High frequency and founder effect of the *CYP3A4*20* loss-of-function allele in the Spanish population classifies CYP3A4 as a polymorphic enzyme

Authors: María Apellániz-Ruiz[1], Lucia Inglada-Pérez[1,2] Maria Eugenia G. Naranjo[3], Lara Sánchez[1], Veronika Mancikova[1], María Currás-Freixes[1], Aguirre A de Cubas[1], Iñaki Comino-Méndez[1], Soumaya Triki[4], Ahmed Rebai (PhD)[4], Mahmood Rasool (PhD)[5], Graciela Moya (MD)[6], Manuela Grazina (PhD)[7], Guiseppe Opocher (MD)[8], Alberto Cascón (PhD)[1,2], Patricia Taboada-Echalar (PhD)[9], Magnus Ingelman-Sundberg (PhD)[10], Angel Carracedo (MD/PhD)[5,9,11], Mercedes Robledo (PhD)[1,2], Adrián Llerena (MD/PhD)[3], Cristina Rodríguez-Antona (PhD)[1,2]

Affiliations:

[1]Hereditary Endocrine Cancer Group, Spanish National Cancer Research Centre (CNIO), Madrid, Spain.

[2] ISCIII Center for Biomedical Research on Rare Diseases (CIBERER), Madrid, Spain.

[3]CICAB Clinical Research Centre at Extremadura University Hospital and Medical School, Badajoz, Spain.

[4]Research Group on Molecular and Cellular Screening Processes, Laboratory of Microorganisms and Biomolecules, Centre of Biotechnology of Sfax, Tunisia.

[5] Center of Excellence in Genomic Medicine Research, King Abdulaziz University, Jeddah, KSA.

[6]Pontificia Universidad Católica Argentina and Genos laboratory, Buenos Aires, Argentina.

[7]Faculty of Medicine CNC-Centre for Neuroscience & Cell Biology, University of Coimbra, Portugal.

[8] Familial Cancer Clinic and Oncoendocrinology, Veneto Institute of Oncology, Padova, Italy

[9]Unidade de Xenética, Instituto de Ciencias Forenses and Departamento de Anatomía Patolóxica e Ciencias Forenses, Facultade de Medicina, Universidade de Santiago de Compostela, Galicia, Spain

[10]Section of Pharmacogenetics, Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden.

[11]Fundación Pública de Medicina Xenómica-SERGAS. Grupo de Medicina Xenómica, CIBERER, IDIS, Santiago de Compostela, Spain.

Corresponding author:

Dr. Cristina Rodríguez-Antona, Spanish National Cancer Research Center (CNIO), Madrid, Spain. Ph. +34 917328000; Fax. +34 912246972; crodriguez@cnio.es

Running title: Worldwide distribution of *CYP3A4*20* allele

ABSTRACT

Cytochrome P450 3A4 (CYP3A4) is a key drug metabolizing enzyme. Loss-of-function variants have been reported as rare events, and the only demonstration of a CYP3A4 protein lacking functional activity is caused by *CYP3A4*20* allele. Here we characterized the world distribution and origin of *CYP3A4*20* mutation. *CYP3A4*20* was determined in more than 4000 individuals representing different populations and haplotype analysis was performed using *CYP3A* polymorphisms and microsatellite markers. *CYP3A4*20* allele was present in 1.2% of the Spanish population (up to 3.8% in specific regions), and all *CYP3A4*20* carriers had a common haplotype. This is compatible with a Spanish founder effect and classifies CYP3A4 as a polymorphic enzyme. This constitutes the first description of a *CYP3A4* loss-of-function variant with high frequency in a population. *CYP3A4*20* results together with the key role of CYP3A4 in drug metabolism, support screening for rare *CYP3A4* functional alleles among subjects with adverse drug events in certain populations.

Keywords: *CYP3A4*20* allele, loss-of-function allele, rare variant, founder effect, *CYP3A4* haplotype

INTRODUCTION

Cytochromes P450 (CYPs) are the most important drug metabolizing enzymes, being CYP1, 2 and 3 families responsible of the biotransformation of ~70-80% of all therapeutic compounds [1, 2]. CYP3A4 is the most abundant P450 enzyme in the human liver and gastrointestinal tract and it is involved in the biotransformation of more than half of all clinically used drugs [1-3]. There is a high variability in CYP3A4 expression (> 100 fold) [1] caused by non-genetic and genetic factors which contributes to unpredictable drug responses and toxicities. Among environmental factors, drug-drug interactions are one of the most studied causes of variation, but gender, hormonal status and age also influence CYP3A4 expression and activity [4, 5]. With respect to genetic factors, twin studies and repeated drug administration approaches have estimated a high degree of heritability in CYP3A4 inter-individual variation [6-8]. In this regard, the Human CYP Allele Nomenclature Database includes 23 different CYP3A4 variant proteins. Two are truncated proteins resulting from rare premature stop codons (*CYP3A4*6* and *CYP3A4**20 alleles) [9, 10], while the rest are low frequency/ rare missense variants, some with reduced enzymatic activity (*CYP3A4*8, *11,*12,*13,*16* and**17* alleles; http://www.cypalleles.ki.se/). In addition, two non-coding SNPs have been related to altered gene expression (e.g. *CYP3A4*1B* and *CYP3A4*22*) [11, 12]. On the whole, although several genetic variants that affect CYP3A4 activity have been described, in general these are rare or low-frequency alleles expected to explain only a small fraction of CYP3A4 phenotypic variability [13].

The *CYP3A4*20* allele, a single base pair (A) insertion causing a frameshift and premature stop codon in the protein (c.1461_1462insA; p.P488Tfs*494), is the only *CYP3A4* gene variant in which lack of enzymatic activity has been demonstrated. This allele was found in an individual of Brazilian descent with a 6-fold increased exposure of a drug metabolized by CYP3A4 and low systemic midazolam clearance [9], and classified as rare, since no

*CYP3A4*20* carriers were found in 428 Germans. This variant has not been described again, and there is no data regarding its possible origin and distribution in different populations. However, we found one *CYP3A4*20* carrier upon whole exome sequencing of Spanish patients with extreme toxicity [14]. This finding triggered this study investigating the *CYP3A4*20* allele distribution in Spain and worldwide. We find that 1 in 82 Spanish individuals carries this allele, and haplotype analyses suggest a Spanish founder effect. This is the first description of a loss-of-function *CYP3A4* allele with a high frequency in a population, and demonstrates the polymorphic nature in this gene.

MATERIALS AND METHODS

DNA from control individuals

DNA isolated from blood samples (FlexiGene DNA Kit, Qiagen) of 1977 Spanish, 450 Portuguese, 478 Italian, 240 Argentinean, 179 Bolivian, 29 Algerian, 95 Libyan, 117 Israeli, 133 Saudi Arabian, 83 Kuwaiti, 186 Pakistani and 108 Chinese controls were collected. All individuals were over 18 years, the collection of samples was approved by local internal ethical review committees and investigations were conducted according to the principles expressed in the Declaration of Helsinki.

Genotyping of SNPs in *CYP3A* genes

In addition to *CYP3A4*20* (rs67666821), the *CYP3A* functional SNPs *CYP3A4*22* (rs35599367), *CYP3A4*1B* (rs2740574), *CYP3A5*3* (rs776746) and *CYP3A7*2* (rs2257401) alleles were selected for genotyping using the KASPar SNP Genotyping System (LGC Genomics, UK) with 15 ng of genomic DNA. All assays included DNA samples with known genotypes and negative controls. The Sequence Detection System ABI PRISM® 7900HT (Applied Biosystems) was employed for fluorescence detection and allele assignment. The accuracy of the genotyping was confirmed by sequencing all *CYP3A4*20* carriers, and a random selection of individuals with different *CYP3A4*22, CYP3A4*1B, CYP3A5*3* and *CYP3A7*2* genotypes. PCR products were purified and Sanger sequencing run on an ABI PRISM 3700 DNA Analyzer capillary sequencer (Applied Biosystems).

Microsatellites markers

For haplotype analysis a panel of four microsatellite markers on chromosome 7q21-22 spanning an interval of 3.2Mb was used: D7S651, D7S2498, D7S2480 and D7S666 (Suppl. Fig. 1). In brief, PCR was carried out using specific primers, and with the forward primers

labeled with 6-Fam fluorochrome. The diluted PCR products were mixed with Hi-Di Formamide and LIZ-500 size standard (Applied Biosystems), separated and detected using an ABI Prism 3100 automatic sequencer (Applied Biosystems), and analyzed by Peak Scanner software version 1 (Applied Biosystems).

Haplotype analysis and dating the origin of *CYP3A4*20* allele

Haplotypes were identified using SNPs (*CYP3A4*20*, *CYP3A4*22*, *CYP3A4*1B*, *CYP3A5*3* and *CYP3A7*2*) and the 4 microsatellite markers in 20 *CYP3A4*20* carriers (19 Spanish and 1 Portuguese) and in 50 Spanish individuals wild type for this variant, using PHASE software [15].

The mutation origin of *CYP3A4*20* variant was calculated using DMLE+ software version 2.3 developed by Reeve and Rannala [16] (http://dmle.org/). This program uses the Markov Chain Monte Carlo algorithm to allow Bayesian estimation of the mutation age based on: the observed haplotypes in variant carriers and unrelated normal individuals, map distances between markers, the position of the mutation relative to the markers and the estimated population growth rate.

Statistical analysis

Hardy-Weinberg equilibrium was tested for the SNPs genotyped and none significantly deviated from expected values [17]. Fisher's exact test was used to examine the association between *CYP3A4*20* allele frequency and country of origin. P values below 0.05 were considered statistically significant. SPSS v19 was used for the statistical analysis.

RESULTS

*CYP3A4*20* allele distribution in different populations

The Exome Variant Server database (EVS) (http://evs.gs.washington.edu/EVS/) suggested that the loss-of-function *CYP3A4*20* allele is rare but detectable in some individuals, mainly of European origin, and our finding of one carrier in a Spanish individual, led us to carry a *CYP3A4*20* allele frequency population study in individuals from European, African and Asian descent (Table 1; Fig.1). In concordance with previous data, no *CYP3A4*20* carriers were found in Italians, Argentineans, Bolivians and individuals from different countries in Africa and Asia. However, the *CYP3A4*20* variant was detected in heterozygosity in 24 Spanish individuals and in 1 Portuguese, revealing that 1.2% and 0.2%, respectively, of these populations carried the variant (Table 1). The unexpected high number of *CYP3A4*20* allele carriers in the Spanish individuals was significantly different from the other populations studied (Fisher exact test $P<0.0001$).

Characterization of *CYP3A4*20* allele frequency in different Spanish regions

To determine *CYP3A4*20* allele distribution within Spain, we collected the place of birth of 1544 individuals among the 1953 genotyped Spanish controls. When comparing the *CYP3A4*20* allele frequency with the region of birth, we found that the variant had the highest frequency in individuals from Castilla y León, Comunidad Valenciana and Extremadura, where we found one heterozygous every 26, 33 and 48 individuals, respectively, revealing that 3.8, 3.0 and 2.1% of these populations carry *CYP3A4*20* (Fig. 1; Suppl. Table 1). In other Spanish regions, the proportion of variant carriers ranged from 1.6 to 0.8%, with the exception of Galicia, where no variant carriers were found (Fig. 1; Suppl. Table 1).

*CYP3A4*20* ancestral haplotype

To investigate whether all occurrences of *CYP3A4*20* allele descended from a single ancestral mutation event or arisen independently, we constructed haplotypes with *CYP3A4*20*, 4 functional SNPs in *CYP3A* locus and 4 microsatellite markers (Suppl. Fig. 1). Haplotype reconstruction in 20 *CYP3A4*20* allele carriers, suggested that all carriers showed a common haplotype (282, 172, A, C, InsA, C, G) that contained this variant and spanned ~700 kb (from microsatellite D7S2480 to SNP *CYP3A5*3*; Fig. 2 and Suppl. Fig. 2). The *CYP3A4*20* haplotype contained wild type alleles for *CYP3A4*22*, *CYP3A4*1B* and *CYP3A7*2*, and carried *CYP3A5*3*, the most common variant in Caucasians. In 50 individuals wild type for *CYP3A4*20*, representing the control Spanish population, 27 different haplotypes existed with frequencies ranging from 18 to 1% (Fig. 2). Furthermore, 4 out of the 100 chromosomes analyzed were predicted to carry the same haplotype as *CYP3A4*20* but without this mutation (282, 172, A, C, -, C, G). Thus, this result suggests a single ancestral allele in which the variant was likely originated.

Age of *CYP3A4*20* variant

The decay of linkage disequilibrium due to recombination can be used to date the age of a mutation. We used the DMLE+ software to estimate the age of *CYP3A4*20* variant using the haplotype data from the 20 CYP3A4*20 carriers and 50 Spanish controls previously studied. The mutation age was estimated to be 51 generations (95% credible interval of 43-60) using an average growth rate of 0.25. Assuming 20 years for a generation, the age of the variant was estimated to be 1020 years old. For growth rates of 0.15 and 0.35, mutation age was estimated to be 82 and 38 generations, respectively.

DISCUSSION

In contrast to the high polymorphic nature of most drug metabolizing enzymes, *CYP3A4* gene exhibits little genetic variability. The Human CYP Allele Nomenclature Database (http://www.cypalleles.ki.se/) includes only two loss-of-function *CYP3A4* alleles (*CYP3A4*6* and *CYP3A4*20*) and the EVS database (http://evs.gs.washington.edu/EVS/) suggests that only 0.2% of Americans carry *CYP3A4* defective variants and 2% missense variants, many of which have unknown functional significance. The *CYP3A4*20* allele, that encodes a truncated protein devoid of catalytic activity [9], is the most common *CYP3A4* defective allele in the EVS database with 0.1% and 0.02% of European Americans and African Americans carriers, respectively (i.e. 8 carriers out of 4127 and 1 carrier out of 2132 individuals), while it was not detected in 428 German individuals [9]. In the present study we found that 1.2% of the Spanish population (24 out of 1977 individuals) carry the *CYP3A4*20* allele, compared with 0.2% in Portugal (1 out of 450) and no carriers in Italy, Argentina, Bolivia, Libya, Algeria, Israel, Kuwait, Saudi Arabia, Pakistan and China (Fig. 1). On the whole, one in 82 Spanish carried this variant. Within Spain, this figure increased to one *CYP3A4*20* carrier every 26 individuals in Castilla y León, and one in 33 in Comunidad Valenciana and one in 48 individuals in Extremadura. These results constitute the first proof that *CYP3A4* loss-of-function alleles can be classified as polymorphisms (i.e. with allele frequencies above 1%) and affect a large number of individuals, in specific populations.

Haplotype analysis suggested that *CYP3A4*20* appeared in a haplotype present in only 4% of chromosomes of the Spanish population, and containing the most common Caucasian variants for the functional *CYP3A* SNPs genotyped (i.e. wild type *CYP3A4*22, CYP3A4*1B* and *CYP3A7*2* and variant *CYP3A5*3* allele, see Fig. 2) [18]. The highest frequency of *CYP3A4*20* in Spain and the infrequent detection outside the Spanish peninsula, together with a 700 Kb haplotype common to all mutation carriers (Suppl. Fig. 2), suggests a recent

occurrence of the mutation. In agreement with this, dating of *CYP3A4*20* mutation, suggested that it appeared about 1000 years ago. Altogether this data is compatible with a single origin of the mutation in Spain, and then spreading to different geographical areas in recent times.

CYP3A4 plays a prominent role in the biotransformation of a broad range of xenobiotics, including many clinical drugs [13], and contributes to the metabolism of endogenous substrates such as vitamine D3, arachidonic acid, bile acids and steroid hormones [2, 19]. It has been suggested that the *CYP3A4* gene allows little variation due to this fact and indeed no single individual being homozygous for defect *CYP3A4* alleles has been described. However, with the exception of drug metabolism impairment, no other major phenotype could be detected in *Cyp3a* knock-out mice [20, 21]. Alternative enzymes may have compensated the effect on the metabolism of endogenous compounds, while the prominent decrease in xenotibitics biotransformation would only manifest after xenobiotics' exposure. By contrast in mice transgenic for *CYP3A4* the females were found to be deficient in lactation, leading to a markedly lower pup survival and the mammary glands of the Tg-CYP3A4 lactating mothers had underdeveloped alveoli with low milk content [22]. Because of the absence of a null phenotype in mice, the small number of individuals expected to be homozygous for *CYP3A4*20* (i.e. one in 4100 individuals in Castilla y León) might not have any clinical manifestation, although a very severe toxicity profile would be expected when exposed to drugs metabolized by this enzyme. *CYP3A4*20* heterozygous carriers, with decreased CYP3A4 activity, may not show an effect when treated with single doses of wide therapeutic index drugs, but may show altered response upon treatment with narrow therapeutic index drugs. This is supported by Westlind-Johnsson et al. [9] that described a 6-fold higher exposure to a drug metabolized by CYP3A4 and low systemic midazolam clearance in an individual heterozygous for *CYP3A4*20* allele [9] and by a 7-fold higher risk of paclitaxel dose reductions due to peripheral neuropathy in *CYP3A4*20* carriers as described by us [14]. We also found that

*CYP3A4*20* is independent of *CYP3A4*22*, an intronic polymorphism robustly associated with a decreased elimination of CYP3A4 substrates and carried by about 5-7 % of Caucasians [12, 23]. Thus, a highly reduced CYP3A4 activity would be expected in individuals carrying both of these two alleles.

In conclusion, this is the first demonstration of a polymorphic nature of *CYP3A4* gene, with 1.2% of Spanish individuals carrying *CYP3A4*20* allele, likely due to a founder effect. Furthermore, the key role of CYP3A4 in drug metabolism and preliminary clinical evidences support an increased risk of unexpected drug responses in *CYP3A4*20* carriers and suggest the importance of implementing *CYP3A4*20* genotyping in the clinic, at least in the Spanish population.

FUNDING

This work was supported by projects from the Spanish Ministry of Economy and Competiveness (grant number SAF2012–35779). Government of Extremadura-AEXCID (13/A001), the RIBEF IberoAmerican Network of Pharmacogenetics and SIFF (www.ribef.com). María Apellániz and Veronika Mancikova are predoctoral fellows of "la Caixa"/ CNIO international PhD programme. Lucía Inglada-Pérez is supported by CIBERER. Maria Currás is a predoctoral fellow supported by Severo Ochoa. Aguirre de Cubas is supported by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 259735.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed by the other authors.

Supplementary information

Supplementary information is available at The Pharmacogenomics Journal's website.

REFERENCES

1. Zanger UM, Schwab M. Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol Ther* 2013; **138:** 103-141.

2. Nebert DW, Russell DW. Clinical importance of the cytochromes P450. *Lancet* 2002; **360:** 1155-1162.

3. Schuetz EG. Lessons from the CYP3A4 promoter. *Mol Pharmacol* 2004; **65:** 279-281.

4. Haddad A, Davis M, Lagman R. The pharmacological importance of cytochrome CYP3A4 in the palliation of symptoms: review and recommendations for avoiding adverse drug interactions. *Support Care Cancer* 2007; **15:** 251-257.

5. Badyal DK, Dadhich AP. Cytochrome P450 and drug interactions. *Indian Journal of Pharmacology* 2001**:** 248-259.

6. Rahmioglu N, Heaton J, Clement G, Gill R, Surdulescu G, Zlobecka K, *et al*. Genetic epidemiology of induced CYP3A4 activity. *Pharmacogenet Genomics* 2011; **21:** 642-651.

7. Ozdemir V, Kalow W, Tang BK, Paterson AD, Walker SE, Endrenyi L, *et al*. Evaluation of the genetic component of variability in CYP3A4 activity: a repeated drug administration method. *Pharmacogenetics* 2000; **10:** 373-388.

8. Penno MB, Dvorchik BH, Vesell ES. Genetic variation in rates of antipyrine metabolite formation: a study in uninduced twins. *Proc Natl Acad Sci U S A* 1981; **78:** 5193-5196.

9. Westlind-Johnsson A, Hermann R, Huennemeyer A, Hauns B, Lahu G, Nassr N, *et al*. Identification and characterization of CYP3A4*20, a novel rare CYP3A4 allele without functional activity. *Clin Pharmacol Ther* 2006; **79:** 339-349.

10. Hsieh KP, Lin YY, Cheng CL, Lai ML, Lin MS, Siest JP, *et al*. Novel mutations of CYP3A4 in Chinese. *Drug Metab Dispos* 2001; **29:** 268-273.

11. Rodriguez-Antona C, Sayi JG, Gustafsson LL, Bertilsson L, Ingelman-Sundberg M. Phenotype-genotype variability in the human CYP3A locus as assessed by the probe drug quinine and analyses of variant CYP3A4 alleles. *Biochem Biophys Res Commun* 2005; **338:** 299-305.

12. Wang D, Guo Y, Wrighton SA, Cooke GE, Sadee W. Intronic polymorphism in CYP3A4 affects hepatic expression and response to statin drugs. *Pharmacogenomics J* 2011; **11:** 274-286.

13. Klein K, Zanger UM. Pharmacogenomics of Cytochrome P450 3A4: Recent Progress Toward the "Missing Heritability" Problem. *Front Genet* 2013; **4:** 12.

14. Apellániz-Ruiz M, Lee M, Sánchez L, Gutiérrez-Gutiérrez G, Calvo I, García-Estévez L, *et al*. Whole-exome sequencing reveals defective *CYP3A4* variants predictive of paclitaxel dose-limiting neuropathy**.** *(Submitted)* 2014.

15. Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 2003; **73:** 1162-1169.

16. Reeve JP, Rannala B. DMLE+: Bayesian linkage disequilibrium gene mapping. *Bioinformatics* 2002; **18:** 894-895.

17. Rodriguez S, Gaunt TR, Day IN. Hardy-Weinberg equilibrium testing of biological ascertainment for Mendelian randomization studies. *Am J Epidemiol* 2009; **169:** 505-514.

18. Ingelman-Sundberg M, Sim SC, Gomez A, Rodriguez-Antona C. Influence of cytochrome P450 polymorphisms on drug therapies: pharmacogenetic, pharmacoepigenetic and clinical aspects. *Pharmacol Ther* 2007; **116:** 496-526.

19. Nakamura H, Nakasa H, Ishii I, Ariyoshi N, Igarashi T, Ohmori S, *et al*. Effects of endogenous steroids on CYP3A4-mediated drug metabolism by human liver microsomes. *Drug Metab Dispos* 2002; **30:** 534-540.

20. van Herwaarden AE, Wagenaar E, van der Kruijssen CM, van Waterschoot RA, Smit JW, Song JY, *et al*. Knockout of cytochrome P450 3A yields new mouse models for understanding xenobiotic metabolism. *J Clin Invest* 2007; **117:** 3583-3592.

21. Hashimoto M, Kobayashi K, Watanabe M, Kazuki Y, Takehara S, Inaba A, *et al*. Knockout of mouse Cyp3a gene enhances synthesis of cholesterol and bile acid in the liver. *J Lipid Res* 2013; **54:** 2060-2068.

22. Yu AM, Fukamachi K, Krausz KW, Cheung C, Gonzalez FJ. Potential role for human cytochrome P450 3A4 in estradiol homeostasis. *Endocrinology* 2005; **146:** 2911-2919.

23. Elens L, van Gelder T, Hesselink DA, Haufroid V, van Schaik RH. CYP3A4*22: promising newly identified CYP3A4 variant allele for personalizing pharmacotherapy. *Pharmacogenomics* 2013; **14:** 47-62.

FIGURE LEGENDS

Figure 1. Geographical distribution of *CYP3A4*20* allele.

World map showing the percentage of *CYP3A4*20* allele carriers in different populations. The Spanish peninsula is shown in greater detail. Number of carriers and individuals studied are presented in Table 1 and in Suppl. Table 1. Concerning USA, the data corresponds to the EVS, the first number refers to African-Americans and the second to European-Americans.

Figure 2. *CYP3A4*20* haplotype analysis.

Haplotypes were identified using PHASE in 20 *CYP3A4*20* carriers and 50 Spanish control individuals with 5 SNPs (*CYP3A4*1B, CYP3A4*22, CYP3A4*20, CYP3A7*2,* and *CYP3A5*3*) and 4 microsatellite markers at 7q (D7S666, D7S2480, D7S2498 and D7S651), spanning an interval of 3.2Mb.

TABLES

Table 1. Distribution of *CYP3A4*20* allele in different populations.

Country	Population	*CYP3A4* wild type homozygous (Nr)	*CYP3A4*20* heterozygous (Nr)	*CYP3A4*20* carriers (%)	Reference
Spain	European	1953	24	1.21	This study
Portugal		449	1	0.22	This study
Italy		478	0	0	This study
Germany		428	0	0	Westlind-Jonhansson [9]
Argentina[a]		240	0	0	This study
Bolivia[b]		179	0	0	This study
European Americans[c]		4119	8	0.19	EVS[b]
Libya	African	95	0	0	This study
Algeria		29	0	0	This study
African Americans[b]		2131	1	0.05	EVS[b]
Israel	Asian	117	0	0	This study
Saudi Arabia		133	0	0	This study
Kuwait		83	0	0	This study
Pakistan		186	0	0	This study
China		108	0	0	This study

[a] Classified as an European population due to the high number of Argentinians of European origin.
[b] In Bolivian population, the European ancestry is 13-21% and the Native American component 77-86%.
[c] Data from Exome Variant Server (http://evs.gs.washington.edu/EVS/)

SUPPLEMENTARY MATERIAL

Supplementary Figure 1. Representation of the microsatellites and SNPs used for haplotype analysis.

Graph indicates the location of microsatellites (D7S651, D7S2498, D7S2480 and D7S666) and SNPs (*CYP3A4*22*, *CYP3A4*1B*, *CYP3A7*2* and *CYP3A5*3*) used for haplotype analysis. Microsatellites location is indicated with respect to *CYP3A4*20*. *CYP3A* locus is magnified indicating the orientation of the genes. Sanger sequencing of a *CYP3A4*20* (c.1461_1462insA) heterozygous variant is shown.

Supplementary Figure 2. Haplotypes of *CYP3A4*20* carriers.

PHASE was used to determine the common haplotype in the 20 *CYP3A4*20* carriers used in the analysis. D7S651 and D7S666 were not amplified for samples 222 and 260. Recurring haplotypes are color shaded.

Supplementary Table 1. Distribution of *CYP3A4*20* allele in Spain.

Region in Spain	***CYP3A4* wild type homozygous (Nr)**	***CYP3A4*20* heterozygous (Nr)**	***CYP3A4*20* carriers (%)**
Castilla León	151	6	3.8
Comunidad Valenciana	129	4	3.0
Extremadura	142	3	2.1
Castilla la Mancha	62	1	1.6
Andalucia	113	1	0.9
Cataluña	126	1	0.8
Comunidad Madrid	508	4	0.8
Galicia	153	0	0.0
Other Spanish regions*****	78	1	0.4
No birth place data	491	3	0.6
TOTAL	**1953**	**24**	**1.2**

*****Other Spanish regions: 57 from Murcia, 5 from Asturias, 4 from Navarra (1 of them *CYP3A4*20* carrier), 4 from País Vasco, 3 from Cantabria, 2 from Aragón, 2 from Ceuta y Melilla, 1 from Canarias and 1 from Islas Baleares.

Figure 1

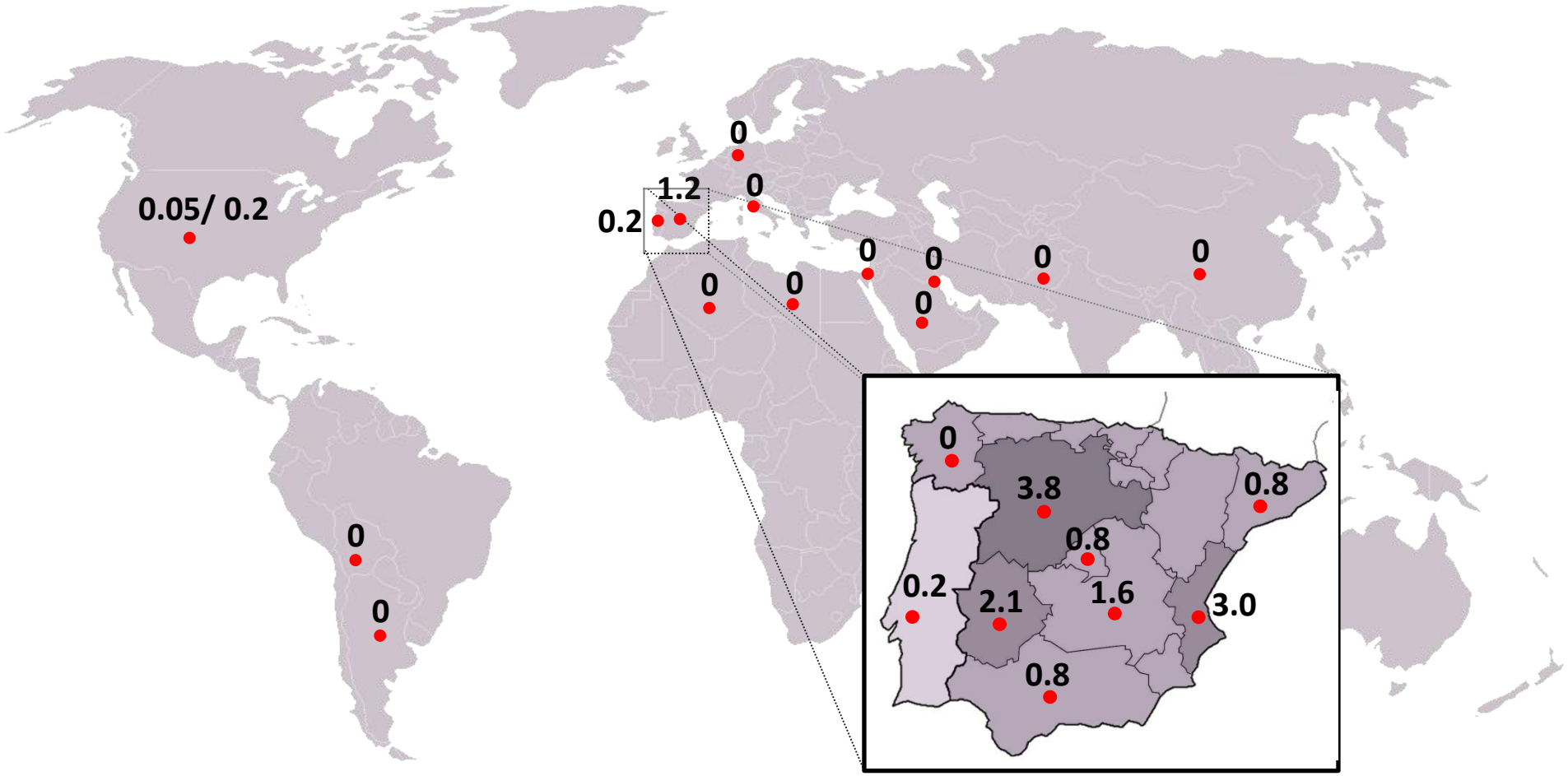
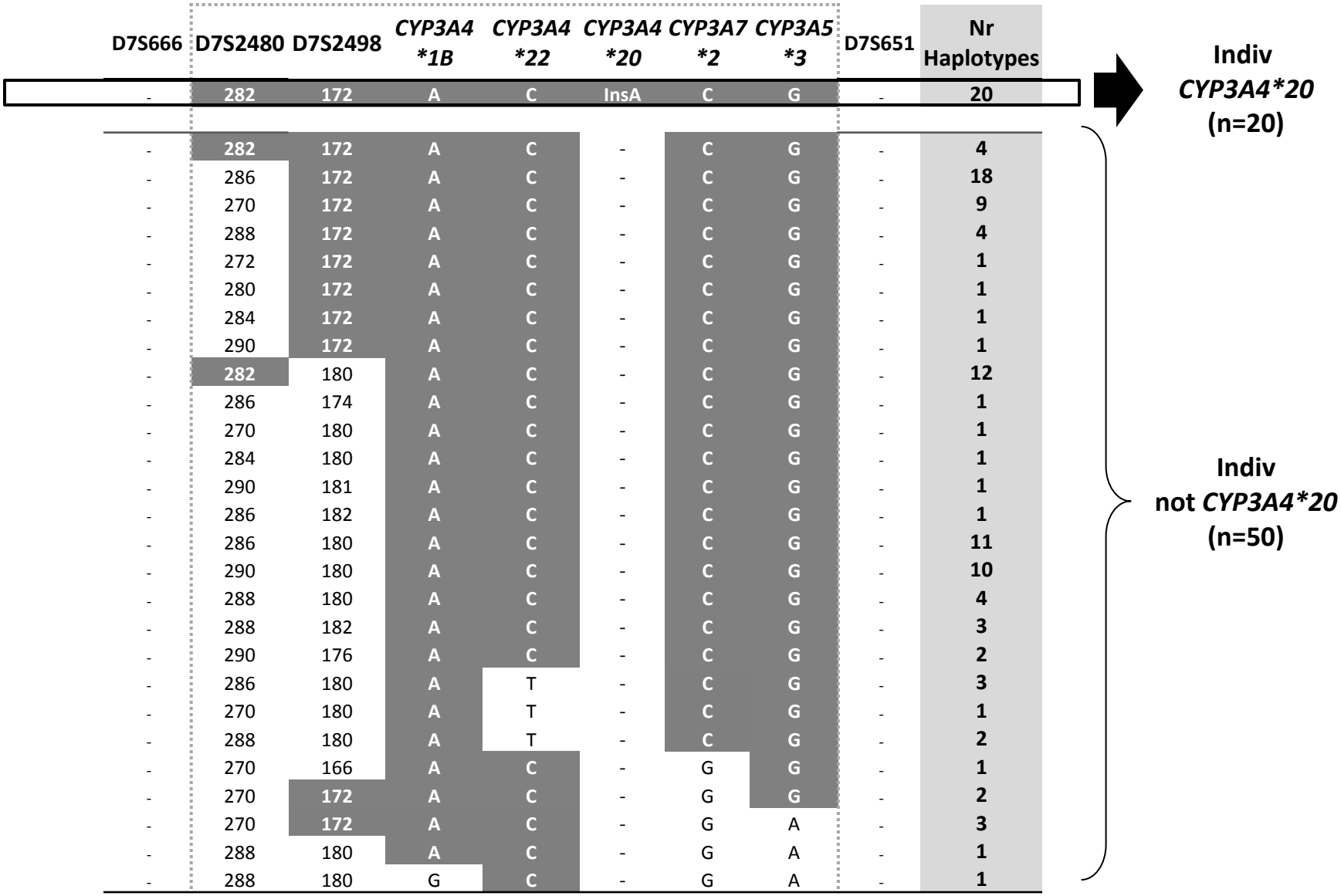


Figure 2



D7S666	D7S2480	D7S2498	*CYP3A4 *1B*	*CYP3A4 *22*	*CYP3A4 *20*	*CYP3A7 *2*	*CYP3A5 *3*	D7S651	Nr Haplotypes
-	282	172	A	C	InsA	C	G	-	20
-	282	172	A	C	-	C	G	-	4
-	286	172	A	C	-	C	G	-	18
-	270	172	A	C	-	C	G	-	9
-	288	172	A	C	-	C	G	-	4
-	272	172	A	C	-	C	G	-	1
-	280	172	A	C	-	C	G	-	1
-	284	172	A	C	-	C	G	-	1
-	290	172	A	C	-	C	G	-	1
-	282	180	A	C	-	C	G	-	12
-	286	174	A	C	-	C	G	-	1
-	270	180	A	C	-	C	G	-	1
-	284	180	A	C	-	C	G	-	1
-	290	181	A	C	-	C	G	-	1
-	286	182	A	C	-	C	G	-	1
-	286	180	A	C	-	C	G	-	11
-	290	180	A	C	-	C	G	-	10
-	288	180	A	C	-	C	G	-	4
-	288	182	A	C	-	C	G	-	3
-	290	176	A	C	-	C	G	-	2
-	286	180	A	T	-	C	G	-	3
-	270	180	A	T	-	C	G	-	1
-	288	180	A	T	-	C	G	-	2
-	270	166	A	C	-	G	G	-	1
-	270	172	A	C	-	G	G	-	2
-	270	172	A	C	-	G	A	-	3
-	288	180	A	C	-	G	A	-	1
-	288	180	G	C	-	G	A	-	1