Genomic Ancestry, *CYP2D6*, *CYP2C9*, and *CYP2C19* Among Latin Americans

Fernanda Rodrigues-Soares[1,2,3,†], Eva M. Peñas-Lledó[1,4,†], Eduardo Tarazona-Santos[1,2,5,6,*], Martha Sosa-Macías[1,7], Enrique Terán[1,8], Marisol López-López[1,9], Idania Rodeiro[1,10], Graciela E. Moya[1,11], Luis R. Calzadilla[1,12], Ronald Ramírez-Roa[1,13], Manuela Grazina[1,14], Francisco E. Estévez-Carrizo[1,15], Ramiro Barrantes[1,16], Adrián LLerena[1,4,*] and RIBEF Ibero-American Network of Pharmacogenetics and Pharmacogenomics[‡]

We present the distribution of *CYP2D6*, *CYP2C9*, and *CYP2C19* variants and predicted phenotypes in 33 native and admixed populations from Ibero-America (*n* > 6,000) in the context of genetic ancestry (*n* = 3,387). Continental ancestries are the major determinants of frequencies of the increased-activity allele *CYP2C19*17* and *CYP2C19* gUMs (negatively associated with Native American ancestry), decreased-activity alleles *CYP2D6*41* and *CYP2C9*2* (positively associated with European ancestry), and decreased-activity alleles *CYP2D6*17* and *CYP2D6*29* (positively associated with African ancestry). For the rare alleles, *CYP2C9*2* and *CYPC19*17*, European admixture accounts for their presence in Native American populations, but rare alleles *CYP2D6*5* (null-activity), *CYP2D6*-multiplication alleles (increased activity), and *CYP2C9*3* (decreased-activity) were present in the pre-Columbian Americas. The study of a broad spectrum of Native American populations from different ethno-linguistic groups show how autochthonous diversity shaped the distribution of pharmaco-alleles and give insights on the prevalence of clinically relevant phenotypes associated with drugs, such as paroxetine, tamoxifen, warfarin, and clopidogrel.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
☑ *CYP2D6*, *CYP2C9*, and *CYP2C19* alleles result in variability in drug response between individuals and populations. However, knowledge about the distribution of pharmaco-alleles in non-European populations is limited.

WHAT QUESTION DID THIS STUDY ADDRESS?
☑ We studied the genetic structure of *CYP2D6*, *CYP2C9*, and *CYP2C19* pharmaco-alleles in 33 Ibero-American populations (*n* > 6,000 individuals) in the context of biogeographic continental ancestry inferred using genetic markers.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?
☑ We show that continental admixture is the major determinant of frequencies of the *CYP2C19*17* allele and CYP2C19 gUMs (negatively associated with Native American ancestry), *CYP2D6*41* and *CYP2C9*2* alleles (associated with European ancestry), and *CYP2D6*17* and *CYP2D6*29* alleles (associated with African ancestry). There is also substantial variation in allele frequencies that depend on autochthonous diversity.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?
☑ Studied *CYP* genes present clinically actionable genotypes for tamoxifen, phenytoin, warfarin, and clopidogrel. Thus, studying how geography and ancestry influence genetic structure contribute to define public health and clinical strategies in a globally diverse context.

[1]RIBEF Ibero-American Network of Pharmacogenetics and Pharmacogenomics, Badajoz, Extremadura, Spain; [2]Departamento de Genética, Ecologia e Evolução, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; [3]Departamento de Patologia, Genética e Evolução, Universidade Federal do Triângulo Mineiro, Uberaba, Minas Gerais, Brazil; [4]Instituto de Investigación Biosanitaria de Extremadura, Universidad de Extremadura, SES, Badajoz, Extremadura, Spain; [5]Universidad Peruana Cayetano Heredia, Lima, Peru; [6]Instituto de Estudos Avançados Transdisciplinares, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; [7]Instituto Politécnico Nacional, Durango, Mexico; [8]Colegio de Ciencias de la Salud, Universidad San Francisco de Quito, Quito, Ecuador; [9]Universidad Autónoma Metropolitana, Ciudad de México, Mexico; [10]Instituto de Ciencias del Mar, La Habana, Cuba; [11]Universidad Católica de Argentina, Buenos Aires, Argentina; [12]Centro Comunitario de Salud Mental, La Habana Vieja, La Habana, Cuba; [13]Universidad Nacional Autónoma de Nicaragua, León, Nicaragua; [14]Faculdade de Medicina & Centro de Neurociências e Biologia Celular, University of Coimbra, Coimbra, Portugal; [15]Universidad de Montevideo, Montevideo, Uruguay; [16]Universidad de Costa Rica, San José, Costa Rica. *Correspondence: Eduardo Tarazona-Santos (edutars@icb.ufmg.br) and Adrián LLerena (allerena@unex.es)

[†]These authors have contributed equally to this paper.

[‡]Consortia affiliations are presented in Text S1.

CYP2D6, CYP2C9, and CYP2C19 are highly polymorphic enzymes involved in the metabolism of many drugs commonly used in clinical practice. They present variants that being related to increased or decreased protein activity (https://www.pharm var.org/) result in variability in drug response between individuals[1–6] and populations.[7–9] These polymorphic enzymes present clinically actionable genotypes for treatments with drugs, such as tamoxifen (used to treat breast cancer), the antidepressants paroxetine and nortriptyline, and the opioid analgesic codeine in the case of *CYP2D6*, the anti-epileptic phenytoin in the case of *CYP2C9*, and for the anticoagulants warfarin (*CYP2C9*) and clopidogrel (*CYP2C19*) (https://cpicpgx.org/genes-drugs/).

Most studies on *CYP2D6*, *CYP2C9*, and *CYP2C19* have been conducted in individuals with predominant European ancestry, which impose a limit for the implementation of global clinical pharmacogenetics in understudied populations, such as Latin Americans.[7–10] The Ibero-American Network of Pharmacogenetics (RIBEF) is focused on the study of Latin American populations, including native and *mestizo*/latino[10] populations (i.e., the product of admixture among Native Americans, Europeans, and Africans[11]). Latin Americans account for 8.4% of the world population (http://data.worldbank. org/region/LAC) and Latin American immigrants and their descendants in the United States (Hispanic/Latino) are the largest and fastest-growing US minority (http://www.census.gov/ population/projections/data/national/2012.html). The association between ancestry and the distribution of pharmacogenetic variants and phenotypes in Latin America remains to be studied at a continental scale. For example, the decreased-activity allele *CYP2C9*2* was the first clinically relevant pharmacogenetic variant found to be less prevalent in Native Americans compared with *mestizo* or European individuals,[8,12] but a survey of Latin American populations with different ancestries, geographic, and ethno-linguistic backgrounds is pending.

Recently, the Consortium of the Ibero-American Network of Pharmacogenetics and Pharmacogenomics (CEIBA-RIBEF) published the largest survey of aggregated frequencies and predicted phenotypes for *CYP2D6*, *CYP2C9*, and *CYP2C19* alleles in North, Central, and South American native and mestizo populations, as well as in Iberians, the source of most of European immigrants into the Americas.[13] The CEIBA-RIBEF report classified populations as native or admixed, based on self-identification or on local consensus. Here, we present the results of the *Mestizo* Pharmacogenetics (MESTIFAR) RIBEF Project,[14] that aims to study the interindividual and interpopulation differences in *CYP2D6*, *CYP2C9*, and *CYP2C19* genotypes and their predicted phenotypes as a function of individual and population biogeographic continental ancestry, inferred using Ancestry Informative Markers (AIMs). We present data for 31 Latin American populations from North America, Central America, and South America, the Caribe, and for two Iberian populations, the region of origin of most immigrants from Europe to Latin America. The studied populations from the Americas are settled in urban and rural environments and presenting different degrees of admixture. We included Native American populations representing the most important continental ethno-linguistic groups, such as Uto-Aztecan, Mayan,

Chibchan, Arawak, and Quechuamaran,[15] therefore, presenting the most extensive and diverse study of *CYP2D6*, *CYP2C9*, and *CYP2C19* variation in the American continent.

RESULTS

The context of admixture in Latin Americans

Figure 1 shows the results of Principal Component (**Figure 1a**) and ADMIXTURE analyses (**Figure 1b**), based on AIM genotypes for 3,387 RIBEF samples (Latin Americans and Iberians) and 119 Yorubas from the 1000 Genomes Project. These analyses identify populations and individuals that are predominantly Native Americans, and illustrates the intensive admixture and the high genetic diversity of Latin American populations in general, and of the RIBEF dataset in particular. Ancestry analysis confirms that populations classified as Native American in Mexico, Costa Rica, and Peru have at least 86% of Native American ancestry (**Figure 1b and Table 1**). The exceptions are the native Mayo population from Mexico (~40% of Old World admixture), and Bribri and Chorotega populations from Costa Rica (European ancestry 18% and 22%, respectively). We studied a broad spectrum of 31 autochthonous and admixed populations distributed across Latin America. Among urban admixed populations, Native American ancestry is predominant in admixed Peruvians (from Lima, 71%), admixed Ecuadorians (from Quito and Guayaquil, 61%), and admixed Mexicans (from *Distrito Federal*, 60%). Interestingly, Lima-Peru admixed have similar Old World ancestry than some populations classified as Native Americans, such as the Mayo from Mexico (61%) and the Costa Rican Chorotega (69%). European ancestry predominates in Argentinians Ashkenazi Jews (88%), Argentinians (82%), Uruguayans (78%), Brazilians (68%), and Cubans (64%), all considered admixed populations. The admixed Nicaraguans and Costa Ricans have the most balanced admixture among the studied populations (Native American, European, and African ancestry, respectively: 41%, 45%, and 14% for Nicaragua, and 40%, 43%, and 17% for Costa Rica). The populations with the highest African ancestry are the Costa Rican blacks (86%) and Cuban blacks (52%), where some individuals reach > 90% of African ancestry.

Association of *CYP2D6* alleles and predicted metabolic phenotypes with genomic ancestry

Table S2 shows *CYP2D6*, *CYP2C9*, and *CYP2C19* allele and predicted phenotype frequencies for all the studied populations. **Table 2 and Figure 2** show the results of regression analysis of *CYP2D6* allele frequencies on population-based ancestry.

For the common normal activity allele *CYP2D6*2* and the rare null-activity allele *5, continental ancestry is not correlated with allele frequencies. Differently, both alleles *CYP2D6*10* (defective, rare outside East Asia[7]) and *35 (normal activity) are correlated with European ancestry, but this ancestry explains a larger amount of allele frequencies variance of the latter ($R^2 > 0.60$; **Table 2 and Figure 2a**) than of the former allele ($R^2 < 0.40$; **Table 2**). The strongest association with continental ancestry is observed for the decreased-activity and rarer alleles *CYP2D6*41*, *17*, and *29* (**Table 2 and Figure 2**). The frequencies of *CYP2D6*41* (range in
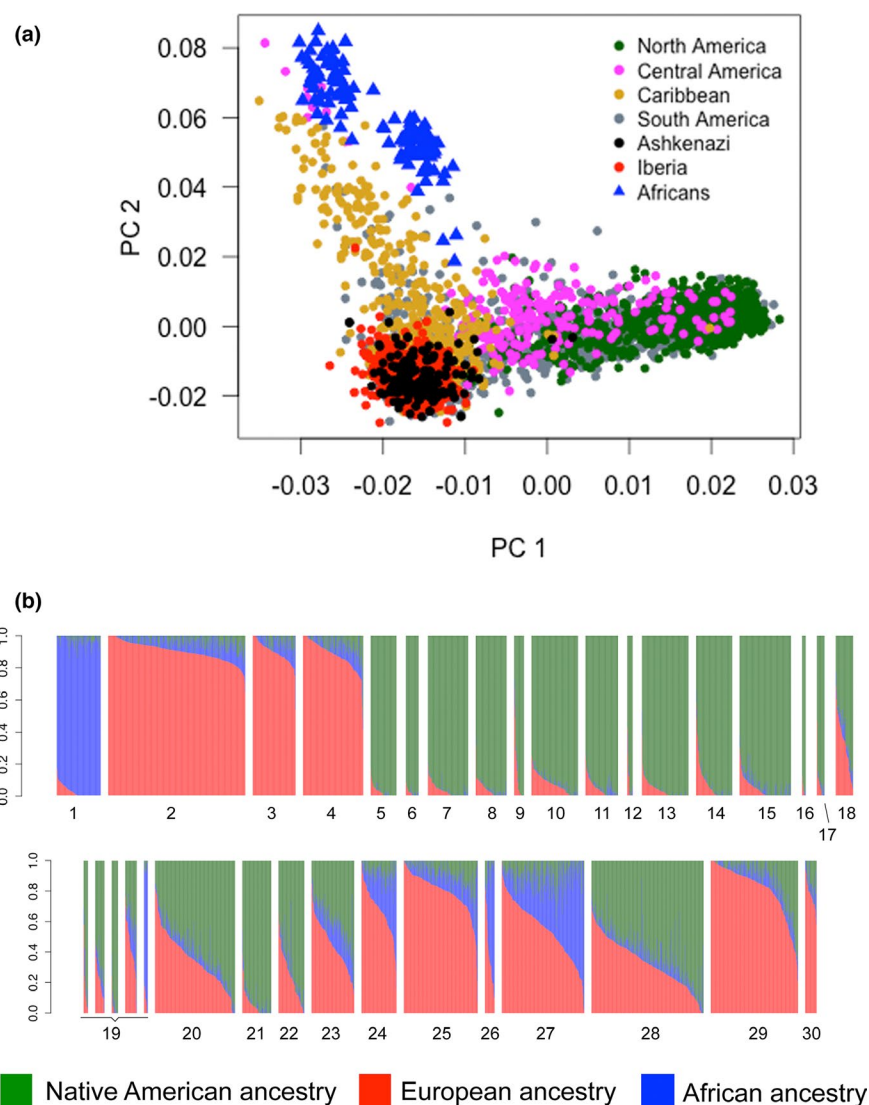
Figure 1 Genomic ancestry of studied individuals and populations and African Yoruba (Nigerians) from the 1000 Genomes Project, based on the 83 ancestry informative markers. (**a**) Principal Component (PC) Analysis representation. Individuals are colored according to geographic regions. (**b**) Vertical bar plots of individual continental ancestry. **1**: African Yoruba from Nigeria, 1000 Genomes Project; **2**: Spaniards from Extremadura; **3**: Portuguese; **4**: Ashkenazi from Argentina; **5**: Ashaninka from Peru; **6**: Shimaa from Peru; **7**: Aymara from Peru; **8**: Mexicaneros from Mexico; **9**: Seris from Mexico; **10**: Tepehuanos from Mexico; **11**: Tarahumaras from Mexico; **12**: Guarijios from Mexico; **13**: Huicholes from Mexico; **14**: Coras from Mexico; **15**: Lacandones from Mexico; **16**: E. Lacandones from Mexico; **17**: Tzeltales from Mexico; **18**: Mayos from Mexico; **19**: Costa Rica populations: 1st: Bribri; 2nd: Chorotega; 3rd: Guaymí; 4th: Admixed population; 5th Afro-Caribbeans; **20**: Admixed from Mexico (DF); **21**: Admixed from Chiapas, Mexico; **22**: Admixed from Peru; **23**: Admixed from Nicaragua; **24**: Admixed from Brazil; **25**: Cuban; **26**: Afro-Caribbeans from Cuba; **27**: Admixed from Cuba; **28**: Admixed from Ecuador; **29**: Admixed from Argentina; and **30**: Admixed from Uruguay.

Ibero-America: 0–16.2%) are well predicted by European ancestry (R^2 = 0.81). In particular, the Ashkenazi Jews from Argentina show the highest frequency of *CYP2D6*41* in Ibero-America (16.2%, **Table S2**, **Figure 2b**).[16] Differently, the frequencies of the decreased-activity alleles *CYP2D6*17* and *29* (ranges in Ibero-America: 0–18% and 0–11%, respectively) are positively correlated and well predicted by African ancestry (R^2 > 0.88; **Figure 2c,d**).

Besides continental ancestry and Old World admixture, the geographic and ethnic diversity of Native American populations contribute to *CYP2D6* variation (**Table S2**). Alleles defining poor and ultra-rapid metabolizer (gUM) individuals

are particularly relevant in clinical pharmacology. *CYP2D6* null alleles *4 and *5 account for most of the predicted poor metabolizer individuals (gPMs). Frequencies of *CYP2D6*4* are positively correlated with European ancestry (**Table 2 and Figure 2e**), but this ancestry only explains around 20% of allele frequency variance. Indeed, there is important variation among Native American populations with very low Old World admixture, with frequencies varying from 0% in Mexicanero from Mexico (North America, 95% of Native American ancestry, Uto-Aztecan linguistic group) to 31% in Costa Rican Bribri from Central America (74% of Native American ancestry,

Table 1 Mean and interindividual SDs of European, African, and Native American ancestries in the studied populations

Country	Population (ethno-linguistic group)[a]	N[b]	European ancestry	SD	African ancestry	SD	Native American ancestry	SD
Argentina	Ashkenazi	163	0.878	0.093	0.060	0.057	0.062	0.073
	Admixed	243	0.814	0.166	0.053	0.051	0.133	0.154
Brazil	Admixed	98	0.672	0.151	0.239	0.146	0.088	0.059
Costa Rica	Admixed	32	0.429	0.167	0.168	0.099	0.403	0.137
	Afro-Caribbeans	11	0.093	0.072	0.863	0.082	0.044	0.035
	Bri Bri (Chibchan)	12	0.182	0.196	0.072	0.081	0.745	0.234
	Chorotega (Otomanguean)	26	0.220	0.112	0.090	0.063	0.690	0.108
	Guaymi (Chibchan)	18	0.030	0.062	0.027	0.045	0.943	0.095
Cuba	Admixed	230	0.551	0.183	0.337	0.204	0.111	0.102
	Afro-Caribbeans	27	0.401	0.226	0.523	0.276	0.076	0.074
	White	206	0.793	0.134	0.114	0.112	0.093	0.081
Ecuador	Admixed	312	0.329	0.156	0.056	0.074	0.615	0.166
Mexico	Admixed Chiapas (Mayan)	81	0.051	0.075	0.022	0.036	0.927	0.095
	Admixed DF	223	0.350	0.173	0.050	0.049	0.601	0.182
	Coras (Uto-Aztecan)	98	0.070	0.113	0.009	0.019	0.920	0.121
	E. Lacandon (Mayan)	10	0.035	0.043	0.010	0.019	0.955	0.038
	Guarijios (Uto-Aztecan)	15	0.078	0.124	0.026	0.037	0.896	0.138
	Huichol (Uto-Aztecan)	126	0.037	0.053	0.004	0.016	0.959	0.062
	Lacandon (Mayan)	139	0.039	0.057	0.021	0.034	0.940	0.073
	Mayos (Uto-Aztecan)	47	0.327	0.165	0.058	0.045	0.615	0.168
	Mexicaneros (Uto-Aztecan)	84	0.037	0.049	0.010	0.016	0.953	0.050
	Seris (Seri)	27	0.129	0.187	0.016	0.030	0.855	0.212
	Tzental (Mayan)	21	0.069	0.104	0.021	0.026	0.910	0.099
	Tepehuano (Uto-Aztecan)	126	0.074	0.058	0.010	0.017	0.916	0.059
	Tarahumara (Uto-Aztecan)	88	0.044	0.054	0.017	0.028	0.939	0.055
Nicaragua	Admixed	119	0.447	0.381	0.138	0.217	0.415	0.416
Peru	Admixed	72	0.241	0.157	0.049	0.079	0.711	0.183
	Ashaninka (Arawak)	70	0.024	0.035	0.004	0.012	0.972	0.036
	Aymara (Quechuamaran)	109	0.032	0.040	0.006	0.016	0.962	0.043
	Shimaa (Arawak)	35	0.018	0.024	0.008	0.013	0.974	0.023
Portugal	White	116	0.896	0.067	0.067	0.063	0.037	0.042
Spain	White	371	0.903	0.058	0.060	0.047	0.036	0.041
Uruguay	Admixed	32	0.775	0.087	0.076	0.067	0.149	0.085

[a]Ethno-linguistic classification of Native American populations (in italics) is according to Campbell.[15] [b]N = Number of individuals genotyped both for CYPs and ancestry.

Chibchan linguistic group). The high percentages of the null-activity alleles *4 and *5 in Bribri account for its highest percentage of CYP2D6 gPMs in our Ibero-American dataset (21.7%).[17] Importantly, although the *4 allele is common and cosmopolitan, allele *5 is rarer worldwide. Our studied populations with > 90% of Native American ancestry host 31 copies of the CYP2D6*5 allele, whereas given its low frequency in Europeans (~2%) we would expect only 2.4 copies if all these alleles would have been the product of European introgression. Thus, we infer that the presence of the rare CYP2D6*5 null allele in the American continent predates the arrival of Iberians ~500 years ago.

The multiplication of CYP2D6 normal alleles (wt, *2 or *35xN) accounts for most of the gUM individuals. This multiplication allele class shows no association with continental ancestry, but present relevant variation among Native American populations, with its frequency ranging from 0% in several Native American populations, to values as higher as 5.9% in the Mexican populations in Cora (92% of Native American ancestry), 6.8% in Tarahumara (94% of Native American ancestry), and 12% in Huichol (23; 96% of Native American ancestry), all belonging to the Uto-Aztecan ethnic group (**Table S2**). Our studied populations with > 90% of Native American ancestry host 87 copies of the CYP2D6-multiplication

Table 2 Linear regression results of the *CYP2D6* allele and phenotype frequencies on the population ancestry estimates

	CYP2D6 alleles										Predicted phenotypes	
	Wt	*2	*4[a]	*5[a]	*10[b]	*17[b]	*29[b]	*35	*41[b]	wt-*2-*35xN	gPM[c]	gUM[c]
Ancestry												
Univariate linear regression												
Native American												
Beta[d]	2.45	0.74	−0.69	−0.12	−0.21	−0.52	−0.35	−0.36	−0.81	−0.02	−3.06	0.17
P	0.001	NS	0.05	NS	0.001	0.01	0.01	0.001	0.001	NS	0.01	NS
R^2	0.63		0.11		0.30	0.24	0.27	0.47	0.59		0.20	
European												
Beta	−2.64	−0.32	1.08	0.09	0.16	0.08	0.06	0.52	1.14	−0.02	0.45	−0.01
P	0.001	NS	0.01	NS	0.01	NS	NS	0.001	0.001	NS	0.001	NS
R^2	0.49		0.20		0.10			0.67	0.81		0.31	
African												
Beta	−2.98	−1.25	−0.20	0.27	0.46	2.11	1.16	0.03	0.15	0.16	0.01	−0.75
P	0.01	0.05	NS	NS	0.001	0.001	0.001	NS	NS	NS	NS	NS
R^2	0.18	0.11			0.31	0.95	0.88					
Multivariate linear regression												
Native American controlling for European												
Beta	2.42	1.20	0.45		−0.45	−2.13	−1.20	0.09	0.11		0.09	
P	0.01	NS	NS		0.001	0.001	0.001	NS	NS		NS	
R^2	0.62				0.38	0.95	0.90					
Native American controlling for African												
Beta	2.46	0.24	−1.15		−0.13	0.07	0.12	−0.52	−1.15		−0.45	
P	0.001	NS	0.01		0.05	NS	0.05	0.001	0.001		0.001	
R^2	0.62				0.38		0.90	0.67	0.81		0.29	
European controlling for Native American												
Beta	−0.04	0.96	1.56		−0.30	−2.20	−1.32	0.61	1.26		0.55	
P	NS	NS	NS		0.001	0.001	0.001	0.001	0.001		0.05	
R^2					0.38	0.95	0.90	0.67	0.81		0.29	
European controlling for African												
Beta	−2.46	−0.24	1.11		0.13	−0.07	−0.12	0.52	1.15		0.45	
P	0.001	NS	0.01		0.05	NS	0.05	0.001	0.001		0.001	
R^2	0.62		0.19		0.38		0.90	0.67	0.81		0.29	
African controlling for Native American												
Beta	0.05	−0.96	−1.56		0.30	2.20	1.32	−0.61	−1.26		−0.55	
P	NS	NS	NS		0.01	0.001	0.001	0.001	0.001		0.05	
R^2					0.38	0.95	0.90	0.67	0.81		0.29	
African controlling for European												
Beta	−2.42	−1.20	−0.45		0.43	2.13	1.20	−0.09	−0.11		−0.09	
P	0.01	0.05	NS		0.001	0.001	0.001	NS	NS		NS	
R^2	0.62	0.09			0.38	0.95	0.90					
Native American controlling for region[e]												
P	< 0.01	< 0.01	NS		NS	NS	NS	NS	NS	< 0.01	NS	< 0.01
Native American controlling for country												
P	< 0.01	NS	< 0.01		< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	NS	< 0.01	NS

gPM, predicted poor metabolizer based on genotype; gUM, predicted ultra-rapid metabolizer based on genotype; NS, not significant. Values significant at 1% are in italic characters.
[a]Null activity allele. [b]Decreased activity allele. Alleles without symbols present normal activity. [c]gPM: predicted poor metabolizer based on genotype. gUM: predicted ultra-rapid metabolizer based on genotype. [d]Beta coefficients are expressed in units of percentage for each 10% change in ancestry. P, significance (values significant at 1% are in italic characters). R^2, variance of allele frequency explained by ancestry. [e]Regions are represented by a categorical variable with the following values: North America, Caribbean, Central America, and South America.
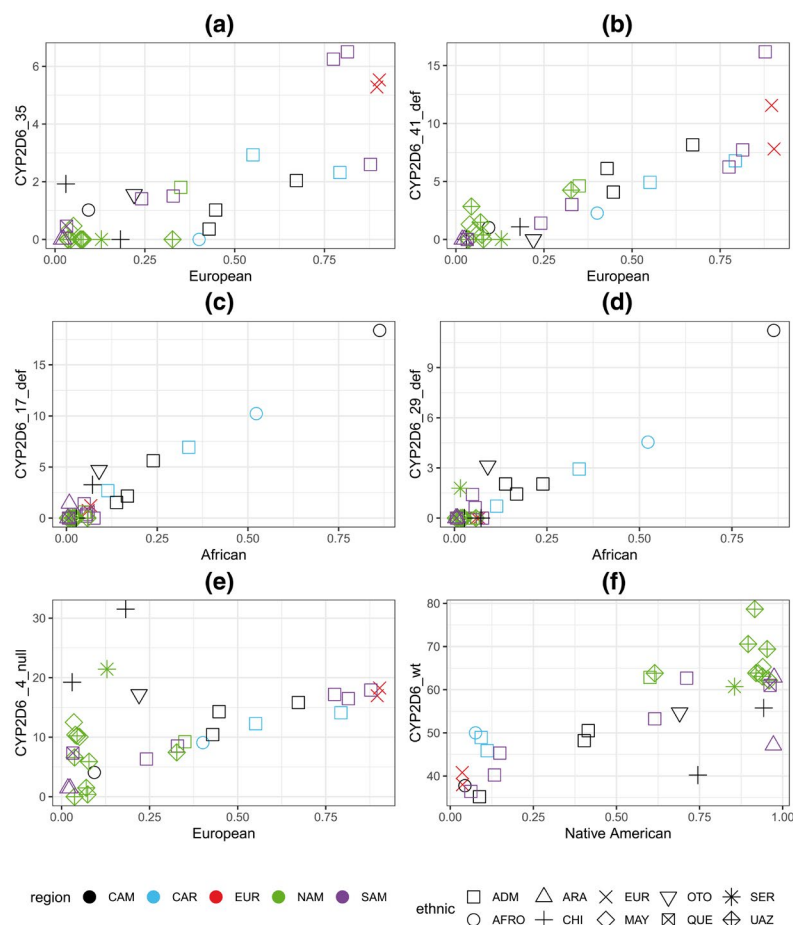
Figure 2 Scatterplot of *CYP2D6* allele frequencies as a function of correlated continental ancestry for populations of the Americas. Populations are categorized by subcontinental regions, admixture, or ethnic classification. Allele classification: def, defective enzymatic activity; null, null enzymatic activity; *wt*, wild type (inferred as normal enzymatic activity).

alleles. Although we would expect only six copies if all these alleles would have been the product of European introgression, we infer that the presence of the *CYP2D6*-multiplication alleles in the American continent predates the arrival of Iberians ~500 years ago. Frequencies of *CYP2D6* gUMs in Native American populations, consistently with the frequency distribution of *CYP2D6*-multiplication alleles, also range from 0% to values as higher as 11.8% in Cora, 12.5% in Tarahumara, 23% in Huichol, 21.3% in Mexican Mayo (61% of Native American ancestry), 17.6% in Guarijio (all Uto-Aztecan groups), and 14.3% in Seri. In summary, among Native American populations, the frequency distributions of *CYP2D6*-multiplication alleles and predicted gUMs are higher in North America (particularly in Uto-Aztecan groups), lower in Central America, and almost null in South America, a pattern that is independent of Old World admixture (**Table S2**).

Association of *CYP2C9* alleles and predicted metabolic phenotypes with genomic ancestry
Table S2 shows the allele and predicted genotype frequencies for *CYP2C9* in Ibero-America. **Table 3 and Figure 3** show the results of regression analysis of *CYP2C9* allele frequencies on population-based ancestry. The frequency of the

decreased-activity allele *CYP2C9*2* (range in Ibero-America: 0–18%) is negatively associated with Native American ancestry and positively associated with European ancestry, which is the best predictor ancestry of the frequency of this allele ($R^2 = 0.89$; **Table 3**, **Figure 3a**). The very low frequencies of *CYP2C9*2* in populations with very low European admixture seem to be the product of European introgression, because we do not observe an excess of copies of this allele among Native American populations with > 90% of autochthonous ancestry, respect to the expected, based solely on European admixture (7 observed vs. 15 expected).

The decreased-activity allele *CYP2C9*3* (frequency range in Ibero-America: 0–11%) is also negatively associated with Native American ancestry and positively associated with European ancestry, but compared with *CYP2C9*2*, European ancestry explains a lower portion of its frequency variance ($R^2 = 0.23$; **Table 3**). This is because Mexican native populations show considerable variation for *CYP2C9*3* allele frequency, that range from 0–10%, independently of European admixture (**Table S2**), in particular due to the frequencies in the range 5.82–10.79% observed in the Uto-Aztecan Guarijio, Tepehuano, and Tarahumara populations. Interestingly, none of the 214 South Amerindians analyzed carried

Table 3 Linear regression parameters of the *CYP2C9* and *CYP2C19* alleles and phenotype frequencies on the population ancestry estimates

	CYP2C9				CYP2C19				
	*3[a]	*2[a]	wt	gPM	*2[b]	*17[c]	wt	gPM	gUM
Univariate linear regression									
Native American									
Beta[d]	−0.42	−1.32	1.77	−0.04	−0.57	−2.19	2.83	−0.16	−3.40
P	0.01	0.001	0.001	0.01	0.05	< 0.001	0.001	NS	0.001
R^2	0.20	0.70	0.69	0.26	0.13	0.94	0.77		0.92
European									
Beta	0.54	1.80	−2.35	0.05	0.42	2.32	−2.81	0.15	3.63
P	0.01	0.001	0.001	0.001	NS	< 0.001	0.001	NS	0.001
R^2	0.23	0.90	0.83	0.29		0.70	0.50		0.70
African									
Beta	0.27	0.46	−0.80	0.02	1.30	2.89	−4.20	0.26	4.27
P	NS	NS	NS	NS	0.05	< 0.001	0.001	NS	0.001
R^2					0.16	0.33	0.35		0.30
Multivariate linear regression									
Native American controlling for European									
Beta	−0.14	−0.01	0.22	−0.01	−1.23	−2.36	3.61		−3.50
P	NS	NS	NS	NS	0.05	< 0.001	0.001		0.001
R^2					0.17	0.94	0.78		0.92
Native American controlling for African									
Beta	−0.52	−1.99	2.34	−0.05	−0.33	−2.13	2.54		−3.37
P	0.01	0.001	0.001	0.01	NS	< 0.001	0.001		0.001
R^2	0.21	0.89	0.83	0.27		0.94	0.78		0.92
European controlling for Native American									
Beta	0.39	−1.78	−2.11	0.04	−0.90	0.25	1.08		−0.13
P	NS	0.001	0.001	NS	NS	NS	NS		NS
R^2		0.89	0.83						
European controlling for African									
Beta	0.53	1.79	−2.34	0.05	0.33	2.13	−2.54		3.37
P	0.01	0.001	0.001	0.01	NS	< 0.001	0.001		0.001
R^2	0.21	0.89	0.83	0.27		0.94	0.78		0.92
African controlling for Native American									
Beta	−0.39	−1.78	2.11	−0.04	0.90	0.23	−1.08		0.13
P	NS	0.001	0.001	NS	NS	NS	NS		NS
R^2		0.89	0.83						
African controlling for European									
Beta	0.14	0.02	−0.23	0.01	1.23	2.36	−3.62		3.50
P	NS	NS	NS	NS	0.05	< 0.001	0.001		0.001
R^2					0.17	0.94	0.78		0.92
Native American controlling for region[e]									
P	< 0.01	< 0.01	NS	NS	NS	< 0.01	NS	< 0.01	< 0.01
Native American controlling for country									
P	< 0.01	< 0.01	< 0.01	NS	NS	< 0.01	NS	NS	< 0.01

gPM, predicted poor metabolizer based on genotype; gUM, predicted ultra-rapid metabolizer based on genotype; NS, not significant.Values significant at 1% are in italic characters.
[a]Decreased activity allele. [b]Null activity allele. [c]Increased activity allele. Alleles without symbols present normal activity. [d]Beta coefficients are expressed in units of percentage for each 10% change in ancestry. P, significance (values significant at 1% are in italic characters). R^2, variance of allele frequency explained by ancestry. [e]Regions are represented by a categorical variable with the following values: North America, Caribbean, Central America, and South America.
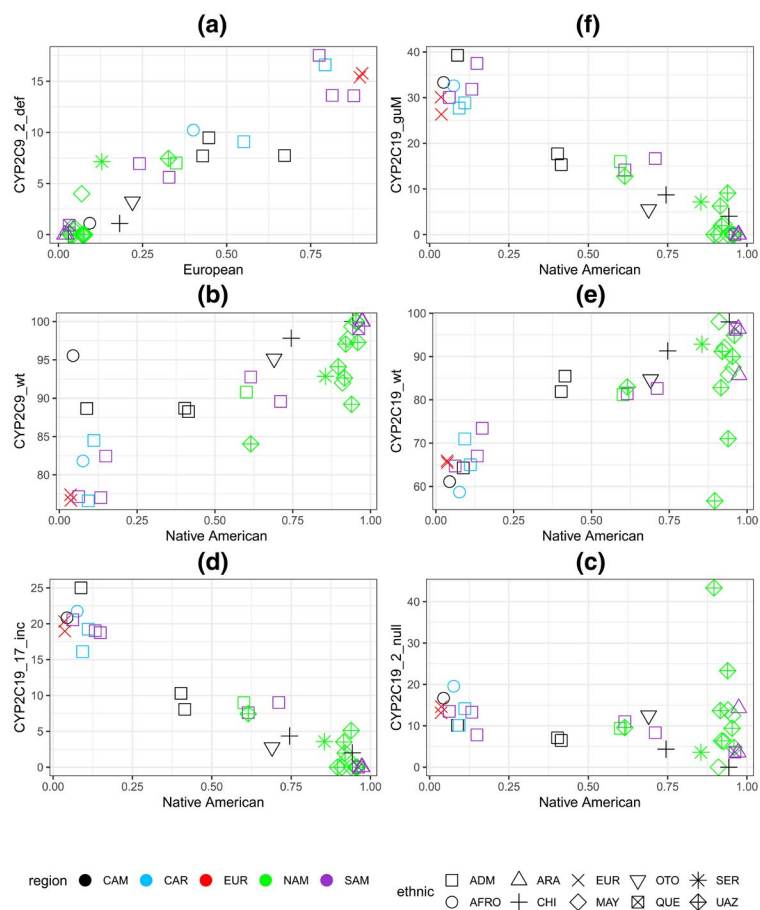
Figure 3 Scatterplot of *CYP2C9* and *CYP2C19* allele frequencies as a function of correlated continental ancestry for populations of the Americas. Populations are categorized by subcontinental regions and ethnic classification. Allele classification: def, defective enzymatic activity; inc, increased enzymatic activity; null, null enzymatic activity; and *wt*, wild type (inferred as normal enzymatic activity). Phenotype classification: gUMs: ultra-rapid metabolizer as inferred from genotype.

the *CYP2C9*3* allele (binomial allele frequency 95% confidence interval = 0–0.007). Because the number of copies of *CYP2C9*3* in Native American populations with > 90% of autochthonous ancestry was more than expected under the hypothesis that all the observed alleles were of European origin (57 vs. 7.5), we infer that *CYP2C9*3* was present in North America before the arrival of Europeans around ~500 years ago. This inference is consistent with its presence in Asiatic populations at a frequency of 3–12%.[8]

The rare null-activity allele *CYP2C9*6* was only found in four chromosomes from admixed Cubans with a range of 31.6–70.9% of African ancestry of the 6,606 (3,525 of them with genomic ancestry estimated) individuals surveyed in the Americas for *CYP2C9*. This result is consistent with the fact that the *CYP2C9*6* allele is rare in most human populations (< 1%),[8] but their frequencies are quite higher in Africans (2.7%).[8] Complementarily to *CYP2C9*2* and *CYP2C9*3* allele frequencies, the *CYP2C9 wt* allele is positively associated with Native American ancestry and negatively associated with European ancestry (**Figure 3b and Table 3**).

Only 7 individuals (0.2%) of the 6,060 analyzed for *CYP2C9* were predicted as gPMs on the basis of genotyping of alleles *CYP2C9*2*, *3*, *6*, and *8*. From these seven individuals, five were found in Cuba (51–81% of individual European ancestry), and

two in the Mexican Tepehuano population (3.5% and 7.9% of individual European ancestry, respectively).

Association of *CYP2C19* alleles and predicted metabolic phenotypes with genomic ancestry

Table S2 shows the allele and predicted genotype frequencies for *CYP2C19* in Ibero-America. **Table 3 and Figure 3** show the results of regression analysis of *CYP2C19* allele frequencies on population-based ancestry. The increased-activity allele *CYP2C19*17* is rare in Native Americans (< 5%) and more common in predominantly European and African-ancestry populations (20–25%). Thus, *CYP2C19*17* is negatively associated with Native American ancestry, which explains a very high proportion of its allele frequency variance in Latin America ($R^2 = 0.94$; **Table 3 and Figure 3d**). *CYP2C19* gUMs are also negatively associated with Native American ancestry, following the pattern of *17* ($R^2 = 0.92$; **Table 3 and Figure 3f**). Because our studied populations with > 90% of Native American ancestry host 25 copies of the *CYP2C19*17* allele, whereas we would expect 23 copies if all these alleles would have been the product of European introgression, we infer that the European immigration in the last ~500 years accounts for the presence of this allele in the Americas.

The frequency of the null-activity allele *CYP2C19*2* is negatively associated with Native American ancestry and positively associated with African ancestry, but continental ancestry explains only a very low fraction of its allele frequency variance ($R^2 = 0.17$; **Table 3 and Figure 3c**). Interestingly, continent-wide Native American populations show high variability in the frequency of *CYP2C19*2* independently of their Old World admixture. Excluding the outlier Guarijio from Mexico (43%), the frequencies of *CYP2C19*2* range from 0% in Guaymi from Costa Rica (Chibchan linguistic group) to 23% in Tarahumara from Mexico (Uto Aztecan linguistic group, both populations with 93–94% of Native American ancestry). Regarding predicted metabolic phenotypes, and mimicking the pattern for the null allele *2*, the frequencies of gPMs are not associated with continental ancestry, but show variable frequencies among Native American populations. Frequencies of 0% are observed in several Native American populations, as well as values higher as 2.8% in Peruvian Shimaa from the Amazon Yunga (Arawak linguistic group), 2.6% in Lacandon from Mexico (Maya linguistic group), 6.8% in Mexican Tarahumara (Uto-Azteca linguistic group), and 17% in Guarijio from Mexico (Uto-Azteca linguistic group), all these populations with > 89% of Native American ancestry.

DISCUSSION

We present the largest study in populations from Ibero-America that analyzes clinically relevant *CYP2D6*, *CYP2C9*, and *CYP2C19* genetic polymorphisms and ancestry biomarkers, including a large set of Native Americans, an ethnic group that is underrepresented in genetic studies. Importantly, the study of pharmacogenetic polymorphisms was contextualized considering continental genetic ancestry estimated on autochthonous and admixed individuals and populations, revealing genetic differences that are dependent on the degree of admixture among Native American, European, and African stocks. The analysis encompasses a vast array of 33 populations that include admixed groups settled in cities of different sizes from North, Central, and South America inhabited by admixed individuals, and a large set of ethno-linguistically diverse Native American populations living in different environments from Mexico (North America), Costa Rica (Central America), and Peru (South America; **Table 1**).

Continental ancestries are the major determinants of variation in frequencies of the following alleles (**Tables 2 and 3**): (i) the increased-activity allele *CYP2C19*17* and consequently *CYP2C19* gUMs, which are negatively associated with Native American ancestry; (ii) decreased-activity alleles *CYP2D6*41* and *CYP2C9*2*, positively associated with European ancestry; and (iii) decreased-activity alleles *CYP2D6*17* and *CYP2D6*29*, positively associated with African ancestry.

Because populations with predominant European ancestry are overrepresented in genetic studies, the allelic spectrum of *CYP* alleles is affected by ascertainment bias.[18] Although most human populations usually share common alleles, rare alleles may be restricted to specific populations, and rare alleles typical of European populations are more likely to be known that rare alleles typical of Native American populations, that may remain undetected by classical pharmacogenetic genotyping methods. For instance, in this study, the null-activity alleles *CYP2D6*3*, *6*, and *CYP2C9*6* were only present at low frequencies in populations with predominant Old World ancestry, and because they were not observed among the large number of predominant Native American individuals and populations, are likely absent in these populations. Differently, although for rare alleles *CYP2C9*2* (decreased-activity) and *CYPC19*17* (increased-activity), European admixture during the last ~500 years account for their presence in Native American populations, for alleles *CYP2D6*5* (null activity), *CYP2D6*-multiplication alleles (increased activity), and *CYP2C9*3* (decreased activity), we inferred that they were present at low frequency in the Americas before the first arrival of Europeans > 500 years ago.

There is a vast set of ethnically diverse Native American populations settled since the first peopling of the continent > 12,000 years ago,[19] living in different environments. In these populations with diverse degrees of isolation, the genetic drift has created considerable genetic diversity between populations. Indeed, indigenous from the Americas are known since the first studies on protein polymorphisms for being settled in the regions of the world with the highest values of interpopulation differentiation.[20] This is exemplified in our dataset by the high frequencies observed for common polymorphisms, such as the null alleles *CYP2D6*4* (and consequently, the high frequencies of *CYP2D6* clinically relevant gPMs phenotype) in the Bribri from Costa Rica and the Seri from Mexico; the null allele *CYP2D6*5* in the Bribri (Costa Rica) and Shimaa (Peruvian Amazon Yunga), the normal allele *CYP2D6*2* in the Ashaninkas (Peruvian Amazon Yunga) as well as *CYP2C19*2* (decreased activity) for the Guarijio from Mexico. In addition, unexpectedly relative high frequencies in small populations are observed for alleles that are generally very rare in Native Americans, such as *CYP2C9*3* (10%) for the Tarahumara from Mexico or *CYP2D6*6* for the Mayo from Mexico (3%). As a more general pattern, the studied Uto-Aztecan populations (particularly Mexicanero, Guarijio, and Tarahumara), one of the largest ethno-linguistic groups of North America, are characterized by relatively high frequencies of the *CYP2D6*-multiplication alleles and the *CYP2D6* clinically relevant gUMs phenotype, decreased-activity allele *CYP2C9*3*, as well as the null-activity allele *CYP2C19*2*, and consequently the clinically relevant *CYP2C19* gPMs phenotype.

The more relevant Clinical Pharmacogenetics Implementation Consortium (CPIC) drug-interactions guidelines (https://cpicpgx.org/genes-drugs/) for *CYP2D6* include the drugs paroxetine (antidepressant), nortriptyline (antidepressant), tamoxifen (used as breast cancer adjuvant therapy), and the analgesic pro-drug codeine. *CYP2D6* gUMs are at higher risk of low plasma concentrations for paroxetine, nortriptyline, and of toxicity induced by codeine, and, therefore, *CYP2D6* gUMs should avoid the use of these drugs. In Latin Americans, we observed that *CYP2D6* gUM individuals are present in admixed populations and North Amerindian populations (particularly in the Uto-Aztecan ethnic group), being almost absent in South Amerindian populations with negligible Old World admixture. Differently, the exceptionally high frequency of gPM individuals in the Chibchan Bribri population from Costa Rica (21%) exemplifies how the genetic drift may lead to high frequencies of a genotype, its strongly

associated metabolic phenotype *CYP2D6* mPMs, as well as of clinically relevant traits, such as higher plasma concentrations and risk of side effects for paroxetine and nortriptyline, or increased likelihood of breast cancer recurrence after therapy with tamoxifen. In the case of *CYP2C19* and the antiplatelet clopidogrel, there is a CPIC recommendation for gPM individuals who, having a higher risk of decreased response, should receive an alternative antiplatelet treatment (e.g., prasugrel or ticagrelor). This is relevant in the context of high variation of the null allele *CYP2C19*2* among Native Americans (independently of Old World admixture), that accounts for most of the *CYP2C19* gPMs observed in our study.

Our study has a limitation shared with most pharmacogenetic studies. *CYP wt* alleles, associated with normal metabolism, are defined by the absence of any of the assayed alleles. *CYP wt* alleles are positively associated with Native American ancestry (**Figures 2 and 3 and Tables 2 and 3**), which could be due, at least in part, to the presence of unknown rare alleles in Native Americans that may be associated with poor or ultra-rapid drug metabolism. In this study, we observed that Uto-Aztecan populations have the highest frequencies of the *CYP2D6*wt* allele. In addition, in classical pharmacogenetics, phenotypes are often predicted on the basis of genotypes, using a classification of alleles based on "activity scores." To assess incongruences in these genotype-based predictions, RIBEF investigators have developed the *CEIBA cocktail*, a "multiplex phenotyping method" for studies on Native Americans and Latin Americans, and reported incongruences in the prediction of the ultra-rapid metabolizer phenotype based solely on genetic data in Mexican Natives, Ecuadorians, and Cubans.[21–23] In addition, a study in Puerto Rico has identified a new rare allele that affect warfarin metabolism in tri-hybrid admixed Puerto Ricans.[24] Thus, further association studies between pharmacokinetic phenotypes and resequencing of *CYP* genes in Native Americans are necessary to discover specific and rare defective-activity or increased-activity alleles in these populations, as well as to better estimate the frequencies of poor and ultra-rapid metabolizer phenotypes.

After *VKORC1* polymorphisms, *CYP2C9* genotypes are consensually included in dosing algorithms for the anticoagulant warfarin, which exemplifies a clinically relevant complex interaction among pharmaco-genotypes, biogeographic ancestry, and ethnic self-classification in the Americas. In the United States, Limdi *et al.*[25] has shown that self-reported race-specific pharmacogenetic algorithms (including *VKORC1*, *CYP2C9*, and *CYP4F2* genotypes) improve the prediction of the warfarin target-range respect to a unique self-reported race-adjusted algorithm. However, this result has not been replicated in a smaller cohort of Brazilians,[26] likely due to a combination of a lower statistical power and the higher degree of admixture of Latin American populations. For Latin American populations, diverse pharmacogenetic algorithms incorporating *CYP2C9* genotypes have been proposed: (i) ignoring self-reported race as predictive variable but with the same predictive power in *white*, *intermediate*, and *black* Brazilians,[27] (ii) including self-reported race of Brazilians as predictive variable,[28] and (iii) including a biogeographic ancestry estimate based on genetic data in Puerto Ricans.[29] Here, we found that the

decreased-activity rare alleles *CYP2C9*2* and **3* are negatively associated with Native American ancestry and positively associated with Europeans (**Table 3**, **Figure 3a**), and the null-activity *CYP2C9*6* allele is absent among Native Americans. Therefore, *CYP2C9* gPM individuals, the most relevant in the case of warfarin-dosing algorithms, are almost absent among Native American populations and only present in admixed Latin American populations.

In conclusion, we have studied the distribution of *CYP2D6*, *CYP2C9*, and *CYP2C19* variants in 33 native and admixed populations from Ibero-America in the context of their biogeographic ancestry inferred using genetic data. This approach allowed to properly characterize Native American populations and to illustrate the intensive history of admixture and the high genetic diversity of Latin American populations in general and of the RIBEF-MESTIFAR dataset in particular. Continental ancestries are the major determinants of variation in frequencies of the increased-activity allele *CYP2C19*17* and consequently *CYP2C19* gUMs, which are negatively associated with Native American ancestry, decreased-activity alleles *CYP2D6*41* and *CYP2C9*2*, positively associated with European ancestry, and decreased-activity alleles *CYP2D6*17* and *CYP2D6*29*, positively associated with African ancestry. Although for rare alleles *CYP2C9*2* (decreased activity) and *CYPC19*17* (increased activity), European admixture during the last ~500 years entirely accounts for their presence in Native American populations, alleles *CYP2D6*5* (null activity), *CYP2D6*-multiplication alleles (increased activity), and *CYP2C9*3* (decreased activity) seem to have been present at low frequency in the Americas before the first arrival of Europeans > 500 years ago. Finally, and independently of Old World admixture, by studying a broad spectrum of Native American populations representative of different ethno-linguistic groups and from the three subcontinents, we show how autochthonous diversity also contribute to define the allele and predicted phenotype frequencies distribution of *CYP2D6*, *CYP2C9*, and *CYP2C19* in the Americas, as well as their clinically relevant associated traits, such as plasma concentrations for paroxetine, nortriptyline, and tamoxifen, toxicity induced by codeine, maintenance of target warfarin dose, and response to treatment with clopidogrel.

METHODS
Subjects, populations, and groups
From the 6,060 individuals genotyped for *CYP2D6*, *CYP2C9*, or *CYP2C19* polymorphisms,[13] 3,387 were analyzed for individual ancestry in the present study, with the following distribution (**Table 1**): (i) 1,051 Native Americans from Mexico, Costa Rica, and Peru, representing North, Central, and South America, respectively—they live in rural populations, are locally recognized as indigenous, and are settled in regions where the population is predominantly indigenous; (ii) 38 Afro-Latin Americans who self-reported as Afro-descendants in Costa Rica and reported to have four black grandparents in Cuba; (iii) 163 self-reported Ashkenazi Jews from Argentina; (iv) 206 Cubans that self-reported to have four white grandparents; (v) 1,442 admixed (*Mestizo*) in the Hispanic America; and (vi) 371 Spaniards and 116 Portuguese. For Native American populations, we followed the ethno-linguistic classification by Campbell[15] (**Table 1**). University of Extremadura and local institutional review boards approved the use of studied sample for the present study.

Continental ancestry analysis

To estimate African, European, and Native American individual ancestry, 83 AIMs[30,31] were genotyped (**Table S1**). Because these AIMs have an average differentiation (F_{CT}[32]) of 44.6% among Africans, Europeans, and Native Americans, which is considerably higher than the average intercontinental differentiation between human populations (12%),[33] they provide a fair assessment of individual continental admixture for Latin American individuals. Indeed, panels with > 80 AIMs provide good estimates of continental admixture in Latin Americans.[34] AIMs genotyping was performed at the Spain National Genotyping Center (CEGEN) from Santiago de Compostela, using the Sequenom (San Diego, CA) platform, and for part of Brazilian and Peruvian samples at the *Laboratório Multiusuário de Genômica* from the Federal University of Minas Gerais using the Bead-Xpress Illumina (San Diego, CA) platform.

The admixture proportions were estimated using the model-based method implemented in the software *Admixture*,[35] assuming a tri-hybrid model ($K = 3$) and performing unsupervised inferences. In addition to the 3,387 RIBEF-MESTIFAR individuals, 119 unrelated African Yoruba individuals from the 1000 Genomes Project were included for these analyses.[36] To represent the genetic structure of the studied individuals, we performed Principal Component Analysis[37] of individual AIMs genotypes.

CYPs genotyping methods and classification in predicted metabolic phenotypes

Genotype analysis, allelic classification, and phenotype predictions were carried out for *CYP2D6* ($n = 5,992$ individuals), *CYP2C9* ($n = 5,609$ individuals), and *CYP2C19* ($n = 5,220$ individuals), alleles as in Naranjo *et al.*[13] Briefly, we used real time-PCR TaqMan (ThermoFisher Scientific, Waltham, MA) allelic discrimination assays to identify the null-enzymatic activity alleles *CYP2D6*3*, **4*, **5*, **6*; *CYP2C9*6*; *CYP2C19*2*, **3*, **4*, **5*; decreased-activity alleles *CYP2D6*10*, **17*, **29*, **41*; *CYP2C9*2*, **3*; normal alleles *CYP2D6*2*, **35*; and the increased activity allele *CYP2C19*17* as well as the multiplications of active *CYP2D6* alleles (https://www.pharmvar.org/). *CYP2D6* allele multiplications (*wtxN*, **2xN*, and **35xN*) are presented as a unique class because their discrimination is only possible if one of these alleles is in a diplotype with **10* or **4*, which occurs in very few samples. The presence of *CYP2D6*, *CYP2C9*, and *CYP2C19 wt* alleles was defined by the absence of the abovementioned genotyped polymorphisms.

The combination of *CYP* alleles in an individual can result in different enzymatic capacities, which allow to predict metabolic phenotype groups: individuals carrying two nonactive *CYP2D6*, *CYP2C9*, or *CYP2C19* alleles are classified as predicted gPMs and are predicted to have no metabolic capacity for these enzymes. Subjects carrying more than two active *CYP2D6* alleles, or the *CYP2C19*17* allele without nonactive *CYP2C19* alleles, have been related to increased enzyme activity and are denominated *CYP2D6* or *CYP2C19* predicted gUMs, respectively (**Table S2**).

Statistical analysis

The frequencies of *CYP2D6*, *CYP2C9*, and *CYP2C19* alleles, as well as those of *CYP2D6*, *CYP2C9*, and *CYP2C19* predicted metabolic phenotypes, were calculated for each population. Hardy-Weinberg equilibrium was tested for each allele in each population using the χ^2 test with Yates' correction. A linear regression (lm function in R environment) was used to describe the dependence of population allele frequencies on Native American, European, and African ancestry at population levels. For each allele, we estimated the linear regression coefficient beta on each continental ancestry, its significance, as well as the percentage of the allele frequencies variance explained by each continental ancestry (R^2).

To infer if specific alleles usually observed in Europeans were also present in the Native Americans before the first arrival of Europeans ~500 years ago, we focused on the 13 populations ($N = 1,067$ individuals) with > 90% of Native American ancestry (mean: 94.4% of Native American ancestry, 4.3% of European ancestry, and 1.3% of African ancestry; **Table S2**). In these populations, we estimated the expected number of copies of the specific allele (EI) under the hypothesis that all copies were introduced by Europeans, using the formula: $EI = f_{eu} \times 0.056 \times 2N$, where f_{eu} is the average allele frequency in Spain and Portugal populations, and 0.056 is the Old World ancestry. We compared EI with the observed copies (O). If O > EI, then Old World introgression cannot account for the copies observed in the 13 Native American populations, and, therefore, this result suggest the presence of the allele in the Americas before the arrival of Europeans ~500 years ago.

CONSORTIA

Consortia affiliations are presented in **Text S1** (**Supplementary file**).

Catalina Altamirano Tinoco, Mayra Álvarez, Angélica Borbón, Carolina Céspedes-Garro, Jesús Cobaleda, Fernando de Andrés, René Delgado, Pedro Dorado, Humberto Fariñas, Verónica Ferreiro, Ingrid Fricke-Galindo, Carlos Galaviz-Hernández, Lourdes Garza-Ocañas, Robert H. Gilman, Francisco Hernández, Gerardo Jiménez-Arce, Helgi Jung-Cook, Ismael Lares-Aseff, Blanca P. Lazalde-Ramos, Marisol López-López, Lucas Michelin, Nancy Monroy-Jaramillo, María Eugenia G. Naranjo, Alberto Ortega-Vázquez, Rocío Ortiz-López, Bárbaro Pérez, Yadira X. Pérez-Páramo, Diadelis Remirez, Augusto Rojas-Martínez, Alba P. Sarmiento, Marilia Scliar, Santiago Terán, and Roxana Zamudio.

SUPPORTING INFORMATION

Supplementary information accompanies this paper on the *Clinical Pharmacology & Therapeutics* website (www.cpt-journal.com).

Table S1. Eighty-three Ancestry Informative Markers genotyped to infer continental ancestry.
Table S2. Allele and predicted-phenotype frequencies for *CYP2D6*, *CYP2C9*, and *CYP2C19* for the 33 studied populations.
Text S1. CEIBA-RIBEF Consortium of coauthors, listed by alphabetical order, with affiliations.

CONFLICT OF INTEREST

The authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

F.R.-S., E.P.-L., E.T.-S., and A.LL. wrote the manuscript. A.LL. designed the research. F.R.-S., E.P.-L., E.T.-S., and A.LL. performed the research. F.R.-S. analyzed the data. M.S.-M., E.T., M.L.-L., I.R., G.E.M., L.R.C., R.R.-R., M.G., F.E.E., and R.B. contributed new reagents/analytical tools.

1. LLerena, A., Berecz, R., de la Rubia, A. & Dorado, P. QTc interval lengthening is related to CYP2D6 hydroxylation capacity and plasma concentration of thioridazine in patients. *J. Psychopharmacol.* **16**, 361–364 (2002).
2. Berecz, R. *et al.* Relationship between risperidone and 9-hydroxy-risperidone plasma concentrations and CYP2D6 enzyme

activity in psychiatric patients. *Pharmacopsychiatry* **35**, 231–234 (2002).

3. Peñas-Lledó, E.M. *et al*. CYP2D6 ultrarapid metabolism and early dropout from fluoxetine or amitriptyline monotherapy treatment in major depressive patients. *Mol. Psychiatry* **18**, 8–9 (2013).

4. Dorado, P., López-Torres, E., Peñas-LLedó, E.M., Martínez-Antón, J. & LLerena, A. Neurological toxicity after phenytoin infusion in a pediatric patient with epilepsy: influence of CYP2C9, CYP2C19 and ABCB1 genetic polymorphisms. *Pharmacogenomics J.* **13**, 359–361 (2013).

5. Peñas-Lledó, E. *et al*. A combined high CYP2D6-CYP2C19 metabolic capacity is associated with the severity of suicide attempt as measured by objective circumstances. *Pharmacogenomics J.* **15**, 172–176 (2015).

6. Jukić, M.M. *et al*. Elevated CYP2C19 expression is associated with depressive symptoms and hippocampal homeostasis impairment. *Mol. Psychiatry* **22**, 1155–1163 (2017).

7. LLerena, A. *et al*. Interethnic variability of CYP2D6 alleles and of predicted and measured metabolic phenotypes across world populations. *Expert Opin. Drug Metab. Toxicol.* **10**, 1569–1583 (2014).

8. Céspedes-Garro, C. *et al*. Worldwide interethnic variability and geographical distribution of CYP2C9 genotypes and phenotypes. *Expert Opin. Drug Metab. Toxicol.* **11**, 1893–1905 (2015).

9. Fricke-Galindo, I. *et al*. Interethnic variation of CYP2C19 alleles, 'predicted' phenotypes and 'measured' metabolic phenotypes across world populations. *Pharmacogenomics J.* **16**, 113–123 (2016).

10. Huddart, R. *et al*. Standardized biogeographic grouping system for annotating populations in pharmacogenetic research. *Clin. Pharmacol. Ther.* **105**, 1256–1262 (2018).

11. Soares-Souza, G., Borda, V., Kehdy, F. & Tarazona-Santos, E. Admixture, genetics and complex diseases in Latin Americans and US Hispanics. *Curr. Genet. Med. Rep.* **6**, 208–223 (2018).

12. Dorado, P. *et al*. CYP2C9 allele frequency differences between populations of Mexican-Mestizo, Mexican-Tepehuano and Spaniards. *Pharmacogenomics J.* **11**, 108–112 (2011).

13. Naranjo, M.-E.G. *et al*. Interethnic variability in *CYP2D6*, *CYP2C9*, and *CYP2C19* genes and predicted drug metabolism phenotypes among 6060 Ibero- and Native Americans: RIBEF-CEIBA Consortium Report on Population Pharmacogenomics. *OMICS* **22**, 575–588 (2018).

14. Sosa-Macías, M. *et al*. Pharmacogenetics and ethnicity: relevance for clinical implementation, clinical trials, pharmacovigilance and drug regulation in Latin America. *Pharmacogenomics* **17**, 1741–1747 (2016).

15. Campbell, L. *American Indian Languages: The Historical Linguistics of Native America. Oxford Studies in Anthropological Linguistics* (Oxford University Press, New York, NY, 1997).

16. Moya, G. *et al*. High frequency of CYP2D6 ultrarapid metabolizer genotypes in an Ashkenazi Jewish population from Argentina. *Pharmacogenomics J.* **17**, 378–381 (2017).

17. Céspedes-Garro, C. *et al*. Relevance of the ancestry for the variability of the drug-metabolizing enzymes CYP2C9, CYP2C19 and CYP2D6 polymorphisms in a multiethnic Costa Rican population. *Rev. Biol. Trop.* **64**, 1067–1076 (2016).

18. Sirugo, G., Williams, S.M. & Tishkoff, S.A. The missing diversity in human genetic studies. *Cell* **177**, 26–31 (2019).

19. Dillehay, T.D. *The Settlement of the Americas: A New Prehistory* (Basic Books, New York, NY, 2000).

20. Cavalli-Sforza, L.L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton University Press, Princeton, NJ, 1996).

21. De Andrés, F., Terán, S., Hernández, F., Terán, E. & LLerena, A. To genotype or phenotype for personalized medicine? CYP450 Drug metabolizing enzyme genotype–phenotype concordance and discordance in the Ecuadorian population *OMICS* **20**, 699–710 (2016).

22. de Andrés, F., Sosa-Macías, M., Ramos, B.P.L., Naranjo, M.-E.G. & LLerena, A. CYP450 genotype/phenotype concordance in Mexican Amerindian indigenous populations–where to from here for global precision medicine? *OMICS* **21**, 509–519 (2017).

23. Dorado, P. *et al*. Lessons from Cuba for global precision medicine: *CYP2D6* genotype is not a robust predictor of CYP2D6 ultrarapid metabolism. *OMICS* **21**, 17–26 (2017).

24. Claudio-Campos, K.I. *et al*. CYP2C9*61, a rare missense variant identified in a Puerto Rican patient with low warfarin dose requirements. *Pharmacogenomics* **20**, 3–8 (2019).

25. Limdi, N.A., Wiener, H., Goldstein, J.A., Acton, R.T. & Beasley, T.M. Influence of CYP2C9 and VKORC1 on warfarin response during initiation of therapy. *Blood Cells Mol. Dis.* **43**, 119–128 (2015).

26. Suarez-Kurtz, G. & Struchiner, C.J. Differential effects of predictors of warfarin dose according to race/color categories in the admixed Brazilian population. *Pharmacogenet. Genomics* **27**, 210–211 (2017).

27. Perini, J.A. *et al*. Pharmacogenetics of warfarin: development of a dosing algorithm for Brazilian patients. *Clin. Pharmacol. Ther.* **84**, 722–728 (2008).

28. Santos, P.C.J.L. *et al*. Development of a pharmacogenetic-based warfarin dosing algorithm and its performance in Brazilian patients: highlighting the importance of population-specific calibration. *Pharmacogenomics* **16**, 865–876 (2015).

29. Duconge, J. & Ruaño, G. Admixture and ethno-specific alleles: missing links for global pharmacogenomics. *Pharmacogenomics* **17**, 1479–1482 (2016).

30. Yaeger, R. *et al*. Comparing genetic ancestry and self-described race in African Americans born in the United States and in Africa. *Cancer Epidemiol. Biomarkers Prev.* **17**, 1329–1338 (2008).

31. Pereira, L. *et al*. Socioeconomic and nutritional factors account for the association of gastric cancer with Amerindian ancestry in a Latin American admixed population. *PLoS One* **7**, e41200 (2012).

32. Excoffier, L., Smouse, P.E. & Quattro, J.M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491 (1992).

33. Barbujani, G. & Colonna, V. Human genome diversity: frequently asked questions. *Trends Genet.* **26**, 285–295 (2010).

34. Galanter, J.M. *et al*. Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. *PLoS Genet.* **8**, e1002554 (2012).

35. Alexander, D.H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

36. Auton, A. *et al*. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).

37. Patterson, N., Price, A.L. & Reich, D. Population structure and Eigen analysis. *PLoS Genet.* **2**, e190 (2006).