, Trotta F, et al. [Medicines use in Italy. 2020 OsMed National Report.]. *Recenti Prog Med* 2021;112:659–67. <https://doi.org/10.1701/3679.36657>.
- [23] Cooper-DeHoff RM, Niemi M, Ramsey LB, Luzum JA, Tarkiainen EK, Straka RJ, et al. The Clinical Pharmacogenetics Implementation Consortium Guideline for SLCO1B1,

- ABCG2, and CYP2C9 genotypes and Statin-Associated Musculoskeletal Symptoms. *Clin Pharmacol Ther* 2022;111:1007–21. <https://doi.org/10.1002/cpt.2557>.
- [24] Lee CR, Luzum JA, Sangkuhl K, Gammal RS, Sabatine MS, Stein CM, et al. Clinical Pharmacogenetics Implementation Consortium Guideline for CYP2C19 Genotype and Clopidogrel Therapy: 2022 Update. *Clin Pharmacol Ther* 2022. <https://doi.org/10.1002/cpt.2526>.
- [25] Quartuccio L, Fabris M, Pontarini E, Salvin S, Zabotti A, Benucci M, et al. The 158VV Fcγ receptor 3A genotype is associated with response to rituximab in rheumatoid arthritis: results of an Italian multicentre study. *Ann Rheum Dis* 2014;73:716–21. <https://doi.org/10.1136/annrheumdis-2012-202435>.
- [26] Rishavy MA, Hallgren KW, Wilson L, Singh S, Runge KW, Berkner KL. Warfarin alters vitamin K metabolism: a surprising mechanism of VKORC1 uncoupling necessitates an additional reductase. *Blood* 2018;131:2826–35. <https://doi.org/10.1182/blood-2017-09-804666>.
- [27] Limdi NA, Wadelius M, Cavallari L, Eriksson N, Crawford DC, Lee M-TM, et al. Warfarin pharmacogenetics: a single VKORC1 polymorphism is predictive of dose across 3 racial groups. *Blood* 2010;115:3827–34. <https://doi.org/10.1182/blood-2009-12-255992>.
- [28] Hershfield MS, Callaghan JT, Tassaneeyakul W, Mushiroda T, Thorn CF, Klein TE, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for human leukocyte antigen-B genotype and allopurinol dosing. *Clin Pharmacol Ther* 2013;93:153–8. <https://doi.org/10.1038/clpt.2012.209>.
- [29] Génin E, Schumacher M, Roujeau J-C, Naldi L, Liss Y, Kazma R, et al. Genome-wide association study of Stevens-Johnson Syndrome and Toxic Epidermal Necrolysis in Europe. *Orphanet J Rare Dis* 2011;6:52. <https://doi.org/10.1186/1750-1172-6-52>.
- [30] Cheng L, Sun B, Xiong Y, Hu L, Gao L, Lv Q, et al. The minor alleles HCP5 rs3099844 A and PSORS1C1 rs3131003 G are associated with allopurinol-induced severe cutaneous adverse reactions in Han Chinese: a multicentre retrospective case-control clinical study. *Br J Dermatol* 2018;178:e191–3. <https://doi.org/10.1111/bjd.16151>.
- [31] García-Nieto PE, Morrison AJ, Fraser HB. The somatic mutation landscape of the human body. *Genome Biol* 2019;20:298. <https://doi.org/10.1186/s13059-019-1919-5>.
- [32] Brazhnik K, Sun S, Alani O, Kinkhabwala M, Wolkoff AW, Maslov AY, et al. Single-cell analysis reveals different age-related somatic mutation profiles between stem and differentiated cells in human liver. *Sci Adv* 2020;6:eaax2659. <https://doi.org/10.1126/sciadv.aax2659>.