UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA GERAL
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA

Tese de Doutorado

Farmacogenética populacional e ancestralidade na América Latina

Autora: Fernanda Rodrigues Soares

Orientador: Dr. Eduardo Martín Tarazona Santos

BELO HORIZONTE

Maio – 2016

Fernanda Rodrigues Soares

Farmacogenética populacional e ancestralidade na América Latina

Tese apresentada ao Programa de Pós-graduação em Genética do Departamento de Biologia Geral do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como pré-requisito para obtenção do Grau de Doutora em Genética.

Orientador: Dr. Eduardo Martín Tarazona Santos

Belo Horizonte

Maio – 2016

043 Soares, Fernanda Rodrigues.
Farmacogenética populacional e ancestralidade na América Latina [manuscrito] / Fernanda Rodrigues Soares. – 2016.

223 f. : il. ; 29,5 cm.

Orientador: Dr. Eduardo Martín Tarazona Santos.
Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas.

1. Farmacogenética - Teses. 2. Genética de populações - Teses. 3. Ancestralidade. 4. Polimorfismo de nucleotídeo único. I. Tarazona Santos, Eduardo Martín. II. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. III. Título.

CDU: 575

Ficha catalográfica elaborada pela Biblioteca do Instituto de Ciências Biológicas da UFMG

Programa de Pós-Graduação em Genética
Departamento de Biologia Geral, ICB
Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627 - C.P. 486 - Pampulha - 31270-901 - Belo Horizonte - MG
e-mail: pg-gen@icb.ufmg.br FAX: (+31) - 3409-2570

"Farmacogenética populacional e ancestralidade na América Latina"

Fernanda Rodrigues Soares

Tese aprovada pela banca examinadora constituída pelos Professores:

Dr. Eduardo Martin Tarazona Santos - Orientador
(UFMG)

Dra. Daniela Rezende Garcia Junqueira
(Evidências em Saúde/MG)

Dra. Maria Bernadete Lovato
UFMG

Dra. Maria Clara Fernandes da S. Malta
HEMOMINAS

Dr. Renan Pedra de Souza
UFMG

Belo Horizonte, 30 de maio de 2016.

AGRADECIMENTOS

Nenhum trabalho é realizado exclusivamente uma pessoa, por isso gostaria de agradecer a várias pessoas que contribuíram diretamente e indiretamente com a tese e meu período de doutorado:

Ao meu orientador, Dr. Eduardo Tarazona, que me orientou não somente nesta tese mas em vários aspectos da minha vida profissional (algumas vezes até mesmo pessoal) e contribuiu de forma extremamente significativa para meu amadurecimento pessoal e científico.

Ao meu orientador no exterior, Dr. Adrián LLerena, por me dar a oportunidade de fazer parte de seu grupo de pesquisa, por toda a confiança em mim depositada e pelos grandes ensinamentos de ciência geral, farmacogenética e escrita de artigos.

À Capes, CNPq e Fapemig, pelo financiamento dos projetos e bolsas usufruídas.

Aos professores do Programa de Pós Graduação em Genética e Bioinformática, pelos ensinamentos ministrados.

Aos amigos do LDGH (e digo amigos porque realmente fiz grandes amigos neste grupo), pelos quase 6 anos de convivência diária e grandes momentos compartilhados, em especial:

À Fernanda Kehdy (Ferdi). Grande amiga e companheira que levarei pelo resto da vida. Não temos só o nome em comum, mas muitas características de personalidade. Você é como uma irmã, e sempre será. Seus conselhos tanto profissionais como pessoais me ajudaram a amadurecer e ver a vida de uma forma mais consciente. Você tem participação fundamental em todo esse processo, praticamente me co-orientou, tanto cientificamente como pessoalmente.

À Roxana, também grande amiga companheira que sempre manterei contato. Compartilhamos não só artigos, mas também a casa e muitos momentos de alegria, tristeza, preocupação e confidências. Sou muito feliz por ter sua amizade.

Ao Gil, amigo de todos os dias e de grandes "*conseeelhos*". Obrigada por sempre me salvar no R e também por todos os assuntos pessoais que você sempre me ajudou a superar.

Ao Mateus, grande parceiro não só de análises, mas também de músicas, viagens e longas e sábias conversas.

À Moara (Momô), que sempre foi tão paciente com todas as minhas dúvidas de bioinformática e tão ouvinte e consoladora com meus problemas. Obrigada por tudo.

Ao Rennan, que compartilha comigo a primeira autoria do segundo manuscrito do Capítulo 3 desta tese e consequentemente compartilhou todo o trabalho, desde 2011. Obrigada por isso e por todas as grandes conversas que tivemos. Você sempre soube exatamente o que eu precisava ouvir com sua grande sabedoria.

À Nath, por ter me incentivado a ter melhores hábitos de vida e também me ajudado com assuntos bioinformáticos e pessoais. Nossa convivência aumentou muito em 2015 e sou muito grata por todas as nossas conversas sempre tão divertidas e que fizeram o ambiente de trabalho se tornar mais alegre.

Ao Wagner, por todos os ensinamentos bioinformáticos desde o básico no mestrado até questões mais complexas no fim do doutorado.

Às "minhas meninas", queridas alunas de iniciação científica, Julia e Poliana, que tiveram participação fundamental no manuscrito do Capítulo 2 desta tese e sempre fizeram seu trabalho com esforço e dedicação, além de terem me proporcionado a experiência de ser orientadora, o que contribuiu muito para meu amadurecimento científico.

A todos os membros do laboratório que fazem/fizeram com que nosso grupo seja tão unido, alegre e único, Giordano, Hanaisa, Camila, Marília, Maíra, Thiago, Victor e Meddly. Obrigada por tudo!

A todos los compañeros de CICAB, Caro, Mariu, Bea, Inés, Áurea, Fernando, Paloma, Paco, Tere, Humberto y Marga. Muchas gracias por todos los momentos, desayunos, paciencia con mi diferente cultura y gran receptividad. Especialmente, agradezco a Carolina Céspedes (Caro), por compartir la primera autoría (y el trabajo) del primer manuscrito del Capítulo 3 de esta tesis y por todo lo que me ha ayudado personalmente en España. Eres una gran amiga que ha hecho tanto por mi que nunca tendré maneras suficientes para agradecerte, especialmente cuando me quedé sin caminar. Creo que no podría quedarme bien en Badajoz por casi un año sin ti y tu ayuda desde el principio. Muchas gracias por todos los momentos compartidos, que siempre estarán conmigo.

Ao Genepop, em especial Lu e Rê, que convivem comigo desde o mestrado sempre tão simpáticas, solícitas e companheiras em todos os momentos.

Aos meus amigos pessoais de Minas, “amigos do Allan”, por todos os momentos de descontração e incentivo, que fizeram minha estadia em Minas mais divertida e agradável, o que impactou consideravelmente na elaboração deste trabalho. Em especial agradeço ao meu primo Lucas Soares, por toda a companhia, parceria musical, grandes conversas e momentos de irmandade.

Ao meu namorado Emerson, pela infinita fonte de paciência, apoio, conselhos, companheirismo, confidências, amor e dedicação oferecidos a mim em todos os momentos. Não consigo imaginar como seria minha vida aqui sem você, especialmente nesta etapa final. Seu apoio foi essencial para a finalização deste trabalho.

À minha família, pai, Ana Paula, Matheus, Vinícius, Jac, Yasmin, vovó Zezé, tia Luzia e família, tia Luiza e família, tio Carlinho e família, tio Taquinho, tia Denise e família, Deolina, Wellington e Cicília, muito obrigada pelo ponto de apoio que tive sempre com vocês, que foi essencial para minha maravilhosa estadia em Minas e consequente elaboração desta tese. Meu avô, antes de partir, questionou meu objeto de pesquisa e se encantou (fato dificílimo de ocorrer) com a farmacogenética e me incentivou muito com a realização do mesmo, obrigada pelo apoio.

À minha mãe, meu espelho e porto seguro, a grande responsável pela pessoa que sou hoje e por todas as minhas conquistas. Por ter me apoiado, me acompanhado em todos os momentos e me colocado pra fazer tantos cursos além do básico para que eu tivesse uma melhor formação durante toda a vida, que foi minha base para conseguir chegar até aqui. Esta tese é dedicada a você.

SUMÁRIO

LISTA DE FIGURAS

Capítulo 1

Genomic ancestry and ethnoracial self-classification based on 5,871 community-dwelling Brazilians (The Epigen Initiative)

Genomic ancestry and ethnoracial self-classification based on 5,871 community-dwelling Brazilians (The Epigen Initiative)

Capítulo 2

Interethnic variability of CYP2D6 alleles and of predicted and measured metabolic phenotypes across world populations

Genetic structure of pharmacogenetic biomarkers in Brazil inferred from a systematic review and population-based cohorts: a RIBEF/EPIGEN-Brazil initiative

Capítulo 3

LISTA DE TABELAS

Capítulo 1

Genomic ancestry and ethnoracial self-classification based on 5,871 community-dwelling Brazilians (The Epigen Initiative)

Capítulo 2

Interethnic variability of CYP2D6 alleles and of predicted and measured metabolic phenotypes across world populations

Genetic structure of pharmacogenetic biomarkers in Brazil inferred from a systematic review and population-based cohorts: a RIBEF/EPIGEN-Brazil initiative

Capítulo 3

LISTA DE ABREVIATURAS

ADME	Absorção, distribuição, metabolização e excreção
AIMs	Marcadores informativos de ancestralidade
AYM	População Aymara, Peru
BRA	População de miscigenados de Minas Gerais, Brasil
CEGEN	Centro Nacional de Genotipado
CEIBA	Consorcio Europeo Ibero Americano de Farmacogenética
CICAB	Centro de Investigación Clínica del Área de Salud Badajoz
CPs	Componentes principais
DMEs	Enzimas metabolizadoras de drogas
DNA	Ácido desoxirribonucléico
EM	*Expectation Maximization*
IBGE	Instituto Brasileiro de Geografia e Estatística
LIM	População de miscigenados de Lima, Peru
LLA	Leucemia linfoide aguda
ML	*Maximum Likelihood*
NGS	*Next Generation Sequencing*
PCA	Análise de componentes principais
PREDICT	*Pharmacogenomic Resource for Enhanced Decisions in Care and Treatment*
REFARGEN	Rede Brasileira de Farmacogenética
RIBEF	Rede Iberoamericana de Farmacogenética e Farmacogenômica
SHI	População Shimaa, Peru
SNP	*Single Nucleotide Polymorphism*

RESUMO

Frequências de polimorfismos nos genes que codificam enzimas metabolizadoras de fármacos variam consideravelmente entre diferentes grupos étnicos. Em populações miscigenadas como as latino-americanas, o conhecimento de características farmacogenéticas de diferentes grupos étnicos pode ser relevante para a implementação de estratégias terapêuticas adaptadas a eles. O objetivo desta tese é verificar a ancestralidade e a frequência das variantes de interesse farmacogenético em populações latino-americanas, visando esclarecer os padrões destas frequências na América Latina. A partir de dois estudos de miscigenação em populações brasileiras inseridas no projeto EPIGEN, concluímos que as coortes do Sudeste/Sul são predominantemente europeias (do norte Europeu/Oriente Médio) e menos africanas (do centro-leste da África), enquanto a coorte do Nordeste possui prevalência de ancestralidade africana (do centro-oeste da África) e menos europeia (da Península Ibérica). Além disso, encontramos que há correlação entre ancestralidade genômica e etnia autodeclarada em brasileiros. Para a caracterização de frequências das principais variantes de interesse farmacogenético em populações mundiais, apresentamos três revisões sistemáticas considerando genes *CYP2D6*, *CYP2C9* e *CYP2C19* constatando que existem padrões mundiais na distribuição de frequências alélicas. No contexto da América Latina, apresentamos uma revisão sistemática de 121 biomarcadores farmacogenéticos no Brasil, verificando que padrões nesta população são complexos. Além disso, apresentamos um estudo com dados originais da Rede Iberoamericana de Farmacogenética, particularmente na Costa Rica, mostrando uma correlação significativa entre ancestralidade genômica e CYPs em três populações deste país. Adicionalmente, a partir de um estudo de admixture mapping que associa ancestralidade ameríndia com risco aumentado para recidiva em leucemia linfoide aguda devido a dois SNPs, verificamos que esta região genômica está diferenciada em populações latino-americanas, especialmente em nativos. Portanto, em um contexto populacional e epidemiológico, as informações de cor/raça e ancestralidade genômica são informativas para inferências farmacogenéticas. No entanto, para aplicação clínica da farmacogenética em latino-americanos, a genotipagem individualizada deve ser realizada pois a resposta não pode ser predita pela ancestralidade.

ABSTRACT

Polymorphism frequencies in genes coding for drug metabolizing enzymes vary among distinct ethnic groups. In admixed populations, such as Latin Americans, the knowledge of pharmacogenetic characteristics for distinct Latin American ethnic groups may be relevant for the implementation of therapeutic strategies adapted to them. The goal of this dissertation is to verify the ancestry and the frequency of pharmacogenetic variants in Latin American populations, aiming to clarify frequency patterns in Latin America. In two admixture studies using Brazilian populations comprising EPIGEN project we concluded that Southeast/South cohorts are predominantly Europeans (from North Europe/Middle East) and less Africans (from Center-East Africa), whereas Northeast cohort is prevalently African (from Center-West Africa) and less European (from Iberic Peninsula). Further, we found a correlation between genomic and self-reported ancestry in Brazilians. To characterize frequencies of the main variants from worldwide populations, we presented three systematic reviews considering CYP2D6, CYP2C9 and CYP2C19 genes, concluding that there are worldwide patterns in allele frequencies distributions. In a Latin American context, we showed in a systematic review using 121 pharmacogenetic biomarkers that patterns in this population are complicated due to its complex demographic history. Moreover, we presented an article using original data from the Iberoamerican Pharmacogenetics Network, particularly in Costa Rica, showing a significant correlation between genomic ancestry and CYPs in three populations from this country. Furthermore, from an admixture mapping study which associated Native American ancestry with increased risk of acute lymphoblastic leukemia relapse due to two SNPs, we verified that this genomic region is differentiated in Latin American populations, especially Native Americans. So, in a population and epidemiologic context, self-reported and genomic ancestry are informative to pharmacogenetic inferences. However, pharmacogenetic clinical implementation in Latin Americans must be performed by individual genotyping because drug response cannot be predicted by ancestry.

INTRODUÇÃO

As populações atuais da América Latina são o resultado de várias contribuições biológicas e culturais de populações parentais ao longo do tempo. Os três maiores grupos populacionais que contribuíram com esta ancestralidade diversa foram europeus, africanos e ameríndios. Por isso, as populações da América Latina podem ser consideradas como tri-híbridas. Outras contribuições menores, com grupos parentais como leste-asiáticos por exemplo, também ocorreram ao longo do tempo e torna a história da América Latina mais complexa (Salzano & Bortolini 2002; Galanter et al. 2012).

As contribuições de populações parentais ao longo do tempo são uma forma de fluxo gênico entre populações que estiveram isoladas por um longo período de tempo. Este processo é denominado miscigenação, e a população híbrida é chamada miscigenada (Tarazona-Santos et al. 2007).

Cada indivíduo, sendo parte de populações miscigenadas ou não, possui uma etnia autodeclarada. Um indivíduo pode se autodeclarar de uma determinada etnia de acordo com características fenotípicas e culturais. De acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE), por exemplo, no Brasil temos classificações de cor ou raça definidas como Branca, Preta, Parda, Amarela ou Indígena (IBGE n.d.). A cor ou raça autodeclarada individual é aquela em que o próprio indivíduo se enquadra em uma destas categorias. Dessa forma, a cor/raça autodeclarada populacional seria a proporção de um conjunto de indivíduos autodeclarados da mesma cor ou raça; por exemplo, a população brasileira possui 47,5% Brancos, 43,4% Pardos, 7,5% Pretos, 1,1% Amarelos e 0,4% Indígenas (IBGE 2010).

Entretanto, para elucidar os processos de miscigenação sem o viés fenotípico e cultural, a origem ancestral de determinadas populações pode ser determinada pela genética, com a produção e análise de dados de DNA. Os dados genéticos das populações miscigenadas são comparados com dados de populações parentais, com auxílio de análises estatísticas, assim, pode-se inferir padrões de migração e acasalamento dessas populações. Apesar de por vezes haver uma associação, a ancestralidade genômica não está sempre correlacionada com a etnia autodeclarada, especialmente em populações com alto índice de miscigenação (Pena et al. 2011; Lima-Costa et al. 2015)

Os estudos tradicionais de inferências de ancestralidade genômica costumavam enfocar apenas a miscigenação populacional. Atualmente, a disponibilidade de um maior número de marcadores genotipados com custos razoáveis permite o estudo da miscigenação de cada indivíduo, cuja distribuição é informativa em relação não unicamente à quantidade de miscigenação, mas à sua dinâmica, ou seja, como ocorreu o processo de miscigenação no tempo.

Existem vários tipos de marcadores moleculares utilizados para estimar a miscigenação individual. Eles podem ser classificados quanto à sua localização: Cromossomos autossômicos (herança biparental), DNA mitocondrial e Cromossomo Y (herança uniparental); e também quanto à natureza molecular: Microssatélites (repetições em tandem de sequências de nucleotídeos), Indels (polimorfismos de inserção e deleção) e SNPs (*Single Nucleotide Polymorphisms* – variação em um único nucleotídeo de DNA).

Os SNPs são os marcadores mais prevalentes do genoma (Anon 2003). Atualmente eles são vastamente utilizados em estudos genéticos pois há um grande número já caracterizado e disponível em grandes bancos de dados, como o Projeto Internacional Hapmap (The International HapMap Consortium, 2010, www.hapmap.org), 1000 Genomes (http://www.1000genomes.org/) e dbSNP (http://www.ncbi.nlm.nih.gov/SNP/). A disponibilidade de milhões de SNPs nestes bancos de dados permite uma maior proximidade à ancestralidade real dos indivíduos inclusive no nível de cromossomos (Via et al. 2009).

Podemos encontrar estudos com estimativas de miscigenação populacionais, individuais e ao longo dos cromossomos, assim como por continentes, regiões subcontinentais e populações. Todas essas estimativas possuem erros embutidos, pois é complicado obter amostras das populações parentais genuinamente representativas de uma população, região ou do mundo (Cavalli-Sforza 1998), além do erro amostral e das diferenças entre as frequências no momento da amostragem e aquelas no passado nas populações parentais e híbridas devido à deriva genética (flutuação aleatória ao acaso das frequências alélicas).

Para estimar as ancestralidades populacionais, individuais e ao longo de cromossomos, há duas principais abordagens de inferência:

A primeira abordagem diz respeito à análise de marcadores informativos de ancestralidade (AIMs). AIMs são marcadores com frequências alélicas altamente diferenciadas entre populações parentais, e por isso são considerados informativos para estimar a ancestralidade de indivíduos miscigenados (Yaeger, Avila-Bront, et al. 2008; Galanter et al. 2012). Estes marcadores existem porque o isolamento geográfico prolongado de uma população gera diferenças em suas frequências alélicas quando comparada a outras populações. Isso ocorre por deriva genética e mutações exclusivas daquela população, que pode gerar alelos específicos na mesma.

Se um número suficiente de AIMs for caracterizado, podemos inferir a ancestralidade em níveis de resolução refinados, como individual e ao longo de cromossomos, não apenas populacional. O número de AIMs necessários para inferir ancestralidade com um determinado nível de acurácia é menor que o número de marcadores selecionados ao acaso, o que facilita e diminui o custo das análises (Rosenberg et al. 2002; Galanter et al. 2012).

AIMs utilizados em populações latino-americanas para estimar as ancestralidades no nível continental visam minimizar a heterogeneidade intracontinental e aumentar a diferença intercontinental (Galanter et al. 2012). Por isso, em estudos com AIMs esperamos sempre que as frequências alélicas dentro de cada grupo parental sejam semelhantes e entre as populações sejam diferentes.

Para populações latino-americanas especificamente, não há um único conjunto de marcadores ótimos, porque a informatividade dos marcadores depende da combinação das frequências alélicas das populações parentais e da proporção de miscigenação nessas populações (Pfaff et al. 2004). Marcadores razoavelmente bons para latino-americanos são, então, aqueles que mostram uma frequência alélica bastante divergente entre populações africanas, europeias e ameríndias.

A segunda abordagem diz respeito aos marcadores *genome-wide*. A disponibilidade de milhões de SNPs em bancos de dados públicos é resultado do decréscimo exponencial no custo da genotipagens devido às novas tecnologias de *arrays* e *next-generation sequencing* (NGS). Com estes dados, é possível inferir a ancestralidade individual, populacional e cromossômica se uma grande densidade de SNPs estiver disponível para as amostras de interesse. A ancestralidade individual e populacional é

inferida da mesma forma que a análise com AIMs, no entanto com uma maior quantidade de marcadores que não são necessariamente AIMs, pois a densidade *genome-wide* possui SNPs diferenciados que naturalmente separam os componentes parentais de ancestralidade.

A miscigenação individual pode ser calculada pelo método de *Maximum Likelihood* (ML). O método de ML é um procedimento estatístico amplamente utilizado em diversas áreas biológicas. Ele consiste em escolher um modelo probabilístico apropriado para ser atribuído aos dados observados, e em estimar os parâmetros desse modelo, o que é feito com a técnica de *Expectation Maximization* (EM) (Do & Batzoglou 2008). No caso de estimativas de ancestralidade individual, este modelo (ML+EM) considera variáveis demográficas, como a adequação ao Equilíbrio de Hardy-Weinberg nas frequências alélicas das populações parentais (K) que deram origem à população estudada (Wollstein & Lao 2015). Quando o número de marcadores e de amostras é amplo, em geral proveniente de NGS, o método de ML, além de ser acrescido do algoritmo EM, não considera o desequilíbrio de ligação, o que otimiza a velocidade computacional. Um exemplo desta adaptação é o programa *ADMIXTURE* (Alexander et al. 2009), utilizado para calcular a miscigenação individual em todos os trabalhos contidos nesta tese.

A ancestralidade cromossômica pode ser feita a partir de uma alta densidade de SNPs, que normalmente é alcançada apenas com abordagem *genome-wide*. A alta densidade de SNPs disponível em cada um dos cromossomos permite que se faça a inferência de componentes parentais de ancestralidade em trechos cromossômicos específicos. Esta abordagem é especialmente útil quando se trata de populações miscigenadas e se deseja saber a ancestralidade de um gene ou trecho de cromossomo específico, que pode não coincidir com a ancestralidade individual pelo fato de um indivíduo miscigenado ser um mosaico dos componentes de suas populações parentais.

A ancestralidade é um fator determinante em muitas áreas de interesse clínico, como por exemplo maior suscetibilidade ao câncer em pessoas de ancestralidade europeia (Fejerman et al. 2010), maior predisposição a asma em indivíduos com alta ancestralidade africana (Flores et al. 2012) e variabilidade em resposta clínica a fármacos em indivíduos de diferentes etnias (Limdi et al. 2015).

A resposta clínica a fármacos varia entre indivíduos devido a fatores como dieta, patologias, idade, sexo e genética. O estudo das variantes genéticas que influenciam a resposta a medicamentos é denominado farmacogenética, e, mais especificamente, a farmacogenética populacional analisa diferenças populacionais dessas variantes genéticas e sua relação com resposta a fármacos.

A variabilidade interindividual na resposta a terapias medicamentosas implica em respostas clínicas reduzidas ou incidência aumentada de reações adversas aos medicamentos. Diferenças de concentração nos níveis plasmáticos de fármacos e sub-produtos do seu metabolismo variam entre indivíduos após a ingestão da mesma dose de um fármaco.

O metabolismo de eliminação dos fármacos ocorre por sistemas enzimáticos. Esta atividade enzimática influi não só em fármacos, mas também em substâncias que se ingerem nos alimentos, que por sua natureza lipofílica são difíceis de eliminar do organismo. Da mesma forma, estas enzimas também são determinantes no metabolismo de substâncias endógenas.

As reações que realizam a eliminação de fármacos podem ser compostas por duas fases consecutivas. O objetivo da fase I consiste em aumentar a polaridade do fármaco e a fase II agrega um composto orgânico que facilite ainda mais sua eliminação.

Embora os polimorfismos genéticos de várias enzimas metabolizadoras de fármacos tenham sido descritos, o processo oxidativo da fase I mediado por enzimas do complexo citocromo p-450 (CYP) parece ser um dos passos mais importantes para muitos medicamentos clinicamente relevantes. É por isso que os polimorfismos genéticos das enzimas CYP foram extensivamente estudados nas últimas décadas (Fricke-Galindo et al. 2015; LLerena et al. 2014; Céspedes-Garro et al. 2015). Um grande número desses CYPs foi descrito (atualmente mais de 50), e dividido em numerosas famílias. Muitos CYPs são sujeitos a polimorfismos genéticos com consequências funcionais (http://www.cypalleles.ki.se). As principais enzimas humanas CYP responsáveis por catalisar a biotransformação de fármacos são membros das famílias CYP1, CYP2, e CYP3. Dessas enzimas, CYP1A2, CYP2C9, CYP2C19, CYP2D6, e CYP3A4 são conhecidas por serem responsáveis por mais de 90% das oxidações que são clinicamente relevantes para terapias medicamentosas humanas (Shimada et al. 1994). CYP2D6 é a

maior determinante de concentração plasmática de vários medicamentos clinicamente importantes, como beta-bloqueadores, antidepressivos e neurolépticos (Llerena et al. 1993; Dorado et al. 2006). CYP2C9 é importante no metabolismo de medicamentos como varfarina, tolbutamida, fenitoína, diclofenaco, losartan, fluoxetina e amitriptilina (Dorado et al. 2003). O polimorfismo genético de CYP2C19 é de importância clínica porque ela catalisa o metabolismo de vários fármacos importantes como omeprazol, clopidogrel, amitriptilina, clomipramina, diazepam, entre outros (Bertilsson 1995).

Para a farmacoterapia individualizada, a habilidade de prever a resposta ao tratamento a partir do conhecimento do genótipo, permitiria maximizar as chances de encontrar um fármaco e uma dose adequados para cada paciente. Essa personalização poderia maximizar a relação eficácia/reações adversas, oferecendo economia no tempo e custo do tratamento e substancialmente melhorando o prognóstico do paciente a longo prazo. Em suma, a farmacogenética pode ser útil na melhora da relação risco/benefício de medicamentos, permitindo a prevenção de reações adversas ou falhas terapêuticas.

A prevalência de polimorfismos nos genes que codificam enzimas metabolizadoras de fármacos varia consideravelmente entre diferentes grupos étnicos (Li, Lou, et al. 2014; Ramos et al. 2014) e isso faz com que a resposta clínica destes grupos seja variável. Por isso, um dos grandes desafios para se alcançar a farmacoterapia individualizada é o fato de que o desenvolvimento de fármacos é realizado principalmente em populações com ancestralidade europeia, mas populações de outras origens também fazem uso destes medicamentos. Por isso, populações de outras etnias, especialmente as miscigenadas, podem ser suscetíveis a falhas terapêuticas ou reações adversas a estes medicamentos.

A heterogeneidade genética é evidente entre latino-americanos. O conhecimento de características farmacogenéticas de diferentes grupos étnicos em latino-americanos pode ser muito relevante para a implementação de estratégias terapêuticas adaptadas a eles. A farmacogenética, em um enfoque populacional, permite o desenvolvimento de recomendações regionais de medicamentos de acordo com fatores étnicos e culturais de cada local.

O fato de que estudos para o desenvolvimento de fármacos seja realizado majoritariamente em populações de ancestralidade europeia faz com que populações da

América Latina, com sua ancestralidade complexa, sejam mais vulneráveis a falhas terapêuticas ou reações adversas. Dados para estas populações ainda são escassos pois estes indivíduos são normalmente negligenciados em estudos de farmacologia, genética e farmacogenética. Por isso, novos dados farmacogenéticos e de ancestralidade para latino-americanos devem ser gerados e os dados já publicados devem ser concatenados para que se extraia o maior nível de informação existente.

Portanto, o objetivo geral desta tese é verificar a ancestralidade e a frequência das variantes de interesse farmacogenético de populações latino-americanas, visando esclarecer se há padrões de frequências de genótipos e fenótipos farmacogenéticos nestas populações.

Esta tese está estruturada em três capítulos de acordo aos objetivos propostos.

Capítulo 1: Ancestralidade no Brasil e na América Latina.

Possui como objetivo inferir a ancestralidade genômica de diferentes populações e sua relação com etnia autodeclarada, explorando diferentes usos dos estudos de ancestralidade em estudos biomédicos.

Objetivo 1.1: Inferir a ancestralidade populacional, individual e cromossômica de três coortes brasileiras em uma abordagem *genome-wide* (artigo publicado);

Objetivo 1.2: Utilizar a ancestralidade individual e populacional *genome-wide* da coorte de Bambuí e verificar se há correlação com a etnia autodeclarada (artigo publicado);

Capítulo 2: Farmacogenética populacional e ancestralidade.

Possui como objetivo verificar as frequências alélicas e fenotípicas dos principais genes farmacogenéticos em diferentes etnias e regiões mundiais, através de revisões sistemáticas.

Objetivo 2.1: Realizar uma revisão sistemática de frequências alélicas, fenótipos preditos e fenótipos metabólicos por grupos étnicos e regiões mundiais em voluntários sadios para os genes *CYP2D6*, *CYP2C9* e *CYP2C19* (artigos publicados);

Objetivo 2.2: Realizar uma revisão sistemática de frequências alélicas, fenótipos preditos e fenótipos metabólicos por grupos étnicos e regiões do Brasil (primeira autoria - a ser submetido).

Capítulo 3: Farmacogenética de genes específicos e ancestralidade.
O objetivo é, com dados originais, avaliar os padrões de ancestralidade e correlacioná-los com genótipos de interesse farmacogenético em populações latino-americanas.

Objetivo 3.1: Correlacionar a ancestralidade individual continental inferida por AIMs com a frequência dos alelos e fenótipos preditos dos genes *CYP2D6*, *CYP2C9* e *CYP2C19* em populações da Costa Rica (artigo aceito - primeira autoria compartilhada);

Objetivo 3.2: Verificar a frequência de alelos e fenótipos preditos dos genes *CYP2D6*, *CYP2C9* e *CYP2C19* em nativos peruanos, peruanos miscigenados e brasileiros miscigenados (dados não publicados);

Objetivo 3.3: Identificar o padrão de genótipos e haplótipos dos genes *PDE4B* e *MYT1L* previamente correlacionados com ancestralidade ameríndia e recidiva de leucemia linfoide aguda (LLA) em populações de nativos brasileiros e peruanos e miscigenados brasileiros e peruanos (primeira autoria - a ser submetido).

CAPÍTULO 1 - Ancestralidade no Brasil e na América Latina

Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations

O consórcio EPIGEN-Brasil é um projeto de genética epidemiológica financiado pelo Ministério da Saúde Brasileiro que envolve cinco centros de pesquisa: Fundação Oswaldo Cruz de Belo Horizonte, Universidade Federal de Pelotas, Universidade Federal de Minas Gerais, Universidade Federal da Bahia e Universidade de São Paulo. O objetivo do projeto é inferir a estrutura populacional e a ancestralidade genômica de 3 coortes brasileiras de base populacional: coorte de crianças de Salvador (Projeto SCAALA com n=1.309; (Barreto et al. 2006)), coorte de idosos de Bambuí (n=1.442; (Lima-Costa et al. 2011)) e coorte de nascidos vivos de Pelotas (n=3.736; (Barros et al. 2006)), além de realizar estudos de associação de varredura genômica (GWAS: *Genome Wide Association Studies*) para uma série de desfechos de interesse biomédico que incluem asma, doença de Chagas, doenças cardiovasculares, doenças psiquiátricas e outras, para as quais existem dados longitudinais. Para isso foi realizada a genotipagem de aproximadamente 4,3 milhões de SNPs (Illumina HumanOmni5 array) para 265 indivíduos e a genotipagem de aproximadamente 2,3 milhões de SNPs (Illumina HumanOmni2.5 array) para 6.487 indivíduos, além do sequenciamento do genoma completo em alta resolução (cobertura média de 42x) de 30 brasileiros, sendo 10 de cada coorte.

Esse grande e miscigenado conjunto de dados permite fazer ainda inferência da dinâmica do processo demográfico de mistura em diferentes regiões do Brasil bem como avaliar o tempo e o modo como mutações clinicamente relevantes chegaram ao Brasil. Além disso, possíveis regiões afetadas pela seleção natural pós-colombiana podem também ser identificadas a partir do excesso ou déficit de ancestralidade europeia, africana ou nativo-americana.

No artigo a seguir, publicado na revista *PNAS* (Kehdy et al. 2015), nosso grupo de pesquisa realizou análises com os dados mencionados visando elucidar a dinâmica da miscigenação brasileira a nível continental e subcontinental. Na primeira etapa do projeto, antes de meu estágio sanduiche, fui responsável pelas inferências de ancestralidade individual em nosso grupo de pesquisa, utilizando a metodologia

implementada no software *ADMIXTURE* (Alexander et al. 2009). Esta metodologia consiste na adequação das populações miscigenadas ao Equilíbrio de Hardy-Weinberg nas frequências alélicas das populações parentais (K) que deram origem à população miscigenada estudada (Wollstein & Lao 2015). Neste trabalho, realizei o primeiro conjunto de análises de ancestralidade, identificando componentes continentais e subcontinentais de ancestralidade.

Com as análises de ancestralidade continental, nós verificamos que a coorte de Salvador é a que possui maior ancestralidade africana (50,8%), seguido de 42,9% de ancestralidade europeia e 6,4% de ancestralidade ameríndia. A coorte de Bambuí apresentou 78,5% de ancestralidade europeia, 14,7% de africana e 6,7% de ameríndia, enquanto a coorte de Pelotas possui 76,1% de ancestralidade europeia, 15,9% de africana e 8% de ameríndia. Portanto, as coortes do Sudeste/Sul são predominantemente europeias, enquanto a coorte do Nordeste apresentou uma prevalência de ancestralidade africana.

As análises de ancestralidade subcontinental mostraram que a ancestralidade europeia no Sudeste e Sul do Brasil é originada de uma região mais ampla, incluindo o norte Europeu e Oriente Médio. Em Salvador, a origem europeia é mais restrita à Península Ibérica. Entretanto, a imigração europeia massiva e a migração interna de descendentes de africanos no Brasil não aboliu dois componentes de ancestralidade africana na estrutura genética dos brasileiros: um associado a populações Yoruba/Mandenka não-Bantu (oeste africano), mais evidente na população de Salvador, e um associado a populações do centro-leste da África/Bantu, mais presente no Sudeste/Sul do Brasil. Portanto, estes resultados ampliam nosso conhecimento da estrutura genética da diáspora africana.

Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations

Fernanda S. G. Kehdy[a,1], Mateus H. Gouveia[a,1], Moara Machado[a,1], Wagner C. S. Magalhães[a,1], Andrea R. Horimoto[b], Bernardo L. Horta[c], Rennan G. Moreira[a], Thiago P. Leal[a], Marilia O. Scliar[a], Giordano B. Soares-Souza[a], Fernanda Rodrigues-Soares[a], Gilderlanio S. Araújo[a], Roxana Zamudio[a], Hanaisa P. Sant Anna[a], Hadassa C. Santos[b], Nubia E. Duarte[b], Rosemeire L. Fiaccone[d], Camila A. Figueiredo[e], Thiago M. Silva[f], Gustavo N. O. Costa[f], Sandra Beleza[g], Douglas E. Berg[h,i], Lilia Cabrera[j], Guilherme Debortoli[k], Denise Duarte[l], Silvia Ghirotto[m], Robert H. Gilman[n,o], Vanessa F. Gonçalves[p], Andrea R. Marrero[k], Yara C. Muniz[k], Hansi Weissensteiner[q], Meredith Yeager[r], Laura C. Rodrigues[s], Mauricio L. Barreto[f], M. Fernanda Lima-Costa[t,2], Alexandre C. Pereira[b,2], Maíra R. Rodrigues[a,2], Eduardo Tarazona-Santos[a,2,3], and The Brazilian EPIGEN Project Consortium[4]

[a]Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, 31270-901, Belo Horizonte, Minas Gerais, Brazil; [b]Instituto do Coração, Universidade de São Paulo, 05403-900, São Paulo, Sao Paulo, Brazil; [c]Programa de Pós-Graduação em Epidemiologia, Universidade Federal de Pelotas, 464, 96001-970 Pelotas, Rio Grande do Sul, Brazil; [d]Departamento de Estatística, Instituto de Matemática, Universidade Federal da Bahia, 40170-110, Salvador, Bahia, Brazil; [e]Departamento de Ciências da Biointeração, Instituto de Ciências da Saúde, Universidade Federal da Bahia, 40110-100, Salvador, Bahia, Brazil; [f]Instituto de Saúde Coletiva, Universidade Federal da Bahia, 40110-040, Salvador, Bahia, Brazil; [g]Department of Genetics, University of Leicester, LE1 7RH, Leicester, United Kingdom; [h]Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO 63110; [i]Department of Medicine, University of California, San Diego, CA 92093; [j]Biomedical Research Unit, Asociación Benéfica Proyectos en Informática, Salud, Medicina y Agricultura (AB PRISMA), 170070, Lima, Peru; [k]Departamento de Biologia Celular, Embriologia e Genética, Universidade Federal de Santa Catarina, 88040-900, Florianópolis, Santa Catarina, Brazil; [l]Departamento de Estatística, Universidade Federal de Minas Gerais, 31270-901, Belo Horizonte, Minas Gerais, Brazil; [m]Dipartimento di Scienze della Vita e Biotecnologie, Università di Ferrara, 44121 Ferrara, Italy; [n]Bloomberg School of Public Health, International Health, Johns Hopkins University, Baltimore, MD 21205; [o]Laboratorio de Investigación de Enfermedades Infecciosas, Universidade Peruana Cayetano Heredia, 15102, Lima, Peru; [p]Department of Psychiatry and Neuroscience Section, Center for Addiction and Mental Health, University of Toronto, Toronto, ON, Canada M5T 1R8; [q]Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, 6020 Innsbruck, Austria; [r]Cancer Genomics Research Laboratory, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD 20850; [s]Department of Infectious Disease Epidemiology, Faculty of Epidemiology, London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom; and [t]Instituto de Pesquisa Rene Rachou, Fundação Oswaldo Cruz, 30190-002, Belo Horizonte, Minas Gerais, Brazil

Edited by Marcus W. Feldman, Stanford University, Stanford, CA, and approved May 27, 2015 (received for review March 8, 2015)

While South Americans are underrepresented in human genomic diversity studies, Brazil has been a classical model for population genetics studies on admixture. We present the results of the EPIGEN Brazil Initiative, the most comprehensive up-to-date genomic analysis of any Latin-American population. A population-based genome-wide analysis of 6,487 individuals was performed in the context of worldwide genomic diversity to elucidate how ancestry, kinship, and inbreeding interact in three populations with different histories from the Northeast (African ancestry: 50%), Southeast, and South (both with European ancestry >70%) of Brazil. We showed that ancestry-positive assortative mating permeated Brazilian history. We traced European ancestry in the Southeast/South to a wider European/Middle Eastern region with respect to the Northeast, where ancestry seems restricted to Iberia. By developing an approximate Bayesian computation framework, we infer more recent European immigration to the Southeast/South than to the Northeast. Also, the observed low Native-American ancestry (6–8%) was mostly introduced in different regions of Brazil soon after the European Conquest. We broadened our understanding of the African diaspora, the major destination of which was Brazil, by revealing that Brazilians display two within-Africa ancestry components: one associated with non-Bantu/western Africans (more evident in the Northeast and African Americans) and one associated with Bantu/eastern Africans (more present in the Southeast/South). Furthermore, the whole-genome analysis of 30 individuals (42-fold deep coverage) shows that continental admixture rather than local post-Columbian history is the main and complex determinant of the individual amount of deleterious genotypes.

Latin America | population genetics | Salvador SCAALA | Bambuí Cohort Study of Ageing | Pelotas Birth Cohort Study

Latin Americans, who are classical models of the effects of admixture in human populations (1, 2), remain underrepresented in studies of human genomic diversity, notwithstanding recent studies (3, 4). Indeed, no large genome-wide study on admixed South Americans has been conducted so far. Brazil is the largest and most populous Latin-American country. Its over 200 million inhabitants are the product of post-Columbian admixture between Amerindians, Europeans colonizers or immigrants, and African slaves (1). Interestingly, Brazil was the destiny of nearly 40% of the African diaspora, receiving seven times more slaves than the United States (nearly 4 million vs. 600,000).

Here, we present results of the EPIGEN Brazil Initiative (https://epigen.grude.ufmg.br), the most comprehensive up-to-date genomic analysis of a Latin-American population. We genotyped nearly 2.2 million SNPs in 6,487 admixed individuals from three population-based cohorts from different regions with distinct demographic and socioeconomic backgrounds and sequenced the whole genome of 30 individuals from these populations at an

Author contributions: E.T.-S. designed research; F.S.G.K., M.H.G., M.M., W.C.S.M., A.R.H., B.L.H., R.G.M., M.L.B., M.F.L.-C., A.C.P., M.R.R., and E.T.-S. performed research; T.P.L., R.Z., R.L.F., C.A.F., T.M.S., G.N.O.C., S.B., D.E.B., L.C., R.H.G., M.Y., L.C.R., M.R.R., and T.B.E.P.C. contributed new reagents/analytic tools; F.S.G.K., M.H.G., M.M., W.C.S.M., A.R.H., R.G.M., T.P.L., M.O.S., G.B.S.-S., F.R.-S., G.S.A., H.P.S.A., H.C.S., N.E.D., G.D., D.D., S.G., V.F.G., A.R.M., Y.C.M., and H.W. analyzed data; F.S.G.K., M.H.G., M.M., W.C.S.M., R.G.M., M.R.R., and E.T.-S. wrote the paper; F.S.G.K. coordinated the ancestry team of the project; W.C.S.M. coordinated the inputation team of the project; A.R.H. coordinated the basic analyses team of the project; B.L.H. coordinated the 1982 Pelotas Birth Cohort; M.L.B. coordinated the SCAALA (Social Changes, Asthma and Allergy in Latin America Program) cohort; M.F.L.-C. coordinated the Bambui cohort; A.C.P. and E.T.-S. supervised the genome analysis group of the project; and M.R.R. coordinated the bioinformatics team of the project.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper have been deposited in the European Nucleotide Archive (PRJEB9080 (ERP010139) Genomic Epidemiology of Complex Diseases in Population-Based Brazilian Cohorts), accession no. EGAS00001001245, under EPIGEN Committee Controlled Access mode.

[1]F.S.G.K., M.H.G., M.M., and W.C.S.M. contributed equally to this work.

[2]M.F.L.-C., A.C.P., M.R.R., and E.T.-S. contributed equally to this work.

[3]To whom correspondence should be addressed. Email: edutars@icb.ufmg.br.

[4]A complete list of the Brazilian EPIGEN Project Consortium can be found in *SI Appendix*.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1504447112/-/DCSupplemental.

Significance

The EPIGEN Brazil Project is the largest Latin-American initiative to study the genomic diversity of admixed populations and its effect on phenotypes. We studied 6,487 Brazilians from three population-based cohorts with different geographic and demographic backgrounds. We identified ancestry components of these populations at a previously unmatched geographic resolution. We broadened our understanding of the African diaspora, the principal destination of which was Brazil, by revealing an African ancestry component that likely derives from the slave trade from Bantu/eastern African populations. In the context of the current debate about how the pattern of deleterious mutations varies between Africans and Europeans, we use whole-genome data to show that continental admixture is the main and complex determinant of the amount of deleterious genotypes in admixed individuals.

average deep coverage of 42× (Fig. 1*B* and *SI Appendix*, sections 1, 2, and 8). By leveraging on a population-based approach, we (*i*) identified and quantified ancestry components of three representative Brazilian populations at a previously unmatched geographic resolution; (*ii*) developed an approximate Bayesian computation (ABC) approach and inferred aspects of the admixture dynamics in Northeastern, Southeastern, and Southern Brazil; (*iii*) elucidated how aspects of the ancestry-related social history of Brazilians influenced their genetic structure; and (*iv*) studied how admixture, kinship, and inbreeding interact and shape the pattern of putative deleterious mutations in an admixed population.

Results and Discussion

Populations, Continental Ancestry, and Population Structure. We studied the following three population-based cohorts (Fig. 1*B*). (*i*) SCAALA (Social Changes, Asthma and Allergy in Latin America Program) (5) (1,309 individuals) from Salvador, a coastal city with 2.7 million inhabitants in Northeastern Brazil that harbors the most conspicuous demographic and cultural African contribution (6). We inferred (7) that this population has the largest African ancestry (50.8%; SE = 0.35) among the EPIGEN populations, with 42.9% (SE = 0.35) and 6.4% (SE = 0.09) of European and Amerindian ancestries, respectively. Notably, this African ancestry is lower than that usually observed in African Americans (8, 9). (*ii*) The Bambuí Aging Cohort Study (10), ongoing in the homonymous city of ∼15,000 inhabitants, in the inland of Southeastern Brazil (1,442 individuals who were 82% of the residents older than 60 y old at the baseline year). We estimated that Bambuí has 78.5% (SE = 0.4) of European, 14.7% (SE = 0.4) of African, and 6.7% (SE = 0.1) of Amerindian ancestries. (*iii*) The 1982 Pelotas Birth Cohort Study (11) (3,736 individuals; 99% of all births in the city at the baseline year). Pelotas is a city in Southern Brazil with 214,000 inhabitants. Ancestry in Pelotas is 76.1% (SE = 0.33) European, 15.9% (SE = 0.3) African, and 8% (SE = 0.08) Amerindian.

By comparing autosomal mtDNA and X-chromosome diversity, we found across the three populations the signature of a historical pattern of sex-biased preferential mating between males with predominant European ancestry and women with predominant African or Amerindian ancestry (12) (*SI Appendix*, sections 6.6 and 6.9, Fig. S12, and Table S18). We determined (13) that individuals from Salvador and Pelotas were, with few exceptions, unrelated and have low consanguinity (Fig. 1*A* and *SI Appendix*, Figs. S1 and S2). Conversely, the Bambuí cohort has the highest family structure and inbreeding [Fig. 1*A* and *SI Appendix*, section 4.1 (discussion about the age structure of this cohort) and Figs. S1 and S2]. Bambuí includes several families with more than five related individuals showing at least one second-degree (or closer) relative. Bambuí mean inbreeding coefficient (0.010; SE = 0.0008) (*SI Appendix*, Fig. S2) is comparable with estimates observed in populations with 15–25% of consanguineous marriages from India (14). Interestingly, inbreeding in Bambuí was correlated with European ancestry ($\rho_{Spearman} = 0.20$; $P < 10^{-15}$). These higher inbreeding and kinship structures are consistent with Bambuí being the smallest and the most isolated of the EPIGEN populations.

Continental genomic ancestry in Latin America (and specifically, in Brazil) is correlated with a set of phenotypes, such as skin color and self-reported ethnicity, and social and cultural features, such as socioeconomic status (15–17). We observed a positive correlation across the three EPIGEN populations between SNP-specific Africans/Europeans F_{ST} (a measurement of informativeness of ancestry) and SNP-specific F_{IT} (a measurement of departure from Hardy–Weinberg equilibrium)

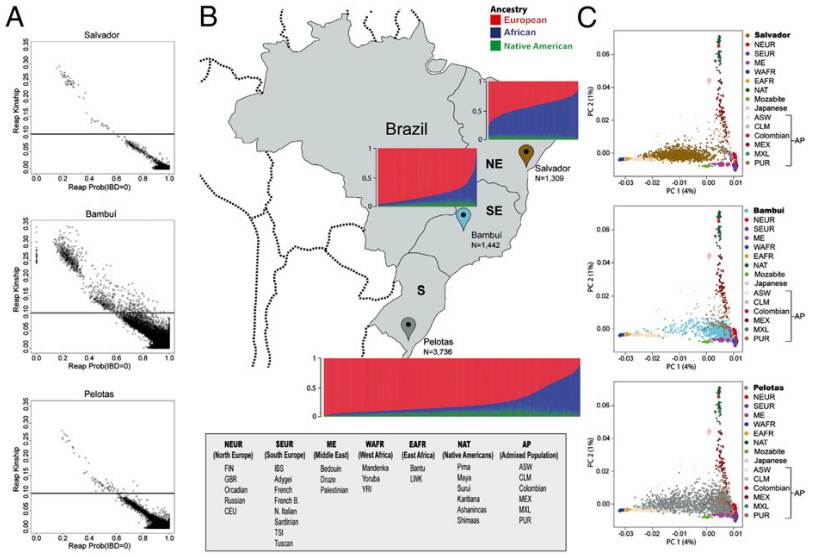

Fig. 1. Continental admixture and kinship analysis of the EPIGEN Brazil populations. (*A*) Kinship coefficient for each pair of individuals and the probability that they share zero identity by descent (IBD) alleles (IBD = 0). Horizontal lines represent a kinship coefficient threshold used to consider individuals as relatives. (*B*) Brazilian regions, the studied populations, and their continental individual ancestry bar plots. *N* represents the numbers of EPIGEN individuals in the Original Dataset (including relatives; detailed in *SI Appendix*, section 6). (*C*) PCA representation, including worldwide populations and the EPIGEN populations, using only unrelated individuals (Dataset U; explained in *SI Appendix*, section 6). The three graphics derive from the same analysis and are different only for the plotting of the EPIGEN individuals. AP, admixed population; ASW, Americans of African ancestry in USA; CEU, Utah residents with Northern and Western European ancestry; CLM, Colombians from Medellin, Colombia; EAFR, east Africa; FIN, Finnish in Finland; French B, Basque; GBR, British in England and Scotland; IBS, Iberian population in Spain; LWK, Luhya in Webuye, Kenya; ME, Middle East; MXL/MEX, Mexican ancestry from Los Angeles; N., (North) Italian; NAT, Native American; NE, northeast; NEUR, north Europe; PC, principal component; PUR, Puerto Ricans from Puerto Rico; S, south; SE, southeast; SEUR, south Europe; TSI, Toscani in Italia; YRI, Yoruba in Ibadan, Nigeira; WAFR, west Africa.

(*SI Appendix*, Fig. S3). This finding indicates that, after five centuries of admixture, Brazilians still preferentially mate with individuals with similar ancestry (and its correlated morphological phenotypes and socioeconomic characteristics), a trend also observed in Mexicans and Puerto Ricans (18). Interestingly, the highest correlations were found in Pelotas and Bambuí, consistent with their higher proportion of individuals with a clearly predominant ancestry (European or African) compared with Salvador (Fig. 1 *B* and *C*). Conversely, in Salvador, despite its highest mean African ancestry, individuals are more admixed (Fig. 1 *B* and *C*), probably because of a combination of a longer history of admixture (see below) and the lower and more homogeneous socioeconomic status of this cohort (5).

Three outcomes illustrate how population subdivision and inbreeding (both partly ancestry-dependent) interact to shape population structure in admixed populations with different sizes (*SI Appendix*, Figs. S1 and S3). First, Bambuí (the smallest city) has the strongest departure from Hardy–Weinberg equilibrium (F_{IT} =0.016; SE = 0.00003) because of both inbreeding (F_{IS} = 0.010; SE = 0.0008) and ancestry-based population subdivision ($\rho_{FIT\text{-}FST} = 0.18$; $P < 10^{-16}$). Second, Pelotas (a medium-sized city; $F_{IT} = 0.012$; SE = 0.00002) has negligible inbreeding (F_{IS} = −0.001; SE = 0.0002) but the strongest ancestry-based population subdivision ($\rho_{FIT\text{-}FST} = 0.38$; $P < 10^{-16}$). Third, the large city of Salvador shows the lowest inbreeding and ancestry-based population subdivision (F_{IT} = −0.003; SE = 0.00002; F_{IS} = −0.001; SE = 0.0003; $\rho_{FIT\text{-}FST} = 0.08$; $P < 10^{-16}$).

Overall, the EPIGEN populations studied by a population-based approach exemplify how ancestry, kinship, and inbreeding may be differently structured in small (Bambuí), medium (Pelotas), and large (Salvador) admixed Latin-American populations. These populations fairly represent the three most populated Brazilian regions (Northeast, Southeast, and South) with their geographic distribution and continental ancestry (Fig. 1) and are good examples of the Latin-American genetic diversity with their ethnic diversity.

Differences in Admixture Dynamics. We estimated the continental origin of each allele for each SNP along each chromosome of the EPIGEN individuals (19) (*SI Appendix*, section 6.7) and calculated the lengths of chromosome segments of continuous specific ancestry (CSSA) (Fig. 2*A*), with distribution that informs how admixture occurred over time. By leveraging on the model by Liang and Nielsen (20) of CSSA, we developed an ABC framework to infer admixture dynamics (*SI Appendix*, section 6.8). We simulated CSSA distributions generated by a demographic history of three pulses of trihybrid admixture that occurred 18–16, 12–10, and 6–4 generations ago, conditioning on the observed current admixture proportions of each of the EPIGEN populations. This demographic model conciliates statistical complexity and the real history of admixture. We inferred the posterior distributions of nine parameters $m_{n,P}$, where m is the proportion of immigrant individuals entering in the admixed population from the n ancestral population (African, European, or Native-American ancestry) in the P admixture pulse.

Interestingly, ABC results (Fig. 2*B*) show that the observed low Native-American ancestry was mostly introduced in different regions of Brazil soon after the European Conquest of the Americas, which is consistent with the posterior depletion of the Native-American population in Brazil. Also, we inferred a predominantly earlier European colonization in the Northeast (Salvador) vs. a more recent immigration in Southeastern and Southern Brazil (Bambuí and Pelotas), consistent with historical records (brasil500anos.ibge.gov.br/). Conversely, African admixture showed a decreasing temporal trend shared by the three EPIGEN populations (21). Complementary explanations are continuous local immigration into the admixed populations from communities with high African ancestry already settled in Brazil [for example, quilombos (i.e., Afro-Brazilian slave-derived communities in Brazil) (22)].

Dissecting European Ancestry. To dissect the ancestry of Brazilians at a subcontinental level, we applied (*i*) the ADMIXTURE method (7) by increasing the number of ancestral clusters (K) that explains the observed genetic structure (*SI Appendix*, Figs. S4 and S5) and (*ii*) the Principal Component Analysis (PCA) (23) (Figs. 1*C* and 3 *B* and *D* and *SI Appendix*, Fig. S6). To study biogeographic ancestry, we excluded sets of relatives that could affect our inferences at the within-continent level (24). We developed a method based on complex networks to reduce the relatedness of the analyzed individuals by minimizing the number of excluded individuals (*SI Appendix*, section 6.1). Using this method, we created the Dataset Unrelated (Dataset U), including 5,825 Brazilians, 1,780 worldwide individuals, and no pair of individuals closer than second-degree relatives. Hereafter, PCA and ADMIXTURE results are relative to Dataset U.

Brazil received several immigration waves from diverse European origins during the last five centuries (brasil500anos.ibge.gov.br/): Portuguese (the first colonizers), who also arrived in large numbers during the last 150 y; Italians (mostly to the South and Southeast); and Germans (mostly to the South). In our PCA representation (Fig. 3*B*), the European component of the genomes of most Brazilians is similar to individuals from the Iberian Peninsula and neighboring regions. The resemblance in within-European ancestry of individuals from Pelotas (South) and Bambuí (Southeast) to central North Europeans and Middle Easters, respectively (Fig. 3*B*), reflects a geographically wider European ancestry of these two populations with respect to Salvador. Considering the total European ancestry estimated by ADMIXTURE, we inferred a higher proportion of North European-associated ancestry in Pelotas (40.2%) than in Bambuí (35.8%) and Salvador (36.7%; $P < 10^{-15}$; Wilcoxon tests) (Fig. 3*A*, red cluster in $K = 7$). We confirmed these results by analyzing a reduced number of SNPs with a larger set of

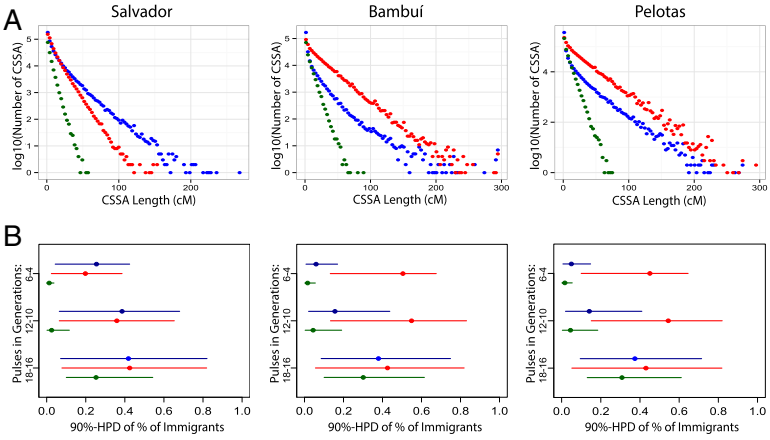

Fig. 2. Distributions of lengths of chromosomal segments of (*A*) CSSA and (*B*) admixture dynamics inferences estimated for three EPIGEN Brazilian populations. (*A*) CSSA lengths were distributed in 50 equally spaced bins per population. Red, blue, and green dots represent a European, an African, and a Native-American CSSA, respectively. (*B*) We inferred the posterior densities of the proportions of immigrants (with respect to the admixed population) from each origin, and we show their 90% highest posterior density (HPD) intervals. Inferences are based on a model of three pulses of admixture (vertical axis) simulated based on the model of CSSAs evolution by Liang and Nielsen (20). Inferences are based on approximate Bayesian computation. Ancestry color codes are red for European, blue for African, and green for Native American.

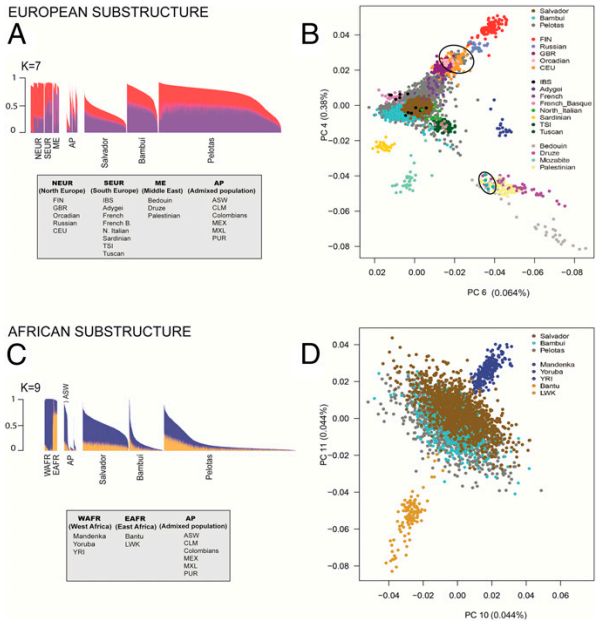

Fig. 3. European and African ancestry clusters in the Brazilian populations. We show (*A* and *C*) relevant ADMIXTURE individual ancestry bar plots and (*B* and *D*) plots of principal components (PCs) that dissect ancestry within (*A* and *B*) Europe and (*C* and *D*) Africa. We performed the analyses using Dataset U (unrelated Brazilians and worldwide individuals). We only plot individuals from relevant ancestral populations. Complete ADMIXTURE and PCA results are represented in *SI Appendix*, section 6 and Figs. S4–S6. Black ellipses in *B* show some individuals from Pelotas (Southern Brazil) clustering with northern European individuals toward the top and individuals from Bambuí (Southeastern Brazil) clustering with Middle Eastern individuals toward the bottom. AP, admixed population; ASW, Americans of African ancestry in USA; CEU, Utah residents with Northern and Western European ancestry; CLM, Colombians from Medellin, Colombia; EAFR, east Africa; FIN, Finnish in Finland; French B, Basque; GBR, British in England and Scotland; IBS, Iberian population in Spain; LWK, Luhya in Webuye, Kenya; ME, Middle East; MXL/MEX, Mexican ancestry from Los Angeles; N., (North) Italian; NAT, Native American; NE, northeast; NEUR, north Europe; PUR, Puerto Ricans from Puerto Rico; S, south; SE, southeast; SEUR, south Europe; TSI, Toscani in Italia; YRI, Yoruba in Ibadan, Nigeira; WAFR, west Africa.

European individuals and populations (25, 26) (*SI Appendix*, section 6.2).

Brazil, the Main Destination of the African Diaspora. African slaves arrived to Brazil during four centuries, whereas most arrivals to the United States occurred along two centuries, and the geographic and ethnic origin of Brazilian slaves differ from Caribbeans and African Americans (27). In fact, the Portuguese Crown imported slaves to Brazil from western and central west Africa (the two are the major sources of the slave trade to all of the Americas) as well as Mozambique. We detected two within-Africa ancestry clusters in the current Brazilian population (Fig. 3*C*, $K = 9$ and *SI Appendix*, section 6.3): one associated with the Yoruba/Mandenka non-Bantu western populations (Fig. 3*C*, blue) and one associated with the Luhya/HGDP (Human Genome Diversity Project) Bantu populations from eastern Africa (Fig. 3*C*, mustard). Interestingly, the proportions of these ancestry clusters, which are present across all of the analyzed African and Latin-American populations, differ across them. The blue cluster in Fig. 3*C* predominates in African Americans and in Salvador, accounting for 83% and 75% of the total African ancestry, respectively (against 17% and 25%, respectively, of the mustard cluster in Fig. 3*C*) (*SI Appendix*, Table S17). Comparatively, the mustard cluster in Fig. 3*C* is more evident in Southeastern and Southern Brazil (36% and 44% of African ancestry in Bambuí and Pelotas, respectively). These results are consistent with the fact that a large proportion of Yoruba slaves arrived in Salvador, whereas the Mozambican Bantu slaves disembarked primarily in Rio de Janeiro in Southeastern Brazil (21). These results show for the first time, to our knowledge, that the genetic structure of Latin Americans reflects a more diversified origin of the African diaspora into the continent. Interestingly, the two within-African ancestry clusters in the Brazilian populations (showing an average F_{ST} of 0.02) are characterized by 3,318 SNPs, with the 10% top F_{ST} values higher than 0.06, and include 38 SNPs that are hits of genome-wide association studies (*SI Appendix*, section 7 and Table S25).

Pattern of Deleterious Variants: Effect of Continental Admixture, Kinship, and Inbreeding. Based on whole-genome data from 30 individuals (10 from each of three EPIGEN populations), we identified putative deleterious nonsynonymous variants (28) (*SI Appendix*, section 8). There are recent interest in and apparently conflicting results on whether Europeans have proportionally more deleterious variants in homozygosis than Africans (29–32). Lohmueller et al. (29) explained these differences as an effect of the Out of Africa bottleneck on current non-African populations. Out of Africa would have enhanced the effect of genetic drift and attenuated the effect of purifying natural selection, preventing, in many instances, the extinction of (mostly weakly) deleterious variants in non-Africans.

We investigated how European ancestry shapes the amount of deleterious variants in homozygosis (a more likely genotype for common/weakly deleterious variants) and heterozygosis in admixed Latin-American individuals. We observed three patterns (Fig. 4). (*i*) Considering all (i.e., weakly and highly) deleterious variants, for a class of individuals with high European ancestry (>65%; from Bambuí and Pelotas), the individual number of deleterious variants in homozygosis is correlated with European ancestry, but importantly, this correlation is not observed among individuals with intermediate European ancestry (from Salvador) (Fig. 4*A*). (*ii*) The individual number of deleterious variants (both all and rare classes) in heterozygosis (Fig. 4 *B* and *D*) decreases linearly with European ancestry, regardless the cohort of origin. This result is also observed for rare deleterious variants in homozygosis, although the pattern is not very clear in this case (Fig. 4*C*). (*iii*) There are no differences in the amount of deleterious variants between individuals from Bambuí and Pelotas. These populations have similar continental admixture proportions and dynamics, but different post-Columbian population sizes and histories of isolation, assortative mating, kinship structure, and inbreeding. Taken together, our results are consistent with the results and evolutionary scenario proposed by Lohmueller et al. (29) and Lohmueller (31), and suggest that, in Latin-American populations, the main determinant of the amount of deleterious variants is the history of continental admixture, although in a more complex fashion than previously thought (pattern *i*). Comparatively, the role of local demographic history seems less relevant.

Conclusion

A thread of historical facts has modeled the genetic structure of Brazilians. Our population-based and fine-scale analyses revealed novel aspects of the genetic structure of Brazilians. In 1870, blacks were the major ethnic group in Brazil (21), but this scenario changed after the arrival of nearly 4 million Europeans during the second one-half of the 19th century and the first one-half of the 20th century. This immigration wave was encouraged by Brazilian officials as a way of "whiting" the population (33), and it transformed Brazil into a predominantly white country, particularly in the Southeast and South. Consistently, (*i*) we observed that larger chromosomal segments of continuous European ancestry in the southeast/south are the signature of this recent European immigration, and (*ii*) we traced the European ancestry in the Southeast/South of Brazil to a wider geographical region (including central northern Europe and the Middle East) than in Salvador (more

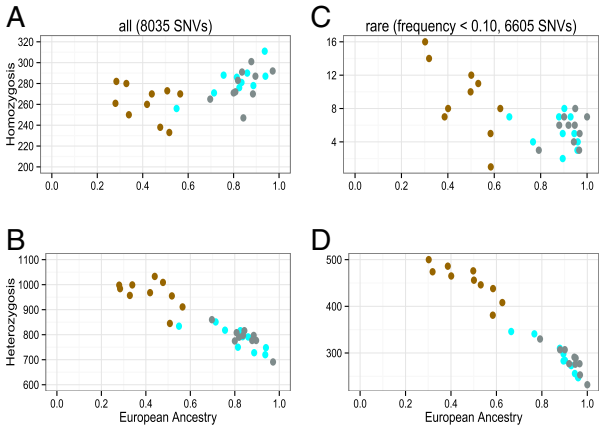

Fig. 4. Individual numbers of genotypes with nonsynonymous deleterious variants in homozygosis and heterozygosis vs. European ancestry based on the whole-genome sequence (42×) of 30 individuals (10 from each population): Salvador (Northeast; brown), Bambuí (Southeast; cyan), and Pelotas (South; gray). Deleterious variants were identified using CONDEL (28) and corrected for the bias reported by Simons et al. (30). Spearman correlation between European ancestry and the number of all deleterious variants in homozygosis for Bambuí and Pelotas individuals was 0.57 ($P = 0.009$). The numbers of genotypes considering all deleterious variants in homozygosis or heterozygosis are in *A* and *B*, respectively, and considering only rare deleterious variants are in *C* (in homozygosis) and *D* (in heterozygosis). SNVs, single nucleotide variants.

restricted to the Iberian Peninsula). However, neither this massive immigration nor the internal migration of black Brazilians have concealed two components of their African ancestry from the genetic structure of Brazilians: one associated with the Yoruba/Mandenka non-Bantu populations, which is more evident in the Northeast (Salvador), and one associated with central east African/Bantu populations, which is more present in the Southeast/South. This result broadens our understanding of the genetic structure of the African diaspora. Furthermore, we showed that positive assortative mating by ancestry is a social factor that permeates the demographic history of Brazilians and also, shapes their genetic structure, with implications for the design of genetic association studies in admixed populations. For instance, because mating by ancestry produces Hardy–Weinberg disequilibrium, filtering SNPs for genome-wide association studies based on the Hardy–Weinberg equilibrium conceals real aspects of the genetic structure of these populations. Finally, in Latin-American populations, the history of continental admixture rather than local demographic history is the main determinant of the burden of deleterious variants, although in a more complex fashion than previously thought. We speculate that future studies on populations from Northern Brazil (including large cities, such as Manaus, next to the Amazon forest) or the Central-West may reveal larger and different dynamics of Amerindian ancestry. Also, fine-scale studies on large urban centers from the Southeast and South of Brazil, such as Rio de Janeiro or Sao Paulo, that have been the destination of migrants from all over the country during the last decades, may show an even more diversified origin of Brazilians, including Japanese ancestry components, for instance, that we did not identify in our study. The EPIGEN Brazil initiative is currently conducting studies to clarify how the genetic variation and admixture interact with environmental and social factors to shape the susceptibility to complex phenotypes and diseases in the Brazilian populations.

Methods

Genotyping and Data Curation. Genotyping was performed by the Illumina facility using the HumanOmni2.5–8v1 array for 6,504 individuals and the HumanOmni5-4v1 array for 270 individuals (90 randomly selected from each cohort). After that, we performed quality control analysis of the data using Genome Studio (Illumina), PLINK (34), GLU (code.google.com/p/glu-genetics/), Eigenstrat (35), and in-house scripts. This study was approved by the Brazilian National Research Ethics Committee (CONEP, resolution 15895).

Whole-Genome Sequencing and Functional Annotation. We randomly selected 10 individuals from each of the three EPIGEN populations. The Illumina facility performed whole-genome sequencing of these individuals from paired-end libraries using the Hiseq 2000 Illumina platform. CASAVA v.1.9 modules were used to align reads and call SNPs and small INDELs (insertion or deletion of bases). Each genome was sequenced, on average, 42 times, with the following quality control parameters: 128 Gb (Gigabase) of passing filter aligned to the reference genome (HumanNCBI37_UCSC), 82% of bases with data quality (QScore) ≥30, 96% of non-N reference bases with a coverage ≥10×, a HumanOmni5 array agreement of 99.53%, and a HumanOmni2.5 array agreement of 99.27%. Functional annotation was performed with ANNOVAR (August 2013 release) with the refGene v.hg19_20131113 reference database in April of 2014. The nonsynonymous variants were predicted to be deleterious using CONDEL v2.0 (cutoff = 0.522) (28), which calculates a consensus score based on MutationAssessor (36) and FatHMM (37). These results were corrected for the bias reported in the work by Simons et al. (30), which evidenced that, when the human reference allele is the derived one, methods that infer deleterious variants tend to underestimate its deleterious effect (*SI Appendix*, section 8).

Relatedness and Inbreeding Analysis. We estimated the kinship coefficients for each possible pair of individuals from each of the EPIGEN populations using the method implemented in the Relatedness Estimation in Admixed Populations (REAP) software (13). It estimates kinship coefficients solely based on genetic data, taking into account the individual ancestry proportion from *K* parental populations and the *K* parental populations allele frequencies per each SNP. For these analyses, we calculated individual ancestry proportion and *K* parental populations allele frequencies per each SNP using the ADMIXTURE software (7) in unsupervised mode assuming three parental populations ($K = 3$). Inbreeding coefficients were also estimated for each individual using REAP. We represented families by networks, which were defined as groups of individuals (vertices) linked by kinship coefficient higher than 0.1 (edges).

***F* Statistics.** The F_{IS} statistic for each population is estimated as the average of the REAP inbreeding coefficients across individuals. For each SNP i and each population, we estimated the departure from Hardy–Weinberg equilibrium as $F_{IT(i)} = (He_i - Ho_i)/He_i$, where Ho_i and He_i are the observed and the expected heterozygosities under Hardy–Weinberg equilibrium for the SNP i, respectively. We estimated the population F_{IT} by averaging $F_{IT(i)}$ across SNPs. We estimated the F_{ST} for each SNP between the YRI and CEU populations using the R package hierfstat (38). The correlation between YRI vs. CEU F_{ST} and F_{IT} values for each SNP was calculated by the Spearman's rank correlation-ρ using the R cor.test function.

Population Structure Analyses. To study population structure, we applied (*i*) the ADMIXTURE method (7), increasing the number of ancestral clusters (*K*) that explains the observed genetic structure from $K = 3$, and (*ii*) PCA (35) (Figs. 1*C* and 3 and *SI Appendix*, section 6 and Figs. S4–S6). To study biogeographic ancestry, we have to exclude sets of relatives that could affect our inferences at within-continental level (24). We conceived and applied a method based on complex networks to reduce the relatedness of the analyzed individuals by minimizing the number of excluded individuals (*SI Appendix*, section 6.1). Applying this method, we created Dataset U, with 5,825 Brazilians, 1,780 worldwide individuals, and no pairs of individuals closer than second-degree relatives (REAP kinship coefficient >0.10) (*SI Appendix*, Table S13). We performed ADMIXTURE analyses with both the Original Dataset and Dataset U (*SI Appendix*, section 6 and Figs. S4 and S5).

PCA and ADMIXTURE analyses were performed with integrated datasets comprising the three cohort-specific EPIGEN working datasets and the public datasets populations described in *SI Appendix*, section 5. For the PCA and ADMIXTURE analyses, we used the SNPs shared by all of these populations, comprising a total of 8,267 samples and 331,790 autosomal SNPs (called the Original Dataset).

Analyses with X-chromosome data used only female samples from the Original Dataset. To perform such analyses, we integrated genotype data of shared SNPs from the X chromosome of EPIGEN female samples (from all three cohorts) and the X chromosome of female samples from the public datasets populations described in *SI Appendix*, section 5. This data integration yielded genotyping data with 5,792 SNPs for 4,192 females.

Local Ancestry Analyses. We inferred chromosome local ancestry using the PCAdmix software (19) and ~2 million SNPs shared by EPIGEN (Original

Dataset) and the 1000 Genomes Project (*SI Appendix*, section 5.2). Considering our SNPs density, we defined a window length of 100 SNPs following the work by Moreno-Estrada et al. (27). PCAdmix infers the ancestry of each window. Local ancestry inferences were performed after linked markers ($r^2 >$ 0.99) were pruned to avoid ancestry misestimating caused by overfitting (4). We considered only the windows in which ancestry was inferred by the forward–backward algorithm with a posterior probability >0.90.

After local ancestry inferences, we calculated the lengths of the chromosomal segments of CSSA for each haplotype from each chromosome from each individual. The distribution of CSSA length was organized in 50 equally spaced bins defined in centimorgans and plotted for each population (Fig. 2*A*).

For the local ancestry analyses, we used phased data from the 1000 Genomes Project populations YRI and LWK (Africans) as well as CEU, FIN, GBR, TSI, and IBS (Europeans), Native-American populations Ashaninka and Shimaa [from the Tarazona–Santos group LDGH (Laboratory of Human Genetic Diversity) dataset], and the three EPIGEN populations (Original Dataset). The SHAPEIT software (39) was used to generate phased datasets.

We estimated admixture dynamics parameters using ABC. We used the model by Liang and Nielsen (20) to simulate CSSA distributions generated by a demographic history of three pulses of trihybrid admixture occurring 18–16, 12–10, and 6–4 recent generations ago conditioned on the observed admixture proportions of the EPIGEN populations. We inferred the posterior distributions of nine parameters $m_{n,P}$ (*SI Appendix*, section 6.8).

Lineage Markers Haplogroups Inferences. We performed mtDNA haplogroup assignments using HaploGrep (40), a web tool based on Phylotree (build 16) for mtDNAhaplogroup assignment. For Y-chromosome data, we inferred haplogroups using an automated approach called AMY tree (41). For Y-chromosome haplogroups, we considered the Karafet tree (42) and more recent studies to describe additional subhaplogroups. By these means, an updated tree was considered based on the information given by The International Society of Genetic Genealogy (ISOGG version 9.43; www.isogg.org).

ACKNOWLEDGMENTS. The authors thank David Alexander and Fernando Levi Soares for technical help and discussion and Rasmus Nielsen and Mason Liang for sharing their software for continuous specific ancestry simulations and feedback on its use. Centro Nacional de Processamento de Alto Desempenho em MG/Financiadora de Estudos e Projetos–Ministério da Ciência, Tecnologia e Inovação, Centro Nacional de Super Computação, and Programa de Desenvolvimento Tecnológico em Insumos para Saúde-Bioinformatics Platform at Fundação Oswaldo Cruz-Minas Gerais provided computational support. The EPIGEN Brazil Initiative is funded by the Brazilian Ministry of Health (Department of Science and Technology from the Secretaria de Ciência, Tecnologia e Insumos Estratégicos) through Financiadora de Estudos e Projetos. The EPIGEN Brazil investigators received funding from the Brazilian Ministry of Education (CAPES Agency), Brazilian National Research Council (CNPq), Pró-Reitoria de Pesquisa from the Universidade Federal de Minas Gerais, and the Minas Gerais State Agency for Support of Research (FAPEMIG).

1. Salzano FM, Freire-Maia N (1967) *Populações Brasileiras; Aspectos Demográficos, Genéticos e Antropológicos* (Companhia Editora Nacional, São Paulo, Brazil).
2. Giolo SR, et al. (2012) Brazilian urban population genetic structure reveals a high degree of admixture. *Eur J Hum Genet* 20(1):111–116.
3. Moreno-Estrada A, et al. (2014) Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science* 344(6189): 1280–1285.
4. Eyheramendy S, Martinez FI, Manevy F, Vial C, Repetto GM (2015) Genetic structure characterization of Chileans reflects historical immigration patterns. *Nat Commun* 6:6472.
5. Barreto ML, et al. (2006) Risk factors and immunological pathways for asthma and other allergic diseases in children: Background and methodology of a longitudinal study in a large urban center in Northeastern Brazil (Salvador-SCAALA study). *BMC Pulm Med* 6:15.
6. Bacelar J (2001) *A Hierarquia sas Raças. Negros e Brancos em Salvador* (Pallas Editora, Rio de Janeiro).
7. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655–1664.
8. Tishkoff SA, et al. (2009) The genetic structure and history of Africans and African Americans. *Science* 324(5930):1035–1044.
9. Bryc K, et al. (2010) Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proc Natl Acad Sci USA* 107(2):786–791.
10. Lima-Costa MF, Firmo JO, Uchoa E (2011) Cohort profile: The Bambui (Brazil) Cohort Study of Ageing. *Int J Epidemiol* 40(4):862–867.
11. Victora CG, Barros FC (2006) Cohort profile: The 1982 Pelotas (Brazil) birth cohort study. *Int J Epidemiol* 35(2):237–242.
12. Salzano FM, Bortolini MC (2002) *The Evolution and Genetics of Latin American Populations* (Cambridge Univ Press, New York).
13. Thornton T, et al. (2012) Estimating kinship in admixed populations. *Am J Hum Genet* 91(1):122–138.
14. Bittles AH (2002) Endogamy, consanguinity and community genetics. *J Genet* 81(3): 91–98.
15. Telles EE (2006) *Race in Another América: The Significance of Skin Color in Brazil* (Princeton Univ Press, Princeton).
16. Lima-Costa MF, et al.; Epigen-Brazil group (2015) Genomic ancestry and ethnoracial self-classification based on 5,871 community-dwelling Brazilians (The Epigen Initiative). *Sci Rep* 5:9812.
17. Ruiz-Linares A, et al. (2014) Admixture in Latin America: Geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genet* 10(9):e1004572.
18. Risch N, et al. (2009) Ancestry-related assortative mating in Latino populations. *Genome Biol* 10(11):R132.
19. Brisbin A, et al. (2012) PCAdmix: Principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. *Hum Biol* 84(4):343–364.
20. Liang M, Nielsen R (2014) The lengths of admixture tracts. *Genetics* 197(3):953–967.
21. Klein HS (2002) *Homo brasilis Aspectos Genéticos, Lingüísticos, Históricos e Socio-antropológicos da Formação do Povo Brasileiro* (FUNPEC-RP, Ribeirão Preto, Brasil), 2nd Ed, pp 93–112.
22. Scliar MO, Vaintraub MT, Vaintraub PM, Fonseca CG (2009) Brief communication: Admixture analysis with forensic microsatellites in Minas Gerais, Brazil: The ongoing evolution of the capital and of an African-derived community. *Am J Phys Anthropol* 139(4):591–595.
23. Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genet* 2(12):e190.
24. Price AL, Zaitlen NA, Reich D, Patterson N (2010) New approaches to population stratification in genome-wide association studies. *Nat Rev Genet* 11(7):459–463.
25. Nelson MR, et al. (2008) The Population Reference Sample, POPRES: A resource for population, disease, and pharmacological genetics research. *Am J Hum Genet* 83(3):347–358.
26. Botigué LR, et al. (2013) Gene flow from North Africa contributes to differential human genetic diversity in southern Europe. *Proc Natl Acad Sci USA* 110(29):11791–11796.
27. Moreno-Estrada A, et al. (2013) Reconstructing the population genetic history of the Caribbean. *PLoS Genet* 9(11):e1003925.
28. González-Pérez A, López-Bigas N (2011) Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet* 88(4):440–449.
29. Lohmueller KE, et al. (2008) Proportionally more deleterious genetic variation in European than in African populations. *Nature* 451(7181):994–997.
30. Simons YB, Turchin MC, Pritchard JK, Sella G (2014) The deleterious mutation load is insensitive to recent population history. *Nat Genet* 46(3):220–224.
31. Lohmueller KE (2014) The distribution of deleterious genetic variation in human populations. *Curr Opin Genet Dev* 29:139–146.
32. Do R, et al. (2015) No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. *Nat Genet* 47(2):126–131.
33. Pena SD, et al. (2011) The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PLoS ONE* 6(2):e17063.
34. Purcell S, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559–575.
35. Price AL, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38(8):904–909.
36. Reva B, Antipin Y, Sander C (2007) Determinants of protein function revealed by combinatorial entropy optimization. *Genome Biol* 8(11):R232.
37. Shihab HA, et al. (2013) Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat* 34(1):57–65.
38. Goudet J (2005) Hierfstat, a package for r to compute and test hierarchical F-statistics. *Mol Ecol Notes* 5(1):184–186.
39. Delaneau O, Marchini J, Zagury JF (2012) A linear complexity phasing method for thousands of genomes. *Nat Methods* 9(2):179–181.
40. Kloss-Brandstätter A, et al. (2011) HaploGrep: A fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum Mutat* 32(1):25–32.
41. Van Geystelen A, Decorte R, Larmuseau MHD (2013) AMY-tree: An algorithm to use whole genome SNP calling for Y chromosomal phylogenetic applications. *BMC Genomics* 14(14):101–112.
42. Karafet TM, et al. (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 18(5):830–838.

Genomic ancestry and ethnoracial self-classification based on 5,871 community-dwelling Brazilians (The Epigen Initiative)

Ainda como resultado das análises de ancestralidade individual que realizei no projeto EPIGEN, outro artigo, publicado na revista *Scientific Reports* (Lima-Costa et al. 2015), foi publicado.

Existem controvérsias na literatura quanto à correlação entre ancestralidade genômica e cor/raça autodeclarada em brasileiros. Com uma abordagem genome-wide (370.539 SNPs) em 5.851 indivíduos das três coortes do EPIGEN (Bambuí, Salvador e Pelotas), neste trabalho identificamos que há correlação entre ancestralidade genômica e etnia autodeclarada em brasileiros, e que indivíduos de Salvador tendem a se autodeclarar brancos mesmo com alta ancestralidade africana, o que sugere que fatores não biológicos também influenciam na autodeclaração.

SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
POPULATION GENETICS
EPIDEMIOLOGY

Received
21 October 2014

Accepted
6 March 2015

Published
27 April 2015

Correspondence and requests for materials should be addressed to M.F.L-C. (lima-costa@cpqrr.fiocruz.br)

* A comprehensive list of authors and affiliations appear at the end of the paper

Genomic ancestry and ethnoracial self-classification based on 5,871 community-dwelling Brazilians (The Epigen Initiative)

M. Fernanda Lima-Costa[1], Laura C. Rodrigues[2], Maurício L. Barreto[3], Mateus Gouveia[4], Bernardo L. Horta[5], Juliana Mambrini[1], Fernanda S. G. Kehdy[4], Alexandre Pereira[6], Fernanda Rodrigues-Soares[4], Cesar G. Victora[5], Eduardo Tarazona-Santos[4] & Epigen-Brazil group*

[1]Fundação Oswaldo Cruz, Instituto de Pesquisas Rene Rachou, Belo Horizonte, Brazil, [2]London School of Hygiene and Tropical Medicine, Department of Infectious Disease Epidemiology, London, UK, [3]Universidade Federal da Bahia, Instituto de Saúde Coletiva, Salvador, Brazil, [4]Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas, Belo Horizonte, Brazil, [5]Universidade Federal de Pelotas, Programa de Pós Graduação em Epidemiologia, Pelotas, Brazil, [6]Universidade de São Paulo, Instituto do Coração, São Paulo, Brazil.

Brazil never had segregation laws defining membership of an ethnoracial group. Thus, the composition of the Brazilian population is mixed, and its ethnoracial classification is complex. Previous studies showed conflicting results on the correlation between genome ancestry and ethnoracial classification in Brazilians. We used 370,539 Single Nucleotide Polymorphisms to quantify this correlation in 5,851 community-dwelling individuals in the South (Pelotas), Southeast (Bambui) and Northeast (Salvador) Brazil. European ancestry was predominant in Pelotas and Bambui (median= 85.3% and 83.8%, respectively). African ancestry was highest in Salvador (median = 50.5%). The strength of the association between the phenotype and median proportion of African ancestry varied largely across populations, with pseudo R^2 values of 0.50 in Pelotas, 0.22 in Bambui and 0.13 in Salvador. The continuous proportion of African genomic ancestry showed a significant S-shape positive association with self-reported Blacks in the three sites, and the reverse trend was found for self reported Whites, with most consistent classifications in the extremes of the high and low proportion of African ancestry. In self-classified Mixed individuals, the predicted probability of having African ancestry was bell-shaped. Our results support the view that ethnoracial self-classification is affected by both genome ancestry and non-biological factors.

Brazil is the 5th most populous nation in the world, with about 200 million inhabitants[1]. Its population originated from three main ancestral roots: African, European and Native American, the latter constituting the autochthonous population. Colonization was predominantly Portuguese. The slave trade of Africans to Brazil was the oldest, the longest-running and the largest in the Americas. Early European colonizers and their descendants brought an estimated of 3.6 million African slaves, seven times more than their counterparts in the United States[2].

Brazil never had segregation laws defining who should belong to an ethnoracial group, as the United States and South Africa had. This was probably a result of the Brazilian elite decision to "whiten" the Brazilian population through miscegenation rather than impose segregation; and ethnoracial classification was left to individual perception[2]. As a consequence, the composition of the Brazilian population is more mixed, and its ethnoracial classification is more complex and fluid than in those countries where segregation was imposed by law[2]. This was to such a degree that it has been questioned whether – and how – ethnoracial classification in Brazil correlates with genomic ancestry. Previous genome studies based on up to a hundred informative markers showed conflicting results on this correlation[3–8].

The Brazilian census adopts a classification based on ethnoracial self-classification with five groups: White, Mixed ("pardo" in official Portuguese), Black, Yellow (Asian) and Indigenous (Native American), the latter two representing less than 1% of the total population[1]. People who self-report as Whites predominate in the South and Southeast, and as Mixed and/or Black in the North and Northeast[1]. Persons self-reporting as Black and Mixed are

more likely to have lower income and education[2,9-11], to report experiencing discrimination[11,12], and have more negative health-related outcomes[11,13-17]. The most plausible explanation for these disparities is the cumulative effect of the lack of social policies to support individuals of African origin and their descendants since the abolition of slavery in 1888[18]. To some extent, recent affirmative action in Brazil, mostly based on ethnoracial self-classification, is supported by this theoretical framework. Thus, the debate over whether ethnoracial self-classification correlates with ancestry has scientific and policy implications.

The Epigen-Brazil initiative is based on three well-defined ongoing population-based cohorts from Brazil's South[19], Southeast[20] and Northeast[21]. We used 370,539 Single Nucleotide Polymorphisms (SNPs) to quantify the association between likelihood of self-classification as White, Mixed and Black and genome-wide based individual proportions of African, European and Native American ancestry in 5,851 participants of these cohorts.

Results

The study included 3,533 individuals from Pelotas (South), 1,442 from Bambuí (Southeast), and 876 from Salvador (Northeast). Self-reported as White predominated in Pelotas (77.5%) and Bambuí (60.6%), while self-reported as Black (43.4%) and Mixed (49.3%) predominated in Salvador. The Pelotas and the Bambuí cohort populations had predominant European ancestry (median= 85.3% and 83.8%, respectively), while African ancestry was the highest in Salvador (median = 50.5%). Native American ancestry was little and relatively uniform in the three sites (~ 5-6%) (Table 1).

Median African, European and Native American individual ancestry across ethnoracial categories are shown in 12 panels in Figure 1. In the joint analysis of the 3 cohorts, as well as within each cohort population, there was a significant increase on the median African ancestry from people self-reporting as White to Mixed and then to Black ($p<0.001$ in Mann Whitney test for differences across ethnoracial categories); median European ancestry decreased in the opposite direction, as expected. It is of note, however, that the distribution of African and European ancestry across ethnoracial categories showed more overlapping in Salvador than in the other sites. With regards to Native American ancestry, there was no clear pattern: in Pelotas, persons self-reported as Mixed and Black had significant higher median of Native American ancestry than Whites; in Bambuí, only persons self-reporting as Mixed showed higher level of Native ancestry, while in Salvador this was true only for persons self-reporting as White.

Ethnoracial self-classification as White, Mixed and Black in each cohort, by quartiles of individual African ancestry are shown in Table 2. Self-reporting as Black were more likely at the highest quartile of African ancestry in Pelotas (83.8%), Bambui (100.0%) and Salvador (97.2%). In contrast, we found a stronger likelyhood of self reporting as White at the highest quartile of African ancestry in Salvador (60.0%) relative to Pelotas (0.7%) and Bambui (0.8%). Results of the quantile regression anlysis showed that the strength of the association between the phenotype and African ancestry varied largely across the 3 sites, with pseudo R^2 values of 0.50 in Pelotas, 0.22 in Bambui and 0.13 in Salvador in the analysis comparing those above/bellow median of African ancestry. The differences across populations remained in the analyses comparing those above/below the 0.75 percentile of African ancestry (pseudo R^2 = 0.64, 0.32 and 0.13, respectively).

The joint analysis and the analysis by cohort population of the predicted probabilities of self-reporting as Black, Mixed and White along the African ancestry continuum are shown in Figure 2. African genomic ancestry showed an S-shape positive association with self-reporting as Black, which was consistent in all populations, whereas the reverse was observed for self-reporting as White. In the joint analysis, as well as for each cohort separately, these trends were statistically significant ($p<0.001$ in Waldś test). The probability of self-reporting as Black increased sharply as the proportion of African ancestry reached about 20% in Pelotas and 40% in Bambuí. The probability of self-reporting as White decreased sharply as the proportion of African ancestry reached about 10%-20% in these two populations. These increase/decrease were smoother in Salvador than in the other two sites. Self-classified Mixed individuals showed a bell-shaped predicted probability of having African ancestry in all sites.

Discussion

This is the first large community-based multicenter study to investigate the association between individual proportions of genome-wide based African, European and Native American ancestries and likelihood of ethnoracial self-classification in Brazil. The key findings are: first, the association between the phenotype and genome ancestry was statistically significant, but the strength of the association varied largely across populations; second: the association between Black and White self-classification with ancestry was most consistent in the extremes of the high and low proportion of African ancestry.

We confirmed previous historical and genetics reports of the largest African ancestry observed in Northeastern, as well as predominant European ancestry in Southeastern and Southern Brazil[2,5,7,22]. Furthermore, the contribution by Native Americans to the studied individuals was consistently small in the three sites. This is also in agreement with genetic reports indicating that Native American ancestry is higher in the North-West Brazil (Amazonia), a region that was not considered in our analysis[7].

In order to examine whether – and how – ethnoracial classification correlates with genomic ancestry, we used three different methods of

Table 1 | Ethnoracial self-classification and median individual proportion of African, European and Native American ancestries in all participants and by cohort population (Epigen-Brazil). (*) P <0.001 for differences across population. Mixed is ''pardo'' in official Portuguese.

	Cohort population			
	Pelotas (South)	Bambui (Southeast)	Salvador (Northeast)	All
	N= 3,533	N=1,442	N=876	N=5,851
Ethnoracial classification, %				
Black	16.6	2.5	49.3 *	18.1
Mixed (''pardo'')[1]	5.8	36.9	43.3	19.1
White	77.5	60.6	7.4	62.9
Genomic ancestry, median (interquartile range)				
African	6.6 (3.8-16.3)	9.6 (4.8-17.5)	50.5 (40.9-60.4) *	9.2 (4.5, 33.8)
European	85.3 (72.8-91.0)	83.8 (74.2-91.2)	42.4 (33.7-52.3) *	82.1 (57.1, 90.1)
Native American	6.3 (3.8-9.6)	5.4 (2.8-8.4)	5.8 (4.2-7.8) *	6.0 (3.7, 9.0)

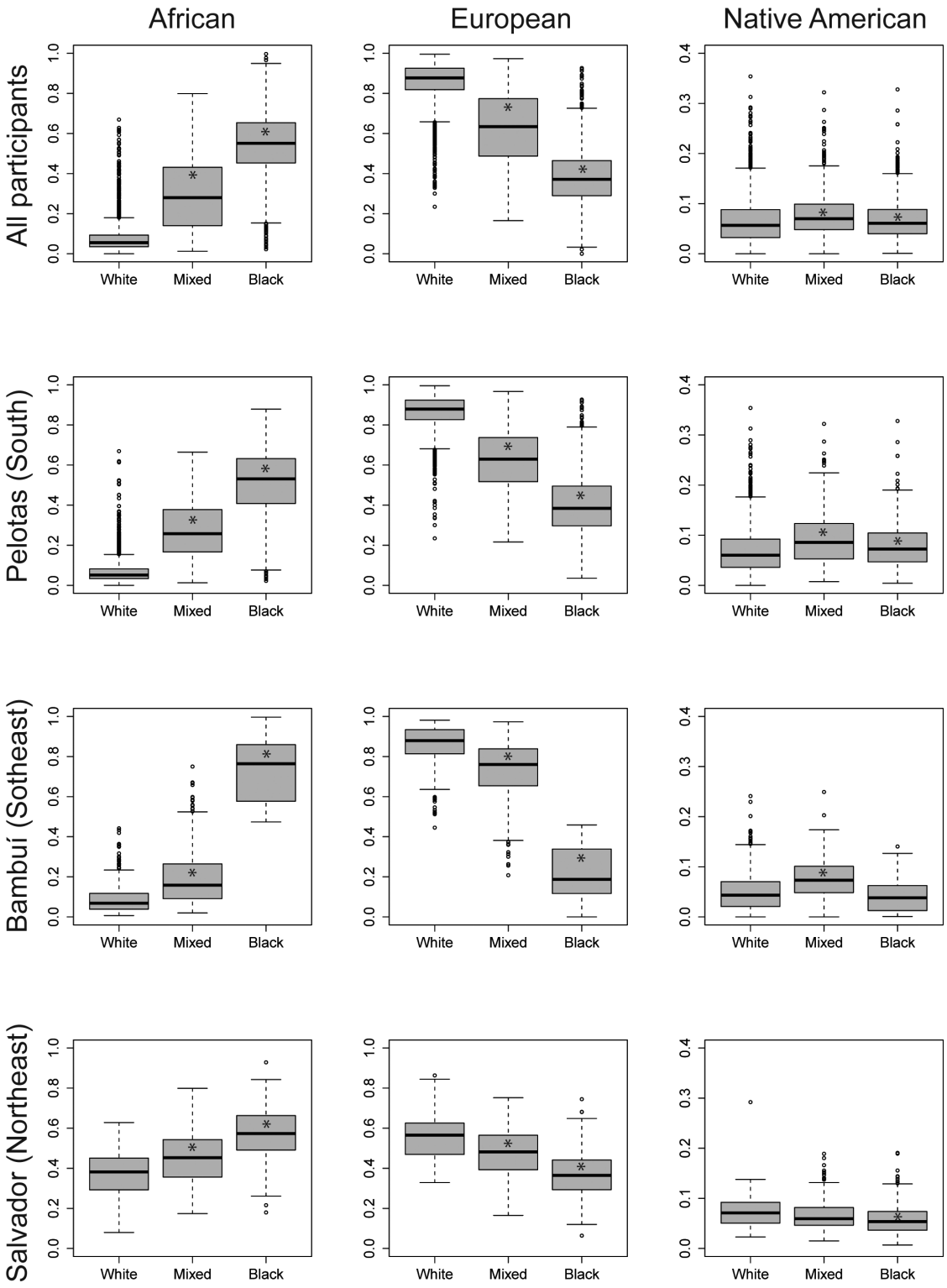

Figure 1 | Box plot contrasting ethnoracial self-classification (White, Mixed and Black) to median individual proportion of genomic African, European and Native American ancestries in all participants, and by cohort population (The Epigen Initiative). Mixed is "pardo" in official Portuguese. (*) p <0,001 for comparisons between each ethnoracial category to White.

Table 2 | Ethnoracial self-classification by quartiles of individual African ancestry, and by cohort population (Epigen-Brazil). B (95% CI): coeficient and 95% confidence intervals estimated by quantile regression. (*) p<0.01; (**) p<0.001. Mixed is ''pardo'' in official Portuguese.

	Quartiles						
	Total	Lowest	2nd	3rd	Highest	B (95% CI)	B (95% CI)
	N	N(%)	N (%)	N (%)	N (%)	(median regression model)	(0.75 regression model)
Pelotas (South)							
White	2739	41.1	39.2	19.0	0.7	1.0	1.0
Mixed (''pardo'')[1]	206	3.4	5.3	58.7	32.5	0.21 (0.20. 0.23) **	0.30 (0.28, 0.31) **
Black	588	0.7	1.4	14.1	83.8	0.48 (0.47, 0.49) **	0.55 (0.54, 0.56) **
						Pseudo R^2 = 0.50	Pseudo R^2 = 0.64
Bambui (Southest)							
White	874	33.4	30.2	35.6	0.8	1.0	1.0
Mixed (''pardo'')	532	6.4	19.7	59.2	14.7	0.09 (0.08. 10.3) **	0.15 (0.13, 0.16) **
Black	36	0	0	0	100.0	0.70 (0.66, 0.74) **	0.73 (0.68, 0.79) **
						Pseudo R^2 = 0.22	Pseudo R^2 = 0.32
Salvador (Northeast)							
White	65	0	1.5	38.5	60.0	1.0	1.0
Mixed (''pardo'')[1]	379	0	0	19.8	80.2	0.07 (0.03, 11.5) *	0.09 (0.05, 0.13) **
Black	432	0	0	2.8	97.2	0.19 (0.15, 0.23) **	0.21 (0.17, 0.25) **
						Pseudo R^2 = 0.13	Pseudo R^2 = 0.13

analyses. The first (a population measure), aimed at assessing how ethnoracial self-classification varied by medians of African, European and Native American ancestry. The other two methods, based on individual level measures, aimed at comparing the likelihood of the self-classification at the same levels of African ancestry across populations, as well as assessing how the relationship between ethnoracial self-classification changed along the proportion of genomic African ancestry continuum. Our results showed statistically significant associations between ancestry and the phenotype both at population and individual levels. However, the extent of overlap of individual proportions of each ancestry across ethnoracial groups was more evident in the Salvador population relative to the other sites. The association between Black and White self-identification with African ancestry continuum scale was S shape in all sites, but smoother in the Salvador population. Further, those who self-identified as Mixed tended to show intermediate proportions of African ancestry in all studied populations. This is in agreement with sociological and demographic conceptions that Mixed ("pardo" in official Portuguese) comprises multiple terms of popular discourse denoting ethnoracial admixture in Brazil[2].

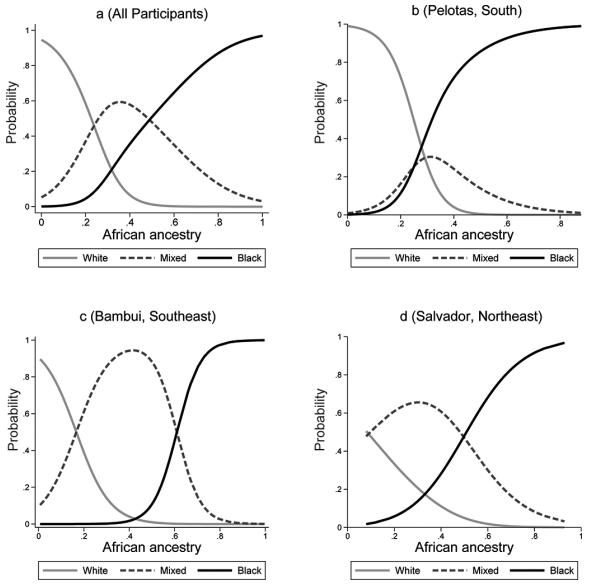

Figure 2 | Predicted probability of ethnoracial self-classification as Black, Mixed and White along the genomic proportion of African ancestry continuum in all participants, and by cohort population (Epigen-Brazil). Mixed is ''pardo'' in official Portuguese.

Previous sociological studies have suggested that ethnoracial self-classification in Brazil may tend to avoid nonwhite, and especially Black, categories since these were often associated with negative characteristics[2]. They suggest that miscegenation tends to shift self-reporting towards White, while segregation – as in the United State – would tend to shift self-reporting towards Black[2]. Our results indicate that avoidance of Black category may not be generalizable for the Brazilian population. In the current study, this effect appears to happen only in individuals from Salvador, where persons at the highest proportion of African ancestry were more likely to call themselves White relative to their counterparts from Pelotas and Bambui.

This study has strengths and limitations. Strengths include the very large number of SNPs used and the use of large community-based samples from different regions of eastern Brazil, as well as the fact that, the same set of reference populations (representing European, African, and Native American individuals) have been used in analyzing the three cohorts; thus, the inferred admixture ratios are comparable among the studied populations. Although the Pelotas and Bambuí cohorts are representative of the general population of their respective areas, in the eligible age groups, the cohort in Salvador oversampled individuals living in poor environments; thus, although there is good internal consistency, the results cannot be interpreted as representing the whole population of this city.

Summarizing, our results respond to three main sociological questions[2] that were not answered yet. They are: first, ethnoracial self-classification in Brazilians is certainly not random with respect to genome individual ancestry; second, the association between ethnoracial self-classification and genome based ancestry is not linear, with most consistent associations in the extremes of the African ancestry continuum scale; third, a tendency to whitening ethnoracial self-identification was found in persons from Salvador (where African ancestry is more common), but not in persons from the remaining two sites (where European ancestry predominates). Our results provides support to the view that ethnoracial self-classification is affected by both genomic ancestry and non-biological factors.

Methods

Cohort designs and ethnoracial self-classification. The 1982 Pelotas birth cohort study was conducted in Pelotas, a city in Brazil's extreme South, near the Uruguay border, with 214 000 urban inhabitants in 1982. Throughout 1982, the three maternity hospitals in the city were visited daily and births were recorded, corresponding to 99.2% of all births in the city. The 5,914 live-born infants whose families lived in the urban area constituted the original cohort. At age of 23 years, 3,736 participants categorized themselves according to the five ethnoracial categories used by the Brazilian census[1], as previously described. The Native American and yellow categories (67 and 64 individuals, respectively) were excluded from the current analyses. Further details are shown in a previous publication[19].

The Bambui cohort study of ageing is ongoing in Bambuí, a city of approximately 15,000 inhabitants, in Minas Gerais State in Southeast Brazil. The population eligible for the cohort study consisted of all residents aged 60 years and over on 1 January 1997, who were identified from a complete census in the city. Of a total of 1,742 older residents, 1,606 constituted the original cohort. At baseline, 1,442 participants categorized themselves into the above mentioned ethnoracial groups[1], according to standard photographs of Brazilians; no individuals categorized themselves as Amerindian or yellow. Further details of the Bambui study can be seen elsewhere[20].

The Salvador-SCAALA project is a longitudinal study involving a sample of 1,445 children aged 4-11 years in 2005, living in Salvador, a city of 2.7 million inhabitants in Northeast Brazil. The population is part of an earlier observational study that evaluated the impact of sanitation on diarrhea in 24 small sentinel-areas selected to represent the population without sanitation in Salvador. In the 2013 follow-up, 879 participants categorized themselves according to the previous mentioned ethnoracial groups[1] and were included in the present analysis; in the same way as in Bambui, no individuals categorized themselves as Amerindian or yellow in Salvador. Further details can be seen elsewhere[21].

Genotyping and external parental populations. The Epigen-Brazil participants were genotyped by the Illumina facility (San Diego, California) using the Omni 2.5M array. We performed the unsupervised tri hybrid (k=3) ADMIXTURE analyses based on 370,539 SNPs shared by samples from the HapMap Project, the Human Genome Diversity Project (HGDP)[23,24] and the Epigen-Brazil study population. As external panels, we used the following HapMap samples: 266 Africans (176 Yoruba in Ibadan, Nigeria [YRI] and 90 Luhya in Webuye, Kenya [LWK]), 262 Europeans (174 Utah residents with Northern and Western European ancestry [CEU] and 88 from Toscans from Italy [TSI]), 170 admixed individuals (77 Mexicans from Los Angeles, California [MEX] and 83 Afro-African from Southwest USA [ASW]), and 93 Native Americans from the HGDP (25 Pima, 22 Karitiana, 25 Maya and 21 Surui). The same set of reference populations was used in analyzing the three cohorts.

Family structure. To assess the familial structure, we estimated kinship coefficients for each possible pair of individuals from each cohort, using the method implemented in the REAP software (Related Estimation in Admixed Populations)[25]. This method was specifically developed to obtain accurate estimations of kinship coefficients in admixed populations, solely using genetic data and without using pedigree information. We considered a pair of individuals as related if the estimated kinship coefficient between them was ≥ 0.1. This cutoff includes second- degree relatives such as a person's uncle/aunt, nephew/niece, grandparent/grandchild or half- sibling, and any closer pair of relatives. Based on this cut-off, we identified set of related individuals (i.e. families) and assigned to each individual a categorical variable that represent his/her family. Because Pelotas and Salvador showed very few families, we decided to exclude related individuals (defined on the basis of the above mentioned cut-off). Therefore, 72 persons from Pelotas and 3 from Salvador were excluded from this analysis because they were related. The Bambuí cohort participants showed an important family structure (885 were related), so excluding them would lead to loss of power and possibly a degree of selection bias, so we opted for keeping related individuals, and undertaking sensitivity analysis to assess the influence of family structure on our results.

Statistical analyses. To take into account the differences across populations, we stratified analyses into the three study areas. To estimate the contribution from Africans, Europeans and Native Americans to the Epigen individuals we used the ADMIXTURE software[26]. We assumed three clusters to mimic the three main components of Brazilian ancestry, and used an unsupervised mode in order to allow the program to identify clusters corresponding to the ancestral populations solely from the genetic structure of our dataset. ADMIXTURE performs a model-based maximum-likelihood estimate of individual ancestry proportions, using an algorithm based on a sequential quadratic programming for block updates, coupled with a novel quasi-Newton acceleration of convergence.

Because the distribution of ancestry proportions was asymmetric, we calculated medians instead of means. Pearson's chi-square test was used to assess statistical significance among frequencies, and Kruskal-Wallis rank test or Mann-Whitney test were used to assess statistical significance of differences among medians, respectively. We compared likelihood of individual self-ethnoracial classification at the same level of African ancestry. We examined this by examining proportions of White, Mixed and Black self-classification by quartiles of African ancestry, calculated for the population as a whole, including the people from the 3 cohorts. Quantile (median and 0.75) regression was used to estimate the strength of these associations[28].

To quantify how the relationship between ethnoracial self-classification changed along the proportion of genomic African ancestry continuum, we fitted a multinomial logistic regression for the joint analysis of the three populations, adjusted for the cohort effect, and plotted the predicted probabilities for the outcome. Similar analyses were performed separated for each cohort population. A generalized Hosmer-Lemenshow goodness-of-fit test was use to assess the adequacy of the above mentioned multinomial models[27].

For the Bambuí cohort, we did a sensitivity analyses to assess the influence of familial structure on our results. We verified this by examining the previous mentioned unadjusted multinomial models relative to a model containing a random effect term for adjustments for family structure[29], and verified that this did not affect our results (not shown). Thus, our analysis were based on all Bambui cohort participants, irrespective of kinship.

The analyses were carried out for pooled men and women, given that in all populations sex showed no statistically significant associations with either ethnoracial classification or genetic ancestry. Furthermore, we excluded age from our analyses for two reasons: first, age distributions were homogeneous in the Pelotas and Salvador cohorts (23 years and 12-22 years, respectively); and, second, age showed no significant associations with ethnoracial self-classification, as well as with genomic ancestry, in the Bambui cohort population, whose age ranged from 60 to 95 years.

Statistical analyses were conducted using STATA 13.0 statistical software (Stata Corporation, College Station, Texas). All p-values were 2-tailed (alpha = 0.05).

Ethics assessment. The Epigen protocol was approved by Brazil's national research ethics committee (CONEP, resolution number 15895, Brasília). The research has been conducted according to the principles expressed in the Declaration of Helsinki. Participants signed an informed consent form and authorized their genotyping.

1. I.B.G.E. (Instituto Brasileiro de Geografia e Estatística). *Atlas do Censo Demográfico de 2010.* Available: http://censo2010.ibge.gov.br/apps/atlas/ Accessed 26 August 2014.
2. Telles, E.E., *Race in Another America: the Significance of Skin Color in Brazil.* (Princeton University Press, Princeton, 2004).
3. Parra, F. C. *et al.* Color and genomic ancestry in Brazilians. *Proc Natl Acad Sci U S A* **100**, 177–82 (2006).
4. Cardena, M. M. *et al.* Assessment of the relationship between self-declared ethnicity, mitochondrial haplogroups and genomic ancestry in Brazilian individuals. *PLoS One* **8**, e62005 (2013).
5. Pena, S. D. *et al.* The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PLoS One* **6**, e17063 (2011).
6. Durso, D. F. *et al.* Association of genetic variants with self-assessed color categories in Brazilians. *PLoS One* **9**, e83296 (2014).
7. Ruiz-Linhares, A. *et al.* Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genet* **10**, e1004572 (2014).
8. Magalhães da Silva, T. *et al.* The correlation between ancestry and color in two cities of Northeast Brazil with contrasting ethnic compositions. *Eur J Hum Genet.* DOI: 10.1038/ejhg.2014.215 (2014) [In press].
9. Chor, D. Health inequalities in Brazil: race matters. *Cad. Saúde Pública.* **29**, 1272–1275 (2013).
10. Travassos, C., Laguardia, J., Marques, P. M., Mota, J. C. & Szwarcwald, C. L. Comparison between two race/skin color classifications in relation to health-related outcomes in Brazil. *Int J Equity Health* **10**, 35 (2011).
11. Perreira, K. M. & Telles, E. E. The color of health: skin color, ethnoracial classification, and discrimination in the health of Latin Americans. *Soc Sci & Med* **116**, 241–250 (2014).
12. Macinko, J., Mullachery, P., Proietti, F. A. & Lima-Costa, M. F. Who experiences discrimination in Brazil? Evidence from a large metropolitan region. *Int J Equity Health* **18**, 80 (2012).
13. Chor, D., Faerstein, E., Kaplan, G. A., Lynch, J. W. & Lopes, C. S. Association of weight change with ethnicity and life course socioeconomic position among Brazilian civil servants. *Int J Epidemio* **33**, 100–6 (2004).
14. Almeida-Filho, N. *et al.* Social inequality and alcohol consumption-abuse in Bahia, Brazil--interactions of gender, ethnicity and social class. *Soc Psychiatry Psychiatr Epidemiol* **40**, 214–22 (2005).
15. Chor, D. & Lima, C. R. Aspectos epidemiológicos das desigualdades raciais em saúde no Brasil. *Cad Saude Publica* **21**, 1586–94 (2005).
16. Horta, B. L., Gigante, D. P., Candiota, J. S., Barros, F. C. & Victora, C. G. Monitoring mortality in Pelotas birth cohort from 1982 to 2006, Southern Brazil. *Rev Saude Publica* **42**, 108–14 (2008).
17. Matijasevich, A. *et al.* Widening ethnic disparities in infant mortality in southern Brazil: comparison of 3 birth cohorts. *Am J Public Health* **98**, 692–68 (2008).
18. Fernandes, F. *O negro no Mundo dos Brancos.* (1972) Available: http://eraju2013.files.wordpress.com/2013/09/fernandes-florestan-o-negro-no-mundo-dos-brancos-1.pdf. Accessed 26 August 2014
19. Victora, C. G. & Barros, F. C Cohort profile: the 1982 Pelotas (Brazil) birth cohort study. *Int J Epidemiol* **35**, 237–42 (2006).
20. Lima-Costa, M. F., Firmo, J. O. & Uchoa, E. Cohort profile: the Bambui (Brazil) Cohort Study of Ageing. *Int J Epidemiol* **40**, 862–7 (2011).
21. Barreto, M. L. *et al.* Risk factors and immunological pathways for asthma and other allergic diseases in children: background and methodology of a longitudinal study in a large urban center in Northeastern Brazil (Salvador-SCAALA Study). *BMC Pulmonary Medicine* **6**, e15 (2006).
22. Santos, N. P. *et al.* Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. *Hum Mutat* **31**, 184–90 (2010).

23. International HapMap 3 Consortium *et al*., Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52-8 (2010).
24. Li, J. Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation, *Science* **319**, 1100–4 (2008).
25. Thornton, T. *et al.* Estimating kinship in admixed populations. *Am J Hum Genet* **91**, 122–38 (2012).
26. Alexander, D. H., Novembre, J., Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655–64 (2009).
27. Fagerland, M. W. & Hosmer, D. W. A generalized Hosmer-Lemenshow goodness-of-fit test for multinomial logistic regression models. *Stata J* **12**, 447-453.
28. Koenker, R. *Quantile Regression* (Cambridge University Press, New York, 2005).
29. McCulloch, C. E., Searle, S. R. & Neuhaus, J. M. *Generalized, Linear and Mixed Models* (2nd Wiley, Hoboken, 2008).

Acknowledgments

This work was supported by the Department of Science and Technology (DECIT, Ministry of Health) and National Fund for Scientific and Technological Development (FNDCT, Ministry of Science and Technology), Funding of Studies and Projects (FINEP, Ministry of Science and Technology, Brazil), Coordination of Improvement of Higher Education Personnel (CAPES, Ministry of Education, Brazil). MFLC, MLB, BLH, ACP, CGV, ETS, CBC, JOAF and SVP are supported by the Brazilian National Research Council (CNPq).

Authors contributions

MFL-C, MLB, BLH, CGV and LCR conceived the study. MFL-C, MLB, BLH, CGV are the cohorts Coordinators, providing samples and data for each cohort. MHG, JM, FSGK and FR-S analyzed the data. ACP and ET-S coordinated the genomic analyses. MFL-C wrote the manuscript. All the authors contributed with discussion on the results and on the manuscript. The Consortiate authors CCC, JSC, GNOC, NE, RLF, CAF, JOAF, ARVRH, TPL, MM, WCSM, IOO, SVP, MRR, HCS and TMS contributed with data, bioinformatic resources or statistical analyses.

Additional information

Competing financial interests: The author(s) declare no competing financial interests.

How to cite this article: Lima-Costa, M.F. *et al.* Genomic ancestry and ethnoracial self-classification based on 5,871 community-dwelling Brazilians (The Epigen Initiative). *Sci. Rep.* 5, 9812; DOI:10.1038/srep09812 (2015).

Consortia

Cibele C. Cesar[1], Jackson S. Conceição[2], Gustavo N.O. Costa[2], Nubia Esteban[3], Rosemeire L. Fiaccone[2], Camila A. Figueiredo[2], Josélia O.A. Firmo[4], Andrea R.V.R. Horimoto[3], Thiago P. Leal[5], Moara Machado[5], Wagner C.S. Magalhães[5], Isabel Oliveira de Oliveira[3], Sérgio V. Peixoto[4], Maíra R. Rodrigues, Hadassa C. Santos[3] & Thiago M. Silva[2]

[1]Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Belo Horizonte, Brazil, [2]Universidade Federal da Bahia, Instituto de Saúde Coletiva, Salvador, Brazil, [3]Universidade de São Paulo, Instituto do Coração, São Paulo, Brazil, [4]Fundação Oswaldo Cruz, Instituto de Pesquisas Rene Rachou, Belo Horizonte, Brazil, [5]Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas, Belo Horizonte, Brazil

Outras aplicações das inferências de ancestralidade em estudos biomédicos

Estudos de ancestralidade genômica possuem aplicações em estudos biomédicos. Foram verificados especialmente três tipos de aplicações.

A primeira aplicação consistiu em determinar um painel mínimo de AIMs para realizar a análise de ancestralidade continental com o menor número de marcadores possível com a mesma qualidade de uma abordagem *genome-wide* (Santos et al. 2015). Esta análise é importante para que não se necessite utilizar uma densidade extremamente alta de SNPs (portanto alto custo) para fazer uma inferência robusta de ancestralidade. Portanto, neste trabalho, também parte do consórcio EPIGEN-Brasil, realizei as análises de ancestralidade individual com quatro grupos candidatos de AIMs, concluindo que a estimativa mais próxima da abordagem *genome-wide* foi composta por 192 AIMs (Anexo A).

É conhecido que a ancestralidade pode atuar como variável confundidora em estudos epidemiológicos (Tarazona-Santos et al. 2007; Pereira et al. 2012). Em qualquer estudo caso-controle, o primeiro passo é realizar uma amostragem da população, onde casos e controles devem ser da mesma população, para que a amostragem seja etnicamente homogênea. Se houver uma amostragem etnicamente diferente entre casos e controles, e houver uma maior prevalência da doença na população de casos amostrada, pode haver uma associação estatística espúria, pois indivíduos com a ancestralidade da população onde a doença for mais comum podem estar super-representados na amostragem de casos e qualquer alelo mais comum nesta população pode apresentar correlação estatística com a doença, mesmo sem ser um alelo de suscetibilidade. Portanto, em todos os estudos de associação, o primeiro passo deve ser estimar a ancestralidade dos indivíduos amostrados, para verificar se casos e controles são etnicamente diferentes e evitar este problema de associação espúria ou controlar o efeito das diferenças de ancestralidade entre casos e controles, especialmente em populações miscigenadas.

Portanto, a segunda aplicação foi utilizar a ancestralidade individual como potencial variável confundidora em estudos de genética epidemiológica. Especificamente, em um trabalho que participei realizando as análises de ancestralidade, foram utilizados 87 AIMs para estimar a ancestralidade individual de populações de nativos peruanos e

miscigenados peruanos a fim de ser um cofator na análise de associação de genes candidatos à susceptibilidade ao câncer gástrico (Zamudio et al. 2015, Anexo B).

A terceira aplicação consiste em estudos de farmacogenética, onde variantes alélicas de interesse clínico relacionadas a resposta a fármacos podem ser relacionadas à ancestralidade. Esta aplicação foi avaliada nos capítulos seguintes desta tese.

CAPÍTULO 2 - Farmacogenética populacional e ancestralidade

Interethnic variability of CYP2D6 alleles and of predicted and measured metabolic phenotypes across world populations

Este capítulo da tese foi desenvolvido no contexto de atividades da *Rede Iberoamericana de Farmacogenética y Farmacogenómica* (RIBEF - http://www.ribef.com/), que consiste em uma colaboração de 41 grupos de pesquisa ao longo da América Latina e Península Ibérica. O fundador e coordenador desta rede (Dr. Adrián Llerena) foi pesquisador visitante da Universidade Federal de Minas Gerais pelo programa Ciência sem Fronteiras e meu orientador no exterior, e colabora ativamente com nosso grupo de pesquisa (Rodeiro et al. 2012; LLerena et al. 2014). O Consórcio CEIBA (Consórcio Europeu Iberoamericano de Farmacogenética) é um dos componentes da RIBEF, cujo objetivo principal é estudar a variabilidade inter-étnica dos genótipos e fenótipos mais relevantes de genes envolvidos no metabolismo de fármacos como os do citocromo p-450 (*CYP2D6, CYP2C9, CYP2C19*) e outros biomarcadores em populações mundiais, especialmente em latino-americanos e europeus da Península Ibérica. Os trabalhos envolvidos nos Capítulo 2 e 3 desta tese (exceto o manuscrito sobre leucemia linfoide aguda) são resultado da colaboração do nosso grupo de pesquisa com o consórcio CEIBA da RIBEF.

A RIBEF complementa os esforços da Rede Brasileira de Farmacogenética (REFARGEN - http://www.refargen.org.br/), um consórcio de grupos de pesquisa brasileiros criado em 2004 com o objetivo de estudar farmacogenética com impacto na saúde de populações (Suarez-Kurtz 2004).

As revisões sistemáticas são de grande importância para a sumarização de respostas a perguntas científicas e resolução de controvérsias na literatura. Este tipo de revisão, diferente das revisões simples de literatura, possui uma pergunta clara e uma metodologia sistemática, seguindo as etapas de um artigo original, eliminando o viés subjetivo de escolha de literatura (Khan et al. 2003). Além disso, as revisões sistemáticas geralmente esgotam as fontes da literatura sobre um determinado tema, de acordo com critérios de busca, de inclusão e de exclusão dos artigos, que devem ser previamente estabelecidos. Dessa forma, a RIBEF possui uma iniciativa de realização de revisões

sistemáticas mundiais com três genes de grande importância farmacogenética: *CYP2D6*, *CYP2C9* e *CYP2C19*.

Na publicação a seguir (LLerena et al. 2014), referente ao gene *CYP2D6*, realizei, juntamente com a então doutoranda María Eugenia G. Naranjo, as etapas de busca, análises, interpretação dos resultados e escrita do manuscrito, bem como as fases de submissão e revisão do artigo. Além desta publicação, duas outras revisões foram publicadas seguindo a mesma metodologia, porém utilizando os genes *CYP2C9* e *CYP2C19* (Céspedes-Garro et al. 2015; Fricke-Galindo et al. 2015). Nestes dois artigos contribuí com a busca, análises e revisão do texto final (Anexos A e B). Estas publicações foram realizadas no período de meu doutorado sanduíche no laboratório do Dr. LLerena (CICAB), Badajoz, Espanha, sede da RIBEF.

EXPERT OPINION

Interethnic variability of *CYP2D6* alleles and of predicted and measured metabolic phenotypes across world populations

Adrián LLerena, Maria Eugenia G Naranjo, Fernanda Rodrigues-Soares, Eva M Penas-LLedó, Humberto Fariñas & Eduardo Tarazona-Santos
[†]*University of Extremadura Hospital, CICAB Clinical Research Center, Badajoz, Spain*

Introduction: The frequency of *CYP2D6* alleles, related to either a lack of or increased enzymatic activity, which may lead to poor metabolism (PM) or ultrarapid metabolism (UM), can vary across ethnic groups and hence across geographic regions.

Areas covered: Worldwide original research papers on *CYP2D6* allelic frequencies, metabolic phenotype frequencies measured with a probe drug, and/or genotype frequencies that studied > 50 healthy volunteers, were included in analyses to describe the distributions of alleles, phenotypes predicted from genotypes (predicted poor metabolizers [gPMs], predicted ultrarapid metabolizers [gUMs]) and metabolic phenotypes (mPMs, mUMs) across ethnic groups and geographic regions. The analysis included 44,572 individuals studied in 172 original research papers.

Expert opinion: As of today, Africa and Asia are under-represented in this area relative to the total number of their inhabitants, so that further studies in these regions are warranted. The *CYP2D6*4* allele frequency was higher in Caucasians, *CYP2D6*10* in East Asians, *CYP2D6*41* and duplication/multiplication of active alleles in Middle Easterns, *CYP2D6*17* in Black Africans and *CYP2D6*29* in African Americans, than in other ethnic groups. Overall, gPMs and mPMs are more frequent among Caucasians, and gUMs among Middle Easterns and Ethiopians. However, mUMs could not be evaluated because only two studies were found presenting this information. Further studies including mUMs are thus warranted. There is a correspondence between gPMs and mPMs, but the few studies of mUMs meant that their relationship with gUMs could not be demonstrated. Finally, evolutionary aspects of the *CYP2D6* allele distribution appear to support the Great Human Expansion model.

Keywords: *CYP2D6*, *CYP2D6*4*, *CYP2D6*10*, *CYP2D6*17*, *CYP2D6*29*, *CYP2D6*41*, ethnicity, poor metabolizers, population pharmacogenetics, ultrarapid metabolizers

Expert Opin. Drug Metab. Toxicol. (2014) ***10****(11):1569-1583*

1. Introduction

CYP2D6 is a drug-metabolizing enzyme of the CYP superfamily involved in the metabolism of around 25% of the drugs commonly used in clinical practice [1]. The *CYP2D6* gene is polymorphic, with over 100 allelic variants described to date [2]. Some *CYP2D6* variants such as *CYP2D6*3*, **4, *5* or **6* are related to lack of protein activity or synthesis. *CYP2D6*10, *17, *29* or **41* are associated with reduced enzyme activity, whereas the duplication/multiplication of active alleles (*CYP2D6*1xN, *2xN*) is related to increased activity [3].

10.1517/17425255.2014.964204 ISSN 1742-5255, e-ISSN 1744-7607

Article highlights.

- CYP2D6 research on healthy voluteers is under-represented in Africa and Asia in relation to the total number of their inhabitants.
- *CYP2D6*4* is more frequent in Europe and Caucasians, *CYP2D6*10* in Asia and East Asians, *CYP2D6*17* and **29* in Africa and Black populations, *CYP2D6*41* and multiplications in Middle East populations.
- gPMs and mPMs are more frequent in Europe and Caucasians.
- The highest frequency of gUMs is found in Middle Easterns and Ethiopians.

This box summarizes key points contained in the article.

Subjects who have null enzyme activity are denominated poor metabolizers (PMs), whereas individuals with increased activity are called ultrarapid metabolizers (UMs). These two metabolic conditions can be evaluated either by verifying the individual genotype, and based on the functional characteristics of each allele, thereby predicting the enzyme hydroxylation capacity (predicted poor metabolizers [gPMs] and predicted ultrarapid metabolizers [gUMs]), or by calculating the metabolic phenotype (mPMs and mUMs) after measuring a substrate test drug/metabolite's concentrations in urine or plasma [4].

Clinically, PM and UM phenotypes are particularly relevant because they may have either increased susceptibility to adverse drug reactions or a lack of response to treatment [5,6]. PMs have no appreciable activity and may present more adverse events when they are treated with CYP2D6 substrates [7-9]. Conversely, CYP2D6 UMs exhibit increased metabolism and have been related to variability in treatment response and susceptibility to disease. For example, UMs are reported to be more likely to not respond to antidepressants [6]. In regard to suicide, in some European populations, UMs are more frequent among those who complete it [10], present a lifetime history of suicide attempts [11], or make a severe suicide attempt [12], although further studies are needed in order to confirm these findings. Moreover, CYP2D6 UMs have also been associated with fatal exposure to morphine (the active metabolite of codeine) in children treated with codeine [13,14].

Previous studies have demonstrated that *CYP2D6* allele frequencies and metabolic phenotypes vary among world populations [15]. For instance, *CYP2D6*10* has been found with a high frequency in Asian populations, whereas *CYP2D6*17* and **29* are more frequent in Africans, and *CYP2D6*4* in Europeans [16,17].

The aim of the present study was to evaluate the variability of the *CYP2D6* allele, the phenotype predicted from genotype, and the measured metabolic phenotype frequencies in healthy volunteers across ethnic groups and main geographical territories. Some evolutionary aspects of the *CYP2D6* allele distribution are also discussed.

2. Review methods

A PubMed literature review was conducted using the terms 'CYP2D6' and the names of the different countries and continents, one by one. The phenotype search was done with the terms 'CYP2D6' and the probe drugs dextromethorphan, debrisoquine, sparteine, metoprolol, codeine and tamoxifen. The inclusion criteria were: i) *CYP2D6* allelic frequencies or measured metabolic phenotypes were reported; ii) the evaluated subjects were healthy volunteers, that is, controls or patients from case-control studies were excluded; and iii) at least 50 individuals were studied with some exceptions in accordance with previous methods [18]. All articles with one or more *CYP2D6* alleles were included in the review.

Finally, 172 original research papers were selected and classified by: i) geographic region; and ii) ethnic group. Six major geographic regions were considered: Africa, America, Asia, Europe, Oceania and the Middle East. The ethnic groups were classified following previous methods [18] with some modifications (**Table 1**): Caucasians were divided into American Caucasian, Scandinavian, Central European, Slav and Mediterranean South-European; Americans were divided into Native American and American Admixed; Africans were divided into North African, Black African and African Admixed. Australian Caucasians and Circumpolar populations (Eskimo plus Lapps from Finland) were added. South Asian, East Asian, African American and Middle Eastern were not subdivided as in previous work [18].

The *CYP2D6* allelic, predicted phenotype and measured metabolic phenotype frequencies of each geographic region and ethnic group were calculated. In order to calculate the phenotype predicted from genotype frequencies, the individuals with zero active *CYP2D6* alleles were classified as gPMs and those with more than two active *CYP2D6* alleles as gUMs, in accordance with a previous method [19-21]. Furthermore, some papers presented *CYP2D6* measured metabolic phenotypes based on more than one probe drug. In these cases, only one result was taken. The choice was made according to the reliability of the drug for the evaluation of the enzyme hydroxylation capacity (debrisoquine=dextromethorphan>sparteine >metoprolol) [22], followed by the highest number of individuals tested for each drug.

The *CYP2D6* allelic, predicted phenotype and measured metabolic phenotype frequencies of each geographic region and ethnic group were compared by a chi-squared test with Yate's correction using GraphPad Software [23]. The comparisons were made between the allelic, predicted phenotype and measured metabolic phenotype frequencies of each geographic region and ethnic group versus a weighted average of the allelic frequencies of the remaining groups. To calculate the weighted average, the number of subjects studied was taken into account. Moreover, the correlation between CYP2D6 gPMs and mPMs were analyzed using SPSS Base 19.0 (SPSS Inc., Chicago, IL).

Table 1. Major *CYP2D6* allele frequencies in different ethnic groups.

CYP2D6 alleles			*4	*10	*17	*29	*41	Duplication/ multiplication of active alleles	Country/zone	Ref.
Black	North Africans (n = 0)								North Africa	
	Black Africans (n = 766)	n	766	571	571	70	70	491	Ethiopia, Ghana, Tanzania,	[25,28,34,42-44]
		%	3.14	5.07	22.36	4.30	0.70	5.8	Zimbabwe, South Africa	
	African Americans	n	1190	867	867	713	462	867	USA, Trinidad-Tobago, Brazil	[19,24,29,45-48]
	(n = 1190)	%	6.94	4.03	17.64	7.45	3.47	2.59		
Middle	Middle Easterns	n	1690	1450	1350	391	686	1250	Iran, Israel, Jordania, Syria,	[16,49-57]
Eastern	(n = 1882)	%	12.30	7.12	2.31	1.01	16.66	6.72	Turkey, USA*	
Asian	South Asians (n = 1253)	n	1253	1253	1147	167	167	253	India, Trinidad-Tobago‡	[29,58-60]
		%	7.86	10.71	0.07	0.90	13.50	1.58		
	East Asians (n = 6894)	n	4003	6109	1321		4023	3356	China, Korea, Japan,	[61-80]
		%	0.42	45.00	0.23		2.78	0.91	Malaysia, Vietnam	
Native	Native Oceanians	n	299	299	239		60	239	Australia, New Zealand	[81,82]
	(n = 299)	%	2.78	1.86	0.20		3.50	0.00		
	Circumpolar populations	n	90	90					Canada	[83]
	(n = 90)	%	8.30	2.20						
	Native Americans	n	2140	1866	1071	785	1086	475	Chile, Colombia, Panama,	[16,84-93]
	(n = 2230)	%	10.09	2.61	0.38	0.13	2.91	2.21	Mexico, USA, Brazil, Argentina, Paraguay, Costa Rica, Venezuela	
Admixed	American Admixed	n	2542	2075	1604	617	633	1716	Colombia, Cuba, USA,	[20,86,87,91-104]
	(n = 2642)	%	12.02	5.11	1.65	0.72	4.50	2.55	Mexico, Nicaragua, Puerto Rico, Brazil, Chile, Costa Rica, Ecuador, Venezuela	
	African Admixed	n	99	99	99	99	99	99	South Africa	[19,105]
	(n = 99)	%	7.07	2.53	12.63	4.55	3.54	2.53		
Caucasian	Australian Caucasians (n = 0)								Australia	
	American Caucasians	n	1337	918	918	439	645	918	USA, Brazil, Cuba	[19,20,24,45,46,48, 101,106]
	(n = 1337)	%	18.48	2.55	0.81	0.22	8.61	2.02		
	Scandinavians (n = 921)	n	921						Sweden, Denmark	[107-109]
		%	26.29							
	Central-Europeans	n	1847	453	360	165	370	565	Germany, Hungary,	[110-116]
	(n = 1847)	%	18.80	2.73	0.14	0.30	8.39	1.68	Netherlands, Switzerland, Austria	
	Mediterranean	n	1938	905	800	122	122	1648	Spain, Italy, Portugal, Greece	[20,53,117-123]
	South-Europeans (n = 1938)	%	17.55	2.43	0.44	0.00	10.20	3.85		
	Slavs (n = 1527)	n	1527	290				713	Poland, Czech Republic,	[124-130]
		%	19.53	4.20				2.81	Estonia, Russia	

*Ashkenazi Jewish from USA.

‡Indian population living in Trinidad-Tobago.

n: Number of individuals studied for each *CYP2D6* allele and ethnic group.

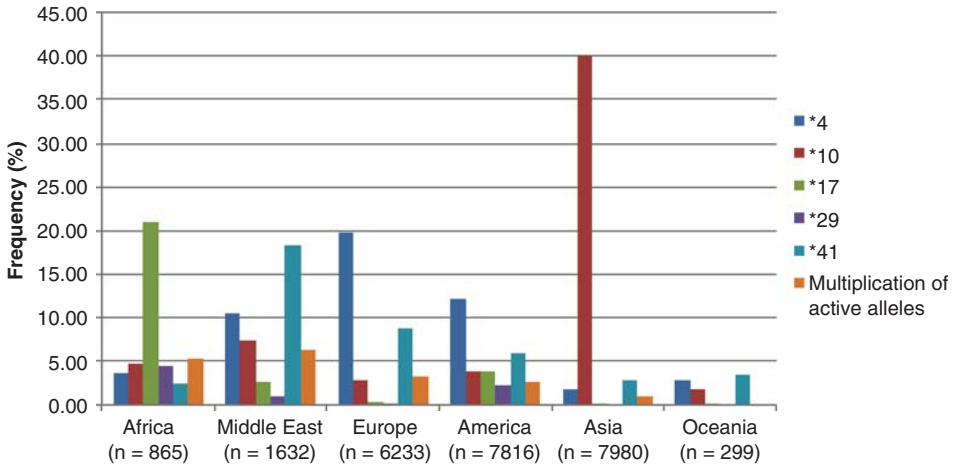

Figure 1. Main *CYP2D6* alleles in different geographic regions. Number of subjects studied (n).

This methodological approach has some limitations. In some cases, the single nucleotide polymorphism (SNP) studied for the *CYP2D6* alleles differed from one study to another. For example, *CYP2D6*2* is identified as 2850C>T plus 4180G>C in some papers and as -1584C>G in others [24,25]. However, in the present review, the *CYP2D6* allele results from the original papers are used independently of the genotyped SNPs. Moreover, the *CYP2D6*1* frequency can vary depending on the number of genotyped SNPs because this allele is estimated by the absence of other *CYP2D6* polymorphisms. For this reason, the *CYP2D6*1* frequency is not calculated in the present review.

In summary, the paper analyses the frequencies that have been reported as to date of *CYP2D6* alleles, phenotypes predicted from genotype and measured metabolic phenotypes by ethnic group and geographic region.

3. CYP2D6 studies across the world population

CYP2D6 has been studied in a total of 44,572 healthy volunteers (6.35×10^{-4} % of the world population), of whom 39.42% were Europeans, 24.05% Asians, 23.27% Americans, 6.65% Africans, 5.49% Middle Easterns and 1.12% Oceanians. Specifically, *CYP2D6* alleles were studied in 24,915 subjects, the CYP2D6 phenotype predicted from genotype was evaluated in 16,478 subjects and the CYP2D6 metabolic phenotype measured with a probe drug was evaluated in 23,925 subjects.

Taking into account the population of each geographic region [26], current knowledge about CYP2D6's world distribution seems to be based on an over-representation of European and American subjects. On the contrary, the African and Asian populations are under-represented, which suggests that additional studies in these populations are needed.

4. *CYP2D6* allele variability across ethnic and geographic groups

It is of note that globally, apart from the *CYP2D6*1* allele, *CYP2D6*2* is the predominant allele (19.93%), followed by *CYP2D6*10* (19.05%) and *CYP2D6*4* (11.32%). Major *CYP2D6* allelic variants by geographic region and ethnic group are shown in **Figure 1** and **Table 1**, respectively.

By ethnic group, the *CYP2D6*4* allele frequency was found to be higher in Caucasians than in the rest ($p < 0.001$), with the highest being in Scandinavians (26.29%). *CYP2D6*10* presented the highest frequency in East Asians (45.00%, $p < 0.001$). *CYP2D6*41* was found to be more predominant in Middle Easterns (16.66%), followed by South Asians (13.50%) and some Caucasian populations (American Caucasians, Central Europeans and Mediterranean-South Europeans; $p < 0.001$). *CYP2D6*17* was more frequent in Black Africans (22.36%) followed by African Americans (17.62%) and African Admixed (12.63%, $p < 0.001$), whereas *CYP2D6*29* was more frequent in African Americans (7.45%, $p < 0.001$) and African Admixed (4.55%, $p = 0.02$) than in other ethnic groups. The duplication/multiplication of active alleles was higher in Middle Easterns (6.72%), Black Africans (5.80%) and Mediterranean-South Europeans (3.85%, $p < 0.001$) than in the rest.

A previous study [18] described *CYP2D6* allele variability among diverse ethnic groups. However, that review used original papers containing not only healthy volunteers but also patients and controls from clinical studies. In order to avoid a potential population selection bias, only randomized healthy volunteers from original papers were included in the present study because a health condition or a selection criterion could both be related to some genetic condition, which could bias the *CYP2D6* frequencies [10,12,27]. Some differences in frequency values were found when comparing the present with

Table 2. Frequencies of CYP2D6 phenotypes predicted from genotype and phenotypes from probe drug measurements in the different ethnic groups.

Ethnic group		Predicted phenotype			Metabolic phenotype			Ref.
		N	gPMs n (%)	gUMs n (%)	N	mPMs n (%)	mUMs n (%)	
Black	North African	-	-	-	906	23 (2.54)	-	[131]
	Black African	718	7 (1.24)	43 (6.74)	1842	78 (4.23)	-	[25,28,34,35,42-44,132-140]
	African American	664	14 (2.11)	22 (4.10)	861	37 (4.30)	1 (0.97)	[24,29,45-48,141-144]
Middle Eastern	Middle Eastern	1780	38 (2.13)	121 (10.54)	816	19 (2.33)	-	[49-52,54-57,145-153]
Asian	South Asian	1328	18 (1.36)	4 (2.40)	1329	50 (3.76)	3 (1.8)	[29,58,59,78,80,149,154-161]
	East Asian	2051	5 (0.26)	28 (2.20)	1913	16 (0.84)	-	[61,68,72,75-77,162-173]
Native	Native Oceanian	239	1 (0.42)	0 (0.00)	100	0 (0.00)	-	[81,145,174]
	Circumpolar populations	90	3 (3.33)	-	407	15 (3.69)	-	[83,175,176]
	Native American*	1339	53 (3.96)	76 (7.67)	588	18 (3.06)	-	[16,84-93,177]
Admixed	American Admixed	2322	61 (2.63)	51 (3.34)	1441	71 (4.93)	3 (1.15)	[20,86,87,91-93,95-104,178]
	African Admixed	99	3 (3.03)	4 (4.04)	-	-	-	[105]
Caucasian	Australian Caucasian	-	-	-	100	6 (6.00)	-	[179]
	American Caucasian	905	52 (5.75)	26 (4.33)	1568	111 (7.08)	7 (5.34)	[20,24,45,46,48,101,106, 141,144,159,171,180]
	Scandinavian	921	80 (8.69)	-	2685	194 (7.23)	-	[107-109,176,181-186]
	Central European	1330	84 (6.32)	12 (2.12)	5717	465 (8.13)	-	[110-116,172,187-194]
	Mediterranean South-European	1165	55 (4.72)	62 (5.82)	2666	143 (5.36)	-	[22,119-123,195-203]
	Slavs	1527	97 (6.35)	26 (3.65)	986	82 (8.32)	-	[124-130,204,205]

N: Total number of individuals studied for predicted or metabolic phenotype. The number of studied subjects differ in gPMs, gUMs, mPMs and mUMs; n: Number of individuals for each phenotype and geographic localization; Empty cells: data not available.

*Original research papers with N < 50 were included for Native Americans.

gPMs: Genotype poor metabolizers; gUMs: Genotype ultrarapid metabolism; mPMs: Metabolic phenotypes poor metabolizers; mUMs: Metabolic phenotypes ultrarapid metabolism.

the previous study [18], for example, *CYP2D6*41* in Black Africans and African Americans, or duplications/multiplications of *CYP2D6* in Oceania. These differences could be due to the inclusion of just articles with > 50 healthy volunteers and/or the calculation of the weighted average by using another method.

By geographic region, *CYP2D6*4* (related to lack of enzyme activity) was higher in Europe (19.70%) than in other geographic groups ($p < 0.001$). Regarding alleles with decreased enzyme activity, *CYP2D6*10* was found to be more predominant in Asia (39.96%, $p < 0.001$) than in the rest, and *CYP2D6*41* presented the highest frequency in the Middle East (18.23%) followed by Europe (8.84%; $p < 0.001$). *CYP2D6*17* and *29* were Commoner in Africa (20.92 and 4.45%, respectively) and America (3.95 and 2.24%, respectively) than in other geographic regions ($p < 0.001$). However, the duplication/multiplication of active alleles (related to increased enzyme activity) was more frequent in the Middle East (6.40%, $p < 0.001$) and Africa (5.25%, $p < 0.001$).

5. Variability of the CYP2D6 phenotype predicted from genotype across ethnic and geographic groups

Table 2 shows the frequencies of gPMs and gUMs among different ethnic groups. The gPMs were more frequent in

Table 3. Frequencies of CYP2D6 phenotypes predicted from genotype and phenotypes from probe drug measurements in the different geographic regions.

Continent	Predicted phenotype from genotype					Phenotype from probe drugs					Ref.
	N	N gPMs	gPMs n (%)	N gUMs	gUMs n (%)	N	N mPMs	mPMs n (%)	N mUMs	mUMs n (%)	
Africa	817	663	10 (1.51)	737	47 (6.38)	2748	2748	101 (3.68)	-	-	[25,28,34,35,42-44,105, 132-140]
America	5737	5737	201 (3.50)	4071	201 (4.94)	4980	4918	265 (5.39)	663	14 (2.11)	[20,24,45,46,46,47,57,83,89-91, 94,98,99,101,103, 106,144,159,171,177,178,180]
Europe	4943	4943	316 (6.39)	2343	100 (4.27)	12,257	12,257	893 (7.29)	-	-	[107-111,113-116,119-130,172, 175,176,181-205]
Oceania	239	239	1 (0.42)	239	0 (0.00)	200	200	6 (3.00)	-	-	[81,174,179]
Middle East	1530	1530	25 (1.63)	898	99 (11.02)	816	816	19 (2.33)	-	-	[49-52,54-56,145-153]
Asia	3212	3212	18 (0.56)	1272	28 (2.20)	2924	2924	44 (1.50)	-	-	[58,59,61,68,72,75-78,80,149, 154-173]
Total world	16478	16,324	571 (3.50)	9560	475 (4.97)	23,925	23,863	1328 (5.57)	663	14 (2.11)	

N: Total number of individuals studied in each group and/or each phenotype; n: Number of individuals for each phenotype and geographic localization; Empty cells: data not available.

gPMs: Genotype poor metabolizers; gUMs: Genotype ultrarapid metabolism; mPMs: Metabolic phenotypes poor metabolizers; mUMs: Metabolic phenotypes ultrarapid metabolism.

Caucasians ($p < 0.001$), especially Scandinavians (8.69%) and the gUMs in Middle Eastern populations (10.54%, $p < 0.001$). It is of note that, among all the populations, Ethiopians presented the highest gUM frequency (28.7%, [28]). Moreover, gPMs showed the lowest frequency in East Asians (0.26%, $p < 0.001$) and no gUMs were detected in Native Oceanians.

As shown in **Table 3** concerning geographic regions, Europe (6.52%, $p < 0.001$) presented the highest frequency of gPMs, whereas gUMs were more frequent in the Middle East region (10.54%, $p < 0.001$). Conversely, the lowest frequencies of gPMs were found in Oceania (0.42%) and Asia (0.54%, $p < 0.001$), and gUMs were absent in Oceania and presented a low frequency in Asia (2.30%, $p < 0.001$).

6. Variability of the CYP2D6 measured metabolic phenotype across ethnic and geographic groups

Table 2 shows the mPM and mUM frequencies by ethnic group. Caucasian populations, except Mediterranean South-Europeans, showed the highest frequency of mPMs (6.00 – 8.13%, $p < 0.001$), while the lowest frequency was found in East Asians (0.84%, $p < 0.001$) and they were absent in Native Oceanians. The mUM frequencies were only available for a few populations (African American, American Admixed, American Caucasian and South Asian) and varied from 0.97% in African-Americans to 5.34% in American Caucasians.

As shown in **Table 3**, by geographic region, the mPMs were more frequent in Europe (7.29%, $p < 0.001$), and the lowest frequency was found in Asia (1.50%, $p < 0.001$). The mUM frequency was only available for the American continents, with a frequency of 2.11%.

7. Relationship between the CYP2D6 phenotype predicted from genotype and the CYP2D6 measured metabolic phenotype

There is a correspondence between gPMs and mPMs. **Figure 2** shows the correlation between the frequencies of *CYP2D6* gPMs and mPMs by ethnic group ($r = 0.869$, $p < 0.001$) and by geographic region ($r = 0.936$, $p = 0.006$). The correlation for UMs could not be analyzed because the mUM frequency was only available for two ethnic groups from America [20,29].

A gPM, or those carrying zero *CYP2D6* active alleles, will produce a non-functional protein, and will always be identified as mPMs in phenotyping studies with probe drugs. However, the genotype will not always predict the phenotype for other individuals. For example, some predicted extensive metabolizers (gEMs) may be phenotypically PMs due to interaction with other CYP2D6 substrates or inhibitors (endobiotic or xenobiotic). Similarly, this discordance appears when identifying mUMs by the use of genetic methods [20,29]. These differences between gUMs and mUMs can be due to *CYP2D6* alleles that are not studied, or to epigenetic, environmental or endogenous factors.

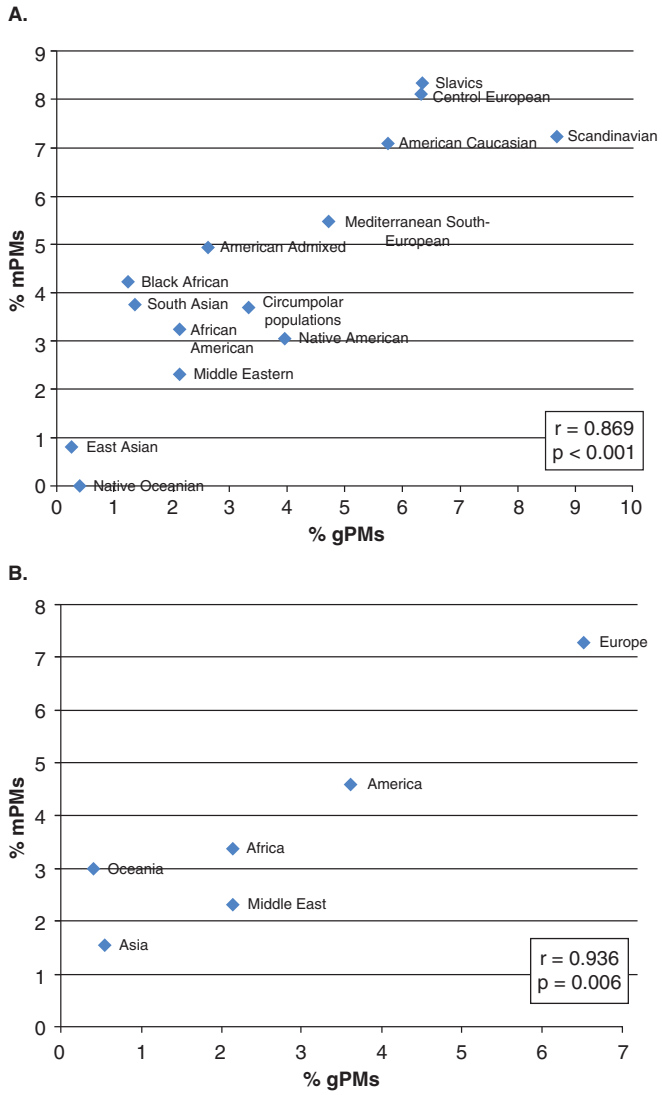

Figure 2. Relationship between the CYP2D6 PMs extrapolated from genotype (gPMs) data and CYP2D6 PM phenotype from probe drugs (mPMs) in different ethnic groups **(A)** and geographic regions **(B)**.
PM: Poor metabolizers.

8. CYP2D6 and human evolutionary history

The global distribution of *CYP2D6* alleles has been shaped by the complex demographic history of the human species [30]. The Great Human Expansion model [30] describes the peopling of the world as a series of founder effects, accompanied by some major bottlenecks such as those of the Out-of-Africa (40 – 60 thousand years ago) and the Peopling of the Americas. Of course, subsequent demographic events (bottlenecks and migration) have also contributed to shaping the allele distribution of *CYP2D6*. In general, the worldwide pattern of genetic diversity of *CYP2D6* is similar to that observed for most of the human genome [16], where most diversity is observed within the various human populations. This statement does not preclude that some alleles do have a peculiar and interesting geographic distribution that is potentially relevant pharmacogenetically and clinically.

Although most *CYP2D6* alleles are shared by most of the world's populations, it is possible to recognize geographical gradients in the distributions of some alleles [16] and to identify alleles that show high frequencies in specific geographic regions. Some combinations of evolutionary factors may have contributed to determining the high frequencies of these alleles in certain regions of the world. For instance, the Out-of-Africa

bottleneck has determined large differences in the frequencies of alleles of several genes between Africans and non-Africans. Also, differentiated patterns of some alleles in some genes may be the product of the first colonization of different geographic regions by modern humans. In a process of geographic population expansion, specifically in the front wave of the expansion, some rare alleles/haplotypes (termed 'surfing' alleles) may by chance become particularly common [31].

CYP2D6 is an enzyme that, throughout our human evolutionary history, has detoxified different compounds in the diet of human populations living in different environments. Given this role, natural selection may have played a part in determining the frequencies of certain alleles in different populations as a way to improve the enzyme's detoxifying properties. In this context, it is intriguing that alleles which present particularly high frequencies in specific regions, such as *CYP2D6*17* in Africa, *CYP2D6*41* in the Middle East, *CYP2D6*4* in Europe and *CYP2D6*10* in East Asia, have a reduced or null enzymatic activity. There is evidence that changes in diet related to the transition from a hunter-gatherer economy to an agriculture-based economy have led to an increase of alleles with slower rates of metabolism [32]. Moreover, based on the capability of *CYP2D6* to detoxify alkaloids and to increase the availability of potential foods, it has been proposed that multiplications of *CYP2D6* active alleles were a product of natural selection in north-east Africa [1]. However, additional signatures of the action of natural selection on the pattern of genetic diversity of *CYP2D6* have yet to be found, and finding them would probably require a more thorough sampling of populations, and re-sequencing or long-range haplotype data that includes the flanking regions of *CYP2D6*.

9. Conclusion

In conclusion, relative to the distribution of the world's population, CYP2D6 research on healthy volunteers has an overrepresentation of subjects from Europe and America. There is a variability of *CYP2D6* allelic frequencies across geographic and ethnic groups: *CYP2D6*10* is more frequent in Asia and in East Asians, *CYP2D6*17* and **29* in Africa and in Black populations, and *CYP2D6*4* in Europe and in Caucasians. In Middle Eastern populations, *CYP2D6*41* and duplications/multiplications of the *CYP2D6* active gene have the highest frequencies. The high frequency of geno- and phenotypically PMs in Europe and in Caucasian populations is confirmed. The highest frequency of gUMs is found in Middle Eastern populations and Ethiopians. The correlation analysis showed that the gPM frequencies corresponded to the mPM frequencies. However, further studies concerning the mUM phenotype in healthy volunteers are warranted. Finally, in the light of these data, natural selection may have played a role in determining the high frequencies of some *CYP2D6* alleles in different populations and geographic regions.

10. Expert opinion

This study has confirmed the high variability of CYP2D6 by updating and summarizing the frequencies of alleles, of phenotypes predicted from genotypes and of measured metabolic phenotypes across different geographic regions and ethnic groups.

An uneven distribution of CYP2D6 studies around the world has become clear in this review. Europe and America have been extensively studied, especially their Caucasian populations. We found only a few studies for the African continent, however. Further studies with a greater number of individuals representing the vast genetic diversity of Africans are needed to clarify the CYP2D6 peculiarities of the different ethnic groups contained in this continent, especially among 'Black Africans' [33].

The mPMs and gPMs are more frequent in Caucasians, which agrees with the higher *CYP2D6*4* frequency in those populations. The highest frequency of gUMs was found in Middle Eastern populations, and, interestingly, this frequency is also high in Ethiopia, whereas it is lower in other African regions [25,34,35]. Ethiopia is one of the African countries nearest to the Middle East, and a constant gene flow between these regions has been extensively documented in population genetics studies [36], so that the high frequency of gUMs may be a characteristic not only of the Middle East, but also of the surrounding region as well.

Concerning CYP2D6 mUMs, only two studies in healthy volunteers evaluated their frequencies in Latino American populations [20,29]. One of these studies found that only 40% of the mUMs were predicted from gUMs [20]. A possible reason could be the absence of a defined antimode that clearly separates mUMs from the rest of the mEMs. Thus, a threshold or cut-off point should be soundly established and standardized to characterize the mUM phenotype with different probe drugs, such as for MRdebrisoquin ≤ 0.1 [20,37] or MRdextromethorphan ≤ 0.0003 [29]. Also, genetic analyses should be improved by taking into account factors other than gene duplication/multiplication. For example, recently we found that the presence of the -1584G allele in the promoter region of the *CYP2D6* gene was related to a high CYP2D6 hydroxylation capacity [38].

The observed variability, in particular the frequencies of gPMs and mPMs and of gUMs and mUMs, can have clinical implications when patients are treated with drugs metabolized by CYP2D6, such as aripiprazole, codeine, haloperidol, metoprolol, nortriptyline or thioridazine. Indeed, the FDA and European Medicines Agency (EMA) have included pharmacogenetic information in the labeling of some drugs, recommending a lower dosage in PM patients being treated with them. They have also restricted the use of codeine in UM children. Therefore, pharmacogenetic programs could be developed especially for ethnic groups or regions in which there is an increased frequency of PMs or UMs, as is the case for the Middle East or Ethiopia where the frequency of gUMs is the highest.

A more systematic and rigorous interpretation of the diversity of CYP2D6 in the context of human evolutionary history is still pending. This would allow clarification of the roles of bottlenecks, human migrations, admixture and natural selection in the current distribution of *CYP2D6* alleles. Another task pending is to integrate the information regarding the distribution and phenotypic effect of CYP2D6 diversity with information derived from current genome sequencing initiatives, such as the 1000 Genomes Project [39], the NHLBI GO Exome Sequencing Project [40], and such functional genomics initiatives as the ENCODE Project [41]. Caution is required to avoid confounding *CYP2D6* regions with highly similar sequences such as those of the *CYP2D7* and *CYP2D8* genes.

In summary, studies of *CYP2D6* genetic polymorphism are under-represented in African and Asian populations. The highest PM frequency is found in Caucasians, whereas gUMs are more frequent in Middle Easterns and Ethiopians. This information could be useful for Drug Regulatory Agencies to apply pharmacogenetic warnings especially in the guidelines for these regions. Moreover, additional studies of the UM phenotype are needed.

Declaration of interest

This study was partially supported by the European Union FEDER/FSE and the Gobierno de Extremadura Consejería de Empleo, Empresa e Innovación (PD10199), and AEXCID Cooperación Extremeña (13IA001) coordinated in the Iberoamerican Network of Pharmacogenetics (SIFF) and Brazilian Government CAPES grants (99999.012350/2013-03 and 110/2012). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Bibliography

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Ingelman-Sundberg M. Genetic polymorphisms of cytochrome. P450 2D6 (CYP2D6): Clinical consequences, evolutionary aspects and functional diversity. Pharmacogenomics J 2005;5:6-13
2. The human cytocrome P450 (CYP) allele nomenclature database [Internet]. 2014. Available from: www.cypalleles.ki.se/ [Cited 15 April 2014]
3. Gaedigk A. Complexities of CYP2D6 gene analysis and interpretation. Int Rev Psychiatry 2013;25:534-53
4. Llerena A, Dorado P, Penas-Lledo EM. Pharmacogenetics of debrisoquine and its use as a marker for CYP2D6 hydroxylation capacity. Pharmacogenomics 2009;10:17-28
5. LLerena A, Berecz R, de la Rubia A, Dorado P. QTc interval lengthening and debrisoquine metabolic ratio in psychiatric patients treated with oral haloperidol monotherapy. Eur J Clin Pharmacol 2002;58:223-4
6. Penas-Lledo EM, Trejo HD, Dorado P, et al. CYP2D6 ultrarapid metabolism and early dropout from fluoxetine or amitriptyline monotherapy treatment in major depressive patients. Mol Psychiatry 2013;18:8-9
7. Berecz R, de la Rubia A, Dorado P, et al. Thioridazine steady-state plasma concentrations are influenced by tobacco smoking and CYP2D6, but not by the CYP2C9 genotype. Eur J Clin Pharmacol 2003;59:45-50
8. LLerena A, de la Rubia A, Berecz R, Dorado P. Relationship between haloperidol plasma concentration, debrisoquine metabolic ratio, CYP2D6 and CYP2C9 genotypes in psychiatric patients. Pharmacopsychiatry 2004;37:69-73
9. Llerena A, Berecz R, Dorado P, de la Rubia A. QTc interval, CYP2D6 and CYP2C9 genotypes and risperidone plasma concentrations. J Psychopharmacol 2004;18:189-93
10. Zackrisson AL, Lindblom B, Ahlner J. High frequency of occurrence of CYP2D6 gene duplication/multiduplication indicating ultrarapid metabolism among suicide cases. Clin Pharmacol Ther 2010;88:354-9
11. Penas-Lledo EM, Dorado P, Aguera Z, et al. High risk of lifetime history of suicide attempts among CYP2D6 ultrarapid metabolizers with eating disorders. Mol Psychiatry 2011;16:691-2
12. Penas-Lledo EM, Blasco-Fontecilla H, Dorado P, et al. CYP2D6 and the severity of suicide attempts. Pharmacogenomics 2012;13:179-84
13. Ciszkowski C, Madadi P, Phillips MS, et al. Codeine, ultrarapid-metabolism genotype, and postoperative death. N Engl J Med 2009;361:827-8
14. Kelly LE, Rieder M, van den Anker J, et al. More codeine fatalities after tonsillectomy in North American children. Pediatrics 2012;129:e1343-7
15. Llerena A, Cobaleda J, Martinez C, Benitez J. Interethnic differences in drug metabolism: influence of genetic and environmental factors on debrisoquine hydroxylation phenotype. Eur J Drug Metab Pharmacokinet 1996;21:129-38
16. Sistonen J, Sajantila A, Lao O, et al. CYP2D6 worldwide genetic variation shows high frequency of altered activity variants and no continental structure. Pharmacogenet Genomics 2007;17:93-101
17. Teh LK, Bertilsson L. Pharmacogenomics of CYP2D6: molecular genetics, interethnic differences and clinical importance. Drug Metab Pharmacokinet 2012;27:55-67
18. Crews KR, Gaedigk A, Dunnenberger HM, et al. Clinical pharmacogenetics implementation consortium guidelines for cytochrome P450 2D6 genotype and codeine therapy: 2014 update. Clin Pharmacol Ther 2014;95:376-82
• **Review about the clinical implication of *CYP2D6* genotype in the codeine therapy. Furthermore, a description of *CYP2D6* alleles in different ethnic groups was performed.**
19. Gaedigk A, Simon SD, Pearce RE, et al. The CYP2D6 activity score: translating genotype information into a qualitative

measure of phenotype.
Clin Pharmacol Ther 2008;83:234-42

- **In this study, the *CYP2D6* activity score, that allow to extrapolate phenotype from genotype, was described.**

20. Llerena A, Dorado P, Ramirez R, et al. CYP2D6 genotype and debrisoquine hydroxylation phenotype in Cubans and Nicaraguans. Pharmacogenomics J 2012;12:176-83

- **In this study the relationship between CYP2D6 predicted phenotype and metabolic phenotypes is evaluated. Moreover, it is one of the two studies that estimates CYP2D6 metabolic UMs.**

21. Hicks JK, Swen JJ, Gaedigk A. Challenges in CYP2D6 phenotype assignment from genotype data: a critical assessment and call for standardization. Curr Drug Metab 2014;15:218-32
22. Frank D, Jaehde U, Fuhr U. Evaluation of probe drugs and pharmacokinetic metrics for CYP2D6 phenotyping. Eur J Clin Pharmacol 2007;63:321-33
23. GraphPad software [Internet]. 2014. Available from: www.graphpad.com [Cited April 2014]
24. Kohlrausch FB, Gama CS, Lobato MI, et al. Molecular diversity at the CYP2D6 locus in healthy and schizophrenic southern Brazilians. Pharmacogenomics 2009;10:1457-66
25. Wennerholm A, Johansson I, Massele AY, et al. Decreased capacity for debrisoquine metabolism among black Tanzanians: analyses of the CYP2D6 genotype and phenotype. Pharmacogenetics 1999;9:707-14
26. World population statistics [Internet]. 2014. Available from: www.worldpopulationstatistics.com [Cited 1 May 2014]
27. Penas-Lledo EM, Dorado P, Aguera Z, et al. CYP2D6 polymorphism in patients with eating disorders. Pharmacogenomics J 2012;12:173-5
28. Aklillu E, Persson I, Bertilsson L, et al. Frequent distribution of ultrarapid metabolizers of debrisoquine in an Ethiopian population carrying duplicated and multiduplicated functional CYP2D6 alleles. J Pharmacol Exp Ther 1996;278:441-6
29. Montane Jaime LK, Lalla A, Steimer W, Gaedigk A. Characterization of the CYP2D6 gene locus and metabolic activity in Indo- and Afro-Trinidadians: discovery of novel allelic variants. Pharmacogenomics 2013;14:261-76

- **This article is one of the two studies that estimate CYP2D6 metabolic UMs.**

30. Henn BM, Cavalli-Sforza LL, Feldman MW. The great human expansion. Proc Natl Acad Sci USA 2012;109:17758-64
31. Excoffier L, Ray N. Surfing during population expansions promotes genetic revolutions and structuration. Trends Ecol Evol 2008;23:347-51
32. Fuselli S, de Filippo C, Mona S, et al. Evolution of detoxifying systems: the role of environment and population history in shaping genetic diversity at human CYP2D6 locus. Pharmacogenet Genomics 2010;20:485-99
33. Campbell MC, Tishkoff SA. The evolution of human genetic and phenotypic variation in Africa. Curr Biol 2010;20:R166-73
34. Griese EU, Asante-Poku S, Ofori-Adjei D, et al. Analysis of the CYP2D6 gene mutations and their consequences for enzyme function in a West African population. Pharmacogenetics 1999;9:715-23
35. Panserat S, Sica L, Gerard N, et al. CYP2D6 polymorphism in a Gabonese population: contribution of the CYP2D6*2 and CYP2D6*17 alleles to the high prevalence of the intermediate metabolic phenotype. Br J Clin Pharmacol 1999;47:121-4
36. Pagani L, Kivisild T, Tarekegn A, et al. Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. Am J Hum Genet 2012;91:83-96
37. Dahl ML, Johansson I, Bertilsson L, et al. Ultrarapid hydroxylation of debrisoquine in a Swedish population. Analysis of the molecular genetic basis. J Pharmacol Exp Ther 1995;274:516-20
38. Llerena A, Dorado P, Ramirez R, et al. CYP2D6 -1584C>G promoter polymorphism and debrisoquine ultrarapid hydroxylation in healthy volunteers. Pharmacogenomics 2013;14:1973-7
39. 1000 genomes [Internet]. 2014. Available from: www.1000genomes.org/ [Cited 26 June 2014]
40. NHLBI grand opportunity exome sequencing project (ESP) [Internet]. 2014. Available from: https://esp.gs.washington.edu/drupal/ [Cited 26 June 2014]
41. National human genome research institute [Internet]. 2014. Available from: www.genome.gov/encode/ [Cited 26 June 2014]
42. Masimirembwa C, Hasler J, Bertilssons L, et al. Phenotype and genotype analysis of debrisoquine hydroxylase (CYP2D6) in a black Zimbabwean population. Reduced enzyme activity and evaluation of metabolic correlation of CYP2D6 probe drugs. Eur J Clin Pharmacol 1996;51:117-22
43. Dodgen TM, Hochfeld WE, Fickl H, et al. Introduction of the AmpliChip CYP450 test to a South African cohort: a platform comparative prospective cohort study. BMC Med Genet 2013;14:20
44. Bathum L, Skjelbo E, Mutabingwa TK, et al. Phenotypes and genotypes for CYP2D6 and CYP2C19 in a black Tanzanian population. Br J Clin Pharmacol 1999;48:395-401
45. Evans WE, Relling MV, Rahman A, et al. Genetic basis for a lower prevalence of deficient CYP2D6 oxidative drug metabolism phenotypes in black Americans. J Clin Invest 1993;91:2150-4
46. Wan YJ, Poland RE, Han G, et al. Analysis of the CYP2D6 gene polymorphism and enzyme activity in African-Americans in southern California. Pharmacogenetics 2001;11:489-99
47. Gaedigk A, Bradford LD, Marcucci KA, Leeder JS. Unique CYP2D6 activity distribution and genotype-phenotype discordance in black Americans. Clin Pharmacol Ther 2002;72:76-89
48. Maciel ME, Oliveira FK, Propst GB, et al. Population analysis of xenobiotic metabolizing genes in south Brazilian Euro- and Afro-descendants. Genet Mol Biol 2009;32:723-8
49. Kouhi H, Hamzeiy H, Barar J, et al. Frequency of five important CYP2D6 alleles within an Iranian population (eastern Azerbaijan).

Genet Test Mol Biomarkers 2009;13:665-70

50. Hashemi-Soteh SM, Sarzare F, Merat F, et al. Frequencies of three CYP2D6 nonfunctional alleles (CYP2D6*3, *4, and *6) within an Iranian population (Mazandaran). Genet Test Mol Biomarkers 2011;15:821-5
51. Luo HR, Aloumanis V, Lin KM, et al. Polymorphisms of CYP2C19 and CYP2D6 in Israeli ethnic groups. Am J Pharmacogenomics 2004;4:395-401
52. Zihlif M, Imraish A, Irshaid YM. Frequency of certain single-nucleotide polymorphisms and duplication of CYP2D6 in the Jordanian population. Genet Test Mol Biomarkers 2012;16:1201-5
53. Fuselli S, Dupanloup I, Frigato E, et al. Molecular diversity at the CYP2D6 locus in the Mediterranean region. Eur J Hum Genet 2004;12:916-24
54. Aynacioglu AS, Sachse C, Bozkurt A, et al. Low frequency of defective alleles of cytochrome P450 enzymes 2C19 and 2D6 in the Turkish population. Clin Pharmacol Ther 1999;66:185-92
55. Aydin M, Hatirnaz O, Erensoy N, Ozbek U. CYP2D6 and CYP1A1 mutations in the Turkish population. Cell Biochem Funct 2005;23:133-5
56. Serin A, Canan H, Alper B, Gulmen M. The frequencies of mutated alleles of CYP2D6 gene in a Turkish population. Forensic Sci Int 2012;222:332-4
57. Scott SA, Edelmann L, Kornreich R, et al. CYP2C9, CYP2C19 and CYP2D6 allele frequencies in the Ashkenazi Jewish population. Pharmacogenomics 2007;8:721-30
58. Ismail R, Teh LK. Genetic polymorphism of CYP2D6: malaysian Indians have the highest frequency for CYP2D6*4 in Asia. Eur J Clin Pharmacol 2001;57:617-18
59. Naveen AT, Adithan C, Soya SS, et al. CYP2D6 genetic polymorphism in South Indian populations. Biol Pharm Bull 2006;29:1655-8
60. Adithan C, Gerard N, Naveen AT, et al. Genotype and allele frequency of CYP2D6 in Tamilian population. Eur J Clin Pharmacol 2003;59:517-20
61. Garcia-Barcelo M, Chow LY, Lam KL, et al. Occurrence of CYP2D6 gene duplication in Hong Kong Chinese. Clin Chem 2000;46:1411-13
62. Garcia-Barcelo M, Chow LY, Chiu HF, et al. Genetic analysis of the CYP2D6 locus in a Hong Kong Chinese population. Clin Chem 2000;46:18-23
63. Ji L, Pan S, Marti-Jaun J, et al. Single-step assays to analyze CYP2D6 gene polymorphisms in Asians: allele frequencies and a novel *14B allele in mainland Chinese. Clin Chem 2002;48:983-8
64. Sheng HH, Zeng AP, Zhu WX, et al. Allelic distributions of CYP2D6 gene copy number variation in the eastern Han Chinese population. Acta Pharmacol Sin 2007;28:279-86
65. Qin S, Shen L, Zhang A, et al. Systematic polymorphism analysis of the CYP2D6 gene in four different geographical Han populations in mainland China. Genomics 2008;92:152-8
66. Zhou Q, Yu XM, Lin HB, et al. Genetic polymorphism, linkage disequilibrium, haplotype structure and novel allele analysis of CYP2C19 and CYP2D6 in Han Chinese. Pharmacogenomics J 2009;9:380-94
67. Shi Y, Xiang P, Li L, Shen M. Analysis of 50 SNPs in CYP2D6, CYP2C19, CYP2C9, CYP3A4 and CYP1A2 by MALDI-TOF mass spectrometry in Chinese Han population. Forensic Sci Int 2011;207:183-7
68. Yin SJ, Ni YB, Wang SM, et al. Differences in genotype and allele frequency distributions of polymorphic drug metabolizing enzymes CYP2C19 and CYP2D6 in mainland Chinese Mongolian, Hui and Han populations. J Clin Pharm Ther 2012;37:364-9
69. Zuo LJ, Guo T, Xia DY, Jia LH. Allele and genotype frequencies of CYP3A4, CYP2C19, and CYP2D6 in Han, Uighur, Hui, and Mongolian Chinese populations. Genet Test Mol Biomarkers 2012;16:102-8
70. Qian JC, Xu XM, Hu GX, et al. Genetic variations of human CYP2D6 in the Chinese Han population. Pharmacogenomics 2013;14:1731-43
71. Kim EY, Lee SS, Jung HJ, et al. Robust CYP2D6 genotype assay including copy number variation using multiplex single-base extension for Asian populations. Clin Chim Acta 2010;411:2043-8
72. Tateishi T, Chida M, Ariyoshi N, et al. Analysis of the CYP2D6 gene in relation to dextromethorphan O-demethylation capacity in a Japanese population. Clin Pharmacol Ther 1999;65:570-5
73. Nishida Y, Fukuda T, Yamamoto I, Azuma J. CYP2D6 genotypes in a Japanese population: low frequencies of CYP2D6 gene duplication but high frequency of CYP2D6*10. Pharmacogenetics 2000;10:567-70
74. Ishiguro A, Kubota T, Sasaki H, et al. Common mutant alleles of CYP2D6 causing the defect of CYP2D6 enzyme activity in a Japanese population. Br J Clin Pharmacol 2003;55:414-15
75. Ryu SW, Kim YJ, Kim E. Mutation analysis of CYP2D6 locus in the Korean population: identification of rare poor metabolizer alleles at the nucleotide level. Mol Cells 1998;8:758-63
76. Lee SY, Sohn KM, Ryu JY, et al. Sequence-based CYP2D6 genotyping in the Korean population. Ther Drug Monit 2006;28:382-7
77. Lee SJ, Lee SS, Jung HJ, et al. Discovery of novel functional variants and extensive evaluation of CYP2D6 genetic polymorphisms in Koreans. Drug Metab Dispos 2009;37:1464-70
78. Teh LK, Ismail R, Yusoff R, et al. Heterogeneity of the CYP2D6 gene among Malays in Malaysia. J Clin Pharm Ther 2001;26:205-11
79. Ismail R, Teh LK, Amir J, et al. Genetic polymorphism of CYP2D6 in Chinese subjects in Malaysia. J Clin Pharm Ther 2003;28:279-84
80. Veiga MI, Asimus S, Ferreira PE, et al. Pharmacogenomics of CYP2A6, CYP2B6, CYP2C19, CYP2D6, CYP3A4, CYP3A5 and MDR1 in Vietnam. Eur J Clin Pharmacol 2009;65:355-63
81. Griese EU, Ilett KF, Kitteringham NR, et al. Allele and genotype frequencies of polymorphic cytochromes P4502D6, 2C19 and 2E1 in Aborigines from Western Australia. Pharmacogenetics 2001;11:69-76
82. Lea RA, Roberts RL, Green MR, et al. Allele frequency differences of cytochrome P450 polymorphisms in a

sample of New Zealand Maori. N Z Med J 2008;121:33-7

83. Jurima-Romet M, Foster BC, Casley WL, et al. CYP2D6-related oxidation polymorphism in a Canadian Inuit population. Can J Physiol Pharmacol 1997;75:165-72

84. Munoz S, Vollrath V, Vallejos MP, et al. Genetic polymorphisms of CYP2D6, CYP1A1 and CYP2E1 in the South-Amerindian population of Chile. Pharmacogenetics 1998;8:343-51

85. Jorge LF, Eichelbaum M, Griese EU, et al. Comparative evolutionary pharmacogenetics of CYP2D6 in Ngawbe and Embera Amerindians of Panama and Colombia: role of selection versus drift in world populations. Pharmacogenetics 1999;9:217-28

86. Sosa-Macias M, Elizondo G, Flores-Perez C, et al. CYP2D6 genotype and phenotype in Amerindians of Tepehuano origin and Mestizos of Durango. Mexico. J Clin Pharmacol 2006;46:527-36

87. Sosa-Macias M, Dorado P, Alanis-Banuelos RE, et al. Influence of CYP2D6 deletion, multiplication, -1584C->G, 31G->A and 2988G–>a gene polymorphisms on dextromethorphan metabolism among Mexican Tepehuanos and Mestizos. Pharmacology 2010;86:30-6

88. Lazalde-Ramos BP, Martinez-Fierro Mde L, Galaviz-Hernandez C, et al. CYP2D6 gene polymorphisms and predicted phenotypes in eight indigenous groups from northwestern Mexico. Pharmacogenomics 2014;15:339-48

89. Fohner A, Muzquiz LI, Austin MA, et al. Pharmacogenetics in American Indian populations: analysis of CYP2D6, CYP3A4, CYP3A5, and CYP2C9 in the Confederated Salish and Kootenai tribes. Pharmacogenet Genomics 2013;23:403-14

90. Bailliet G, Santos MR, Alfaro EL, et al. Allele and genotype frequencies of metabolic genes in Native Americans from Argentina and Paraguay. Mutat Res 2007;627:171-7

91. Céspedes-Garro C, Jiménez-Arce G, Naranjo MEG, et al. Ethnic background and CYP2D6 genetic polymorphisms in Costa Ricans. Int J Trop Biol Conserv 2014;62:1659-71

92. Griman P, Moran Y, Valero G, et al. CYP2D6 gene variants in urban/admixed and Amerindian populations of Venezuela: pharmacogenetics and anthropological implications. Ann Hum Biol 2012;39:137-42

93. Salazar-Flores J, Torres-Reyes LA, Martinez-Cortes G, et al. Distribution of CYP2D6 and CYP2C19 polymorphisms associated with poor metabolizer phenotype in five Amerindian groups and western Mestizos from Mexico. Genet Test Mol Biomarkers 2012;16:1098-104

94. Isaza CA, Henao J, Lopez AM, Cacabelos R. Isolation, sequence and genotyping of the drug metabolizer CYP2D6 gene in the Colombian population. Methods Find Exp Clin Pharmacol 2000;22:695-705

95. Mendoza R, Wan YJ, Poland RE, et al. CYP2D6 polymorphism in a Mexican American population. Clin Pharmacol Ther 2001;70:552-60

96. Lopez M, Guerrero J, Jung-Cook H, Alonso ME. CYP2D6 genotype and phenotype determination in a Mexican Mestizo population. Eur J Clin Pharmacol 2005;61:749-54

97. Luo HR, Gaedigk A, Aloumanis V, Wan YJ. Identification of CYP2D6 impaired functional alleles in Mexican Americans. Eur J Clin Pharmacol 2005;61:797-802

98. Casner PR. The effect of CYP2D6 polymorphisms on dextromethorphan metabolism in Mexican Americans. J Clin Pharmacol 2005;45:1230-5

99. Contreras AV, Monge-Cazares T, Alfaro-Ruiz L, et al. Resequencing, haplotype construction and identification of novel variants of CYP2D6 in Mexican Mestizos. Pharmacogenomics 2011;12:745-56

100. Orengo-Mercado C, Nieves B, Lopez L, et al. Frequencies of functional polymorphisms in three pharmacokinetic genes of clinical interest within the admixed Puerto Rican population. J Pharmacogenomics Pharmacoproteomics 2013;4:1000113

101. Silveira Vda S, Canalle R, Scrideli CA, et al. Polymorphisms in genes encoding drugs and xenobiotic metabolizing enzymes in a Brazilian population. Biomarkers 2009;14:111-17

102. Roco A, Quinones L, Agundez JA, et al. Frequencies of 23 functionally significant variant alleles related with metabolism of antineoplastic drugs in the Chilean population: comparison with Caucasian and Asian populations. Front Genet 2012;3:229

103. Dorado P, Heras N, Machin E, et al. CYP2D6 genotype and dextromethorphan hydroxylation phenotype in an Ecuadorian population. Eur J Clin Pharmacol 2012;68:637-44

104. Grimán P, Morán Y, Camargo M, Chiurillo MA. Caracterización de variantes alélicas del citocromo CYP2D6 en la población de la región centrooccidental de Venezuela. Acta Biológica Colombiana 2009;14:195-202

105. Gaedigk A, Coetsee C. The CYP2D6 gene locus in South African coloureds: unique allele distributions, novel alleles and gene arrangements. Eur J Clin Pharmacol 2008;64:465-75

106. Chou WH, Yan FX, Robbins-Weilert DK, et al. Comparison of two CYP2D6 genotyping methods and assessment of genotype-phenotype relationships. Clin Chem 2003;49:542-51

107. Zackrisson AL, Lindblom B. Identification of CYP2D6 alleles by single nucleotide polymorphism analysis using pyrosequencing. Eur J Clin Pharmacol 2003;59:521-6

108. Yamada H, Dahl ML, Lannfelt L, et al. CYP2D6 and CYP2C19 genotypes in an elderly Swedish population. Eur J Clin Pharmacol 1998;54:479-81

109. Halling J, Petersen MS, Damkier P, et al. Polymorphism of CYP2D6, CYP2C19, CYP2C9 and CYP2C8 in the Faroese population. Eur J Clin Pharmacol 2005;61:491-7

110. Griese EU, Zanger UM, Brudermanns U, et al. Assessment of the predictive power of genotypes for the in-vivo catalytic function of CYP2D6 in a German population. Pharmacogenetics 1998;8:15-26

111. Bock KW, Schrenk D, Forster A, et al. The influence of environmental and genetic factors on CYP2D6, CYP1A2 and UDP-glucuronosyltransferases in man using sparteine, caffeine, and paracetamol as probes. Pharmacogenetics 1994;4:209-18

112. Stamer UM, Bayerer B, Wolf S, et al. Rapid and reliable method for cytochrome P450 2D6 genotyping. Clin Chem 2002;48:1412-17

113. Beer B, Erb R, Pitterl F, et al. CYP2D6 genotyping by liquid chromatography-electrospray ionization mass spectrometry. Anal Bioanal Chem 2011;400:2361-70

114. Tamminga WJ, Wemer J, Oosterhuis B, et al. The prevalence of CYP2D6 and CYP2C19 genotypes in a population of healthy Dutch volunteers. Eur J Clin Pharmacol 2001;57:717-22

115. Rideg O, Haber A, Botz L, et al. Pilot study for the characterization of pharmacogenetically relevant CYP2D6, CYP2C19 and ABCB1 gene polymorphisms in the Hungarian population. Cell Biochem Funct 2011;29:562-8

116. Rebsamen MC, Desmeules J, Daali Y, et al. The AmpliChip CYP450 test: cytochrome P450 2D6 genotype assessment and phenotype prediction. Pharmacogenomics J 2009;9:34-41

117. Crescenti A, Mas S, Gasso P, et al. Simultaneous genotyping of CYP2D6*3, *4, *5 and *6 polymorphisms in a Spanish population through multiplex long polymerase chain reaction and minisequencing multiplex single base extension analysis. Clin Exp Pharmacol Physiol 2007;34:992-7

118. Menoyo A, del Rio E, Baiget M. Characterization of variant alleles of cytochrome CYP2D6 in a Spanish population. Cell Biochem Funct 2006;24:381-5

119. Correia C, Santos P, Coutinho AM, Vicente AM. Characterization of pharmacogenetically relevant CYP2D6 and ABCB1 gene polymorphisms in a Portuguese population sample. Cell Biochem Funct 2009;27:251-5

120. Albuquerque J, Ribeiro C, Naranjo MEG, et al. Characterization of CYP2D6 genotypes and metabolic profiles in the Portuguese population: pharmacogenetic implications. Personalized Medicine 2013;10:709-18

121. Scordo MG, Caputi AP, D'Arrigo C, et al. Allele and genotype frequencies of CYP2C9, CYP2C19 and CYP2D6 in an Italian population. Pharmacol Res 2004;50:195-200

122. Arvanitidis K, Ragia G, Iordanidou M, et al. Genetic polymorphisms of drug-metabolizing enzymes CYP2D6, CYP2C9, CYP2C19 and CYP3A5 in the Greek population. Fundam Clin Pharmacol 2007;21:419-26

123. Riccardi LN, Lanzellotto R, Luiselli D, et al. CYP2D6 genotyping in natives and immigrants from the Emilia-Romagna region (Italy). Genet Test Mol Biomarkers 2011;15:801-6

124. Gawronska-Szklarz B, Wojcicki M, Kuprianowicz A, et al. CYP2D6 and GSTM1 genotypes in a Polish population. Eur J Clin Pharmacol 1999;55:389-92

125. Niewinski P, Orzechowska-Juzwenko K, Hurkacz M, et al. CYP2D6 extensive, intermediate, and poor phenotypes and genotypes in a Polish population. Eur J Clin Pharmacol 2002;58:533-5

126. Bozina N, Granic P, Lalic Z, et al. Genetic polymorphisms of cytochromes P450: CYP2C9, CYP2C19, and CYP2D6 in Croatian population. Croat Med J 2003;44:425-8

127. Buzkova H, Pechandova K, Slanar O, Perlik F. Frequency of single nucleotide polymorphisms of CYP2D6 in the Czech population. Cell Biochem Funct 2008;26:76-81

128. Marandi T, Dahl ML, Kiivet RA, et al. Debrisoquin and S-mephenytoin hydroxylation phenotypes and CYP2D6 genotypes in an Estonian population. Pharmacol Toxicol 1996;78:303-7

129. Gaikovitch EA, Cascorbi I, Mrozikiewicz PM, et al. Polymorphisms of drug-metabolizing enzymes CYP2C9, CYP2C19, CYP2D6, CYP1A1, NAT2 and of P-glycoprotein in a Russian population. Eur J Clin Pharmacol 2003;59:303-12

130. Marandi T, Dahl ML, Rago L, et al. Debrisoquine and S-mephenytoin hydroxylation polymorphisms in a Russian population living in Estonia. Eur J Clin Pharmacol 1997;53:257-60

131. Attitallah S, Berard M, Belkahia C, et al. Similarities and/or dissimilarities of CYP2D6 polymorphism in three Tunisian ethnic groups: arabs, Berbers, Numides. Therapie 2000;55:355-60

132. Mbanefo C, Bababunmi EA, Mahgoub A, et al. A study of the debrisoquine hydroxylation polymorphism in a Nigerian population. Xenobiotica 1980;10:811-18

133. Iyun AO, Lennard MS, Tucker GT, Woods HF. Metoprolol and debrisoquin metabolism in Nigerians: lack of evidence for polymorphic oxidation. Clin Pharmacol Ther 1986;40:387-94

134. Sommers DK, Moncrieff J, Avenant JC. Absence of polymorphism of sparteine oxidation in the South African Venda. Hum Exp Toxicol 1991;10:175-8

135. Sommers DK, Moncrieff J, Avenant J. Non-correlation between debrisoquine and metoprolol polymorphisms in the Venda. Hum Toxicol 1989;8:365-8

136. Sommers DK, Moncrieff J, Avenant J. Metoprolol alpha-hydroxylation polymorphism in the San Bushmen of Southern Africa. Hum Toxicol 1989;8:39-43

137. Sommers DK, Moncrieff J, Avenant J. Polymorphism of the 4-hydroxylation of debrisoquine in the San Bushmen of Southern Africa. Hum Toxicol 1988;7:273-6

138. Woolhouse NM, Andoh B, Mahgoub A, et al. Debrisoquin hydroxylation polymorphism among Ghanaians and Caucasians. Clin Pharmacol Ther 1979;26:584-91

139. Woolhouse NM, Eichelbaum M, Oates NS, et al. Dissociation of co-regulatory control of debrisoquin/phenformin and sparteine oxidation in Ghanaians. Clin Pharmacol Ther 1985;37:512-21

140. Nsabiyumva F, Furet Y, Autret E, et al. Oxidative polymorphism of dextromethorphan in a Burundi population. Eur J Clin Pharmacol 1991;41:75-7

141. Gaedigk A, Ndjountche L, Divakaran K, et al. Cytochrome P4502D6 (CYP2D6) gene locus heterogeneity: characterization of gene duplication events. Clin Pharmacol Ther 2007;81:242-51

142. He N, Daniel HI, Hajiloo L, Shockley D. Dextromethorphan O-demethylation polymorphism in an African-American population. Eur J Clin Pharmacol 1999;55:457-9

143. Marinac JS, Foxworth JW, Willsie SK. Dextromethorphan polymorphic hepatic oxidation (CYP2D6) in healthy black American adult subjects. Ther Drug Monit 1995;17:120-4

144. Relling MV, Cherrie J, Schell MJ, et al. Lower prevalence of the debrisoquin oxidative poor metabolizer phenotype in American black versus white subjects. Clin Pharmacol Ther 1991;50:308-13

145. Al-Dosari MS, Al-Jenoobi FI, Alkharfy KM, et al. High prevalence of CYP2D6*41 (G2988A) allele in Saudi Arabians. Environ Toxicol Pharmacol 2013;36:1063-7

146. McLellan RA, Oscarson M, Seidegard J, et al. Frequent occurrence of CYP2D6 gene duplication in Saudi Arabians. Pharmacogenetics 1997;7:187-91

147. Qumsieh RY, Ali BR, Abdulrazzaq YM, et al. Identification of new alleles and the determination of alleles and genotypes frequencies at the CYP2D6 gene in Emiratis. PLoS One 2011;6:e28943

148. Afshar M, Rouini M, Ala S. Dextromethorphan metabolic phenotyping in an Iranian population. Eur J Clin Pharmacol 2005;60:849-54

149. Evans DA, Krahn P, Narayanan N. The mephenytoin (cytochrome P450 2C 19) and dextromethorphan (cytochrome P450 2D6) polymorphisms in Saudi Arabians and Filipinos. Pharmacogenetics 1995;5:64-71

150. al-Hadidi HF, Irshaid YM, Rawashdeh NM. Metoprolol alpha-hydroxylation is a poor probe for debrizoquine oxidation (CYP2D6) polymorphism in Jordanians. Eur J Clin Pharmacol 1994;47:311-14

151. Irshaid YM, al-Hadidi HF, Rawashdeh NM. Dextromethorphan O-demethylation polymorphism in Jordanians. Eur J Clin Pharmacol 1993;45:271-3

152. Basci NE, Bozkurt A, Isimer A, Kayaalp SO. Correlation between the metabolic ratios of debrisoquine and metoprolol in Turkish subjects. Pharmacol Toxicol 1994;75:62-4

153. Islam SI, Idle JR, Smith RL. The polymorphic 4-hydroxylation of debrisoquine in a Saudi Arab population. Xenobiotica 1980;10:819-25

154. Weerasuriya K, Jayakody RL, Smith CA, et al. Debrisoquine and mephenytoin oxidation in Sinhalese: a population study. Br J Clin Pharmacol 1994;38:466-70

155. Mamidi RN, Satyavageeswaran S, Vakkalanka SV, et al. Polymorphism of dextromethorphan oxidation in South Indian subjects. Clin Pharmacol Ther 1999;66:193-200

156. Lamba V, Lamba JK, Dilawari JB, Kohli KK. Genetic polymorphism of CYP2D6 in North Indian subjects. Eur J Clin Pharmacol 1998;54:787-91

157. Abraham BK, Adithan C, Shashindran CH, et al. Genetic polymorphism of CYP2D6 in a Keralite (South India) population. Br J Clin Pharmacol 2000;49:285-6

158. Abraham BK, Adithan C, Kiran PU, et al. Genetic polymorphism of CYP2D6 in Karnataka and Andhra Pradesh population in India. Acta Pharmacol Sin 2000;21:494-8

159. Straka RJ, Hansen SR, Walker PF. Comparison of the prevalence of the poor metabolizer phenotype for CYP2D6 between 203 Hmong subjects and 280 white subjects residing in Minnesota. Clin Pharmacol Ther 1995;58:29-34

160. Ismail R, Hussein A, Teh LK, Nizam Isa M. CYP2D6 phenotypes among Malays in Malaysia. J Clin Pharm Ther 2000;25:379-83

161. Wanwimolruk S, Patamasucon P, Lee EJ. Evidence for the polymorphic oxidation of debrisoquine in the Thai population. Br J Clin Pharmacol 1990;29:244-7

162. Johansson I, Yue QY, Dahl ML, et al. Genetic analysis of the interethnic difference between Chinese and Caucasians in the polymorphic metabolism of debrisoquine and codeine. Eur J Clin Pharmacol 1991;40:553-6

163. Cai WM, Chen B, Liu YX, Chu X. Dextromethorphan metabolic phenotyping in a Chinese population. Zhongguo Yao Li Xue Bao 1997;18:441-4

164. Kubota T, Yamaura Y, Ohkawa N, et al. Frequencies of CYP2D6 mutant alleles in a normal Japanese population and metabolic activity of dextromethorphan O-demethylation in different CYP2D6 genotypes. Br J Clin Pharmacol 2000;50:31-4

165. Lane HY, Deng HC, Huang SM, et al. Low frequency of dextromethorphan O-demethylation deficiency in a Chinese population. Clin Pharmacol Ther 1996;60:696-8

166. Wang SL, Huang JD, Lai MD, et al. Molecular basis of genetic variation in debrisoquin hydroxylation in Chinese subjects: polymorphism in RFLP and DNA sequence of CYP2D6. Clin Pharmacol Ther 1993;53:410-18

167. Lee EJ, Jeyaseelan K. Frequency of human CYP2D6 mutant alleles in a normal Chinese population. Br J Clin Pharmacol 1994;37:605-7

168. Roh HK, Dahl ML, Johansson I, et al. Debrisoquine and S-mephenytoin hydroxylation phenotypes and genotypes in a Korean population. Pharmacogenetics 1996;6:441-7

169. Chida M, Yokoi T, Nemoto N, et al. A new variant CYP2D6 allele (CYP2D6*21) with a single base insertion in exon 5 in a Japanese population associated with a poor metabolizer phenotype. Pharmacogenetics 1999;9:287-93

170. Lou YC, Ying L, Bertilsson L, Sjoqvist F. Low frequency of slow debrisoquine hydroxylation in a native Chinese population. Lancet 1987;2:852-3

171. Nakamura K, Goto F, Ray WA, et al. Interethnic differences in genetic polymorphism of debrisoquin and mephenytoin hydroxylation between Japanese and Caucasian populations. Clin Pharmacol Ther 1985;38:402-8

172. Ishizaki T, Eichelbaum M, Horai Y, et al. Evidence for polymorphic oxidation of sparteine in Japanese subjects. Br J Clin Pharmacol 1987;23:482-5

173. Cai WM, Chen B, Zhang WX. Frequency of CYP2D6*10 and *14 alleles and their influence on the metabolic activity of CYP2D6 in a healthy Chinese population. Clin Pharmacol Ther 2007;81:95-8

174. Wanwimolruk S, Bhawan S, Coville PF, Chalcroft SC. Genetic polymorphism of debrisoquine (CYP2D6) and proguanil (CYP2C19) in South Pacific Polynesian populations. Eur J Clin Pharmacol 1998;54:431-5

175. Brosen K. Sparteine oxidation polymorphism in Greenlanders living in Denmark. Br J Clin Pharmacol 1986;22:415-19

176. Arvela P, Kirjarinta M, Kirjarinta M, et al. Polymorphism of debrisoquine hydroxylation among Finns and Lapps. Br J Clin Pharmacol 1988;26:601-3

177. Arias TD, Jorge LF, Inaba T. No evidence for the presence of poor metabolizers of sparteine in an

Amerindian group: the Cunas of Panama. Br J Clin Pharmacol 1986;21:547-8

178. Estevez F, Giusti M, Parrillo S, Oxandabarat J. Dextromethorphan O-demethylation polymorphism in the Uruguayan population. Eur J Clin Pharmacol 1997;52:417-18
179. Peart GF, Boutagy J, Shenfield GM. Debrisoquine oxidation in an Australian population. Br J Clin Pharmacol 1986;21:465-71
180. Turgeon J, Evans WE, Relling MV, et al. Phenotypic debrisoquine 4-hydroxylase activity among extensive metabolizers is unrelated to genotype as determined by the Xba-I restriction fragment length polymorphism. Br J Clin Pharmacol 1991;32:283-8
181. Madsen H, Nielsen KK, Brosen K. Imipramine metabolism in relation to the sparteine and mephenytoin oxidation polymorphisms–a population study. Br J Clin Pharmacol 1995;39:433-9
182. Pedersen RS, Damkier P, Brosen K. Tramadol as a new probe for cytochrome P450 2D6 phenotyping: a population study. Clin Pharmacol Ther 2005;77:458-67
183. Steiner E, Bertilsson L, Sawe J, et al. Polymorphic debrisoquin hydroxylation in 757 Swedish subjects. Clin Pharmacol Ther 1988;44:431-5
184. Drohse A, Bathum L, Brosen K, Gram LF. Mephenytoin and sparteine oxidation: genetic polymorphisms in Denmark. Br J Clin Pharmacol 1989;27:620-5
185. Sanz EJ, Villen T, Alm C, Bertilsson L. S-mephenytoin hydroxylation phenotypes in a Swedish population determined after coadministration with debrisoquin. Clin Pharmacol Ther 1989;45:495-9
186. Brosen K, Otton SV, Gram LF. Sparteine oxidation polymorphism in Denmark. Acta Pharmacol Toxicol (Copenh) 1985;57:357-60
187. Tamminga WJ, Wemer J, Oosterhuis B, et al. CYP2D6 and CYP2C19 activity in a large population of Dutch healthy volunteers: indications for oral contraceptive-related gender differences. Eur J Clin Pharmacol 1999;55:177-84
188. Paar WD, Schuhler H, Fimmers R, Dengler HJ. Sparteine oxidation polymorphism: phenotyping by measurement of sparteine and its dehydrometabolites in plasma. Eur J Clin Pharmacol 1989;36:555-60
189. Leclercq V, Desager JP, van Nieuwenhuyze Y, Harvengt C. Prevalence of drug hydroxylator phenotypes in Belgium. Eur J Clin Pharmacol 1987;33:439-40
190. Szorady I, Santa A. Drug hydroxylator phenotype in Hungary. Eur J Clin Pharmacol 1987;32:325
191. Evans DA, Mahgoub A, Sloan TP, et al. A family and population study of the genetic polymorphism of debrisoquine oxidation in a white British population. J Med Genet 1980;17:102-5
192. Schmid B, Bircher J, Preisig R, Kupfer A. Polymorphic dextromethorphan metabolism: co-segregation of oxidative O-demethylation with debrisoquin hydroxylation. Clin Pharmacol Ther 1985;38:618-24
193. Dick B, Kupfer A, Molnar J, et al. Drug hydroxylation disorders (debrisoquin type) in a random sample of the Swiss population. Schweiz Med Wochenschr 1982;112:1061-7
194. Zimmermann T, Schlenk R, Pfaff G, et al. Prediction of phenotype for dextromethorphan O-demethylation by using polymerase chain reaction in healthy volunteers. Arzneimittelforschung 1995;45:41-3
195. Funck-Brentano C, Boelle PY, Verstuyft C, et al. Measurement of CYP2D6 and CYP3A4 activity in vivo with dextromethorphan: sources of variability and predictors of adverse effects in 419 healthy subjects. Eur J Clin Pharmacol 2005;61:821-9
196. Kimiskidis VK, Niopas I, Firinidis PD, et al. Frequency distribution of dextromethorphan O-demethylation in a Greek population. Int J Clin Pharmacol Ther 2005;43:150-3
197. Llerena A, Edman G, Cobaleda J, et al. Relationship between personality and debrisoquine hydroxylation capacity. Suggestion of an endogenous neuroactive substrate or product of the cytochrome P4502D6. Acta Psychiatr Scand 1993;87:23-8
198. Spina E, Campo GM, Avenoso A, et al. CYP2D6-related oxidation polymorphism in Italy. Pharmacol Res 1994;29:281-9
199. Jacqz E, Dulac H, Mathieu H. Phenotyping polymorphic drug metabolism in the French Caucasian population. Eur J Clin Pharmacol 1988;35:167-71
200. Larrey D, Amouyal G, Tinel M, et al. Polymorphism of dextromethorphan oxidation in a French population. Br J Clin Pharmacol 1987;24:676-9
201. Freche JP, Dragacci S, Petit AM, et al. Development of an ELISA to study the polymorphism of dextromethorphan oxidation in a French population. Eur J Clin Pharmacol 1990;39:481-5
202. Henthorn TK, Benitez J, Avram MJ, et al. Assessment of the debrisoquin and dextromethorphan phenotyping tests by Gaussian mixture distributions analysis. Clin Pharmacol Ther 1989;45:328-33
203. Benitez J, LLerena A, Cobaleda J. Debrisoquin oxidation polymorphism in a Spanish population. Clin Pharmacol Ther 1988;44:74-7
204. Kunicki PK, Sitkiewicz D, Pawlik A, et al. Debrisoquine hydroxylation in a Polish population. Eur J Clin Pharmacol 1995;47:503-5
205. Skretkowicz J, Wojtczak A, Rychlik-Sych M. Polymorphism of dextromethorphan oxidation in Polish population. Acta Pol Pharm 2008;65:611-15

Affiliation

Adrián LLerena[†1], Maria Eugenia G Naranjo[1], Fernanda Rodrigues-Soares[2,3], Eva M Penas-LLedó[1], Humberto Fariñas[1] & Eduardo Tarazona-Santos[2]
[†]Author for correspondence
[1]University of Extremadura Hospital, CICAB Clinical Research Center, Badajoz, 06080, Spain
E-mail: allerena@unex.es
[2]Universidade Federal de Minas Gerais, Departamento de Biologia Geral, Av. Antônio Carlos, 6627 – Pampulha, Caixa Postal 486, Belo Horizonte, 1270-901, Brazil
[3]Capes foundation, Ministry of Education of Brazil, Brasília – DF 70.040-020, Brazil

Genetic structure of pharmacogenetic biomarkers in Brazil inferred from a systematic review and population-based cohorts: a RIBEF/EPIGEN-Brazil initiative

Além das revisões mundiais dos genes *CYP2D6*, *CYP2C9* e *CYP2C19*, a iniciativa de revisões sistemáticas da RIBEF incluiu também revisões divididas por regiões da América Latina englobando os biomarcadores de interesse farmacogenético de acordo a recomendações da *Food and Drug Administration* (FDA) e *European Medicines Agency* (EMA). Duas revisões com esta metodologia foram publicadas: sobre o México (Fricke-Galindo et al. 2016) e sobre a América Central (Céspedes-Garro et al. 2014); nesta última me enquadro como autora consorciada (Anexo C).

Nosso grupo de pesquisa, através da iniciativa da RIBEF, foi convidado a publicar na *Drug Metabolism and Personalized Medicine* uma revisão sistemática sobre farmacogenética no Brasil. Este trabalho consiste em recalcular as frequências alélicas, genotípicas e fenotípicas de todos os estudos realizados com voluntários sadios no Brasil utilizando os 121 biomarcadores da FDA e EMA. Neste estudo, fiquei responsável por organizar todos os passos da metodologia sistemática e realizar a busca, coleta de dados, análise dos dados e escrita do manuscrito. Coordenei o trabalho de duas alunas de Iniciação Científica, Julia Sampaio e Poliana Andrade, que participaram das etapas de busca e análise de dados, e juntas realizamos a discussão de diferentes etapas do trabalho. Adicionalmente fizemos análises com os dados do projeto Epigen e o manuscrito foi submetido à *The Pharmacogenomics Journal*.

Genetic structure of pharmacogenetic biomarkers in Brazil inferred from a systematic review and population-based cohorts: a RIBEF/EPIGEN-Brazil initiative

Fernanda Rodrigues-Soares Ph.D.[1,2*], Fernanda SG Kehdy Ph.D.[1,3,*], Julia Sampaio-Coelho[1], Poliana XC Andrade[1], Carolina Céspedes-Garro Ph.D.[4], Camila Zolini B.Sc.[1,5], Marla M Aquino B.Sc.[1], Mauricio L Barreto Ph.D.[6,7], Bernardo L Horta Ph.D.[8], Maria Fernanda Lima-Costa Ph.D.[9], Alexandre C Pereira Ph.D.[10], Adrián LLerena Ph.D.[11,12¶], Eduardo Tarazona-Santos Ph.D.[1,¶*]

[1] Universidade Federal de Minas Gerais, Departamento de Biologia Geral, Belo Horizonte, MG, Brazil.

[2] Gerência de Malária, Fundação de Medicina Tropical Dr. Heitor Vieira Dourado, Manaus, AM, Brazil.

[3] Laboratório de Hanseníase, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, RJ, Brazil.

[4] Education and Research Department, Genetics Section, School of Biology, University of Costa Rica, San José, Costa Rica.

[5] Beagle, Belo Horizonte, Minas Gerais, Brazil.

[6] Instituto de Saúde Coletiva, Universidade Federal da Bahia, 40110-040, Salvador, BA, Brazil.

[7] Center for Data and Knowledge Integration for Health, Institute Gonçalo Muniz, Fundação Oswaldo Cruz, Salvador, BA, Brazil.

[8] Programa de Pós-Graduação em Epidemiologia, Universidade Federal de Pelotas, Pelotas, RS, Brazil.

[9] Instituto de Pesquisa Rene Rachou, Fundação Oswaldo Cruz, Belo Horizonte, MG, Brazil.

[10] Instituto do Coração, Universidade de São Paulo, São Paulo, SP, Brazil

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

[12] Centro de Investigación Biomédica en Red: Salud Mental, CIBERSAM, Madrid, Spain.

[*] These authors equally contributed to the present manuscript.

[¶] These authors should be considered Senior Authors of the present manuscript.

Running Title: Pharmacogenetics in Brazil

Corresponding author: Eduardo Tarazona-Santos Ph.D. Av. Antônio Carlos, 6627. Belo Horizonte, MG, Brazil, e-mail: edutars@icb.ufmg.br, Tel: +55 31 34092597

Abstract (215 words, max 250)

We present allele frequencies involving 39 pharmacogenetic biomarkers studied in Brazil, and their distribution on self-reported race/color categories that: (1) involve a mix of perceptions about ancestry, morphological traits and cultural/identity issues, being social constructs pervasively used in Brazilian society and medical studies; (2) are associated with disparities in access to health services, as well as in their representation in genetic studies and (3), as we report here, explain a larger portion of the variance of pharmaco-allele frequencies than geography. We integrated a systematic review of studies on healthy volunteers (years 1968-2017) and the analysis of allele frequencies on three population-based cohorts from northeast, southeast and south, the most populated regions of Brazil. Cross-validation of results from these both approaches suggest that, despite methodological heterogeneity of the 120 studies conducted on 51,747 healthy volunteers, allele frequencies estimates from systematic review are reliable. We report differences in allele frequencies between color categories that persist despite the homogenizing effect of >500 years of admixture. Among clinically relevant variants: *CYP2C9*2* (null), *CYP3A5*3* (defective), *SCLO1B1-rs4149056(C)* and *VKORC1-rs9923231(A)* are more frequent in Whites than in Blacks. Brazilian Native Americans show lower frequencies of *CYP2C9*2*, *CYP2C19*17* (increased activity), and higher of *SCLO1B1-rs4149056(C)* than other Brazilian populations. We present the most current and informative database of pharmaco-allele frequencies in Brazilian healthy volunteers.

other categories, while Browns present intermediate African and European ancestry respect to Blacks and Whites[2]. Also, Black and Mixed people are more likely to have lower income and education and more negative health-related outcomes[9]. Most biomedical studies use the IBGE classification as a proxy of biogeographic ancestry.

Brazilian admixed populations present predominant European and African ancestries, typically with <10% of Native American ancestry[1,7]. Importantly, self-reported Black or White individuals are also admixed, with different proportions of continental ancestry[2]. Because in the admixed Brazilian population mating is not ancestry-independent, but characterized by a degree of ancestry-based assortative mating that maintain a population structure[1], a relevant question is: for which pharmacogenetic variants do we observe significant differences in allele frequencies between *color* classes, notwithstanding the homogenizing effect of more than 500 years of admixture?

The US Food and Drug Administration (FDA) and the European Medicines Agency (EMA) recognize 121 pharmacogenetic biomarkers[10,11] (Table S1, Supplementary Information: Regulatory Agencies Documents List). In some instances, such as the risk of carriers of the *HLA-B*1502* allele to develop Steven Johnson syndrome when treated with antiepileptic agents, the FDA and EMA have released alerts regarding population differences in the frequency of this allele in Europeans, African and Asian descendants. Brazilian regulatory agencies (*Agência Nacional de Vigilância Sanitária*, ANVISA, www.portal.anvisa.gov.br and *Agência Nacional de Saúde Suplementar*, ANS, www.ans.gov.br) have issued pharmacogenetic recommendations for only 15 of the 121 FDA/EMA biomarkers: ABCG2, BCR-ABL, BRAF, CFTR,

EGFR, ERBB2, Coagulation Factor V, G6PD, KRAS, POLG, SLCO1B1, TP53 as well as the enzymes CYP2C19, UGT1A1 and TPMT, which define well established Clinically Actionable Genotypes (CAG, *in sensu* Van Driest et al. 2014)[12].

The aim of this study is to estimate allele frequencies of pharmacogenetic biomarkers, as well as predicted and metabolic phenotypes in Brazilians, considering the effects of geography and *color* categories. We performed a systematic review of studies on healthy volunteers, as a part of an international effort of the Iberoamerican Network of Pharmacogenetics (RIBEF, http://www.ribef.com)[13,14]. We complement the systematic review with data on 55 pharmacogenetic SNPs genotyped in three population-based cohorts from the northeast, southeast and south of Brazil studied by the EPIGEN-Brazil Initiative[1] (https://epigen.grude.ufmg.br). Our analysis supplements the efforts of the REFARGEN (Brazilian Pharmacogenomics Network, www.refargen.org), a nationwide consortium of Brazilian research groups[15].

Materials and Methods

Systematic review

The literature search on Pubmed (updated to April 2017) combined "Brazil" with each of the biomarkers listed in Table S1, which were extracted from FDA and EMA regulatory reports[16,17] and the Pharmacogenomics Knowledgebase[18], as previously reported[14]. Another search was performed using the term "Brazil" combined with each of the following probe drugs: debrisoquine, sparteine, metoprolol, dextromethorphan, mephenytoin, omeprazole, tolbutamide, losartan, diclofenac, isoniazid and caffeine.

Studies included in the review should have reported allele, genotype or phenotype frequencies in Brazilian healthy volunteers for at least one of the biomarkers listed in Table S1. Controls from case-control studies were excluded because they are selected to match characteristics of cases, not being necessarily representative of the general population. Studies using tumor cells were not considered for this review. To avoid redundancies, from a set of articles using the same samples, we only considered the article presenting the highest number of individuals studied for a specific allele. Figure 2 summarizes the search and inclusion criteria.

Articles comprising G6PD were considered even if they presented only frequencies for the enzymatic test, which is the standard procedure to evaluate the deficiency for this enzyme, reflecting the genetic condition[19]. The review for *HLA-A* and *HLA-B* were restricted to the alleles with pharmacogenetic relevance according to Whirl-Carrillo et al.[18]. The frequencies of wild-type variants are typically estimated as 1 minus the sum of the frequencies of the other studied polymorphisms. However, the tested polymorphisms may vary across different studies. Thus, wild-type allele frequencies are not presented in this review.

The predicted phenotype was inferred from genotypes for *CYP2D6*, *CYP2C9, CYP2C19, CYP3A5* and *NAT2* as detailed in the Supplementary Information.

To calculate the allele, genotype and phenotype frequencies for the regions and *color* categories, the average of the frequencies, weighted for sample size, were calculated. We considered alleles as rare or common, if their average frequencies were <5% or ≥5%, respectively. For a pairwise comparison of the *color* categories and regions, a χ^2

test was performed. P-values <0.05 were considered statistically significant.

Integration with EPIGEN-Brazil data

EPIGEN-Brazil genotyped 6,487 individuals from three population-based cohorts, using the Illumina-Omni 2.5M SNPs array[1]. Twenty-two of the 39 pharmacogenetic biomarkers considered in this review were also included in the EPIGEN-Brazil dataset, encompassing 55 pharmaco-alleles from Table S1, which were analyzed for this article. We estimated allele frequencies for these SNPs in the south (Pelotas, n=3,736), southeast (Bambuí, n=1,442) and northeast (Salvador, n=1,309) cohorts and stratifying by self-reported *color* for each cohort: Whites, Mixed, Blacks and Not Informed. For 28 of the 55 shared pharmaco-alleles, Spearman´s correlation was estimated between the allele frequencies of the cohorts and those estimated by the systematic review. To verify the effect of *color* and geographical region on the variance of the allelic frequencies of the 55 pharmacogenetic variants in the EPIGEN-Brazil cohorts, a multivariate distance matrix regression (MDMR) [20] was performed.

Results

We found 120 articles published between 1968 and 2016 including 51,747 healthy volunteers. Only 39 (32%) out of the 121 biomarkers searched were examined (Table S1). The most studied biomarker (considering the number of articles) was the deletion of *GSTM1* (24 articles) and in terms of number of studied individuals is G6PD (18,751 individuals), followed by *GSTM1* (7,874 individuals). Intensive agriculture is one of the most important economic activities in Brazil, and most of *GSTM1* studies analyze subjects exposed to organophosphate pesticides, which are toxic to humans,

mutagenic agents, and are detoxified by GSTM1 in individuals that are not homozygous for its deletion[21,22].

Only eleven out of the 120 articles presented biogeographic ancestry data of individuals inferred from genetic markers[23–33]. Only five studies presented measured metabolic phenotype data, one study for CYP2D6[34] and four studies for CYP2C19[34–37], which is critical from a clinical perspective because frequently, predicted gUM are not real ultra-rapid metabolizers, as many factors such as diseases, diet, life style or other concomitant medications may affect the metabolic phenotype[38].

Allele frequencies estimated from a systematic review and from population-based cohorts provide complementary information. For 33 genes, our systematic review allowed to summarize data from more than one study on specific alleles (Table S1), and to compare frequencies across *color* and geographic categories, or to increase sample sizes, rendering frequency estimates more accurate, even if methodological heterogeneity among studies may be a caveat. On the other hand, population-based cohorts that include thousands of individuals (such as the EPIGEN-Brazil cohorts) provide more accurate estimates of allele frequencies in a specific population, particularly for rare alleles (Table S3). The use of both approaches allowed to cross-validate allele frequencies estimates. The high correlation observed for the 28 pharmacogenetic allele frequencies shared by both approaches (Figure S1, correlation coefficients always >0.90) suggests that our systematic review provides accurate estimates of allele frequencies in Brazilian *color* categories. In addition, the MDMR (multivariate distance matrix regression) on the population-based cohort data showed that in Brazil, for pharmacogenetic alleles, *color* explains (2.82%, estimated as a

pseudoR2 by MDMR) approximately four times more of the genetic variance than geographic region (0.65%). The fact that together, these two factors have an omnibus effect (5.16%), that is greater than the sum of individual effects suggests that interaction of color and geography explains part of the variance (1-2%) of allele frequencies for pharmaco-alleles (Table S4). This finding is consistent with the following results from the systematic review: 50.4% (316 out of 627) *versus* 21.5% (69 out of 321) of allele frequency comparisons between *color* categories and regions, respectively, were significantly different.

Here in the Main Text we summarize the most relevant results involving differences in clinically relevant allele frequencies between White and Black Brazilians, as well as alleles associated with Native American ancestry. The entire set of allele frequencies based on the systematic review for all the 39 studied biomarkers is presented in Tables S5-S8. Table S9 shows the allele frequencies that significantly differ between geographic regions when *color* categories were not considered. In addition to EPIGEN-Brazil cohorts allele frequencies (Table S3), the African, European and Native American ancestries estimated for the EPIGEN-Brazil cohorts, stratified by *color* categories are in Table S10, confirming that self-reported Black and White populations are admixed. Table 1 summarizes the allele frequencies that fit at least one of the following conditions: (1) significantly differ in one *color* category respect to all the others, (2) define Clinically Actionable Genotypes (CAGs), or (3) significantly differ between Whites and Blacks.

Pharmacogenetic biomarkers and differences between Blacks and Whites

We report differences in allele frequencies between *color* categories that persist despite the homogenizing effect of >500 years of admixture (Table 1 and S3), that is not completely ancestry-independent, but characterized by a degree of a persistent ancestry-based assortative mating in Brazilian populations[1]. In particular:

(1) CYP2D6, CYP2C9, and CYP3A5 are among the most relevant Phase I drug-metabolizing enzymes (DMEs) and some of their alleles present significant differences between Whites and Blacks (Table 1). *CYP3A5* presents CAGs for the immunosuppressive tacrolimus, used after organ transplantation. Indeed, the Clinical Pharmacogenetics Implementation Consortium (CPIC) guidelines for tacrolimus recommends to increase the initial dose by 1.5-2 times in patients who are intermediate or extensive metabolizers[12]. In Brazil, the sum of frequencies of *CYP3A5* gIM and gEM are 34% for Whites and 65% for Blacks (Table S6).

(2) Among Phase II enzymes (involved in conjugation reactions that inactivates excreted metabolites), White Brazilians show higher frequencies than Blacks for the common allele *UGT1A1*93* (associated with irinotecan toxicity in cancer patients[39]) and the deletion of *GSTM1*.

(3) *SLCO1B1* encodes a transporter protein (OATP1B1) that is involved in cellular influx of endogenous and xenobiotic substances, such as statins. White Brazilians have higher frequencies than Blacks for the common allele *SLCO1B1 521T>C* (rs4149056, which in homozygosis is a CAG for simvastatin toxicity).

(4) VKORC1 is an enzyme that activates vitamin K, a cofactor for blood coagulation, and is used as target of the anticoagulant warfarin. Allele -1639 G>A (rs9923231) reduces gene expression and lead to fewer functional copies of VKORC1 enzyme[40]. Consequently, carriers of this polymorphism require lower doses of the considered therapeutic dose of warfarin (because they present fewer enzymes). If these

individuals are treated with normal dose regimen, warfarin blood levels will increase and induce potential hemorrhagic events. This allele is a biomarker for the prescription of a lower dose of warfarin[41], and, combined with -5723T>G (rs2884737), both are associated with individual European genomic ancestry in admixed individuals[32] (Table 1).

(5) Red blood cells with G6PD deficiency become more susceptible to oxidative stress when exposed to oxidant drugs[42], and individuals with G6PD deficiency show hemolytic anemia when treated with the antimalarial primaquine. As *G6PD* is on the X chromosome, males are hemizygous; whereas women, due to X inactivation, may be mosaics of cells expressing normal and deficient alleles. Whites have a lower frequency of G6PD deficiency than Mixed and Blacks (>10%)[43–45] (Table S8), consistent with a previous review in Latin American populations[46].

Biomarkers in Brazilian Native Americans

Twenty-six of the 39 genes studied in Brazilians have been tested in at least one Native American population. Native American populations are typically tested for fewer alleles than Whites/Mixed/Black populations, and the number of studies per gene is lower in Native Americans (Table S5-S8). Importantly, Native Americans are composed by small groups that are scattered across Brazil, and their allele frequencies may be severely affected by genetic drift[47]. Therefore, allele frequencies reported in a specific study may not be representative of all Brazilian Native Americans. CYP2C9 metabolizes drugs such as the oral anticoagulant warfarin and the antidepressant fluoxetine, and the frequency of the warfarin clinically actionable phenotype gPM is low (≤ 3%) for all *color* categories, matching frequencies in worldwide populations[48] (Table 2). Notably, the genotyping of 222 Native Americans did not reveal any

CYP2C9 gPM individual, reflecting the low frequencies of the defective alleles **2* and **3* in this population. The following clinically relevant alleles are more common in Brazilian Native Americans than in other *color* categories (Table 1, Supplementary Tables S5-S8): *ABCB1 3435C>T* (50%, defective transport of digitalis derivatives such as digoxin), *SCLO1B1 521T>C* (28%, CAG for simvastatin toxicity) and *SCLO1B1 89595T>C* (26%). The following clinically relevant alleles are less common in Native Americans than in other ethnic groups: decreasing activity *CYP2C9*2* (rare in Native Americans and common elsewhere), increasing activity *CYP2C19*17* (11%), *NAT2*13* (21%), deletion of *GSTM1* (25%), deletion of *GSTT1* (19%) and *CCR5 D32* (0.1%). For both genders, and differently from Whites, Mixed and Blacks, no Native American individual was found to be G6PD deficient (n = 153)[49], corroborating previous research[50], which is important because the Amazon region, where several individuals with Native American ancestry are settled, presents a high incidence of malaria[51].

Discussion

We analyzed pharmacogenetic biomarkers in Brazilians integrating two approaches: (1) a systematic review of the literature following the RIBEF methodology[13,14,38]; and (2) the analysis of allele frequencies from the three largest Brazilian population-based cohorts. Population-based cohorts are rarely studied in pharmacogenetic studies and, in this case, provide robust allele frequency estimates from the three most populated Brazilian regions: south (Pelotas), southeast (Bambuí) and northeast (Salvador). A limitation of the EPIGEN-Brazil dataset is that pharmacogenetic variants are poorly represented in the generalist Illumina-Omni2.5 array. Only 22 of the 39 pharmacogenetic biomarkers considered are included in the EPIGEN-Brazil dataset,

encompassing 55 pharmaco-alleles. Important pharmacogenes such as *CYP2D6*, are not represented and some ethno-specific variants such as *CYP2C9*5*, **6*, **8*, **11* were not tested as a part of the EPIGEN-Brazil cohorts.

In this systematic review, there are methodological issues shared with other pharmacogenetic systematic reviews: (1) Articles that used alternative names for some of the included biomarkers may have been missed. Here we used the RIBEF methodology that only uses the most commonly used biomarker name. (2) There is heterogeneity in ethnic classification. Although IBGE uses a specific self-reported individual racial classification, pharmacogenetics and population genetics studies vary in their operational definitions of individuals and populations. For most studies, the IBGE *color* categories (or synonyms) are followed, but for many studies, there are no explicit statements that *color* assignment was obtained via individual self-reporting. (3) Frequently, only allele frequencies but not genotype frequencies are reported, preventing prediction of phenotype frequencies. (4) There is heterogeneity in the methods used to genotype different markers. More importantly, the unambiguous rs nomenclature for SNPs (http://www.ncbi.nlm.nih.gov/snp) is not always followed in pharmacogenetic studies. In this review, we made a special effort to track the rs code for the variants that defined pharmacogenetic alleles and to consistently match the classical pharmacogenetics nomenclature with the rs nomenclature across the text, tables and supplementary information (Table S2). Moreover, while several pharmaco-alleles are defined by a single SNP, others are haplotypes defined by a combination of SNP alleles in linkage disequilibrium (Table S2), adding an additional layer of heterogeneity because different research groups may genotype different number or combinations of these SNPs to define these haplotypes. Despite the heterogeneity in

ethnic classification, genotyping methods and pharmaco-alleles operational definitions, the very high correlation of allele frequencies obtained in this systematic review and those obtained from the population-based cohorts suggests that pharmaco-allele frequencies, estimated in the present study from 120 articles published since 1968, are currently the most complete source of this type of information for the Brazilian populations.

Although most pharmacogenetic studies still use *color* categories as a proxy for ancestry, during the last few years, both the REFARGEN and RIBEF initiatives[27,52] have used genetic markers to infer the biogeographic ancestry of Latin American individuals. Genomic-based inferences treat ancestry as a quantitative variable, being more objective, informative and accurate than *color* categories, which are affected by semantic differences within and among regions[7]. While we recommend to use ancestry informative markers to infer individual biogeographic ancestry in pharmacogenetic studies, the use of *color* categories is also informative in different contexts, notwithstanding some criticisms[7,53]. Here we use race/*color* categories as a social construct that is pervasively and historically used in Brazil and other countries such as the US, and that is correlated with biogeographic ancestry and morphological traits[54]. In terms of population genetics theory, both EPIGEN-Brazil and systematic review data show that c*olor* categories influence the variance of pharmaco-allele frequencies more than geographic regions (which would correspond to a geographical definition of populations, Table S4), supporting the decision of stratifying allele frequencies by *color* categories for this review. This approach has conceptual and methodological advantages over alternative approaches that ignore race/*color* categories, enabling the use of data from a plethora of studies, as well as to focus on

populations such as Black Brazilians, which are associated with disparities in access to health services and are under-represented in genetic studies.

In conclusion, by combining a literature systematic review and studying the largest Brazilian population-based cohorts, we generated the most current and informative database of pharmaco-allele frequencies in Brazilian healthy volunteers. Although we identified methodological issues that might affect the integration of different pharmacogenetic studies, our combined approach and posterior cross-validation show that, despite heterogeneity in ethnic classification, genotyping methods and pharmaco-allele definitions, a literature systematic review in a country such as Brazil, accurately estimate allele frequencies. There is a need of studies that actually measure functional and metabolic phenotypes. Although the introduction of biogeographic ancestry using genomic data is a welcomed trend in pharmacogenetics, the use of *color* categories is informative. We found that *color* categories are four times more determinant of the variance of pharmaco-allele frequencies than geographic regions. Therefore, presenting data stratified by race/*color* categories allows to focus on populations such as Black Brazilians, that have worse access to health services and are under-represented in genetic studies, thereby making pharmacogenetics more inclusive.

Acknowledgements

We thank RIBEF/MESTIFAR members for interactions, to Daniela Junqueira, Maria Clara da Silva, Maria B Lovato and Renan P Souza for discussions. Funding**:** CAPES Foundation from the Brazilian Ministry of Education, Brazilian National Research Council (CNPq), Minas Gerais State Agency for Research (FAPEMIG). The

EPIGEN-Brazil Initiative was funded by the Brazilian Ministry of Health (Secretaria de Ciência, Tecnologia e Insumos Estratégicos [SCTIE]), through Financiadora de Estudos e Projetos (FINEP).

Conflict of interest

The authors declare no competing interests.

References

1 Kehdy FSG, Gouveia MH, Machado M, Magalhães WCS, Horimoto AR, Horta BL *et al.* Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proc Natl Acad Sci* 2015; **112**: 8696–8701.

2 Lima-Costa MF, Rodrigues LC, Barreto ML, Gouveia M, Horta BL, Mambrini J *et al.* Genomic ancestry and ethnoracial self-classification based on 5,871 community-dwelling Brazilians (The Epigen Initiative). *Sci Rep* 2015; **5**: 9812.

3 Ramos E, Doumatey A, Elkahloun AG, Shriner D, Huang H, Chen G *et al.* Pharmacogenomics, ancestry and clinical decision making for global populations. *Pharmacogenomics J* 2014; **14**: 217–22.

4 Li J, Zhang L, Zhou H, Stoneking M, Tang K. Global patterns of genetic diversity and signals of natural selection for human ADME genes. *Hum Mol Genet* 2011; **20**: 528–40.

5 Limdi NA, Brown TM, Yan Q, Thigpen JL, Shendre A, Liu N *et al.* Race influences warfarin dose changes associated with genetic factors. *Blood* 2015; **126**: 539–45.

6 Suarez-Kurtz G, Struchiner CJ. Differential effects of predictors of warfarin dose according to race/color categories in the admixed Brazilian population.

Pharmacogenet Genomics 2017; **27**: 210–211.

7 Pena SDJ, Di Pietro G, Fuchshuber-Moraes M, Genro JP, Hutz MH, Kehdy FDSG *et al.* The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PLoS One* 2011; **6**: e17063.

8 Osório RG. *O sistema classificatório de 'cor ou raça' do IBGE*. 2003http://www.ipea.gov.br/portal/images/stories/PDFs/TDs/td_0996.pdf (accessed 7 Mar2016).

9 Telles EE. *Race in another America : the significance of skin color in Brazil*. Princeton University Press, 2004.

10 Zineh I, Pacanowski MA. Pharmacogenomics in the assessment of therapeutic risks versus benefits: inside the United States Food and Drug Administration. *Pharmacotherapy* 2011; **31**: 729–35.

11 Chen JJ, Lin W-J, Chen H-C. Pharmacogenomic biomarkers for personalized medicine. *Pharmacogenomics* 2013; **14**: 969–80.

12 Van Driest SL, Shi Y, Bowton EA, Schildcrout JS, Peterson JF, Pulley J *et al.* Clinically actionable genotypes among 10,000 patients with preemptive pharmacogenomic testing. *Clin Pharmacol Ther* 2014; **95**: 423–31.

13 Céspedes-Garro C, Naranjo M-EG, Ramírez R, Serrano V, Fariñas H, Barrantes R *et al.* Pharmacogenetics in Central American healthy volunteers: interethnic variability. *Drug Metabol Drug Interact* 2014; : 1–13.

14 Fricke-Galindo I, Jung-Cook H, LLerena A, López-López M. Interethnic variability of pharmacogenetic biomarkers in Mexican healthy volunteers: a report from the RIBEF (Ibero-American Network of Pharmacogenetics and Pharmacogenomics). *Drug Metab Pers Ther* 2016. doi:10.1515/dmpt-2015-0030.

15 Suarez-Kurtz G. Pharmacogenomics in admixed populations: the Brazilian pharmacogenetics/pharmacogenomics network--REFARGEN. *Pharmacogenomics J* 2004; **4**: 347–8.

16 FDA. Table of Pharmacogenomic Biomarkers in Drug Labeling. http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm (accessed 18 Apr2016).

17 Medicines Agency E. Reflection paper on methodological issues associated with pharmacogenomic biomarkers in relation to clinical development and patient selection. 2011.http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2011/07/WC500108672.pdf (accessed 9 Apr2017).

18 Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF *et al.* Pharmacogenomics knowledge for personalized medicine. *Clin Pharmacol Ther* 2012; **92**: 414–7.

19 Santana MS, de Lacerda MVG, Barbosa M das GV, Alecrim WD, Alecrim M das GC. Glucose-6-phosphate dehydrogenase deficiency in an endemic area for malaria in Manaus: a cross-sectional survey in the Brazilian Amazon. *PLoS One* 2009; **4**: e5259.

20 Zapala MA, Schork NJ. Statistical properties of multivariate distance matrix regression for high-dimensional data analysis. *Front Genet* 2012; **3**: 190.

21 Bhinder P, Chaudhry A. Mutagenicity Assessment of Organophosphates using Polymerase Chain Reaction-Restriction Fragment Length Polymorphism Assay. *Toxicol Int* 2013; **20**: 254–60.

22 Muranli FDG, Kanev M, Ozdemir K. Genotoxic effects of diazinon on human peripheral blood lymphocytes / Genotoksično djelovanje diazinona na limfocite

ljudske periferne krvi. *Arch Ind Hyg Toxicol* 2015; **66**: 153–8.

23 Suarez-Kurtz G, Genro JP, de Moraes MO, Ojopi EB, Pena SDJ, Perini JA *et al.* Global pharmacogenomics: Impact of population diversity on the distribution of polymorphisms in the CYP2C cluster among Brazilians. *Pharmacogenomics J* 2012; **12**: 267–76.

24 Suarez-Kurtz G, Perini JA, Bastos-Rodrigues L, Pena SDJ, Struchiner C. Impact of population admixture on the distribution of the CYP3A5*3 polymorphism. *Pharmacogenomics* 2007; **8**: 1299–306.

25 Suarez-Kurtz G, Vargens DD, Struchiner CJ, Bastos-Rodrigues L, Pena SDJ. Self-reported skin color, genomic ancestry and the distribution of GST polymorphisms. *Pharmacogenet Genomics* 2007; **17**: 765–71.

26 Suarez-Kurtz G, Vargens DD, Santoro AB, Hutz MH, de Moraes ME, Pena SDJ *et al.* Global pharmacogenomics: distribution of CYP3A5 polymorphisms and phenotypes in the Brazilian population. *PLoS One* 2014; **9**: e83472.

27 Suarez-Kurtz G, Amorim A, Damasceno A, Hutz MH, de Moraes MO, Ojopi EB *et al.* VKORC1 polymorphisms in Brazilians: comparison with the Portuguese and Portuguese-speaking Africans and pharmacogenetic implications. *Pharmacogenomics* 2010; **11**: 1257–67.

28 Bonifaz-Peña V, Contreras A V., Struchiner CJ, Roela RA, Furuya-Mazzotti TK, Chammas R *et al.* Exploring the Distribution of Genetic Markers of Pharmacogenomics Relevance in Brazilian and Mexican Populations. *PLoS One* 2014; **9**: e112640.

29 Friedrich DC, Genro JP, Sortica VA, Suarez-Kurtz G, de Moraes ME, Pena SDJ *et al.* Distribution of CYP2D6 alleles and phenotypes in the Brazilian population. *PLoS One* 2014; **9**: e110691.

30 Vargens DD, Almendra L, Struchiner CJ, Suarez-Kurtz G. Distribution of the GNB3 825C>T polymorphism among Brazilians: impact of population structure. *Eur J Clin Pharmacol* 2008; **64**: 253–6.

31 Soares RAG, Santos PCJL, Machado-Coelho GLL, do Nascimento RM, Mill JG, Krieger JE *et al.* CYP2C9 and VKORC1 polymorphisms are differently distributed in the Brazilian population according to self-declared ethnicity or genetic ancestry. *Genet Test Mol Biomarkers* 2012; **16**: 957–63.

32 Sortica V de A, Ojopi EB, Genro JP, Callegari-Jacques S, Ribeiro-Dos-Santos A, de Moraes MO *et al.* Influence of genomic ancestry on the distribution of SLCO1B1, SLCO1B3 and ABCB1 gene polymorphisms among Brazilians. *Basic Clin Pharmacol Toxicol* 2012; **110**: 460–8.

33 Santos KT, De Freitas RGA, Manta FSN, De Carvalho EF, Silva DA. Ser49Gly polymorphism in the β-adrenergic receptor 1 gene in a population sample from Rio de Janeiro state, Brazil, stratified by self-identified skin color and genetic ancestry. *Mol Med Rep* 2015; **12**: 1591–7.

34 Rocha A, Coelho EB, Sampaio SA, Lanchote VL. Omeprazole preferentially inhibits the metabolism of (+)-(S)-citalopram in healthy volunteers. *Br J Clin Pharmacol* 2010; **70**: 43–51.

35 Cassiano NM, Oliveira R V., Bernasconi GCR, Cass QB. Enantioselective disposition of omeprazole, pantoprazole, and lansoprazole in a same Brazilian subjects group. *Chirality* 2012; **24**: 289–293.

36 Rocha A, Coelho EB, Moussa SAP, Lanchote VL. Investigation of the in vivo activity of CYP3A in Brazilian volunteers: comparison of midazolam and omeprazole as drug markers. *Eur J Clin Pharmacol* 2008; **64**: 901–6.

37 Cass QB, Lima V V, Oliveira R V, Cassiano NM, Degani ALG, Pedrazzoli J.

Enantiomeric determination of the plasma levels of omeprazole by direct plasma injection using high-performance liquid chromatography with achiral-chiral column-switching. *J Chromatogr B Analyt Technol Biomed Life Sci* 2003; **798**: 275–81.

38 LLerena A, Naranjo MEG, Rodrigues-Soares F, Penas-LLedó EM, Fariñas H, Tarazona-Santos E. Interethnic variability of CYP2D6 alleles and of predicted and measured metabolic phenotypes across world populations. *Expert Opin Drug Metab Toxicol* 2014; **10**: 1569–83.

39 Côté J-F, Kirzin S, Kramar A, Mosnier J-F, Diebold M-D, Soubeyran I *et al.* UGT1A1 polymorphism can predict hematologic toxicity in patients treated with irinotecan. *Clin Cancer Res* 2007; **13**: 3269–75.

40 Yuan H-Y, Chen J-J, Lee MTM, Wung J-C, Chen Y-F, Charng M-J *et al.* A novel functional VKORC1 promoter polymorphism is associated with inter-individual and inter-ethnic differences in warfarin sensitivity. *Hum Mol Genet* 2005; **14**: 1745–1751.

41 Perini JA, Struchiner CJ, Silva-Assunção E, Santana ISC, Rangel F, Ojopi EB *et al.* Pharmacogenetics of warfarin: development of a dosing algorithm for brazilian patients. *Clin Pharmacol Ther* 2008; **84**: 722–8.

42 Santana MS, Monteiro WM, Siqueira AM, Costa MF, Sampaio V, Lacerda M V. *et al.* Glucose-6-phosphate dehydrogenase deficient variants are associated with reduced susceptibility to malaria in the Brazilian Amazon. *Trans R Soc Trop Med Hyg* 2013; **107**: 301–306.

43 Egesie OJ, Joseph DE, Isiguzoro I, Egesie UG. Glucose-6-phosphate dehydrogenase (G6PD) activity and deficiency in a population of Nigerian males resident in Jos. *Niger J Physiol Sci* 2008; **23**: 9–11.

44 Müller SA, Amoah SKB, Meese S, Spranger J, Mockenhaupt FP. High prevalence of anaemia among African migrants in Germany persists after exclusion of iron deficiency and erythrocyte polymorphisms. *Trop Med Int Heal* 2015; **20**: 1180–1189.

45 Nguetse CN, Meyer CG, Adegnika AA, Agbenyega T, Ogutu BR, Kremsner PG *et al.* Glucose-6-phosphate dehydrogenase deficiency and reduced haemoglobin levels in African children with severe malaria. *Malar J* 2016; **15**: 346.

46 Monteiro WM, Val FFA, Siqueira AM, Franca GP, Sampaio VS, Melo GC *et al.* G6PD deficiency in Latin America: systematic review on prevalence and variants. *Mem Inst Oswaldo Cruz* 2014; **109**: 553–68.

47 Callegari-Jacques SM, Tarazona-Santos EM, Gilman RH, Herrera P, Cabrera L, dos Santos SEB *et al.* Autosome STRs in native South America-Testing models of association with geography and language. *Am J Phys Anthropol* 2011; **145**: 371–81.

48 Céspedes-Garro C, Fricke-Galindo I, Naranjo M-EG, Rodrigues-Soares F, Fariñas H, de Andrés F *et al.* Worldwide interethnic variability and geographical distribution of CYP2C9 genotypes and phenotypes. *Expert Opin Drug Metab Toxicol* 2015; **11**: 1893–905.

49 Saldanha PH, Lebensztajn B, Itskan SB. Activity of glucose-6-phosphate dehydrogenase among Indians living in a malarial region of Mato Grosso and its implication to the Indian-mixed populations in Brazil. *Hum Hered* 1976; **26**: 241–51.

50 Salzano FM, Callegari-Jacques SM. *South American Indians: A Case Study in Evolution*. Clarendon Press: Oxford, 1988.

51 Coimbra CEA, Santos RV, Welch JR, Cardoso AM, de Souza MC, Garnelo L *et al.* The First National Survey of Indigenous People's Health and Nutrition in Brazil: rationale, methodology, and overview of results. *BMC Public Health* 2013; **13**: 52.

52 Céspedes-Garro C, Rodrigues-Soares F, Jimenez-Arce G, Naranjo M-EG, Tarazona-Santos E, Fariñas H *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* 2016; **in press**.

53 Barbujani G, Pigliucci M. Human races. *Curr Biol* 2013; **23**: R185-7.

54 Ruiz-Linares A, Adhikari K, Acuña-Alonzo V, Quinto-Sanchez M, Jaramillo C, Arias W *et al.* Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. *PLoS Genet* 2014; **10**: e1004572.

55 Vianna-Jorge R, Perini JA, Rondinelli E, Suarez-Kurtz G. CYP2C9 genotypes and the pharmacokinetics of tenoxicam in Brazilians. *Clin Pharmacol Ther* 2004; **76**: 18–26.

56 Vargens DD, Petzl-Erler M-L, Suarez-Kurtz G. Distribution of CYP2C polymorphisms in an Amerindian population of Brazil. *Basic Clin Pharmacol Toxicol* 2012; **110**: 396–400.

57 Perini JA, Vargens DD, Santana ISC, Moriguchi EH, Ribeiro-Dos-Santos AKC, Tsutsumi M *et al.* Pharmacogenetic polymorphisms in Brazilian-born, first-generation Japanese descendants. *Braz J Med Biol Res* 2009; **42**: 1179–84.

58 Kohlrausch FB, Carracedo Á, Hutz MH. Characterization of CYP1A2, CYP2C19, CYP3A4 and CYP3A5 polymorphisms in South Brazilians. *Mol*

Biol Rep 2014; **41**: 1453–60.

59 Suarez-Kurtz G, Sortica VA, Vargens DD, Bruxel EM, Petzl-Erler M-L, Petz-Erler M-L *et al.* Impact of population diversity on the prediction of 7-SNP NAT2 phenotypes using the tagSNP rs1495741 or paired SNPs. *Pharmacogenet Genomics* 2012; **22**: 305–9.

Figure legends

Figure 1. Self-reported race/*color* distribution in major geographical regions of Brazil and in the entire country. Data from Brazilian Institute of Geography and Statistics (IBGE, https://sidra.ibge.gov.br/Tabela/136). *Color* is the Portuguese word used by IBGE for self-reported racial categories.

Figure 2. Prisma flowchart of the methodology for articles searching and inclusion criteria. Exclusion of duplicates: if an article was included more than once because it was found in more than one biomarker search, only one copy was maintained. Exclusion of duplicated samples: Only one of two or more articles using the same samples was included.

Tables and Figures

Table 1. Pharmaco-allele frequencies that fit at least one of the following conditions: (1) significantly differ in one *color* category respect to all the others (in bold), (2) define Clinically Actionable Genotypes (CAGs, Van Driest et al. (2014)[12], denoted by the ⊠ symbol), or (3) significantly differ between Whites and Blacks (underlined).

Gene and allele	SNPdb code	Allele	Allele frequencies				
			White[a]	Mixed	Black[a]	Native American	Asian descendant
Phase I Drug Metabolizing Enzymes							
*CYP2D6*4⌀*	haplotype[&]	-	**0.16**	0.12	0.10	-	-
*CYP2D6*2*	haplotype		0.24	0.24	0.17	-	-
*CYP2D6*17↓*	haplotype	-	**0.02**	0.06	**0.11**	-	-
*CYP2D6*29↓*	haplotype	-	**0.01**	0.04	0.06	-	-
*CYP2D6*41↓*	haplotype		0.08	0.02	0.01	-	-
*CYP2C9*2↓* ⊠	rs1799853	T	**0.12**	0.09	0.07	**0.03**	**0.00**
*CYP2C9*3↓*	haplotype		0.05	0.05	0.03	0.01	0.04
*CYP2C19*2⌀* ⊠	haplotype	-	0.15	0.14	0.16	0.11	**0.27**
*CYP2C19*3⌀* ⊠	rs4986893	A	0.00	0.00	0.00	0.00	**0.14**
*CYP2C19*17↑*	haplotype	-	0.18	0.19	0.20	**0.11**	-
*CYP2B6*5*	rs3211371	T	**0.21**	0.14	0.13	-	-
*CYP2B6*6↓*	haplotype	-	0.30	0.30	**0.37**	-	-
*CYP2B6*9*	haplotype	-	**0.39**	**0.33**	**0.46**	-	-
*CYP3A5*3↓* ⊠	haplotype	-	**0.79**	**0.64**	**0.50**	-	-
*CYP3A5*6⌀*	rs10264272	T	**0.01**	**0.04**	**0.09**	-	-
*CYP3A5*7↓*	rs41303343	insA	**0.01**	**0.03**	**0.07**	-	-
*CYP3A4*1B*	rs2740574	C	**0.10**	-	**0.38**	-	-
*CYP2E1*5B*	haplotype	-	0.08	0.08	0.06	**0.19**	-
*CYP1A1*2A*	rs4646903	G	0.15	-	**0.30**	0.11	-
*CYP2C8*2.˙*	rs11572103	A	0.03	**0.06**	**0.11**	0.01	-
*CYP2C8*3.˙*	haplotype	-	**0.13**	0.08	0.06	**0.00**	-
Phase II Drug Metabolizing Enzymes							
*NAT2*6↓*	rs1799930	A	0.27	0.29	0.24	**0.12**	-
*NAT2*12*	rs1208	G	0.44	0.46	**0.32**	0.44	-
*NAT2*13*	rs1041983	T	0.33	0.39	0.41	**0.21**	-
*NAT2*14↓*	rs1801279	A	0.02	0.02	**0.07**	0.01	-
*UGT1A1*36↑*	rs8175347	$[TA]_5$	0.01	-	0.06	0	-
*UGT1A1*60*	rs4124874	G	0.52	0.57	0.67	-	-
*UGT1A1*93*	rs10929302	A	0.22	0.21	**0.11**	-	-
GSTT1deleted⌀	-		0.21	0.23	0.26	0.19	-
GSTP1 Ile	rs1695		0.69	-	0.62	0.60	-
GSTP1 Val↓	rs1695		0.31	-	0.38	0.40	-
GSTM1 deletion	-	-	0.53	0.44	0.34	**0.25**	-
Transporters							
ABCB1/MDR1 3435C>T	rs1045642	G	0.44	0.38	0.33	**0.51**	-
ABCB1/MDR1 2677G>T/A	rs2032582	C/T	0.42	0.32	**0.23**	-	-
ABCB1/MDR1 1236 C>T	rs1128503	G	0.40	0.35	0.31	-	-
SLCA4 S	-	-	-	-	-	**0.75**	-
SCLO1B1 388A>G	rs2306283	G	**0.50**	0.59	0.65	-	-

SCLO1B1 521T>C ⊠	rs4149056	C	0.14	0.14	**0.09**	**0.28**	-
SCLO1B1 89595T>C	rs4363657	C	0.15	0.18	0.11	**0.26**	-
Other biomarkers							
*HLA-A*0201*	-	-	-	-	-	**0.04**	-
VKORC1 H2	haplotype	-	0.24	0.23	**0.08**	-	-
VKORC -1639G>A 3673G>A ⊠	rs9923231	A/G/T	**0.46**	0.38	**0.24**	0.39	**0.88**
VKORC1 -5723T>G 5808T>G	rs2884737	C	**0.25**	0.13	0.13	-	**0.003**
VKORC1 6853G>C	rs8050894	G/T	0.40	0.39	0.36	-	**0.88**
VKORC1 9104G>A	rs7294	T	0.38	0.37	0.38	-	**0.12**
LDLR A	-	-	0.47	0.39	**0.27**	-	-
LDLR B	-		0.53	0.61	0.73	-	-
*CFTR KM19*2*	-	-	-	-	-	**0.04**	-
MTHFR 677 C>T	rs1801133	A	0.31	0.24	0.23	-	-
MTHFR 1298A>C (1286 A>C)	rs1801131	G	0.29	0.22	0.17	-	-
MTHFR 1317T>C	rs4846051	A	0.96	0.88	**0.68**	-	-
CCR5 D32	rs333	Indel	0.07	0.06	0.03	**0.001**	0.00
ProteinC 1641A>G	-	-	0.38	-	0.42	**0.77**	-
ProteinC 1476A>G	-	-	0.40	-	0.41	**0.15**	-
SOD 247T>G	rs4880	G	-	-	0.49	**0.32**	-
GNB3 825C>T	rs5443	T	0.35	**0.49**	**0.61**	-	-
*eTP53 BstUI*A1(Arg72Pro)*	rs1042522	C	0.31	-	**0.52**	**0.15**	-
*TP53 16bp*A2*	-	-	**0.12**	-	**0.24**	**0.00**	-
*TP53 MspI*A1*	-	-	**0.10**	-	**0.23**	**0.01**	-

↓ Decreased activity allele
↑ Increased activity allele
∅ Null activity allele
∴ Decreased in vitro activity
⊠ These alleles define Clinically Actionable Genotypes according to Van Driest et al. (2014)[12].
& pharmaco-allele defined by a set of SNPs in linkage disequilibrium that form a haplotype. SNPs defining the haplotype are in Table S3.

Table 2. Predicted and measured phenotype frequencies in Brazilian population.

***CYP2D6* predicted phenotype**						
Population		**gPM**	**gIM**	**gEM**	**gUM**	**References**
White	**n**	327	327	327	327	[29]
	f	0.03	0.05	0.88	0.04	
Mixed	**n**	336	336	336	336	[29]
	f	0.03	0.06	0.88	0.03	
Black	**n**	324	324	324	324	[29]
	f	0.02	0.10	0.83	0.05	
***CYP2D6* measured phenotype**						
Population		**mPM**	**mIM**	**mEM**	**mUM**	**References**
White	**n**	8	8	8	8	[34]
	f	0.00	0.00	1.00	0.00	

***CYP2C9* predicted phenotype**					
Population		**gPM**	**gIM**	**gEM**	**References**
White	**n**	744	744	744	[31,55]
	f	0.03	**0.27**	**0.7**	
Mixed	**n**	431	431	431	[31,55]
	f	0.02	0.22	0.76	
Black	**n**	194	194	194	[31,55]
	f	0.03	0.15	0.82	
Native American	**n**	222	222	222	[31,56]
	f	0	0.1	0.9	
Asian descendants	**n**	64	64	64	[57]
	f	0	0.02	0.98	

***CYP2C19* predicted phenotype**					
Population		**gPM**	**gEM**	**gUM**	**References**
White	**n**	101	101	101	[36,58]
	f	**0.01**	**0.94**	0.05	
Black	**n**	87	87	87	[58]
	f	**0.03**	**0.97**	0	
Asian descendants	**n**	59	59	59	[57]
	f	**0.22**	**0.78**	0	
***CYP2C19* measured phenotype**					
Population		**mPM**	**mEM**	**mUM**	**References**
White	**n**	9	9	9	[36]
	f	0	1	0	
Not informed	**n**	28	28	28	[34,35,37]
	f	0.07	0.93	0	

***CYP3A5* predicted phenotype**					
Population		**gPM**	**gIM**	**gEM**	**References**
White	**n**	498	498	498	[24,26,58]
	f	**0.66**	**0.30**	**0.04**	
Mixed	**n**	703	703	703	[24,26]
	f	**0.44**	0.46	**0.10**	
Black	**n**	491	491	491	[24,26,58]
	f	**0.36**	0.45	**0.20**	
***NAT2* predicted phenotype**					
Population		**Slow**	**Intermediate**	**Rapid**	**References**
White	**n**	136	136	136	[59]
	f	0.6	0.35	0.05	
Mixed	**n**	107	107	107	[59]
	f	0.55	0.38	0.07	
Black	**n**	118	118	118	[59]
	f	0.47	0.45	0.08	
Native American	**n**	88	88	88	[59]
	f	0.51	0.44	0.05	

Tables and Figures

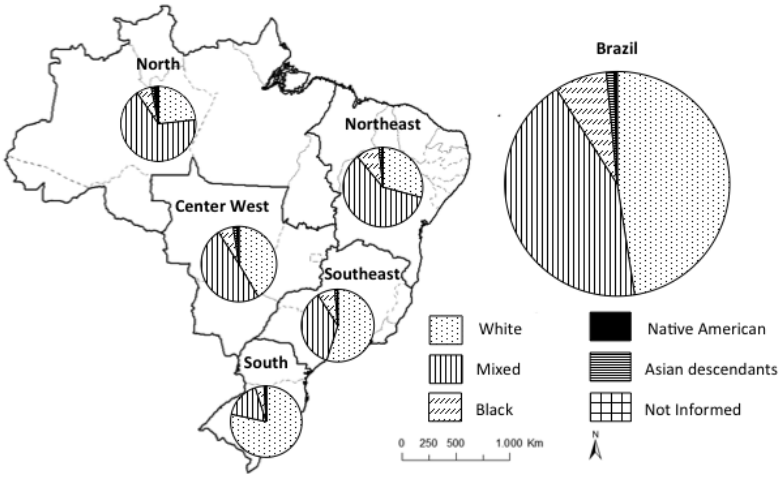

Figure 1.

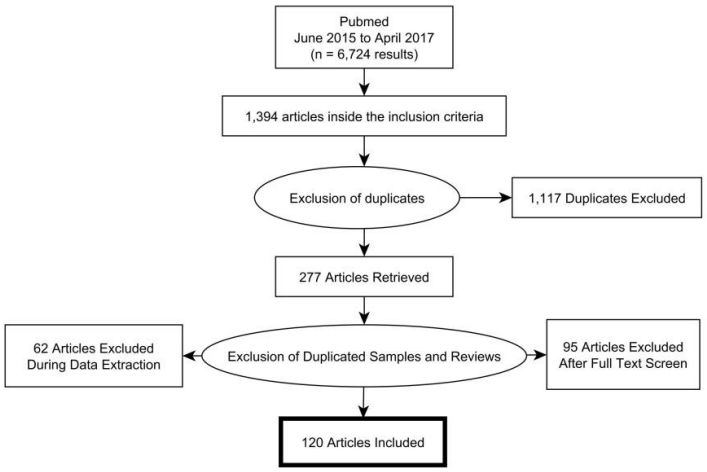
Pubmed
June 2015 to April 2017
(n = 6,724 results)
1,394 articles inside the inclusion criteria
Exclusion of duplicates
1,117 Duplicates Excluded
277 Articles Retrieved
62 Articles Excluded
During Data Extraction
Exclusion of Duplicated Samples and Reviews
95 Articles Excluded
After Full Text Screen
120 Articles Included

Figure 2.

SUPPLEMENTARY INFORMATION

Genetic structure of pharmacogenetic biomarkers in Brazil inferred from a systematic review and population-based cohorts: a RIBEF/EPIGEN-Brazil initiative

Supplementary Information: Regulatory Agencies Documents List

EMA. Scientific guideline [Internet]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2011/07/WC500108672.pdf

FDA. Table of Pharmacogenomic Biomarkers in Drug Labeling [Internet]. [cited 2016 Apr 18]. Available from: http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm

FDA. Carbamazepine Prescribing Information [Internet]. Available from: http://www.accessdata.fda.gov/drugsatfda_docs/label/2009/021710s006lbl.pdf

FDA. Collection of Race and Ethnicity Data in Clinical Trials [Internet]. 2005. Available from: http://www.fda.gov/downloads/regulatoryinformation/guidances/ucm126396.pdf

ANVISA. Dipirona monoidratada [Internet]. [cited 2016 Apr 18]. Available from: http://www.anvisa.gov.br/datavisa/fila_bula/frmVisualizarBula.asp?pNuTransacao=8986942013&pIdAnexo=1838911

ANVISA. Rosuvastatina Cálcica [Internet]. [cited 2016 Apr 18]. Available from: http://www.anvisa.gov.br/datavisa/fila_bula/frmVisualizarBula.asp?pNuTransacao=5776982014&pIdAnexo=2127095

ANVISA. Mesilato de Imanitibe [Internet]. [cited 2016 Apr 18]. Available from: http://portal.anvisa.gov.br/wps/wcm/connect/787ce200474589669361d73fbc4c6735/BM%5B26128-1-0%5D.pdf?MOD=AJPERES

ANVISA. Bissulfato de Clopidogrel [Internet]. [cited 2016 Apr 18]. Available from: http://www.anvisa.gov.br/datavisa/fila_bula/frmVisualizarBula.asp?pNuTransacao=10564832013&pIdAnexo=1903240

ANVISA. Citalopram [Internet]. [cited 2016 Apr 18]. Available from: http://portal.anvisa.gov.br/wps/wcm/connect/fe89fa804879a9c1afbfaffa35813921/Lundbeck_Brasil_Associacao_do_Cipramil_citalopram_com_prolongamento_de_QT_Dose_Dependente.pdf?MOD=AJPERES

ANVISA. Cetuximabe [Internet]. [cited 2016 Apr 18]. Available from: http://portal.anvisa.gov.br/wps/wcm/connect/855e0600469bec31aef6fe2e64280806/Erbitux_DHCP_05+Dez+2014.pdf?MOD=AJPERES

ANVISA. Valproato de Sódio [Internet]. [cited 2016 Apr 18]. Available from: http://www.anvisa.gov.br/datavisa/fila_bula/frmVisualizarBula.asp?pNuTransacao=401042014&pIdAnexo=1940350

ANVISA. Mercaptopurina [Internet]. [cited 2016 Apr 18]. Available from: http://www.anvisa.gov.br/datavisa/fila_bula/frmVisualizarBula.asp?pNuTransacao=4953992013&pIdAnexo=1662336

ANVISA. Azatioprina [Internet]. [cited 2016 Apr 18]. Available from: http://www.anvisa.gov.br/datavisa/fila_bula/frmVisualizarBula.asp?pNuTransacao=1015052015&pIdAnexo=2435974

ANVISA. Tioguanina [Internet]. [cited 2016 Apr 18]. Available from:

http://www.anvisa.gov.br/datavisa/fila_bula/frmVisualizarBula.asp?pNuTransacao=2493222013&pIdAnexo=1554582

ANVISA. Cloridrato de irinotecano tri-hidratado [Internet]. [cited 2016 Apr 18]. Available from: http://www.anvisa.gov.br/datavisa/fila_bula/frmVisualizarBula.asp?pNuTransacao=4775692015&pIdAnexo=2655379

Methodological details

Definition of predicted phenotypes

Carriers of two inactive *CYP2D6*, *CYP2C9, CYP2C19* or *CYP3A5* alleles or *CYP2C9*2/*2, *2/*3* or **3/*3* were classified as predicted poor metabolizers (gPMs)(1–5), while individuals with more than two *CYP2D6* active copies or *CYP2C19*1/*17* or **17/*17* genotypes were classified as predicted ultra-rapid metabolizers (gUMs)(3). For *CYP2D6, CYP2C9* and *CYP3A5*, genotypes containing one normal allele and one reduced or null activity allele were grouped as predicted intermediate metabolizers (gIMs). Individuals with two copies of normal activity CYP alleles were predicted as extensive metabolizers (gEMs). Individuals with two *NAT2* alleles related with decreased activity were classified as predicted slow acetilators (gSAs)(6).

Statistical analysis details

To compare the *color* categories and regions pairwise, χ^2 test was performed using *chisq.test* R function.

The Spearman correlation (for non-parametric data) was estimated between EPIGEN-Brazil cohorts allele frequencies and those estimated by the systematic review, using the R *cor.test (method="spearman")* function.

The Multivariate Distance Matrix Regression (MDMR) was performed using *mdmr* function from R MDMR-package applied to a (1-IBS)_ distance matrix (7).

MDMR is a two-step approach that consists of i) to compute the distance between all pairs of individuals with respect to Y variables of interest and construct a *NXN* matrix (D) whose elements reflect these distances; and ii) to test the association between predictors (X) and D by decomposing the sums of squares of the distance matrix into a portion attributable to regression to X and a portion due to residual. For this analysis, the genetic distance was based on IBS (Identical by State) allele sharing information across 5934 individuals from EPIGEN-Brazil initiative (Whites, Mixed and Blacks from Salvador, Bambuí and Pelotas cohorts) genotyped for 55 pharmacogenetics variants. From IBS information a (1-IBS) distance matrix (5934 X 5934) was constructed using PLINK software and the association between individuals *Color* and geographic region that they were from and the distance matrix was performed using *mdmr function* from R MDMR package (https://cran.r-project.org/web/packages/MDMR/vignettes/mdmr-vignette.html).

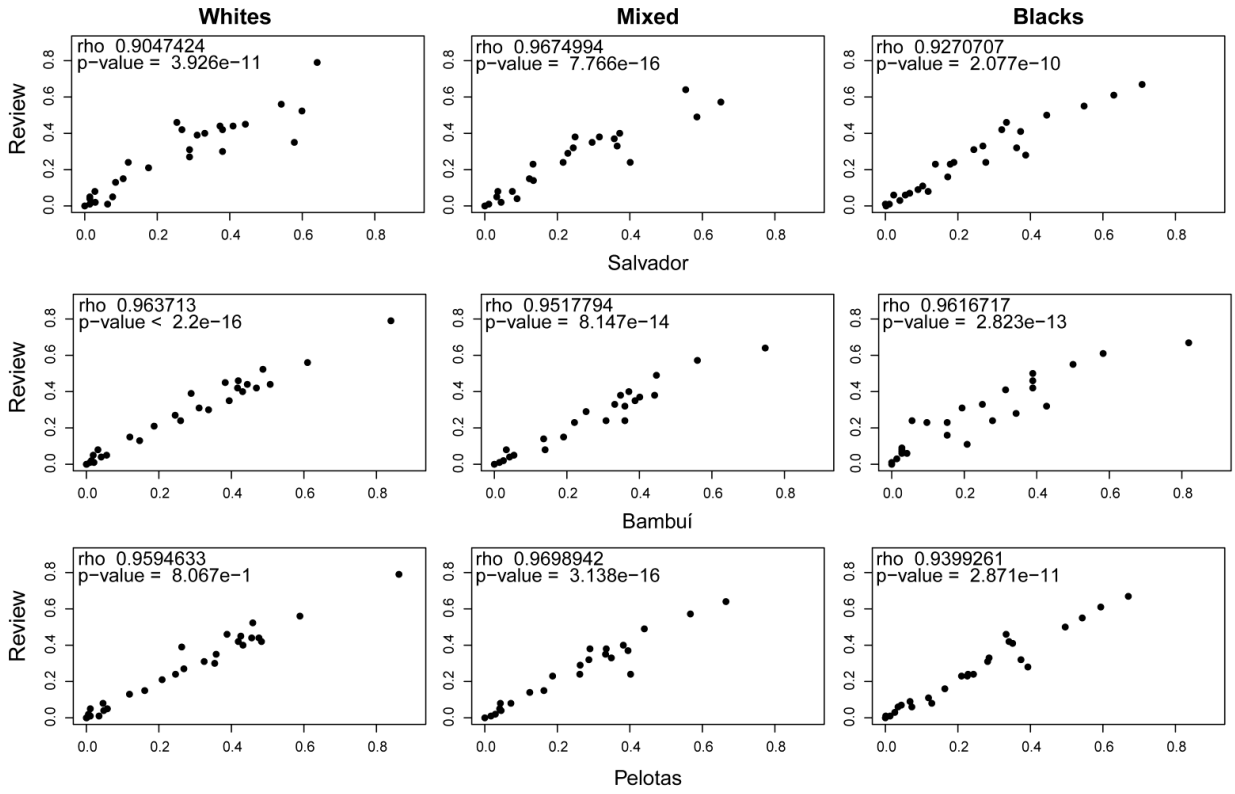

Figure S1. Scatterplot of allele frequencies of 28 pharmacogenetics variants estimated in the systematic review and in three Brazilian population-based cohorts, stratified by *Color* categories.

Table S1. Pharmacogenetic biomarkers searched in this study according to the European Medicines Agency, the Food and Drug Administration and the Pharmacogenomics Knowledgebase CPIC gene/drug pairs. Adapted from Fricke-Galindo et al (2016) (1). Biomarkers in bold are those with sufficient data and analyzed in this study. Ns = Number of studies.

Biomarker	Name	Ns	Biomarker	Name	Ns
ABCB1	***ATP binding cassette subfamily B member 1***	**4**	*FDPS*	*farnesyl diphosphate synthase*	0
ABCC1	*ATP binding cassette subfamily C member 1*	0	*FIP1L1-PDGFRA*	*FIP1L1/PDGFRA fusion gene*	0
ABCC2	*ATP binding cassette subfamily C member 2*	0	*FLOT1*	*flotillin 1*	0
ABCC4	*ATP binding cassette subfamily C member 4*	0	***G6PD****	***glucose-6-phosphate dehydrogenase***	**9**
*ABCG2**	*ATP binding cassette subfamily G member 2 (Junior blood group)*	0	*GGCX*	*gamma-glutamyl carboxylase*	0
ABL2	*ABL proto-oncogene 2, non-receptor tyrosine kinase*	0	***GNB3***	***G protein subunit beta 3***	**4**
ADD1	*adducin 1*	0	*GRIK4*	*glutamate ionotropic receptor kainate type subunit 4*	0
ADORA2A	*adenosine A2a receptor*	0	*GSTA1*	*glutathione S-transferase alpha 1*	0
ADRB1	***adrenoceptor beta 1***	**1**	***GSTM1***	***glutathione S-transferase mu 1***	**24**
ADRB2	***adrenoceptor beta 2***	**0**	***GSTP1***	***glutathione S-transferase pi 1***	**7**
ALK	*anaplastic lymphoma receptor tyrosine kinase*	0	***GSTT1***	***glutathione S-transferase theta 1***	**19**
ANKK1	*ankyrin repeat and kinase domain containing 1*	0	*HERG1*	*human ether-a-go-go related gene 1*	0
ASL	*argininosuccinate lyase*	0	***HLA-A***	***major histocompatibility complex, class I, A***	**3**
ASS1	*argininosuccinate synthase 1*	0	***HLA-B***	***major histocompatibility complex, class I, B***	**3**
ATIC	*5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase*	0	*HMGCR*	*3-hydroxy-3-methylglutaryl-CoA reductase*	0
*BCR-ABL**	*BCR/ABL fusion gene*	0	*HPRT1*	*hypoxanthine phosphoribosyltransferase 1*	0
*BRAF**	*B-Raf proto-oncogene, serine/threonine kinase*	0	*HTR2A*	*5-hydroxytryptamine receptor 2A*	0
BTG3	*BTG family member 3*	0	*HTR2C*	*5-hydroxytryptamine receptor 2C*	0
C11orf65	*chromosome 11 open reading frame 65*	0	*IFNL3*	*interferon, lambda 3*	0
CACNB2	*calcium voltage-gated channel auxiliary subunit beta 2*	0	*IL2RA*	*interleukin 2 receptor subunit alpha*	0
CALU	*Calumenin*	0	*ITPA*	*inosine triphosphatase*	0
CBR3	*carbonyl reductase 3*	0	*KRAS**	*kirsten rat sarcoma viral oncogene homolog*	0
CCR5	***C-C motif chemokine receptor 5 (gene/pseudogene)***	**7**	***LDLR***	***low density lipoprotein receptor***	**4**
CES1	*carboxylesterase 1*	0	*LTC4S*	*leukotriene C4 synthase*	0
CFTR*	***cystic fibrosis transmembrane conductance regulator***	**4**	*MS4A1*	*membrane spanning 4-domains A1*	0
CKIT	*KIT proto-oncogene receptor tyrosine kinase*	0	***MTHFR***	***methylenetetrahydrofolate reductase (NAD(P)H)***	**11**
COAGULATION FII	***coagulation factor II***	**1**	***MTRR***	***5-methyltetrahydrofolate-homocysteine methyltransferase reductase***	**5**

COAGULATION FV*	***coagulation factor V***	**2**	*NAGS*	*N-acetylglutamate synthase*	0
COL22A1	*collagen type XXII alpha 1*	0	*NAT1*	*N-acetyltransferase 1 (arylamine N-acetyltransferase)*	0
COMT	***catechol-O-methyltransferase***	**5**	***NAT2***	***N-acetyltransferase 2 (arylamine N-acetyltransferase)***	**3**
COQ2	*coenzyme Q2, polyprenyltransferase*	0	*OPRM1*	*opioid receptor, mu 1*	0
CPS1	*carbamoyl-phosphate synthase 1*	0	*OTC*	*ornithine carbamoyltransferase*	0
CRHR1	*corticotropin releasing hormone receptor 1*	0	*PDGFRA*	*platelet derived growth factor receptor alpha*	0
CRHR2	*corticotropin releasing hormone receptor 2*	0	*PDGFRB*	*platelet derived growth factor receptor beta*	0
CYB5R1	*cytochrome b5 reductase 1*	0	*PGR*	*progesterone receptor*	0
CYB5R2	*cytochrome b5 reductase 2*	0	*PLM-RARA*	*PML/RARA fusion gene*	0
CYB5R3	*cytochrome b5 reductase 3*	0	*POLG**	*polymerase (DNA) gamma, catalytic subunit*	0
CYB5R4	*cytochrome b5 reductase 4*	0	***PROTEIN C***	***protein C, inactivator of coagulation factors Va and VIIIa***	**1**
CYP1A1	***cytochrome P450 family 1 subfamily A member 1***	**3**	*RYR1*	*ryanodine receptor 1*	0
CYP1A2	***cytochrome P450 family 1 subfamily A member 2***	**2**	*SCN1A*	*sodium voltage-gated channel alpha subunit 1*	0
CYP2A7P1	*cytochrome P450 family 2 subfamily A member 7 pseudogene 1*	0	*SERPINC1*	*serpin peptidase inhibitor, clade C (antithrombin), member 1*	0
CYP2B6	***cytochrome P450 family 2 subfamily B member 6***	**1**	*SLC22A1*	*solute carrier family 22 member 1*	0
CYP2C19*	***cytochrome P450 family 2 subfamily C member 19***	**10**	*SLC22A2*	*solute carrier family 22 member 2*	0
CYP2C8	***cytochrome P450 family 2 subfamily C member 8***	**3**	*SLC22A6*	*solute carrier family 22 member 6*	0
CYP2C9	***cytochrome P450 family 2 subfamily C member 9***	**6**	***SLC6A4***	***solute carrier family 6 member 4***	**3**
CYP2D6	***cytochrome P450 family 2 subfamily D member 6***	**5**	***SLCO1B1****	***solute carrier organic anion transporter family member 1B1***	**2**
CYP2E1	***cytochrome P450 family 2 subfamily E member 1***	**5**	*SLCO2B1*	*solute carrier organic anion transporter family member 2B1*	0
CYP3A4	***cytochrome P450 family 3 subfamily A member 4***	**2**	***SOD2***	***superoxide dismutase 2, mitochondrial***	**4**
CYP3A5	***cytochrome P450 family 3 subfamily A member 5***	**4**	*SULT1A1*	*sulfotransferase family 1A member 1*	0
CYP4F2	*cytochrome P450 family 4 subfamily F member 2*	0	*TMEM43*	*transmembrane protein 43*	0
del (5q)	*chromosome 5q deletion*	0	*TNFRSF8*	*tumor necrosis factor receptor superfamily member 8*	0
DPYD	*dihydropyrimidine dehydrogenase*	0	***TP53****	***tumor protein p53***	**7**
DRD1	*dopamine receptor D1*	0	***TPMT****	***thiopurine S-methyltransferase***	**4**
DRD2	***dopamine receptor D2***	**1**	***UGT1A1****	***UDP glucuronosyltransferase family 1 member A1***	**3**
EGF	*epidermal growth factor*	0	*UGT1A4*	*UDP glucuronosyltransferase family 1 member A4*	0
*EGFR**	*epidermal growth factor receptor*	0	*UGT2B15*	*UDP glucuronosyltransferase family 2 member B15*	0
EPHX1	*epoxide hydrolase 1*	0	*UMPS*	*uridine monophosphate synthetase*	0
*ERBB2**	*erb-b2 receptor tyrosine kinase 2*	0	***VKORC1***	***vitamin K epoxide reductase complex subunit 1***	**5**
ERCC1	*excision repair cross-complementation group 1*	0	***XRCC1***	***X-ray repair complementing defective repair in Chinese hamster cells 1***	**3**

ESR1	*estrogen receptor 1*	0	*YEATS4*	*YEATS domain containing 4*	0
FCGR3A	*Fc fragment of IgG receptor IIIa*	0			

Biomarkers signed with * have recommendations for pharmacogenetics testing issued by Brazilian regulatory agencies.

Table S2. Pharmacogenetic alleles and rs correspondence following PharmGKB.

CYP2D6								
	****2***	rs1058164-G	rs16947-A	rs1135840-G				
	****3***	rs35742686-delT						
	****4***	rs1065852-A	rs28371703-T	rs28371704-C	rs28371705-C	rs1058164-G	rs3892097-T	rs1135840-G
	****9***	rs5030656-delCTT						
	****10***	rs1065852-A	rs1058164-G	rs1135840-G				
	****17***	rs28371706-A	rs1058164-G	rs16947-A	rs1135840-G			
	****29***	rs61736512-T	rs1058164-G	rs16947-A	rs59421388-T	rs1135840-G		
	****35***	rs769258-T	rs1058164-G	rs16947-A	rs1135840-G			
	****41***	rs28735595-C rs75276289-G	rs28624811-A rs28695233-G	rs28633410-T rs1081000-C	rs1080995-G rs1058164-G	rs1080996-T rs16947-A	rs74644586-C rs28371725-T	rs76312385-G rs1135840-G
CYP2C9								
	****2***	rs1799853-T						
	****3***	rs1057910-C						
	****5***	rs28371686-G						
	****6***	rs9332131-delA						
	****8***	rs7900194-A						
	****11***	rs28371685-T						
CYP2C19								
	****2***	rs4244285-A	rs12769205-R					
	****3***	rs4986893-A						
	****17***	rs12248560-T	rs11188072-T	rs3758581-G	rs17885098-T			
CYP2B6								
	****2***	rs8192709-T						
	****5***	rs3211371-T						
	****6***	rs3745274-T	rs2279343-G					
	****9***	rs3745274-T	rs8192719-T					
CYP3A5								
	****1***	rs55817950-G rs10264272-C	rs28383468-G rs28383479-C	rs28383469-del rs41303343-del	rs72552791-T rs28365083-G	rs776746-T rs41279854-A	rs56244447-A rs28365085-A	rs56411402-T rs15524-G
	****3***	rs776746-C	rs15524-A					
	****6***	rs10264272-T						
	****7***	rs41303343-insA						
CYP3A4								
	****1B***	rs2740574-C						
	****20***	rs67666821-insT						
CYP2E1								
	****1***	rs2070676-C	rs72559710-G	rs55897648-G	rs6413419-G	rs3813867-G	rs2031920-C	

	5B	rs3813867-C	rs2031920-T				
CYP1A1							
	2A	rs4646903-G					
	2C	rs1048943-C					
	3	rs1800031-G					
	4	rs1799814-T					
CYP1A2							
	1A	rs2069514-G rs138652540-C rs72547515-G rs45486893-C	rs35694136-T rs72547512-G rs72547516-A rs72547517-G	rs2069526-T rs72547513-C rs149928755-C rs143193369-C	rs12720461-C rs17861157-C rs55889066-G rs2470890-T	rs762551-C rs35796837-G rs56107638-G	rs56160784-C rs56276455-G rs28399424-C	rs72547511-C rs2472304-G rs144148965-G
	1C	rs2069514-A						
	1F	rs762551-A						
	1K	rs2069526-G	rs12720461-T	rs762551-A				
	1M	rs2069526-T	rs12720461-C	rs762551-A				
CYP2C8								
	2	rs11572103-A						
	3	rs11572080-T	rs10509681-C					
	4	rs1058930-C						
NAT2								
	5	rs1801280-C						
	6	rs1799930-A						
	7	rs1799931-A						
	11	rs1799929-T						
	12	rs1208-G						
	13	rs1041983-T						
	14	rs1801279-A						
TPMT								
	238G>C	rs1800462-G						
	460G>A	rs1800460-T						
	719A>G	rs1142345-C						
	146T>C	rs72552740-G						
	539A>T	rs75543815-A						
UGT1A1								
	28	rs8175347-[TA]7						
	36	rs8175347-[TA]5						
	37	rs8175347-[TA]8						
	60	rs4124874-G						
	93	rs10929302-A						
GSTP1								
	Val	rs1695-G						

ABCB1/MDR1					
129T>C	rs3213619-G				
1236C>T	rs1128503-G				
3435C>T	rs1045642-G				
2677G>T/A	rs2032582-C/T				
SLCO1B1					
388A>G	rs2306283-G				
463C>A	rs11045819-A				
521T>C	rs4149056-C				
89595T>C	rs4363657-C				
VKORC1					
H2	rs7196161-G rs17708472-G	rs17880887-G rs9934438-A	rs17881535-C rs8050894-G	rs9923231-T rs2359612-A	rs2884737-C rs7294-C
H7	rs7196161-A rs17708472-G	rs17880887-G rs9934438-G	rs17881535-C rs8050894-C	rs9923231-C rs2359612-G	rs2884737-A rs7294-T
H8	rs7196161-A rs17708472-G	rs17880887-T rs9934438-G	rs17881535-C rs8050894-C	rs9923231-C rs2359612-G	rs2884737-A rs7294-T
H9	rs7196161-A rs17708472-A	rs17880887-T rs9934438-G	rs17881535-G rs8050894-C	rs9923231-C rs2359612-G	rs2884737-A rs7294-C
3673G>A 1639G>A	rs9923231-A/G/T				
5723T>G	rs2884737-C				
6853G>C	rs8050894-G/T				
9104G>A	rs7294-T				
CFTR					
DF508	rs113993960-delCTT	rs199826652-delTCT			
ADRB2					
Arg16Gly	rs1042713-A				
Gln27Glu	rs1042714-C/T				
Thr164Ile	rs1800888-T				
CCR5					
D32	rs333-indel				
COMT					
Met	rs4680-A				
F2					
20210G>A	rs1799963-A				
F5					
Arg506Gln	rs6025-C				
GNB3					
825C>T	rs5443-T				
814G>A	rs5442-A				
MTHFR					

677C>T	rs1801133-A
1286A>C	rs1801131-G
1317T>C	rs4846051-A
MTRR	
66A>G	rs1801394-G
SOD2	
47T>C	rs4880-G
TP53	
BstUI*A1 (Arg72Pro)	rs1042522-C
R337H	rs9894946-T
XRCC1	
Arg194Trp	rs1799782-A
Gln399Arg	rs25487-C

Table S3. Allele frequencies of pharmacogenetic markers in EPIGEN-Brazil population-based cohorts (stratified by *Color*/Race). The Table is a Supplementary Table available as a .txt file in the Journal website.

Table S4. MDMR summary table. Estimate of the effect of Geographic regions and *Color* on the genetic background distances using 55 pharmacogenetics variants in three population based cohorts from EPIGEN Brazil initiative.

	Statistic	df	Pseudo R2	p-value
Omnibus Effect	0.05446	4	0.0516	< 1e-20
Geographic Region	0.00685	2	0.0065	< 1e-20
Color	0.02971	2	0.0282	< 1e-20

Note: Genetic distances are based on (1-IBS) allele-sharing information ((1-IBS) matrix) across 5,934 individuals from EPIGEN Brazil initiative.

Omnibus effect, effect of all predictors (Geographic region and *Color*) on distance matrix; df, numerator degrees of freedom for the corresponding effect; Pseudo R2, size of the corresponding effect on the distance matrix.

Table S5. *CYP2D6, CYP2C9* and *CYP2C19* allele and phenotype frequencies in Brazilian populations. Bold values represent significant interethnic difference (p<0.01). **n** = number of individuals tested for each allele, **f** = frequency. Wild type alleles are not presented in this table.

		***CYP2D6* alleles**											
Population		****2***	****3∅***	****4∅***	****5∅***	****9↓***	****10↓***	****17↓***	****29↓***	****35***	****41↓***	**Multiplication↑**	**References**
White	**n**	431	342	546		342	342	342	342	342	89		(8–10)
	f	0.24	0.01	**0.16**		0.01	0.15	**0.02**	**0.01**	0.04	0.08		
Mixed	**n**	440	350	440		350	350	350	350	350	90		(8,10)
	f	0.21	0.01	0.12		0.01	0.12	0.06	0.04	0.03	0.02		
Black	**n**	431	342	627		342	342	342	342	342	89		(8–10)
	f	0.17	0.01	0.10		0.01	0.12	**0.11**	0.06	0.02	0.01		
Diverse[1]	**n**	1020	1020	1020	1020	1020	1020	1020	1020	1020	1020	1020	(11)
	f	0.21	0.01	0.09	0.05	0.01	**0.02**	0.06	0.03	0.03	0.05	0.01	

		***CYP2D6* predicted phenotype**				
Population		**gPM**	**gIM**	**gEM**	**gUM**	**References**
White	**n**	327	327	327	327	(11)
	f	0.03	0.05	0.88	0.04	
Mixed	**n**	336	336	336	336	(11)
	f	0.03	0.06	0.88	0.03	
Black	**n**	324	324	324	324	(11)
	f	0.02	0.10	0.83	0.05	

		***CYP2D6* measured phenotype**				
Population		**mPM**	**mIM**	**mEM**	**mUM**	**References**
White	**n**	8	8	8	8	(12)
	f	0.00	0.00	1.00	0.00	

		***CYP2C9* alleles**						
Population		****2↓***	****3↓***	****5↓***	****6∅***	****8↓***	****11↓***	**References**
White	**n**	1086	1086	342			342	(1,8,13)

		***CYP2C9* predicted phenotype**			
Population		**gPM**	**gIM**	**gEM**	**References**
White	**n**	744	744	744	(1,13)

	f	**0.12**	0.05	0			0.01			**f**	0.03	**0.27**	**0.70**	
Mixed	**n**	781	781	594			594	(1,8,13)	**Mixed**	**n**	431	431	431	(1,13)
	f	0.09	0.05	0			0			**f**	0.02	0.22	0.76	
Black	**n**	536	536	342			342	(1,8,13)	**Black**	**n**	194	194	194	(1,13)
	f	0.07	0.03	0.01			0.01			**f**	0.03	0.15	0.82	
Native American	**n**	312	312	90	90	90	90	(1,14)	**Native American**	**n**	222	222	222	(1)
	f	**0.03**	0.01	0	0	0.04	0			**f**	0.00	0.10	0.90	
Asian descendants	**n**	64	64					(15)	**Asian descendants**	**n**	64	64	64	(15)
	f	**0.00**	0.04							**f**	0.00	0.02	0.98	

***CYP2C19* alleles**					
Population		****2∅***	****3∅***	****17↑***	**References**
White	**n**	1058	434	966	(8,16–18)
	f	0.15	0	0.18	
Mixed	**n**	665	350	665	(8,18)
	f	0.14	0	0.19	
Black	**n**	528	429	441	(8,16,18)
	f	0.16	0	0.20	
Native American	**n**	272	90	273	(14,18)
	f	0.11	0	**0.11**	
Asian descendants	**n**	59	59		(15)
	f	**0.27**	**0.14**		
Not informed	**n**			9	(12,19,20)
	f			0.39	

***CYP2C19* predicted phenotype**					
Population		**gPM**	**gEM**	**gUM**	**References**
White	**n**	101	101	101	(16,17)
	f	**0.01**	**0.94**	0.05	
Black	**n**	87	87	87	(16)
	f	**0.03**	**0.97**	0	
Asian descendants	**n**	59	59	59	(15)
	f	**0.22**	**0.78**	0	

***CYP2C19* measured phenotype**					
Population		**mPM**	**mEM**	**mUM**	**References**
White	**n**	9	9	9	(17)
	f	0	1	0	
Not informed	**n**	28	28	28	(12,19,20)
	f	0.07	0.93	0	

↓ Decreased activity allele
↑ Increased activity allele
∅ Null activity allele
Alleles without symbols are normal activity
[1] Diverse refers to more than one ethnicities together, not separated by the original articles

Table S6. Other *CYP* allele (and phenotype when available) frequencies in Brazilian populations. **n** = number of individuals tested for each allele, **f** = frequency. Wild type alleles are not presented in this table.

***CYP2B6* alleles**						
Population		****2***	****5***	****6↓***	****9***	**References**
White	**n**	342	342	342	342	(8)
	f	0.07	**0.21**	0.30	**0.39**	
Mixed	**n**	350	350	350	350	(8)
	f	0.10	0.14	0.30	**0.33**	
Black	**n**	342	342	342	342	(8)
	f	0.08	0.13	**0.37**	**0.46**	

***CYP3A5* alleles**					
Population	**n**	****3↓***	****6∅***	****7↓***	**References**
White	**n**	840	742	650	(5,8,16,21)
	f	**0.79**	**0.01**	**0.01**	
Mixed	**n**	757	646	646	(5,8,21)
	f	**0.64**	**0.04**	**0.03**	
Black	**n**	833	734	647	(5,8,16,21)
	f	**0.50**	**0.09**	**0.07**	

***CYP3A5* predicted phenotype**					
Population	***n***	**gPM**	**gIM**	**gEM**	**References**
White	**n**	498	498	498	(5,16,21)

	f	**0.66**	**0.30**	**0.04**	
Mixed	**n**	703	703	703	(5,21)
	f	**0.44**	0.46	**0.10**	
Black	**n**	491	491	491	(5,16,21)
	f	**0.36**	0.45	**0.20**	

***CYP3A4* alleles**				
Population	**n**	****1B***	****20∅***	**References**
White	**n**	92	92	(16)
	f	0.10	0.01	
Black	**n**	87	87	(16)
	f	**0.38**	0.01	
Not informed	**n**	201	87	(22)
	f	0.14	0.01	

***CYP2E1* alleles**				
Population	**n**	****5B***	****6***	**References**
White	**n**	472	151	(9,23,24)
	f	0.08	0.08	
Mixed	**n**	149	63	(23,24)
	f	0.08	0.08	
Black	**n**	271	75	(24)
	f	0.06	0.07	
Native American	**n**	190		(25)
	f	0.19		
Admixed	**n**	114	114	(26)
	f	0.16	0.10	

***CYP1A1* alleles**						
Population	**n**	****2A***	****2C***	****3***	****4***	**References**
White	**n**	115				(9)
	f	0.15				
Black	**n**	196				(9)
	f	**0.30**				
Native American	**n**	190	190			(25)
	f	0.11	0.21			
Not informed	**n**		189	195	190	(27)
	f		0.16	0.00	0.04	

***CYP1A2* alleles**						
Population	**n**	****1C↓***	****1F↑***	****1K↓***	****1M***	**References**
White	**n**		92	92		(16)
	f		0.68	0.00		
Black	**n**		87	87		(16)
	f		0.64	0.01		
Not informed	**n**	196	190		186	(27)
	f	0.17	0.28		0.50	

***CYP2C8* alleles**					
Population	**n**	****2.∴***	****3.∴***	****4.∴***	**References**
White	**n**	342	342	342	(8)
	f	0.03	**0.13**	0.04	
Mixed	**n**	350	350	350	(8)
	f	**0.06**	0.08	0.04	
Black	**n**	342	342	342	(8)
	f	**0.11**	0.06	0.02	
Native American	**n**	90	90	90	(14)
	f	0.01	**0.00**	0.00	

↓ Decreased activity allele
↑ Increased activity allele
∅ Null activity allele
∴ Decreased in vitro activity
Alleles without symbols are of normal activity

Table S7. Phase II biomarkers frequencies (allele and predicted phenotype, when available). **n** = number of individuals tested for each allele, **f** = frequency.

		***NAT2* alleles**									
Population		****4***	****5↓***	****6↓***	****7↓***	****11***	****12***	****13***	****14↓***	***rs1495741***	**References**
White	**n**		136	164	164	164	164	136	164	136	(28,29)
	f		0.45	0.27	0.06	0.42	0.44	0.33	0.02	0.20	
Mixed	**n**		107	107	107	107	107	107	107	105	(28)
	f		0.40	0.29	0.03	0.36	0.46	0.39	0.02	0.26	
Black	**n**		118	243	243	242	243	118	243	101	(28,29)
	f		0.32	0.24	0.12	0.36	**0.32**	0.41	**0.07**	0.36	
Native American	**n**	117	205	235	235	118	201	88	118	81	(28–30)
	f	0.34	0.39	**0.08**	0.23	0.49	0.26	**0.21**	0.01	0.38	

		***NAT2* – predicted phenotype**			
Population		**Slow**	**Intermediate**	**Rapid**	**References**
White	**n**	136	136	136	(28)
	f	0.60	0.35	0.05	
Mixed	**n**	107	107	107	(28)
	f	0.55	0.38	0.07	
Black	**n**	118	118	118	(28)
	f	0.47	0.45	0.08	
Native American	**n**	88	88	88	(28)
	f	0.51	0.44	0.05	

		TPMT					
Population		**238G>C↓**	**460G>A↓**	**719A>G↓**	**146T>C↓**	**539A>T↓**	**References**
White	**n**	423	423	423			(8,31)
	f	0.002	0.01	0.01			
Mixed	**n**	350	350	350			(8)
	f	0.01	0.01	0.04			
Black	**n**	342	342	342			(8)
	f	0.01	0.01	0.02			
Diverse[1]	**n**	200	**200**	200			(32)
	f	0.01	**0.05**	0.04			
Not informed	**n**	406	406	406	202	202	(31,33)
	f	0.01	0.02	0.01	0.00	0.00	

		UGT1A1					
Population		****28↓***	****36↑***	****37↓***	****60***	****93***	**References**
White	**n**	71	71	71	89	89	(10,34)
	f	0.32	0.01	0.01	0.52	0.22	
Mixed	**n**				90	90	(10)
	f				0.57	0.21	
Black	**n**	54	54	54	89	89	(10,34)
	f	0.41	0.06	0.01	0.67	**0.11**	
Native American	**n**	32	32	32			(34)
	f	0.33	0	0			
Not informed	**n**	177					(34)
	f	0.25					

		GSTM1				
Population		***Present***	***Deleted∅***	****A***	****B***	**References**
White	**n**	1796	1796	220	220	(9,23,35–40)
	f	0.47	0.53	0.59	0.41	
Mixed	**n**	427	427	71	71	(23,36,37,40)

	f	0.56	0.44	0.66	0.34	
Black	**n**	1140	1140	62	62	(9,35–38,40–42)
	f	0.66	0.34	0.73	0.27	
Native American	**n**	360	360			(25,35,38,42)
	f	0.75	**0.25**			
River-dwellers	**n**	443	443			(43,44)
	f	0.64	0.36			
Not Informed	**n**	2000	2000			(26,27,35,45–53)
	f	0.58	0.40			
Diverse[1]	**n**	1571	1571			(39,41,42)
	f	0.49	0.51			

*GSTM1 *A* and **B* frequencies are given considering only individuals with "Present" genotypes.

GSTT1

Population		***Present***	***Deleted∅***	**References**
White	**n**	1161	1161	(9,35,36,38–40)
	f	0.79	0.21	
Mixed	**n**	343	343	(36,40)
	f	0.77	0.23	
Black	**n**	973	973	(9,35,36,38,40,41)
	f	0.74	0.26	
Native American	**n**	360	360	(25,35,38,42)
	f	0.81	0.19	
River-dwellers	**n**	443	443	(43,44)
	f	0.60	**0.40**	
Not Informed	**n**	1666	1666	(26,27,35,45,47–49,51,52)
	f	0.80	0.20	
Diverse[1]	**n**	1571	1571	(39,41,42)
	f	0.75	0.25	

GSTP1

Population		***Ile***	***Val↓***	**References**
White	**n**	147	147	(9,38)
	f	0.69	0.31	
Black	**n**	336	336	(9,38)
	f	0.62	0.38	
Native American	**n**	238	238	(38,54)
	f	0.60	0.40	
River-dwellers	**n**	409	409	(43,44)
	f	0.57	0.43	
Not Informed	**n**	407	407	(45,47,54)
	f	0.72	0.28	

↓ Decreased activity allele
↑ Increased activity allele
∅ Null activity allele
∴ Decreased in vitro activity
Alleles without symbols are of normal activity
[1] Diverse refers to more than one ethnicity together, not separated by the original articles

Table S9 Allele frequencies by Brazilian regions. Only the frequencies that vary significantly among regions are shown.

Genes/Regions	North (N)	Northeast (NE)	Center West (CW)	Southeast (SE)	South (S)	Significant differences	References
ABCB1							
2677 G>T/A	0.376	0.356		0.286	0.264	N-SE, N-S, NE-S	(18,55,60)
3435 C>T	0.447				0.357	N-S	
CYP1A1							(9,25,27)
***2A**	0.808				0.246	N-S	
CYP2C9							
gEM				0.792	0.984	SE-S	(1,13,15)
gIM				0.192	0.016	SE-S	
CYP2C19							(15,16,18)
***2**				0.104	0.183	SE-S	
CYP2D6							(10,11)
***41**		0.079		0.044		NE-SE	
GNB3							(75,107,108)
C825T				0.482	0.334	SE -S	
GSTP1							(9,38,43–45,47,54)
Val	0.401				0.318	N-S	
CYP2E1						CW-S, CW-SE, CW-N	(9,23–25)
***5B**	0.019		0.216	0.078	0.069		
CYP3A5							(5,16,21)
***3**	0.686			0.611		N-SE	
***7**	0.010			0.048	0.049	N-S, N-SE	
GSTM1						CW-NE, CW-SE, CW-N, SE-NE, SE-N, SE-S, S-NE	(9,25,27,35,36,38–51)
Deletion	0.663	0.687	0.559	0.511	0.573		
GSTT1						CW-SE, CW-S, NE-SE, NE-S, N-SE, N-S	(9,25,27,35,36,38–45,47–49,51)
Deletion	0.339	0.248	0.257	0.215	0.196		
HLA-A							(63,65)
***0201**	0.271		0.109	0.241		CW-N, CW-SE	
MTHFR							(48,76,77,79,82,89–92,123)
C677T	0.407	0.251		0.320		N-NE, SE-NE	
SLCO1B1							(55)
388A>G	0.524			0.617	0.628	N-SE, N-S	

Table S10. African, European and Native American ancestry estimated in three Brazilian population-based cohorts from EPIGEN-Brazil initiative, stratified by *color* categories.

	Salvador (NE)								
	Whites (N=71)			Mixed (N=581)			Blacks (N=238)		
Ancestry	**African**	**European**	**Native American**	**African**	**European**	**Native American**	**African**	**European**	**Native American**
Average	0.360	0.570	0.070	0.482	0.456	0.061	0.600	0.348	0.052
Median	0.352	0.567	0.063	0.490	0.451	0.056	0.607	0.341	0.051
IQR	0.168	0.170	0.050	0.175	0.168	0.037	0.180	0.153	0.035
	Bambuí (SE)								
	Whites (N=874)			**Mixed (N=532)**			**Blacks (N=36)**		
Ancestry	**African**	**European**	**Native American**	**African**	**European**	**Native American**	**African**	**European**	**Native American**
Average	0.087	0.863	0.051	0.194	0.730	0.076	0.735	0.222	0.043
Median	0.068	0.879	0.043	0.158	0.760	0.073	0.764	0.187	0.038
IQR	0.079	0.121	0.050	0.174	0.184	0.053	0.295	0.227	0.050
	Pelotas (S)								
	Whites (N=2794)			**Mixed (N=209)**			**Blacks (N=602)**		
Ancestry	**African**	**European**	**Native American**	**African**	**European**	**Native American**	**African**	**European**	**Native American**
Average	0.068	0.864	0.068	0.276	0.627	0.097	0.514	0.407	0.079
Median	0.052	0.879	0.060	0.264	0.627	0.086	0.532	0.383	0.072
IQR	0.049	0.098	0.056	0.213	0.222	0.071	0.227	0.201	0.058

References

1. Soares RAG, Santos PCJL, Machado-Coelho GLL, do Nascimento RM, Mill JG, Krieger JE, et al. CYP2C9 and VKORC1 polymorphisms are differently distributed in the Brazilian population according to self-declared ethnicity or genetic ancestry. Genet Test Mol Biomarkers. 2012;16(8):957–63.
2. Céspedes-Garro C, Fricke-Galindo I, Naranjo M-EG, Rodrigues-Soares F, Fariñas H, de Andrés F, et al. Worldwide interethnic variability and geographical distribution of CYP2C9 genotypes and phenotypes. Expert Opin Drug Metab Toxicol. 2015 Nov 23;11(12):1893–905.
3. Fricke-Galindo I, Céspedes-Garro C, Rodríguez-Soares F, Naranjo MEG, Delgado Á, de Andrés F, et al. Interethnic variation of CYP2C19 alleles, "predicted" phenotypes and "measured" metabolic phenotypes across world populations. Pharmacogenomics J. 2016 Oct 27;16(2):113–23.
4. LLerena A, Naranjo MEG, Rodrigues-Soares F, Penas-LLedó EM, Fariñas H, Tarazona-Santos E. Interethnic variability of CYP2D6 alleles and of predicted and measured metabolic phenotypes across world populations. Expert Opin Drug Metab Toxicol. 2014 Nov;10(11):1569–83.
5. Suarez-Kurtz G, Vargens DD, Santoro AB, Hutz MH, de Moraes ME, Pena SDJ, et al. Global pharmacogenomics: distribution of CYP3A5 polymorphisms and phenotypes in the Brazilian population. PLoS One. 2014 Jan;9(1):e83472.
6. Sabbagh A, Langaney A, Darlu P, Gérard N, Krishnamoorthy R, Poloni ES. Worldwide distribution of NAT2 diversity: implications for NAT2 evolutionary history. BMC Genet.

2008 Jan;9:21.

7. Zapala MA, Schork NJ. Statistical properties of multivariate distance matrix regression for high-dimensional data analysis. Front Genet. 2012;3:190.

8. Suarez-Kurtz G, Pena SDJ, Struchiner CJ, Hutz MH. Pharmacogenomic Diversity among Brazilians: Influence of Ancestry, Self-Reported Color, and Geographical Origin. Front Pharmacol. 2012;3:191.

9. Maciel ME, Oliveira FK, Propst GB, da Graça Bicalho M, Cavalli IJ, Ribeiro EM de SF. Population analysis of xenobiotic metabolizing genes in South Brazilian Euro and Afro-descendants. Genet Mol Biol. 2009 Oct;32(4):723–8.

10. Bonifaz-Peña V, Contreras A V., Struchiner CJ, Roela RA, Furuya-Mazzotti TK, Chammas R, et al. Exploring the Distribution of Genetic Markers of Pharmacogenomics Relevance in Brazilian and Mexican Populations. Gong Y, editor. PLoS One. 2014;9(11):e112640.

11. Friedrich DC, Genro JP, Sortica VA, Suarez-Kurtz G, de Moraes ME, Pena SDJ, et al. Distribution of CYP2D6 alleles and phenotypes in the Brazilian population. PLoS One. 2014;9(10):e110691.

12. Rocha A, Coelho EB, Sampaio SA, Lanchote VL. Omeprazole preferentially inhibits the metabolism of (+)-(S)-citalopram in healthy volunteers. Br J Clin Pharmacol. 2010 Feb 24;70(1):43–51.

13. Vianna-Jorge R, Perini JA, Rondinelli E, Suarez-Kurtz G. CYP2C9 genotypes and the pharmacokinetics of tenoxicam in Brazilians. Clin Pharmacol Ther. 2004 Jul;76(1):18–26.

14. Vargens DD, Petzl-Erler M-L, Suarez-Kurtz G. Distribution of CYP2C polymorphisms

in an Amerindian population of Brazil. Basic Clin Pharmacol Toxicol. 2012 Apr;110(4):396–400.

15. Perini JA, Vargens DD, Santana ISC, Moriguchi EH, Ribeiro-Dos-Santos AKC, Tsutsumi M, et al. Pharmacogenetic polymorphisms in Brazilian-born, first-generation Japanese descendants. Braz J Med Biol Res. 2009 Dec;42(12):1179–84.
16. Kohlrausch FB, Carracedo Á, Hutz MH. Characterization of CYP1A2, CYP2C19, CYP3A4 and CYP3A5 polymorphisms in South Brazilians. Mol Biol Rep. 2014 Mar;41(3):1453–60.
17. Rocha A, Coelho EB, Moussa SAP, Lanchote VL. Investigation of the in vivo activity of CYP3A in Brazilian volunteers: comparison of midazolam and omeprazole as drug markers. Eur J Clin Pharmacol. 2008 Sep;64(9):901–6.
18. Santos PCJL, Soares RAG, Santos DBG, Nascimento RM, Coelho GLLM, Nicolau JC, et al. CYP2C19 and ABCB1 gene polymorphisms are differently distributed according to ethnicity in the Brazilian general population. BMC Med Genet. 2011 Jan;12:13.
19. Cassiano NM, Oliveira R V., Bernasconi GCR, Cass QB. Enantioselective disposition of omeprazole, pantoprazole, and lansoprazole in a same Brazilian subjects group. Chirality. 2012 Apr 17;24(4):289–93.
20. Cass QB, Lima V V, Oliveira R V, Cassiano NM, Degani ALG, Pedrazzoli J. Enantiomeric determination of the plasma levels of omeprazole by direct plasma injection using high-performance liquid chromatography with achiral-chiral column-switching. J Chromatogr B Analyt Technol Biomed Life Sci. 2003 Dec 25;798(2):275–81.
21. Suarez-Kurtz G, Perini JA, Bastos-Rodrigues L, Pena SDJ, Struchiner C. Impact of

population admixture on the distribution of the CYP3A5*3 polymorphism. Pharmacogenomics. 2007 Oct;8(10):1299–306.

22. Monteiro SCM, de Sousa IH, Belfort IKP, Nunes JD, Penha BAS, dos Santos M, et al. Genetic variability of CYP3A4 in a heterogeneous Brazilian population from Maranhão. Genet Mol Res. 2016 Feb 19;15(1).

23. Gattás GJ, Soares-Vieira JA. Cytochrome P450-2E1 and glutathione S-transferase mu polymorphisms among Caucasians and mulattoes from Brazil. Occup Med (Lond). 2000 Sep;50(7):508–11.

24. Rossini A, Lima SS, Rapozo DCM, Faria M, Albano RM, Pinto LFR. CYP2A6 and CYP2E1 polymorphisms in a Brazilian population living in Rio de Janeiro. Braz J Med Biol Res. 2006;39(2):195–201.

25. Gaspar PA, Hutz MH, Salzano FM, Hill K, Hurtado AM, Petzl-Erler ML, et al. Polymorphisms ofCYP1a1,CYP2e1,GSTM1,GSTT1, andTP53 genes in Amerindians. Am J Phys Anthropol. 2002 Nov 3;119(3):249–56.

26. Mitri S, Fonseca ASA, Otero UB, Tabalipa MM, Moreira JC, Sarcinelli P de N. Metabolic Polymorphisms and Clinical Findings Related to Benzene Poisoning Detected in Exposed Brazilian Gas-Station Workers. Int J Environ Res Public Health. 2015 Jul 21;12(7):8434–47.

27. Abnet CC, Fagundes RB, Strickland PT, Kamangar F, Roth MJ, Taylor PR, et al. The influence of genetic polymorphisms in Ahr, CYP1A1, CYP1A2, CYP1B1, GST M1, GST T1 and UGT1A1 on urine 1-hydroxypyrene glucuronide concentrations in healthy subjects from Rio Grande do Sul, Brazil. Carcinogenesis. 2007 Jan;28(1):112–7.

28. Suarez-Kurtz G, Sortica VA, Vargens DD, Bruxel EM, Petzl-Erler M-L, Petz-Erler M-L,

et al. Impact of population diversity on the prediction of 7-SNP NAT2 phenotypes using the tagSNP rs1495741 or paired SNPs. Pharmacogenet Genomics. 2012 Apr;22(4):305–9.

29. Talbot J, Magno LA V, Santana CVN, Sousa SMB, Melo PRS, Correa RX, et al. Interethnic diversity of NAT2 polymorphisms in Brazilian admixed populations. BMC Genet. 2010 Jan;11:87.

30. Bisso-Machado R, Ramallo V, Paixão-Côrtes VR, Acuña-Alonzo V, Demarchi DA, Sandoval JRS, et al. NAT2 gene diversity and its evolutionary trajectory in the Americas. Pharmacogenomics J. 2016 Nov 27;16(6):559–65.

31. Reis M, Santoro A, Suarez-Kurtz G. Thiopurine methyltransferase phenotypes and genotypes in Brazilians. Pharmacogenetics. 2003 Jun;13(6):371–3.

32. Gastal GR, Moreira S, Noble CF, Ferreira LE, França PHC de, Pinho M. Toxicity of azathioprine: why and when? analysis of the prevalence of polymorphism in Joinville, SC, Brazil. Arq Gastroenterol. 2012 Jan;49(2):130–4.

33. Boson WL, Romano-Silva MA, Correa H, Falcão RP, Teixeira-Vidigal P V, De Marco L. Thiopurine methyltransferase polymorphisms in a Brazilian population. Pharmacogenomics J. 2003 Jan;3(3):178–82.

34. Fertrin KY, Gonçalves MS, Saad STO, Costa FF. Frequencies of UDP-glucuronosyltransferase 1 (UGT1A1) gene promoter polymorphisms among distinct ethnic groups from Brazil. Am J Med Genet. 2002 Mar 1;108(2):117–9.

35. Arruda VR, Grignolli CE, Gonçalves MS, Soares MC, Menezes R, Saad ST, et al. Prevalence of homozygosity for the deleted alleles of glutathione S-transferase mu (GSTM1) and theta (GSTT1) among distinct ethnic groups from Brazil: relevance to

environmental carcinogenesis? Clin Genet. 1998 Sep;54(3):210–4.

36. Gattás GJF, Kato M, Soares-Vieira JA, Siraque MS, Kohler P, Gomes L, et al. Ethnicity and glutathione S-transferase (GSTM1/GSTT1) polymorphisms in a Brazilian population. Braz J Med Biol Res. 2004 Apr;37(4):451–8.

37. Hatagima A, Marques CFS, Krieger H, Feitosa MF. Glutathione S-transferase M1 (GSTM1) and T1 (GSTT1) polymorphisms in a Brazilian mixed population. Hum Biol. 2004 Dec;76(6):937–42.

38. Magno LA V, Talbot J, Talbot T, Borges Santos AM, Souza RP, Marin LJ, et al. Glutathione s-transferase variants in a brazilian population. Pharmacology. 2009 Jan;83(4):231–6.

39. Schreiber R, Mill JG, Krieger JE, Pereira AC, Nadruz W. Association between glutathione S-transferase M1 polymorphism and urinary sodium excretion in a Brazilian population. Am J Hypertens. 2013 Aug;26(8):1024–9.

40. Suarez-Kurtz G, Vargens DD, Struchiner CJ, Bastos-Rodrigues L, Pena SDJ. Self-reported skin color, genomic ancestry and the distribution of GST polymorphisms. Pharmacogenet Genomics. 2007 Sep;17(9):765–71.

41. Hiragi C de O, de Oliveira SF, Hatagima A, Ferreira LB, Grisolia CK, Klautau-Guimarães M de N. Glutathione S-transferase M1 and T1 polymorphisms in Brazilian African descendants. Hum Biol. 2007 Feb;79(1):131–40.

42. de Oliveira Hiragi C, Miranda-Vilela AL, Rocha DMS, de Oliveira SF, Hatagima A, de Nazaré Klautau-Guimarães M. Superoxide dismutase, catalase, glutathione peroxidase and gluthatione S-transferases M1 and T1 gene polymorphisms in three Brazilian population groups. Genet Mol Biol. 2011 Jan;34(1):11–8.

43. Barcelos GRM, Grotto D, de Marco KC, Valentini J, Lengert A van H, de Oliveira AÁS, et al. Polymorphisms in glutathione-related genes modify mercury concentrations and antioxidant status in subjects environmentally exposed to methylmercury. Sci Total Environ. 2013 Oct 1;463–464:319–25.

44. de Oliveira AÁS, de Souza MF, Lengert A van H, de Oliveira MT, Camargo RB de OG, Braga GÚL, et al. Genetic polymorphisms in glutathione (GSH-) related genes affect the plasmatic Hg/whole blood Hg partitioning and the distribution between inorganic and methylmercury levels in plasma collected from a fish-eating population. Biomed Res Int. 2014 Jan;2014:940952.

45. Da Silva FR, Da Silva J, Allgayer M da C, Simon CF, Dias JF, dos Santos CEI, et al. Genotoxic biomonitoring of tobacco farmers: Biomarkers of exposure, of early biological effects and of susceptibility. J Hazard Mater. 2012 Jul 30;225–226:81–90.

46. Lima Sombra CM, Coêlho Cavalcanti B, de Moraes MO, Santos S, Ribeiro-Dos-Santos A, Rodríguez Burbano R, et al. Genetic biomonitoring of inhabitants exposed to uranium in the north region of Brazil. Ecotoxicol Environ Saf. 2011 Jul;74(5):1402–7.

47. da Silva J, Moraes CR, Heuser VD, Andrade VM, Silva FR, Kvitko K, et al. Evaluation of genetic damage in a Brazilian population occupationally exposed to pesticides and its correlation with polymorphisms in metabolizing genes. Mutagenesis. 2008 Sep;23(5):415–22.

48. Miranda-Vilela AL, Akimoto AK, Alves PCZ, Ferreira LB, Lordelo GS, Melo JGM, et al. Evidence for an association between haptoglobin and MnSOD (Val9Ala) gene polymorphisms in essential hypertension based on a Brazilian case-control study. Genet Mol Res. 2010 Jan;9(4):2166–75.

49. Kato M, Loomis D, Brooks LM, Gattas GFJ, Gomes L, Carvalho AB, et al. Urinary biomarkers in charcoal workers exposed to wood smoke in Bahia State, Brazil. Cancer Epidemiol Biomarkers Prev. 2004 Jun;13(6):1005–12.

50. Grégio D'Arce LP, Cólus IM. Cytogenetic and molecular biomonitoring of agricultural workers exposed to pesticides in Brazil. Teratog Carcinog Mutagen. 2000 Jan;20(3):161–70.

51. Godoy FR, Costa EOA, da Silva Reis AA, Batista MP, de Melo AV, Gonçalves MW, et al. Do GSTT1 and GSTM1 polymorphisms influence intoxication events in individuals occupationally exposed to pesticides? Environ Sci Pollut Res Int. 2014 Mar;21(5):3706–12.

52. Franco FC, Alves AA, Godoy FR, Avelar JB, Rodrigues DD, Pedroso TMA, et al. Evaluating genotoxic risks in Brazilian public health agents occupationally exposed to pesticides: a multi-biomarker approach. Environ Sci Pollut Res Int. 2016 Oct 13;23(19):19723–34.

53. Rocha A V, Rita Cardoso B, Zavarize B, Almondes K, Bordon I, Hare DJ, et al. GPX1 Pro198Leu polymorphism and GSTM1 deletion do not affect selenium and mercury status in mildly exposed Amazonian women in an urban population. Sci Total Environ. 2016 Nov 15;571:801–8.

54. Lima Junior MM de, Reis LO, Ferreira U, Cardoso UO, Barbieri RB, Mendonça GB de, et al. Unraveling Brazilian Indian population prostate good health: clinical, anthropometric and genetic features. Int Braz J Urol. 2015 Jan;41(2):344–52.

55. Sortica V de A, Ojopi EB, Genro JP, Callegari-Jacques S, Ribeiro-Dos-Santos A, de Moraes MO, et al. Influence of genomic ancestry on the distribution of SLCO1B1,

SLCO1B3 and ABCB1 gene polymorphisms among Brazilians. Basic Clin Pharmacol Toxicol. 2012 May;110(5):460–8.

56. Bisso-Machado R, Ramallo V, Tarazona-Santos E, Salzano FM, Bortolini MC, Hünemeier T. Brief communication: 5-HTTLPR genetic diversity and mode of subsistence in Native Americans. Am J Phys Anthropol. 2013 Jul;151(3):492–4.

57. Lage GM, Malloy-Diniz LF, Matos LO, Bastos MAR, Abrantes SSC, Corrêa H. Impulsivity and the 5-HTTLPR polymorphism in a non-clinical sample. PLoS One. 2011 Jan;6(2):e16927.

58. Mendes DC, Silva TF, Barros L de O, de Oliveira MVM, Vieira LT, Haikal DS, et al. Analysis of the normative conditions of oral health, depression and serotonin-transporter-linked promoter region polymorphisms in an elderly population. Geriatr Gerontol Int. 2013 Jan;13(1):98–106.

59. Santos PCJL, Soares RAG, Nascimento RM, Machado-Coelho GLL, Mill JG, Krieger JE, et al. SLCO1B1 rs4149056 polymorphism associated with statin-induced myopathy is differently distributed according to ethnicity in the Brazilian general population: Amerindians as a high risk ethnic group. BMC Med Genet. 2011 Jan;12:136.

60. Barcelos GRM, Souza MF de, Oliveira AÁS de, Lengert A van H, Oliveira MT de, Camargo RB de OG, et al. Effects of genetic polymorphisms on antioxidant status and concentrations of the metals in the blood of riverside Amazonian communities co-exposed to Hg and Pb. Environ Res. 2015 Apr;138:224–32.

61. Calado RT, Falcão RP, Garcia AB, Gabellini SM, Zago MA, Franco RF. Influence of functional MDR1 gene polymorphisms on P-glycoprotein activity in CD34+ hematopoietic stem cells. Haematologica. 2002 Jun;87(6):564–8.

62. Braun-Prado K, Vieira Mion AL, Farah Pereira N, Culpi L, Petzl-Erler ML. HLA class I polymorphism, as characterised by PCR-SSOP, in a Brazilian exogamic population. Tissue Antigens. 2000 Nov;56(5):417–27.

63. Lázaro AM, Moraes ME, Marcos CY, Moraes JR, Fernández-Viña MA, Stastny P. Evolution of HLA-class I compared to HLA-class II polymorphism in Terena, a South-American Indian tribe. Hum Immunol. 1999 Nov;60(11):1138–49.

64. Nunes K, Piovezan B, Torres MA, Pontes GN, Kimura L, Carnavalli JEP, et al. Population variation of HLA genes in rural communities in Brazil, the Quilombos from the Vale do Ribeira, São Paulo - Brazil. Hum Immunol. 2016 Jun;77(6):447–8.

65. Fabreti-Oliveira RA, Nascimento E, Fonseca CG, Santos MA. The heterogeneous HLA genetic composition of the Brazilian population and its relevance to the optimization of hematopoietic stem cell donor recruitment. Tissue Antigens. 2014 Aug;84(2):187–97.

66. Soares-Vieira JA, Billerbeck AEC, Iwamura ESM, Otto PA. Brazilian Population Data on the Polymerase Chain Reaction-Based Loci LDLR, GYPA, HBGG, D7S8, and Gc. Am J Forensic Med Pathol. 2003 Sep;24(3):283–7.

67. Soares-Vieira JA, Billerbeck AEC, Iwamura ESM, Gattás GJF, Bydlowski SP, Otto PA. Distribution of HLA-DQA1, LDLR, GYPA, HBGG, D7S8 and Gc alleles in a Black population sample from São Paulo, Brazil. J Forensic Sci. 2004 Nov;49(6):1374–5.

68. Heller AH, Salzano FM, Barrantes R, Krylov M, Benevolenskaya L, Arnett FC, et al. Intra- and intercontinental molecular variability of an Alu insertion in the 3' untranslated region of the LDLR gene. Hum Biol. 2004 Aug;76(4):591–604.

69. Mattevi VS, Coimbra CE, Santos R V, Salzano FM, Hutz MH. Association of the low-density lipoprotein receptor gene with obesity in Native American populations. Hum

Genet. 2000 May;106(5):546–52.

70. Lanes AM, Louro LS, Ventorim DP, Stur E, Garcia FM, Agostini LP, et al. ΔF508 mutation screening of healthy individuals from two populations in Espírito Santo State, Brazil. Genet Mol Res. 2016 Dec 19;15(4).

71. Raskin S, Petzl-Erler ML, Phillips JA, Pereira-Ferrari L, Probst CM, Faucz FR, et al. Cystic fibrosis gene variability in two southern Brazilian Amerindian populations: analysis of the deltaF508 mutation and the KM19 and XV2C haplotypes. Hum Biol. 2007 Feb;79(1):79–91.

72. Heckmann MIO, Mendes-Júnior CT, Tada MS, Santos MG, Cabello GMK, Salzano FM, et al. CFTR haplotype distribution in the Brazilian Western Amazonian region. Hum Biol. 2005 Aug;77(4):499–508.

73. Moura Costa FM, Santana MA, Lemos ACM, Galvão-Castro B, Acosta AX. Low frequency of the deltaAF508 mutation of the CFTR gene in a highly admixed population in Bahia, Brazil. Hum Biol. 2007 Jun;79(3):293–7.

74. Angeli CB, Kimura L, Auricchio MT, Vicente JP, Mattevi VS, Zembrzuski VM, et al. Multilocus analyses of seven candidate genes suggest interacting pathways for obesity-related traits in Brazilian populations. Obesity (Silver Spring). 2011 Jun;19(6):1244–51.

75. Mattevi VS, Zembrzuski VM, Hutz MH. Impact of variation in ADRB2, ADRB3, and GNB3 genes on body mass index and waist circumference in a Brazilian population. Am J Hum Biol. 2006 Jan;18(2):182–6.

76. Stur E, Silveira AN, Selvatici LS, Alves LNR, de Vargas Wolfgramm E, Tovar TT, et al. Polymorphism analysis of MTHFR, factor II, and factor V genes in the Pomeranian population of Espirito Santo, Brazil. Genet Test Mol Biomarkers. 2012 Mar;16(3):219–

22.

77. Barbosa PR, Stabler SP, Machado ALK, Braga RC, Hirata RDC, Hirata MH, et al. Association between decreased vitamin levels and MTHFR, MTR and MTRR gene polymorphisms as determinants for elevated total homocysteine concentrations in pregnant women. Eur J Clin Nutr. 2008 Aug;62(8):1010–21.

78. Franco RF, Araújo AG, Guerreiro JF, Elion J, Zago MA. Analysis of the 677 C-->T mutation of the methylenetetrahydrofolate reductase gene in different ethnic groups. Thromb Haemost. 1998 Jan;79(1):119–21.

79. Pereira AC, Schettert IT, Morandini Filho AAF, Guerra-Shinohara EM, Krieger JE. Methylenetetrahydrofolate reductase (MTHFR) c677t gene variant modulates the homocysteine folate correlation in a mild folate-deficient population. Clin Chim Acta. 2004 Feb;340(1–2):99–105.

80. Saliba LF, Reis RS, Brownson RC, Hino AA, Tureck L V, Valko C, et al. Obesity-related gene ADRB2, ADRB3 and GHRL polymorphisms and the response to a weight loss diet intervention in adult women. Genet Mol Biol. 2014 Mar;37(1):15–22.

81. Pereira AC, Floriano MS, Mota GFA, Cunha RS, Herkenhoff FL, Mill JG, et al. Beta2 adrenoceptor functional gene variants, obesity, and blood pressure level interactions in the general population. Hypertension. 2003 Oct;42(4):685–92.

82. Tavares EF, Vieira-Filho JPB, Andriolo A, Perez ABA, Vergani N, Sañudo A, et al. Serum total homocysteine levels and the prevalence of folic acid deficiency and C677T mutation at the MTHFR gene in an indigenous population of Amazonia: the relationship of homocysteine with other cardiovascular risk factors. Ethn Dis. 2004 Jan;14(1):49–56.

83. Santos KT, De Freitas RGA, Manta FSN, De Carvalho EF, Silva DA. Ser49Gly

polymorphism in the β-adrenergic receptor 1 gene in a population sample from Rio de Janeiro state, Brazil, stratified by self-identified skin color and genetic ancestry. Mol Med Rep. 2015 Jul 24;12(1):1591–7.

84. Boldt ABW, Culpi L, Tsuneto LT, Souza IR, Kun JFJ, Petzl-Erler ML. Analysis of the CCR5 gene coding region diversity in five South American populations reveals two new non-synonymous alleles in Amerindians and high CCR5*D32 frequency in Euro-Brazilians. Genet Mol Biol. 2009 Jan;32(1):12–9.

85. Muxel SM, Borelli SD, Amarante MK, Voltarelli JC, Aoki MN, de Oliveira CEC, et al. Association study of CCR5 delta 32 polymorphism among the HLA-DRB1 Caucasian population in Northern Paraná, Brazil. J Clin Lab Anal. 2008 Jan;22(4):229–33.

86. Vargas AE, Marrero AR, Salzano FM, Bortolini MC, Chies JAB. Frequency of CCR5delta32 in Brazilian populations. Brazilian J Med Biol Res = Rev Bras Pesqui médicas e biológicas / Soc Bras Biofísica . [et al]. 2006 Mar;39(3):321–5.

87. Grimaldi R, Shindo N, Acosta AX, Dourado I, Brites C, de Melo Carvalho O, et al. Prevalence of the CCR5Delta32 mutation in Brazilian populations and cell susceptibility to HIV-1 infection. Hum Genet. 2002 Jul;111(1):102–4.

88. de Farias JD, Santos MG, de França AK, Delani D, Tada MS, Casseb AA, et al. Distribution of the CCR5delta32 allele (gene variant CCR5) in Rondônia, Western Amazonian region, Brazil. Genet Mol Biol. 2012 Jan;35(1):27–31.

89. Aléssio ACM, Annichino-Bizzacchi JM, Bydlowski SP, Eberlin MN, Vellasco AP, Höehr NF. Polymorphisms in the methylenetetrahydrofolate reductase and methionine synthase reductase genes and homocysteine levels in Brazilian children. Am J Med Genet A. 2004 Jul 30;128A(3):256–60.

90. Barnabé A, Aléssio ACM, Bittar LF, de Moraes Mazetto B, Bicudo AM, de Paula E V, et al. Folate, vitamin B12 and Homocysteine status in the post-folic acid fortification era in different subgroups of the Brazilian population attended to at a public health care center. Nutr J. 2015 Jan;14:19.

91. Couto FD, Adorno EV, Menezes JF, Moura Neto JP, Rêgo MAV, Reis MG dos, et al. C677T polymorphism of the MTHFR gene and variant hemoglobins: a study in newborns from Salvador, Bahia, Brazil. Cad saúde pública. Jan;20(2):529–33.

92. Arruda ITS de, Persuhn DC, Oliveira NFP de. The MTHFR C677T polymorphism and global DNA methylation in oral epithelial cells. Genet Mol Biol. 2013 Dec;36(4):490–3.

93. Leboute AP, de Carvalho MW, Simões AL. Absence of the deltaccr5 mutation in indigenous populations of the Brazilian Amazon. Hum Genet. 1999 Nov;105(5):442–3.

94. Passos GA, Picanço VP. Frequency of the delta ccr5 deletion allele in the urban Brazilian population. Immunol Lett. 1998 Apr;61(2–3):205–7.

95. Lage GM, Miranda DM, Romano-Silva MA, Campos SB, Albuquerque MR, Corrêa H, et al. Association between the catechol-O-methyltransferase (COMT) Val158Met polymorphism and manual aiming control in healthy subjects. PLoS One. 2014 Jan;9(6):e99698.

96. Malloy-Diniz LF, Lage GM, Campos SB, de Paula JJ, de Souza Costa D, Romano-Silva MA, et al. Association between the Catechol O-methyltransferase (COMT) Val158met polymorphism and different dimensions of impulsivity. PLoS One. 2013 Jan;8(9):e73509.

97. Soeiro-De-Souza M. Association of the COMT Met158 allele with trait impulsivity in healthy young adults. Mol Med Rep. 2013 Feb 21;7(4):1067–72.

98. Soeiro-De-Souza MG, Bio DS, David DP, Missio G, Lima B, Fernandes F, et al. Gender effects of the COMT Val 158 Met genotype on verbal fluency in healthy adults. Mol Med Rep. 2013 Sep;8(3):837–44.

99. Mendes CP, Hackel C, Arruda VR, Annichino-Bizzacchi JM. Determination of the allele frequencies of three polymorphisms in the promoter region of the human protein C gene in three Brazilian ethnic groups. Hum Hered. 1999 Jan;49(1):27–30.

100. Hutz MH, Almeida S, Coimbra CEA, Santos R V., Salzano FM. Haplotype and allele frequencies for three genes of the dopaminergic system in South American Indians. Am J Hum Biol. 2000 Sep;12(5):638–45.

101. Franco RF, Elion J, Tavella MH, Santos SE, Zago MA. The prevalence of factor V Arg306-->Thr (factor V Cambridge) and factor V Arg306-->Gly mutations in different human populations. Thromb Haemost. 1999 Feb;81(2):312–3.

102. Capeleto D, Barbisan F, Azzolin V, Dornelles EB, Rogalski F, Teixeira CF, et al. The anti-inflammatory effects of resveratrol on human peripheral blood mononuclear cells are influenced by a superoxide dismutase 2 gene polymorphism. Biogerontology. 2015 Oct;16(5):621–30.

103. Barbisan F, Motta J de R, Trott A, Azzolin V, Dornelles EB, Marcon M, et al. Methotrexate-related response on human peripheral blood mononuclear cells may be modulated by the Ala16Val-SOD2 gene polymorphism. PLoS One. 2014 Jan;9(10):e107299.

104. Dornelles EB, Goncalves BD, Schott KL, Barbisan F, Unfer TC, Glanzner WG, et al. Cytotoxic effects of moderate static magnetic field exposure on human periphery blood mononuclear cells are influenced by Val16Ala-MnSOD gene polymorphism. Environ Sci

Pollut Res. 2017 Feb 21;24(5):5078–88.

105. Costa KA, Guillo LA. TP53 codon 72 polymorphism in pigmentary phenotypes. J Biosci. 2012 Mar;37(1):33–9.

106. Gaspar PA, Hutz MH, Salzano FM, Weimer TA. TP53 polymorphisms and haplotypes in South Amerindians and neo-Brazilians. Ann Hum Biol. 2001 Jan;28(2):184–94.

107. Danoviz ME, Pereira AC, Mill JG, Krieger JE. HYPERTENSION, OBESITY AND GNB3 GENE VARIANTS. Clin Exp Pharmacol Physiol. 2006 Mar;33(3):248–52.

108. Vargens DD, Almendra L, Struchiner CJ, Suarez-Kurtz G. Distribution of the GNB3 825C>T polymorphism among Brazilians: impact of population structure. Eur J Clin Pharmacol. 2008 Mar;64(3):253–6.

109. Jacovas VC, Rovaris DL, Peréz O, de Azevedo S, Macedo GS, Sandoval JR, et al. Genetic Variations in the TP53 Pathway in Native Americans Strongly Suggest Adaptation to the High Altitudes of the Andes. Roemer K, editor. PLoS One. 2015 Sep 18;10(9):e0137823.

110. Tagliani-Ribeiro A, Paskulin DD, Oliveira M, Zagonel-Oliveira M, Longo D, Ramallo V, et al. High twinning rate in Cândido Godói: a new role for p53 in human fertility. Hum Reprod. 2012 Sep;27(9):2866–71.

111. Palmero EI, Schüler-Faccini L, Caleffi M, Achatz MIW, Olivier M, Martel-Planche G, et al. Detection of R337H, a germline TP53 mutation predisposing to multiple cancers, in asymptomatic women participating in a breast cancer screening program in Southern Brazil. Cancer Lett. 2008 Mar 8;261(1):21–5.

112. Kimura L, Angeli CB, Auricchio MTBM, Fernandes GR, Pereira AC, Vicente JP, et al. Multilocus family-based association analysis of seven candidate polymorphisms with

essential hypertension in an african-derived semi-isolated brazilian population. Int J Hypertens. 2012 Jan;2012:859219.

113. Rohr P, da Silva J, Erdtmann B, Saffi J, Guecheva TN, Antônio Pêgas Henriques J, et al. BER gene polymorphisms (OGG1 Ser326Cys and XRCC1 Arg194Trp) and modulation of DNA damage due to pesticides exposure. Environ Mol Mutagen. 2011 Jan;52(1):20–7.

114. Lewgoy F, Salzano FM. G-6-PD deficiency gene dynamics in a Brazilian population. Acta Genet Med Gemellol (Roma). 1968 Oct;17(4):595–606.

115. Saldanha PH, Nóbrega FG, Maia JC. Distribution and heredity of erythrocyte G6PD activity and electrophoretic variants among different racial groups at São Paulo, Brazil. J Med Genet. 1969 Mar;6(1):48–54.

116. Santana MS, de Lacerda MVG, Barbosa M das GV, Alecrim WD, Alecrim M das GC. Glucose-6-phosphate dehydrogenase deficiency in an endemic area for malaria in Manaus: a cross-sectional survey in the Brazilian Amazon. PLoS One. 2009 Jan;4(4):e5259.

117. Brito MAM, Peixoto HM, Almeida ACG de, Oliveira MRF de, Romero GAS, Moura-Neto JP, et al. Validation of the rapid test Carestart(tm) G6PD among malaria vivax-infected subjects in the Brazilian Amazon. Rev Soc Bras Med Trop. 2016 Aug;49(4):446–55.

118. Salzano FM, Lewgoy F, Tondo C V, da Rocha FJ. G-6-PD deficiency and abnormal hemoglobins in a Brazilian population. Acta Genet Med Gemellol (Roma). 1968 Oct;17(4):607–12.

119. Saldanha PH, Lebensztajn B, Itskan SB. Activity of glucose-6-phosphate dehydrogenase among Indians living in a malarial region of Mato Grosso and its implication to the

Indian-mixed populations in Brazil. Hum Hered. 1976 Jan;26(4):241–51.

120. Castro S, Weber R, Dadalt V, Tavares V, Giugliani R. Prevalence of G6PD deficiency in newborns in the south of Brazil. J Med Screen. 2006 Jan;13(2):85–6.

121. Compri MB, Saad ST, Ramalho AS. [Genetico-epidemiological and molecular investigation of G-6-PD deficiency in a Brazilian community]. Cad saúde pública. Jan;16(2):335–42.

122. Santana MS, Monteiro WM, Siqueira AM, Costa MF, Sampaio V, Lacerda M V., et al. Glucose-6-phosphate dehydrogenase deficient variants are associated with reduced susceptibility to malaria in the Brazilian Amazon. Trans R Soc Trop Med Hyg. 2013 Mar 10;107(5):301–6.

123. Barbosa PR, Stabler SP, Trentin R, Carvalho FR, Luchessi AD, Hirata RDC, et al. Evaluation of nutritional and genetic determinants of total homocysteine, methylmalonic acid and S-adenosylmethionine/S-adenosylhomocysteine values in Brazilian childbearing-age women. Clin Chim Acta. 2008 Feb;388(1–2):139–47.

CAPÍTULO 3 – Farmacogenética e ancestralidade na América Latina

Os estudos apresentados neste capítulo consistem em análises de dados originais de genótipos de interesse farmacogenético e ancestralidade continental. As análises destes dados serão completadas e submetidas para publicação.

Ancestralidade e farmacogenética de CYPs no Brasil e Peru (dados não publicados)

A RIBEF, como uma grande rede de colaboração internacional, possui dados de genótipos de *CYP2D6*, *CYP2C9* e *CYP2C9* para voluntários sadios da América do Norte (Monterrey, México-DF e Chiapas do México), América Central (Costa Rica e Nicarágua), Caribe (Cuba), Costa do Pacífico da América do Sul (Equador e Peru), Costa Atlântica da América do Sul (Argentina, Brasil e Uruguai), Portugal e Espanha. Portanto, há dados de genotipagem de CYPs para 5.486 indivíduos de 19 populações ao longo de 12 países da ibero-américa e de 87 AIMs para 3.387 desses indivíduos (Figura 1). A Figura 1 mostra a estruturação populacional das amostras CEIBA usando os dados dos 87 AIMs genotipados realizada através da análise de componentes principais. A Análise de Componentes Principais (PCA) é uma técnica estatística de redução da dimensionalidade que transforma as variáveis originais em novas variáveis (componentes principais), fazendo com que a primeira nova variável, ou componente, seja responsável pela maior parte da variação encontrada no conjunto de dados, a segunda variável seja responsável pela segunda maior variação encontrada, e assim sucessivamente, até que toda a variação dos dados seja elucidada com a propriedade que os componentes são independentes. A PCA identifica padrões ocultos nas variáveis originais, neste caso a partir dos genótipos dos indivíduos (Cavalli-Sforza 1998).

O gráfico desta análise ilustra claramente os componentes de miscigenação para cada população, a nível individual, assim como a distribuição no espaço da variação de ancestralidade para todos indivíduos de cada população.

Espanhóis, portugueses e Ashkenazi da Argentina possuem ancestralidade predominantemente europeia, como esperado. Grupos ameríndios do México e Peru apresentam uma predominância do componente ameríndio, exceto para a população de Maias do México (33% de ancestralidade europeia). Os ameríndios da Costa Rica apresentam componente europeu ao redor de 20%, exceto os Guaymí, com praticamente 100% de componente ameríndio. Os afro-caribenhos

da Costa Rica possuem componente africano comparável às populações Yoruba, Nigéria, coletados do HapMap, portanto praticamente 100% africanos. Mexicanos miscigenados de Chiapas possuem maior componente ameríndio e, portanto, são menos miscigenados que os mexicanos de DF. Em geral, miscigenados do México, Peru e Equador apresentam um componente ameríndio considerável, maior que o europeu. Entretanto, miscigenados da Argentina, Brasil e Uruguai são predominantemente europeus. Os grupos autodeclarados caucásicos e afro-caribenhos de Cuba apresentam um gradiente de componentes africanos e europeus, com diferentes proporções dos mesmos. Os miscigenados deste país são autodeclarados miscigenados com maior componente africano do consórcio CEIBA.

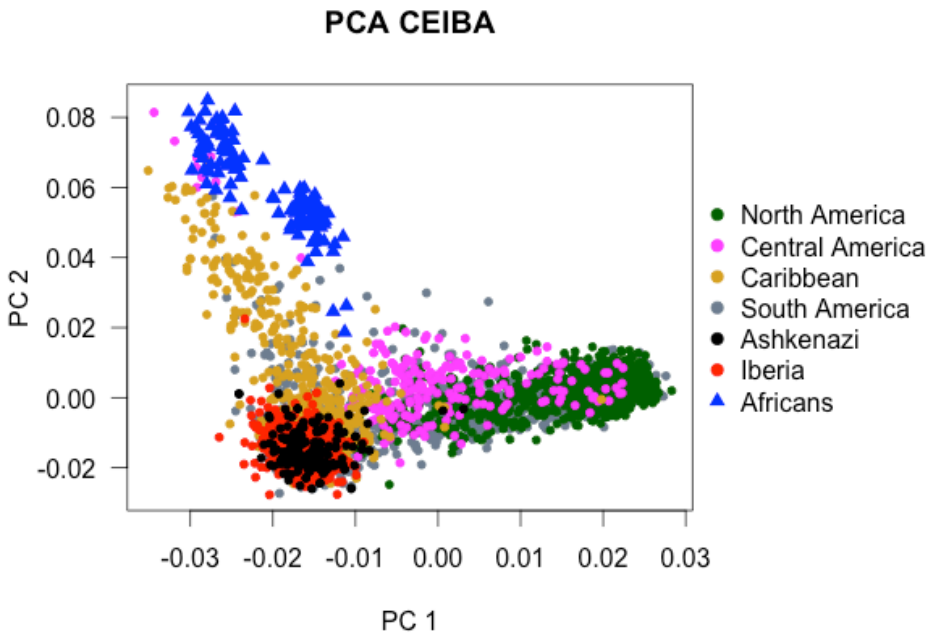

Figura 1: Análise de componentes principais para os 3.387 indivíduos de 12 países ibero-americanos para os 87 AIMs genotipados pelo consórcio CEIBA da RIBEF.

O objetivo deste trabalho é estudar a relevância do processo de miscigenação (especificamente ancestralidade individual nativo-americana, europeia e africana) para a distribuição de genótipos e fenótipos de enzimas metabolizadoras de drogas (DMEs) na América Latina.

Algumas populações latino-americanas, em particular nativos, são normalmente negligenciadas em estudos genéticos, especialmente quando tratamos de populações de nativos americanos. Dessa forma, este estudo visa verificar polimorfismos genéticos importantes na resposta a terapias medicamentosas nestes grupos étnicos.

Estes dados estão em fase de análise e escrita e serão publicados em breve. Nesta tese apresentamos resultados apenas das amostras fornecidas pelo nosso grupo de pesquisa: 98 brasileiros mineiros, 78 peruanos miscigenados de Lima, Peru, e 214 peruanos nativos (70 Ashaninka e 35 Shimaa da Amazônia e 109 Aymara dos Andes Centrais).

Neste trabalho, fui responsável pela preparação das amostras de DNA para serem genotipadas, análise de ancestralidade individual e CYPs e participo da escrita do manuscrito (em andamento).

Metodologia

Fontes de dados e amostras

As amostras do nosso grupo de pesquisa consistem em 390 indivíduos de 5 populações, parte genotipada pela RIBEF e parte terceirizada pela Universidade de Minnesota, EUA: SHI: Shimaas do departamento de Cuzco, Peru (n = 35); AYM: Aymaras do departamento de Puno, Peru (n = 109); ASH: Ashaninkas do departamento de Junin, Peru (n = 70); LIM: Miscigenados de Lima, Peru (n = 78); BRA: Miscigenados de Minas Gerais, Brasil (n = 98). Inserimos também, para os mesmos SNPs, 209 indivíduos da população YRI (Yoruba em Ibadan, Nigéria) do banco de dados Hapmap (http://hapmap.ncbi.nlm.nih.gov/) e SPA: Espanhóis de Extremadura, Espanha (n = 386, dados RIBEF) para efeitos de comparação. Após a junção dos dados, analisamos um total de 985 indivíduos de 7 populações. O projeto conta com aprovação do CONEP-UFMG e dos Comitês de Ética das outras instituições envolvidas.

Genotipagem

A genotipagem dos genes *CYP2D6*, *CYP2C9* e *CYP2C19* foi realizada na Universidade de Extremadura, no laboratório do Prof. Adrián Llerena com a metodologia de PCR em tempo real. A partir do arquivo de genótipos individuais dos três genes, geramos uma tabela de informações e

número de amostras por população para melhor organização e visualização dos dados e inferirmos o *score* de atividade enzimática individual (fenótipo predito) de acordo com os genótipos.

Genotipagem *CYP2D6*: A análise de genótipos para as variantes alélicas *CYP2D6*2*, **3, *4, *6, *10, *17, *29, *35* e **41* foi feita com DNA genômico usando o kit comercial TaqMan® de acordo com as instruções do fabricante, com *primers* específicos para os polimorfismos. Os genótipos foram determinados pela presença de SNPs representativos dos haplótipos (alelos) de interesse. Para detectar a presença de *CYP2D6*5* ou duplicação do gene, foi feita uma *long range (XL)-PCR* como descrito anteriormente (Dorado et al. 2005).

Genotipagem *CYP2C9*: A genotipagem para os alelos *CYP2C9*2, *3* e **6* foi feita com o kit comercial TaqMan® determinando os genótipos pela presença de SNPs representativos dos alelos de interesse, de modo similar ao *CYP2D6*.

Genotipagem *CYP2C19*: A genotipagem foi realizada para os alelos usando TaqMan® e, assim como *CYP2D6* e *CYP2C9*, os genótipos foram determinados pela presença de SNPs representativos dos alelos de interesse.

Fenótipos preditos para *CYP2D6*, *CYP2C19* e *CYP2C19*: Para inferir o fenótipo predito para os três genes, um "score" de atividade foi determinado para cada alelo de acordo com publicações prévias (LLerena et al. 2014; Céspedes-Garro et al. 2015; Fricke-Galindo et al. 2015). Indivíduos com nenhuma cópia ativa dos genes *CYP2D6*, *CYP2C19* e *CYP2C19* foram classificados como metabolizadores lentos preditos (*predicted poor metabolizers*, gPMs). Indivíduos com mais de uma cópia ativa do gene *CYP2D6* ou com os genótipos **1/*17* ou **17/*17* para o gene *CYP2C19* foram denominados metabolizadores ultra-rápidos preditos (*predicted ultrarapid metabolizers*, gUMs).

Ancestralidade

Etnia autodeclarada: Todos os indivíduos do consórcio CEIBA possuem etnia autodeclarada. Indivíduos com quatro avós europeus, africanos ou ameríndios foram classificados como europeus, africanos ou ameríndios, respectivamente; caso contrário foram classificados como miscigenados.

Ancestralidade genômica: Um conjunto de 3.386 indivíduos do consórcio CEIBA foram analisados para a inferência da ancestralidade europeia, africana e ameríndia a nível individual e populacional utilizando 87 AIMs selecionados dos 103 AIMs propostos por (Yaeger, Avila-bront, et

al. 2008). Os indivíduos foram genotipados no Centro Nacional de Genotipado (CEGEN, Santiago de Compostela, Espanha) usando a plataforma SequenomiPLEX (San Diego, CA, USA). Para estimar os componentes de ancestralidade europeu, africano e ameríndio para cada indivíduo, utilizamos o método de genética de populações implementado no software *ADMIXTURE* (Alexander et al. 2009). Esta metodologia adequa as frequências da população de interesse ao Equilíbrio de Hardy-Weinberg de acordo com as frequências alélicas das populações parentais (K) que deram origem à população estudada (Wollstein & Lao 2015).

Além disso, foi realizada uma análise de componentes principais (PCA) com o software *EIGENSOFT* (Price et al. 2006) para verificar a estruturação populacional destas amostras. A PCA auxilia na redução da dimensionalidade dos dados, permitindo uma clara visualização dos componentes de miscigenação para cada população.

A análise de ancestralidade individual e de componentes principais (PCA) foi realizada de acordo com (Kehdy et al. 2015).

O controle de qualidade dos genótipos foi realizado pelo Equilíbrio de Hardy-Weinberg com o módulo QC do software GLU (https://code.google.com/p/glu-genetics/), que gera um valor de significância para cada SNP estudado.

Resultados e Discussão

Ancestralidade individual e populacional

A ancestralidade populacional (média da ancestralidade individual) dos indivíduos do estudo está representada na Tabela 1.

Os espanhóis de Extremadura (SPA) possuem componente europeu elevado, como esperado (média de 90%). Os valores de ancestralidade africana e nativo-americana nessas populações não necessariamente significam que os indivíduos possuem esta ancestralidade. Pode ser meramente um ruído nos dados devido ao fato de estarmos utilizando poucos marcadores.

Tabela 1: Populações do estudo com os respectivos valores médios de ancestralidade ameríndia, europeia e africana, em ordem crescente de ancestralidade ameríndia.

Populações	**Abreviatura**	**Miscigenação**	**País**	**Nativo**	**Europeu**	**Africano**
Yoruba	YRI	Africanos	Nigéria	0.026	0.028	0.946
Extremeños	SPA	Europeus	Espanha	0.049	0.900	0.051
Brasileiros	BRA	Miscigenados	Brasil	0.086	0.684	0.229
Lima	LIM	Miscigenados	Peru	0.710	0.243	0.045
Aymaras	AYM	Nativos	Peru	0.937	0.058	0.006
Ashaninka	ASH	Nativos	Peru	0.975	0.020	0.005
Shimaa	SHI	Nativos	Peru	0.976	0.019	0.005

Evidências dos nossos resultados de médias populacionais (Tabela 1), miscigenação individual (Figura 2) e PCA (Figura 3), mostram que as populações nativas do Peru (ASH, SHI e AYM) possuem pouca ou quase nenhuma miscigenação. Seu componente ameríndio varia de 0,78 a 0,99, portanto não há indivíduos com menos de 78% deste componente na amostragem dessas populações.

A população miscigenada LIM possui componente ameríndio expressivo (Tabela 1, Figura 2, Figura 3) quando comparada à população brasileira BRA, que foi coletada em locais de alto histórico indígena de Minas Gerais. Ao contrário do esperado, os indivíduos BRA apresentaram um componente europeu mais acentuado e um componente africano relevante. Inesperadamente, em função do registro histórico, o componente ameríndio foi extremamente baixo, com média de 8% e máximo de 29,7% (Tabela 1, Figura 2, Figura 3).

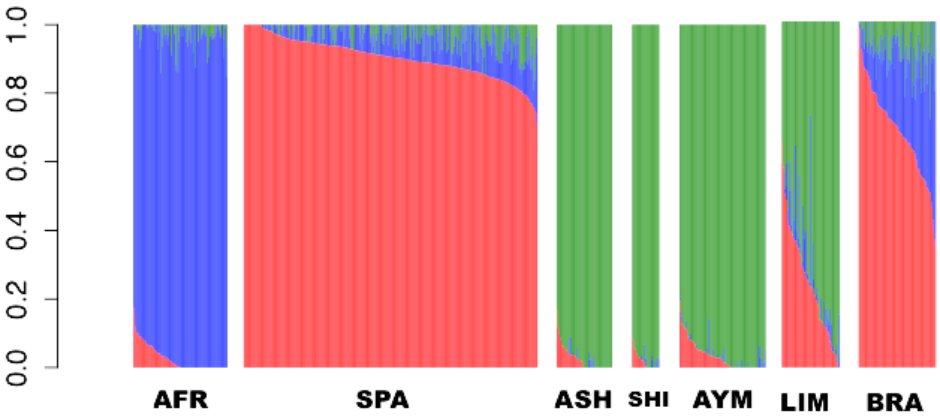

Figura 2: Ancestralidade individual das amostras do nosso grupo de pesquisa e populações parentais. Barras verticais brancas separam as diferentes populações. Cada barra vertical colorida representa um indivíduo, onde as cores representam as seguintes ancestralidades: Verde – Nativo-americana, Azul – Africana, Vermelho – Europeia. **YRI**: Yoruba em Ibadan, Nigéria (Hapmap); **SPA**: Espanhóis de Extremadura, Espanha; **ASH**: Ashaninkas do departamento de Junin, Peru; **SHI**: Shimaas do departamento de Cuzco, Peru; **AYM**: Aymaras do departamento de Puno, Peru; **LIM**: Miscigenados de Lima, Peru; **BRA**: Miscigenados de Minas Gerais, Brasil.

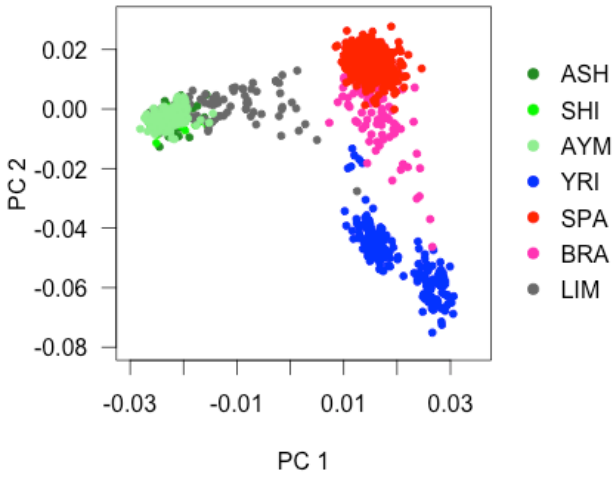

Figura 3: Análise de componentes principais a partir dos dados de genótipos dos indivíduos das 7 populações estudadas mais as duas parentais. O primeiro (abscissa) e o segundo (ordenada) componente principal são mostrados, e representam respectivamente 28,8% e 10,3% da variância total. As siglas das populações correspondem às siglas da Figura 2.

Frequências alélicas de CYPs

Os resultados de frequências alélicas para o gene *CYP2D6* correspondem ao esperado, comparando com populações de etnia similar mundiais (Tabela 2, (LLerena et al. 2014)).

Entretanto, curiosamente a população Shimaa possui uma alta porcentagem do alelo *CYP2D6*5* (8,6%), que corresponde à deleção do gene. Amostras de nativos peruanos são genotipadas para genes CYPs pela primeira vez neste estudo.

Resultados de frequências alélicas para o *CYP2C9* foram de acordo a populações mundiais de mesma etnia (Tabela 3, (Céspedes-Garro et al. 2015)). No entanto, para o gene *CYP2C19,* dois resultados inesperados para a população de nativos peruanos: nenhuma destas populações apresentou o alelo rápido **17* (Tabela 4), enquanto em populações mundiais ele aparece com uma frequência média de 11% em nativos (Fricke-Galindo et al. 2015). Além disso, Ashaninkas e Aymaras apresentaram uma frequência abaixo do esperado para o alelo de atividade diminuída **2*.

Estes resultados podem ser um primeiro indício de que populações nativas peruanas possuem baixa probabilidade de apresentar reações adversas ou falha terapêutica para fármacos metabolizados pela CYP2C19.

Clinicamente, os dois grupos de fenótipos preditos mais importantes são os metabolizadores lentos (gPMs – *poor metabolizers*) e os metabolizadores ultrarrápidos (gUMs). Estes dois grupos são os mais suscetíveis a sofrerem com as reações adversas e com a ineficácia dos tratamentos, portanto devem ser analisados com maior nível de detalhamento.

De acordo com os nossos dados, a frequência de indivíduos gUM e gPM nas amostras estudadas é praticamente nula em nativos peruanos, exceto 2,8% dos Shimaa serem gPMs para o *CYP2C19* e 1% dos Aymara serem gUMs para o *CYP2D6* (Tabela 5). Curiosamente, os brasileiros miscigenados apresentaram uma frequência de *CYP2C19* gUMs (39,3%) mais alta que a maior média até hoje registrada para voluntários sadios (36%) mundialmente, em indivíduos do Oriente Médio (Fricke-Galindo et al. 2015). Em contraste, não há nativos peruanos com essa condição, contrastando também com a média mundial de Nativos de 21%.

Tabela 2: Frequências alélicas do gene *CYP2D6* para as populações estudadas.

Country	**Self-reported Ancestry**	**n**	***wt***	****2***	****3***	****4***	****5***	****6***	****10***	****17***	****29***	****35***	****41***	***wt,*2, *35 xN***	****4xN***
Brazil	Admixed	98	35.20	25.00	1.02	15.82	3.06	1.02	1.02	5.61	2.04	2.04	8.16	0.00	
Peru	Admixed	71	62.68	20.42	0.00	6.34	2.82	0.00	0.00	1.41	1.41	1.41	1.41	2.11	
	Ashaninka	70	47.14	51.43	0.00	1.43	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
	Aymara	109	61.01	29.82	0.00	7.34	0.46	0.00	0.00	0.00	0.00	0.46	0.00	0.46	0.46
	Shimaa	35	62.86	25.71	0.00	1.43	8.57	0.00	0.00	1.43	0.00	0.00	0.00	0.00	

Tabela 3: Frequências alélicas do gene *CYP2C9* para as populações estudadas.

Country	Self-reported Ancestry	n	*wt*	**2*	**3*	**5*	**6*	**8*
Brazil	Admixed	97	88.66	7.73	3.61	-	0.00	-
Peru	Admixed	72	89.58	6.94	3.47	-	0.00	-
	Ashaninka	70	100.00	0.00	0.00	-	0.00	-
	Aymara	109	99.08	0.92	0.00	-	0.00	-
	Shimaa	35	100.00	0.00	0.00	-	0.00	-

Tabela 4: Frequências alélicas do gene *CYP2C19* para as populações estudadas.

Country	Self-reported Ancestry	n	*wt*	**2*	**3*	**4*	**5*	**17*
Brazil	Admixed	84	64.29	10.12	0.00	0.60	0.00	25.00
Peru	Admixed	72	82.64	8.33	0.00	0.00	0.00	9.03
	Ashaninka	70	96.43	3.57	0.00	0.00	0.00	0.00
	Aymara	109	96.33	3.67	0.00	0.00	0.00	0.00
	Shimaa	35	85.71	14.29	0.00	0.00	0.00	0.00

Tabela 5: Frequências de fenótipos preditos para os genes *CYP2D6, CYP2C9* e *CYP2C19* nas populações estudadas.

Country	Self-reported Ancestry	CYP2D6			CYP2C9		CYP2C19		
		n	gPMs	gUMs	n	gPMs	n	gPMs	gUMs
Brazil	Admixed	98	1.02	0.00	97	0.00	84	1.19	39.29
Peru	Admixed	71	0.00	2.82	72	0.00	72	0.00	16.67
	Ashaninka	70	0.00	0.00	70	0.00	70	0.00	0.00
	Aymara	109	0.00	0.92	109	0.00	109	0.00	0.00
	Shima	35	0.00	0.00	35	0.00	35	2.86	0.00

Conclusão

A população brasileira BRA apresentou um valor mais alto que o esperado de *CYP2C19* gUMs. Com um maior número amostral e estudos adicionais poderemos definir o padrão da população brasileira para este fenótipo e inferir futuras implicações clínicas. Em contraste, populações de nativos peruanos possuem menos alelos de importância clínica que o esperado, o que pode ser um indicativo que estes indivíduos carecem menos de intervenções farmacogenéticas. Entretanto, um número maior de estudos e amostras é necessário para confirmar estes resultados.

Relevance of the ancestry for the variability of the Drug-Metabolizing Enzymes CYP2C9, CYP2C19 and CYP2D6 polymorphisms in a multiethnic Costa Rican population

O artigo a seguir foi realizado por iniciativa da RIBEF e está aceito pela *Revista de Biología Tropical*, da Costa Rica, indexada nas Bases de Dados Pubmed, Scielo e Google acadêmico. Neste estudo encontramos correlação estatística significativa de alguns alelos de *CYP2D6* e *CYP2C19* com ancestralidade africana e ameríndia como variáveis contínuas. Além disso, as frequências alélicas de *CYP2D6*, *CYP2C9* e *CYP2C9* variaram entre os diferentes grupos étnicos da Costa Rica: ameríndios, miscigenados e afro-caribenhos.

Neste trabalho realizei as análises de ancestralidade e a correlação entre ancestralidade com genótipos e alelos dos genes *CYP2D6*, *CYP2C9* e *CYP2C9*, além de ter participado na escrita e revisão do manuscrito. Por isso, compartilho a primeira autoria com a Dra. Carolina Céspedes-Garro, da Universidade da Costa Rica.

UCR
REVISTA DE BIOLOGÍA TROPICAL
INTERNATIONAL JOURNAL OF TROPICAL BIOLOGY AND CONSERVATION

March 02, 2016
RBT-026-2016

TO WHOM IT MAY CONCERN

The manuscript: **"Relevance of the ancestry for the variability of the Drug-Metabolizing Enzymes *CYP2C9, CYP2C19* and *CYP2D6* polymorphisms in a multiethnic Costa Rican population"** by **Carolina Céspedes-Garro, Fernanda Rodrigues-Soares, Gerardo Jiménez-Arce, María-Eugenia G. Naranjo, Eduardo Tarazona-Santos, Humberto Fariñas, Ramiro Barrantes & Adrián Llerena** has been accepted for publication in the International Journal of Tropical Biology and Conservation / Revista de Biología Tropical. It will be published on-line within the next weeks and will later appear printed in volume of 64 (3) September 2016.

PDF page proofs will be e-mailed to main author when available. After publication of the paper, the main author will receive PDF reprints and one free copy of the printed issue where the paper was published.

Sincerely,

Julián Monge-Nájera
Director

UNIVERSIDAD DE COSTA RICA
REVISTA DE BIOLOGIA TROPICAL

UCR
Tels. (506) 2511-8982 • (506) 2511-8983 • biologiatropical@ucr.ac.cr • http://revistas.ucr.ac.cr/index.php/rbt/index • www.biologiatropical.ucr.ac.cr

Relevance of the ancestry for the variability of the Drug-Metabolizing Enzymes *CYP2C9, CYP2C19* and *CYP2D6* polymorphisms in a multiethnic Costa Rican population

Carolina Céspedes-Garro[1,2†], Fernanda Rodrigues-Soares[3†], Gerardo Jiménez-Arce[2], María-Eugenia G. Naranjo[1], Eduardo Tarazona-Santos[3,4], Humberto Fariñas[1], Ramiro Barrantes[2]*, Adrián LLerena[1] & CEIBA.FP Consortium of the Ibero-American Network of Pharmacogenetics & Pharmacogenomics RIBEF[a]

1. CICAB Clinical Research Center, Extremadura University Hospital and Medical School, 06080 Badajoz, Spain; carolina.cespedesgarro@ucr.ac.cr, megonzalezn@unex.es, humberto.farinas@ses.juntaextremadura.net, allerena@unex.es.
2. Genetics Section, School of Biology, University of Costa Rica, 2060 San Pedro, San José, Costa Rica; carolina.cespedesgarro@ucr.ac.cr, gerardo.jimenez@ucr.ac.cr, ramiro.barrantes@ucr.ac.cr
3. Departamento de Biologia Geral. Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, Brazil; fernandasoares@ufmg.br, edutars@icb.ufmg.br
4. Asociación Benéfica PRISMA, Lima 32, Lima, Perú; edutars@icb.ufmg.br
†Co-first autorship
*Correspondence author

Abstract: CYP2C9, CYP2C19 and CYP2D6 metabolize around 40 % of drugs and their genes vary across populations. The Costa Rican population has a trihybrid ancestry and its key geographic location turns it into a suitable scenario to evaluate interethnic differences across populations. This study aims to describe the diversity of *CYP2C9*, *CYP2C19* and *CYP2D6* polymorphisms in Costa Rican populations in the context of their ancestry. A total of 448 healthy individuals were included in the study: Bribri (n = 47), Cabecar (n = 27), Maleku (n = 16), Guaymi (n = 30), Huetar (n = 48), Chorotega (n = 41), Admixed/Mestizos from the Central Valley/Guanacaste (n = 189), and Afro-Caribbeans (n = 50) from Limón. *CYP2C9* (alleles **2*, **3*, **6*) and *CYP2C19* (**2*, **3*, **4*, **5*, **17*) genotypes were determined by Real-Time PCR. African, European and Native American ancestry were inferred using 87 ancestry informative markers. The frequency of the decreased activity allele *CYP2C9*2* is lower in the self-reported Amerindian groups compared to the admixed population, and the highest frequencies of *CYP2C19*2* (null activity) and the *CYP2C19*17* (increased activity) were found in the self-reported Afro-Caribbean population. Moreover, a frequency of 0.7 % *CYP2C9* gPMs in the Admixed population and a variable frequency of *CYP2C19* gUMs (0.0 - 32.6 %, more prevalent in Afro-Caribbeans) in Costa Rican populations, was found. Finally, the following alleles were positively correlated with genomic African ancestry and negatively correlated with genomic Native American ancestry: *CYP2D6*5* (null activity), *CYP2D6*17* (decreased activity), *CYP2D6*29* (decreased activity) and *CYP2C19*17* (increased activity). No correlation for *CYP2C9* polymorphisms and genomic ancestry was found. Further studies assessing the *CYP2C9* and *CYP2C19* sequence in these populations, preferentially by sequencing these genes, are warranted.

Key words: *CYP2C9*; *CYP2C19*; *CYP2D6*; Costa Rica; Amerindian; Afro-Caribbean; genomic ancestry.

Cytochrome P450 enzymes (CYPs) are involved in the phase I metabolism of endobiotics and xenobiotics (i.e. drugs). Thus, the activity of these enzymes is related to the plasmatic levels of active drug in patients, as well as to their therapeutic effect. In the CYP2C subfamily of drug-metabolizing enzymes (DMEs), CYP2C9 and CYP2C19 are encoded by the *CYP2C* gene cluster in 10q24, and are polymorphic, presenting interethnic variability ("CYP2C19 allele nomenclature," 2015; Sistonen et al., 2009).

CYP2C9 is involved in the metabolism of drugs such as warfarin, losartan, fluoxetine and non-steroidal anti-inflammatory drugs. Around 60 *CYP2C9* gene variants have been described ("CYP2C9 allele nomenclature," 2015), which explain a considerable proportion of variability in the drug metabolism. Some *CYP2C9* alleles have been related to a null or decreased hydroxylation capacity, such as *CYP2C9*3* allele that dramatically reduces the enzyme activity (Ingelman-Sundberg, Sim, Gomez, & Rodriguez-Antona, 2007). Thus, the carriers of two *CYP2C9*3*, as well as, alleles with null capacity are predicted to be poor metabolizers (gPMs) and to suffer adverse drug reactions (ADRs) (Yang et al., 2013).

CYP2C19 is responsible for the metabolism of antidepressants and proton pump inhibitors, among other drugs. A total of 34 *CYP2C19* allelic variants that affect enzyme activity have been described, from null (i.e. *CYP2C19*2, *3, *4, *5*) to increased activity (i.e. *CYP2C19*17*) ("CYP2C19 allele nomenclature," 2015). Individuals carriers of two inactive *CYP2C19* alleles are predicted to be gPMs, while carriers of **1/*17* or **17/*17* genotypes are predicted ultrarapid metabolizers (gUMs). In pharmacological treatment, ADRs for both metabolic groups have been shown. Drug Regulatory Agencies report *CYP2C19* as a pharmacogenetic biomarker for 16 drugs (Center for Drug Evaluation and Research, 2015) and that *CYP2C19* status of patients might predict clinical outcomes (Altar et al., 2015; Niu et al., 2015; Tabata et al., 2015).

CYP2D6 metabolizes a wide range of drugs such as antidepressants, antiarrythmics, antipsychotics, and antihistamines. More than 100 allelic variants have been described for this gene, some of which have been related to null, decreased, normal and increased enzyme activity ("CYP2D6 allele nomenclature," 2015). CYP2D6 gPMs and gUMs have been related to clinical outcomes in pharmacological therapy (Rolla et al., 2014; Seripa et al., 2015; Youngster et al., 2014) and Drug Regulatory Agencies report *CYP2D6* as a pharmacogenetic biomarker for forty drugs (Center for Drug Evaluation and Research, 2015).

Interethnic differences in such cytochrome P450 genetic polymorphisms are partially responsible for the variations among populations in drug disposition. The trihybrid

ancestry of the Costa Rican population (Segura-Wang, Raventós, Escamilla, & Barrantes, 2010), and the key geographic location of the country makes Costa Ricans fairly representative of the human genetic diversity in Central America.

The CEIBA.FP Consortium of the Ibero-American Network of Pharmacogenetics & Pharmacogenomics (RIBEF) has carried out studies in different Latin American populations (Dorado et al., 2012a; Dorado, Gallego, Peñas-LLedó, Terán, & LLerena, 2014), contributing to increase the pharmacogenetic knowledge of these neglected populations. Nevertheless, this is the first report of the *CYP2C* subfamily in a Costa Rican population including groups from different ethnic backgrounds. The present study aims to estimate the allele frequencies of *CYP2C9, CYP2C19* and *CYP2D6* polymorphisms in Costa Rican populations with different ancestry backgrounds.

MATERIALS AND METHODS

Subjects: The study comprised 448 healthy individuals, of which 385 were previously studied for the *CYP2D6* gene (Céspedes-Garro, Jiménez-Arce, Naranjo, Barrantes, & Llerena, 2014a). The following Native American Chibchan populations were analyzed: Bribri (n = 47), Cabecar (n = 27), Maleku (n = 16), Guaymi (n = 30) and Huetar (n = 48). An Oto-Manguean Mesoamerican Amerindian group: Chorotega (n = 41) was also included. Moreover, Admixed/Mestizos from the Central Valley and Guanacaste (n = 189), and Afro-Caribbeans (n = 50) from Limón were included (Céspedes-Garro, Jiménez-Arce, et al., 2014a). The number of analyzed subjects varied according to the *CYP2C* gene and genomic ancestry analyses (Tables 1, 2 and 3).

All DNA samples were obtained from a DNA biobank of the School of Biology of the University of Costa Rica. The samples were collected and stored after approval from review boards of the University of Costa Rica. Further information of collection and demographic data is available elsewhere (Azofeifa et al., 2004; Barrantes et al., 1990; Barrantes, Smouse, Neel, Mohrenweiser, & Gershowitz, 1982). The inclusion criteria of individuals in each group were previously defined (Céspedes-Garro, Jiménez-Arce, et al., 2014a).

Methods: The strategy followed in the study has been designed by the CEIBA-MESTIFAR Project (Sosa-Macias et al., 2015).

Genomic ancestry: A total of 87 ancestry informative markers (AIMs) selected from 103 previously proposed (Yaeger et al., 2008) were genotyped to infer African, European and Native American ancestry at individual and population levels. The SequenomiPLEX platform (San Diego, CA, USA) genotyping (Pereira et al., 2012) was performed at the Centro Nacional de Genotipado (CEGEN, Santiago de Compostela, Spain). In the analyses, data from 119 Yoruba unrelated individuals from Ibadan, Nigeria (YRI) and 60 Utah residents with European ancestry from the CEPH collection (CEU) from The International HapMap Consortium (2010) were included as parental populations. The Expectation Maximization method implemented in the

software *Admixture* (Alexander, Novembre, & Lange, 2009) was used to estimate ancestry, assuming three parental populations (k = 3).

***CYP2C9* and *CYP2C19* genotyping:** Genotyping for the *CYP2C9*2* (rs1799853), **3* (rs1057910), **6* (rs9332131) and *CYP2C19*2* (rs4244285), **3* (rs4986893), **4* (rs28399504), **5* (rs28399504) and **17* (rs12248560) alleles was carried out on genomic DNA using TaqMan assays as previously described (Dorado et al., 2012b; Llerena et al., 2014a; Peñas-Lledó et al., 2014). Chromosomes lacking the above-mentioned alleles/SNPs were classified as *CYP2C9*1* and *CYP2C19*1*.

***CYP2D6* genotyping:** Data on *CYP2D6*2* (rs1080985), **3* (rs35742686), **4* (rs1065852, rs3892097), **6* (rs5030655), **10* (rs1065852), **17* (rs28371706), **29* (rs59421388), **35* (rs1080985 and rs769258), **41* (rs28371725), *CYP2D6*5*, *CYP2D6*1xN*, **2xN*, **4xN* and **10xN* alleles were published for the Costa Rican population (Céspedes-Garro, Jiménez-Arce, et al., 2014a; Céspedes-Garro, Naranjo, et al., 2014b).

Predicted hydroxylation capacity group: To infer metabolic phenotype from the genotypes, zero value was assigned to *CYP2C9*3* and **6* and *CYP2C19*2*, **3*, **4* and **5* variants, 0.5 value to *CYP2C9*2*, one to *CYP2C9/19*1*, and two to *CYP2C19*17* (Peñas-Lledó et al., 2014). Individuals with activity score values equal to zero were classified as gPMs, and individuals with activity score higher than two were classified as gUMs (for *CYP2C19*) (Peñas-Lledó et al., 2014). An activity score was adapted for CYP2D6 (Gaedigk et al., 2008; LLerena et al., 2012). Previously reported data on *CYP2D6* have been analyzed together with original ancestry information of the subjects.

The differences in *CYP2C9* and *CYP2C19* allele frequencies among populations were compared using the Fisher's exact test (alpha = 0.05). Hardy-Weinberg equilibrium for alleles was determined using a contingency table X^2 statistic with Yate's correction. Statistical analyses were performed using the STATISTICA 4.3 (StatSoft, Tulsa, OK, USA) and GraphPad Prism 3.02 (GraphPad Software, San Diego, CA, USA).

The correlation between individual ancestry and the number of copies of a specific allele in each individual was estimated using the Spearman's rank correlation, with the R cor.test command (R Foundation, 2015).

RESULTS

Genomic ancestry: The studied individuals and populations from Costa Rica encompass a wide spectrum of continental ancestry and self-reported admixed individuals from the Central Valley and Guanacaste showed many of the possible combinations of European, African and Native American admixture (Fig. 1). Moreover, the three self-reported Native groups mostly have Native American ancestry, with low African ancestry (< 9 %) and European ancestry that range from 3 % (in Guaymí) to 22 % in the Chorotega. The Afro-Caribbean population has a very

high African ancestry (86 %), with all the individuals showing more than 76 % of African ancestry (Table 1 and Fig. 1).

***CYP2C9* and *CYP2C19* alleles and predicted metabolic phenotypes:** *CYP2C9* and *CYP2C19* genotype frequencies fit the Hardy-Weinberg equilibrium for all the studied populations.

Consistently with previous studies, the wild-type *CYP2C9*1* is modal in all populations (Table 2). The decreased-activity *CYP2C9*2* allele frequency was higher in the Admixed and Huetar populations (7-8 %) than in the Afro-Caribbean, Bribri, Cabecar, Maleku and Guaymi (< 1.1 %; $p < 0.05$). No differences were found in the frequency of the decreased-activity *CYP2C9*3* allele across the different groups. The null-activity *CYP2C9*6* variant was not detected in the Costa Rican populations. Moreover, only one admixed subject was a CYP2C9 predicted poor metabolizer (gPMs).

For *CYP2C19,* both the null-activity *CYP2C19*2* allele and the increased-activity *CYP2C19*17* allele were more common in the Afro-Caribbean population than in the other Costa Rican groups ($p < 0.05$) (Table 3). These allele distributions suggest that almost a third of the subjects in the Afro-Caribbean population (15 out of 46) are gUMs for CYP2C19, more than in any other Costa Rican group ($p < 0.05$).

The null-activity *CYP2C19*3* and *CYP2C19*5* alleles were not present in any of the studied populations, and *CYP2C19*4* was only present in heterozygosis in two subjects of the admixed population. No CYP2C19 gPMs were found in the entire Costa Rican sample.

Relationship between *CYP2* genes and genomic ancestry: For Costa Rica, a correlation between Native American and African ancestry with *CYP2C19* and *CYP2D6* variant alleles was found. The following alleles were positively correlated with African ancestry and negatively correlated with Native American ancestry: *CYP2D6*5* ($p < 0.01$), *CYP2D6*17* ($p < 0.01$), *CYP2D6*29* ($p = 0.027$ for African and $p = 0.045$ for Native American) and *CYP2C19*17* ($p < 0.01$). We found no correlation for *CYP2C9* polymorphisms and ancestry.

DISCUSSION

This is the first study on *CYP2C9* and *CYP2C19* in a multiethnic Costa Rican population, and it has been contextualized estimating the genomic ancestry of these populations.

The observed low frequency of the decreased-activity allele *CYP2C9*2* in the most Amerindian populations from Costa Rica (Bribri, Cabecar, Maleku and Guaymí) (Azofeifa, Ruiz, & Barrantes, 2001), is consistent with other studies on North- and South- Amerindians, which reported frequencies from 0 to 4.8 % for this allele (Céspedes-Garro et al., 2015). Noteworthy, the high *CYP2C9*2* frequency in the Huetar (8.3 %) is similar to that of the Costa Rican admixed population (7.7 %), in

agreement with reports that estimate that the Huetar groups have European and African admixture as high as 3.9 to 32.9 % (Barrantes, 1993; Santos, Ward, & Barrantes, 1994; Bieber, Bieber, Rodewald, & Barrantes, 1996; Azofeifa et al., 2001; Ruiz-Narváez et al., 2005).

The frequency of *CYP2C9*2* for the admixed Costa Rican population is similar to those reported for other Latin American admixed populations from Brazil, Chile, Ecuador, Mexico and Hispanics from United States of America (Céspedes-Garro et al., 2015).

The very low frequencies of CYP2C9 gPMs in the Costa Rican populations (predicted by surveying the presence of the **3* and **6* allele), is also in agreement with other studies in diverse Latin American and Native American populations, including US-Hispanics (Céspedes-Garro et al., 2015). This frequency can indicate that Costa Rican populations are not susceptible to adverse reactions of CYP2C9-metabolized drugs (warfarin, losartan, diclofenac) due to genetic factors.

Regarding *CYP2C19* in Native Costa Ricans, the frequency of the null-activity allele *CYP2C19*2* varies from 0 to 12.5 %. The *CYP2C19*2* frequency in the Chorotega tribe (12.5 %) was similar to that reported for Ameridian populations from Brazil (10.4 and 11.1 %) (Santos et al., 2011; Vargens, Petzl-Erler, & Suarez-Kurtz, 2012). The low frequencies for this allele in Bribri, Maleku and Guaymi populations (4.3, 0 and 0 %, respectively) are similar to those from the Purépechas, Tzotziles, Tojolabales and Tzeltales Mexican Amerindian tribes (5.4, 5.6, 3.6 and 0 %, respectively) (Salazar-Flores et al., 2012).

As previously reported, *CYP2C19*3* frequency is rare outside Eastern Asia and Melanesia (Sistonen et al., 2009); for this reason, the lack of this allele in Costa Rican populations was expected.

The increased-activity *CYP2C19*17* allele frequency of the Afro-Caribbean population is consistent with its high frequency in the West African Gambia population (23.0 %) (Janha et al., 2014), which is in accordance to the predominant Western African origin of the African diaspora to the Americas (Madrigal, 2006). The ascertainment of CYP2C19 gUMs by genotyping of the *CYP2C19*17* allele has only been performed in other two Latin American admixed populations from Brazil and Ecuador (26.8 and 41.4 %, respectively) (Santos et al., 2011; Vicente et al., 2014), and both frequencies are higher than that of the Costa Rican admixed population (17.7 %; $p < 0.05$). This result suggest that Costa Rican populations are less susceptible to therapeutic failure or adverse reactions in therapies based on drugs metabolized by CYP2C19, such as omeprazole and clopidogrel.

A correlation among *CYP2D6*5*, **17* and **29* alleles with ancestry was also found in this study, consistently with higher frequencies of **17* and **29* in African populations (LLerena et al., 2014b). The null-activity *CYP2D6*4* allele is associated to European ancestry (LLerena et al., 2014b), and is supposed to be a marker of this ancestry. However, in our multiethnic sample of individuals, the *CYP2D6*4* allele is also

present in 23 individuals with more than 30 % of Native American ancestry and less than 60 % of European ancestry. Furthermore, the highest frequencies of *CYP2D6*4* worldwide are in the Chibchan groups (e.g. Bari from Venezuela - 42.5 % - and Bribri and Cabecar from Costa Rica - 31.9 and 26.8 %, respectively) (Céspedes-Garro, Jiménez-Arce, et al., 2014a; Céspedes-Garro, Naranjo, et al., 2014b). Altogether, our results suggest that the **4* allele may be also common in Native American populations.

A limitation of this study is the low number of individuals in some of the populations, mainly Amerindians. However, considering that Amerindians are under-represented in pharmacogenetics surveys, important information is provided. Another limitation, shared with most pharmacogenetics studies, is that we genotyped specific SNPs that define specific alleles (haplotypes) and classified as wild type (*CYP2C9*1* or *CYP2C19*1*) the individuals that do not carry these alleles. However, it is unknown if populations that are under-represented in pharmacogenetics studies, such as Native Americans, may present variants that were not genotyped, or that were even unknown and that may alter enzymatic activity. Thus, further studies assessing the *CYP2C9* and *CYP2C19* sequence in an unbiased fashion, preferentially by sequencing these genes, would be necessary.

ACKNOWLEDGEMENTS

CCG was supported by a fellowship of the University of Costa Rica in the PhD program of the University of Extremadura. The study is part of the Research Program entitled "Genética, Ecología y Salud en los Amerindios de Costa Rica" (N°742-93-903) and the project N° 742-90-416 of the University of Costa Rica. The research was supported by a grant from Junta de Extremadura, Cooperación Extremeña AEXCID 13IA001. ET-S and FRS were supported by the CAPES Agency of the Brazilian Ministry of Education. The project was coordinated in the CEIBA.FP Consortium of the Ibero-American Network of Pharmacogenetics & Pharmacogenomics (RIBEF).

Resumen

CYP2C9, CYP2C19 y CYP2D6 metabolizan aproximadamente el 40 % de los fármacos y los genes que las codifican varían en las distintas poblaciones humanas. La población costarricense posee ancestría trihíbrida y su posición geográfica estratégica la convierten en un escenario idóneo para evaluar la variabilidad interétnica en sus poblaciones multiétnicas. El presente estudio tiene como objetivo describir la diversidad de los polimorfismos *CYP2C9*, *CYP2C19* y *CYP2D6* en las poblaciones costarricenses en el contexto de su ancestría. 448 individuos sanos fueron incluidos: Bribri (n=47), Cabecar (n=27), Maleku (n=16), Guaymi (n=30), Huetar (n=48), Chorotega (n=41), mestizos del Valle Central y Guanacaste (n=189) y afrocaribeños de Limón (n=50). Los genotipos *CYP2C9* (alelos **2*, **3*, **6*) y *CYP2C19* (**2*, **3*, **4*, **5* y **17*) fueron determinados mediante PCR tiempo real. Las ancestrías africana, europea y nativa americana fueron inferidas usando 87 marcadores informativos de ancestría. La frecuencia del alelo de actividad disminuida *CYP2C9*2* fue menor en los grupos autodefinidos de amerindios que en la población

mestiza y las frecuencias más altas de *CYP2C19*2* (actividad nula) y *CYP2C19*17* (actividad incrementada) se encontraron en la población autodefinida afrocaribeña. Asimismo, se encontró una frecuencia de gPMs *CYP2C9* de 0.7 % en la población mestiza y una frecuencia variable de gUMs *CYP2C19* (0.0 a 32.6 %, más prevalente en afrocaribeños) en las poblaciones costarricenses. Por último, los siguientes alelos fueron positivamente correlacionados con la ancestría africana y negativamente con la ancestría nativa americana: *CYP2D6*5* (actividad nula), *CYP2D6*17*, *CYP2D6*29* (ambos de actividad disminuida) y *CYP2C19*17* (actividad incrementada). No se encontró correlación entre los polimorfismos *CYP2C9* y la ancestría. Se requieren estudios posteriores que evalúen la secuencia de *CYP2C9* y *CYP2C19* en estas poblaciones, preferiblemente mediante la secuenciación de estos genes.

Palabras claves: *CYP2C9*, *CYP2C19*, *CYP2D6*, Costa Rica, amerindios, afrocaribeños, ancestría genética.

References

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655-1664. http://doi.org/10.1101/gr.094052.109

Altar, C. A., Carhart, J. M., Allen, J. D., Hall-Flavin, D. K., Dechairo, B. M., & Winner, J. G. (2015). Clinical validity: Combinatorial pharmacogenomics predicts antidepressant responses and healthcare utilizations better than single gene phenotypes. *The Pharmacogenomics Journal*, *15*(5), 443-451. http://doi.org/10.1038/tpj.2014.85

Azofeifa, J., Hahn, M., Ruiz, E., Hummerich, L., Morales, A. I., Jiménez, G., & Barrantes, R. (2004). The STR polymorphism (AAAAT)n within the intron 1 of the tumor protein 53 (TP53) locus in 17 populations of different ethnic groups of Africa, America, Asia and Europe. *Revista de Biologia Tropical*, *52*(3), 645-657.

Azofeifa, J., Ruiz, E., & Barrantes, R. (2001). Blood group, red cell, and serum protein variation in the Cabecar and Huetar, two Chibchan Amerindian tribes of Costa Rica. *American Journal of Human Biology : The Official Journal of the Human Biology Council*, *13*(1), 57-64. http://doi.org/10.1002/1520-6300(200101/02)13:1<57::AID-AJHB1007>3.0.CO;2-0

Barrantes, R. (1993). Diversidad genética y mezcla racial en los amerindios de Costa Rica y Panamá. *Revista de Biologia Tropical*, *41*(3 A), 379-384.

Barrantes, R., Smouse, P. E., Mohrenweiser, H. W., Gershowitz, H., Azofeifa, J., Arias, T. D., & Neel, J. V. (1990). Microevolution in lower Central America: genetic characterization of the Chibcha-speaking groups of Costa Rica and Panama, and a consensus taxonomy based on genetic and linguistic affinity. *American Journal of Human Genetics*, *46*(1), 63-84.

Barrantes, R., Smouse, P. E., Neel, J. V, Mohrenweiser, H. W., & Gershowitz, H. (1982). Migration and genetic infrastructure of the central American Guaymi and their affinities with other tribal groups. *American Journal of Physical Anthropology*, *58*(2), 201-214. http://doi.org/10.1002/ajpa.1330580213

Bieber, H., Bieber, S. W., Rodewald, A., & Barrantes, R. (1996). Microevolution and genetic affinities among six Amerindian tribes of lower Central America:

comparative genetic study of serum proteins. *Human Biology*, *68*(6), 929-953.

Center for Drug Evaluation and Research. (2015). Genomics - Table of Pharmacogenomic Biomarkers in Drug Labeling. Retrieved July 21, 2015, from http://www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics/ucm083378.htm

Céspedes-Garro, C., Fricke-Galindo, I., Naranjo, M.-E. G., Rodrigues-Soares, F., Fariñas, H., de Andrés, F., … Llerena, A. (2015). Worldwide interethnic variability and geographical distribution of CYP2C9 genotypes and phenotypes. *Expert Opinion on Drug Metabolism & Toxicology*, *11*(12),1-13. http://dx.doi.org/10.1517/17425255.2015.1111871.

Céspedes-Garro, C., Jiménez-Arce, G., Naranjo, M.-E. G., Barrantes, R., & Llerena, A. (2014). Ethnic background and CYP2D6 genetic polymorphisms in Costa Ricans. *Revista de Biología Tropical*, *62*(4), 1659-1671.

Céspedes-Garro, C., Naranjo, M.-E. G., Ramírez, R., Serrano, V., Fariñas, H., Barrantes, R., & LLerena, A. (2014). Pharmacogenetics in Central American healthy volunteers: interethnic variability. *Drug Metabolism and Personalized Therapy*, *30*(1), 19-31. http://doi.org/10.1515/dmdi-2014-0025

CYP2C19 allele nomenclature. (2015). Retrieved June 4, 2015, from http://www.cypalleles.ki.se/cyp2c19.htm

CYP2C9 allele nomenclature. (2015). Retrieved June 3, 2015, from http://www.cypalleles.ki.se/cyp2c9.htm

CYP2D6 allele nomenclature. (2015). Retrieved June 3, 2015, from http://www.cypalleles.ki.se/cyp2d6.htm

Dorado, P., Beltrán, L. J., Machín, E., Peñas-Lledó, E. M., Terán, E., & Llerena, A. (2012a). Losartan hydroxylation phenotype in an Ecuadorian population: influence of CYP2C9 genetic polymorphism, habits and gender. *Pharmacogenomics*, *13*(15), 1711-1717. http://doi.org/10.2217/pgs.12.160

Dorado, P., Gallego, A., Peñas-LLedó, E., Terán, E., & LLerena, A. (2014). Relationship between the CYP2C9 IVS8-109A>T polymorphism and high losartan hydroxylation in healthy Ecuadorian volunteers. *Pharmacogenomics*, *15*(11), 1417-1421. http://doi.org/10.2217/pgs.14.85

Dorado, P., Machín, E., De Andrés, F., Naranjo, M. E. G., Peñas-Lledó, E. M., & Llerena, A. (2012b). Development of a HPLC method for the determination of losartan urinary metabolic ratio to be used for the determination of CYP2C9 hydroxylation phenotypes. *Drug Metabolism and Drug Interactions*, *27*(4), 217-223. http://doi.org/10.1515/dmdi-2012-0018

Gaedigk, A., Simon, S. D., Pearce, R. E., Bradford, L. D., Kennedy, M. J., & Leeder, J. S. (2008). The CYP2D6 activity score: translating genotype information into a qualitative measure of phenotype. *Clinical Pharmacology and Therapeutics*, *83*(2), 234-242. http://doi.org/10.1038/sj.clpt.6100406

Ingelman-Sundberg, M., Sim, S. C., Gomez, A., & Rodriguez-Antona, C. (2007). Influence of cytochrome P450 polymorphisms on drug therapies: pharmacogenetic, pharmacoepigenetic and clinical aspects. *Pharmacology & Therapeutics*, *116*(3), 496-526. http://doi.org/10.1016/j.pharmthera.2007.09.004

Janha, R. E., Worwui, A., Linton, K. J., Shaheen, S. O., Sisay-Joof, F., & Walton, R. T. (2014). Inactive alleles of cytochrome P450 2C19 may be positively selected in human evolution. *BMC Evolutionary Biology*, *14*, 71. http://doi.org/10.1186/1471-2148-14-71

Llerena, A., Alvarez, M., Dorado, P., González, I., Peñas-LLedó, E., Pérez, B., … Calzadilla, L. R. (2014). Interethnic differences in the relevance of CYP2C9 genotype and environmental factors for diclofenac metabolism in Hispanics from Cuba and Spain. *The Pharmacogenomics Journal*, *14*(3), 229-234. http://doi.org/10.1038/tpj.2013.28

LLerena, A., Dorado, P., Ramírez, R., González, I., Álvarez, M., Peñas-LLedó, E. M., … Calzadilla, L. R. (2012). CYP2D6 genotype and debrisoquine hydroxylation phenotype in Cubans and Nicaraguans. *The Pharmacogenomics Journal*, *12*(2), 176-183. http://doi.org/10.1038/tpj.2010.85

LLerena, A., Naranjo, M. E. G., Rodrigues-Soares, F., Penas-LLedó, E. M., Fariñas, H., & Tarazona-Santos, E. (2014). Interethnic variability of CYP2D6 alleles and of predicted and measured metabolic phenotypes across world populations. *Expert Opinion on Drug Metabolism & Toxicology*, *10*(11), 1569-1583. http://doi.org/10.1517/17425255.2014.964204

Madrigal, L. (2006). *Human biology of Afro-Caribbean populations*. Cambridge studies in biological and evolutionary anthropology. New York, USA: Cambridge University Press. http://doi.org/10.1017/CBO9780511542497

Niu, X., Mao, L., Huang, Y., Baral, S., Li, J.-Y., Gao, Y., … Hu, B. (2015). CYP2C19 polymorphism and clinical outcomes among patients of different races treated with clopidogrel: A systematic review and meta-analysis. *Journal of Huazhong University of Science and Technology. Medical Sciences = Hua Zhong Ke Ji Da Xue Xue Bao. Yi Xue Ying De Wen Ban = Huazhong Keji Daxue Xuebao. Yixue Yingdewen Ban*, *35*(2), 147-156. http://doi.org/10.1007/s11596-015-1404-7

Peñas-Lledó, E., Guillaume, S., Naranjo, M. E. G., Delgado, A., Jaussent, I., Blasco-Fontecilla, H., … LLerena, A. (2014). A combined high CYP2D6-CYP2C19 metabolic capacity is associated with the severity of suicide attempt as measured by objective circumstances. *The Pharmacogenomics Journal*, *15*(2), 172-176. http://doi.org/10.1038/tpj.2014.42

Pereira, L., Zamudio, R., Soares-Souza, G., Herrera, P., Cabrera, L., Hooper, C. C., … Tarazona-Santos, E. (2012). Socioeconomic and nutritional factors account for the association of gastric cancer with Amerindian ancestry in a Latin American admixed population. *PloS One*, *7*(8), e41200. http://doi.org/10.1371/journal.pone.0041200

R Foundation. (2015). R: The R Project for Statistical Computing. Retrieved May 6, 2015, from http://www.r-project.org/

Rolla, R., Gramaglia, C., Dalò, V., Ressico, F., Prosperini, P., Vidali, M., … Zeppegno, P. (2014). An observational study of Venlafaxine and CYP2D6 in clinical practice. *Clinical Laboratory*, *60*(2), 225-231.

Ruiz-Narváez, E. A., Santos, F. R., Carvalho-Silva, D. R., Azofeifa, J., Barrantes, R., & Pena, S. D. J. (2005). Genetic variation of the Y chromosome in Chibcha-

speaking Amerindians of Costa Rica and Panama. *Human Biology*, *77*(1), 71-91. http://doi.org/10.1353/hub.2005.0032

Salazar-Flores, J., Torres-Reyes, L. A., Martínez-Cortés, G., Rubi-Castellanos, R., Sosa-Macías, M., Muñoz-Valle, J. F., … Rangel-Villalobos, H. (2012). Distribution of CYP2D6 and CYP2C19 polymorphisms associated with poor metabolizer phenotype in five Amerindian groups and western Mestizos from Mexico. *Genetic Testing and Molecular Biomarkers*, *16*(9), 1098-1104. http://doi.org/10.1089/gtmb.2012.0055

Santos, M., Ward, R. H., & Barrantes, R. (1994). mtDNA variation in the Chibcha Amerindian Huetar from Costa Rica. *Human Biology*, *66*(6), 963-977.

Santos, P. C. J. L., Soares, R. A. G., Santos, D. B. G., Nascimento, R. M., Coelho, G. L. L. M., Nicolau, J. C., … Pereira, A. C. (2011). CYP2C19 and ABCB1 gene polymorphisms are differently distributed according to ethnicity in the Brazilian general population. *BMC Medical Genetics*, *12*, 13. http://doi.org/10.1186/1471-2350-12-13

Segura-Wang, M., Raventós, H., Escamilla, M., & Barrantes, R. (2010). Assessment of genetic ancestry and population substructure in Costa Rica by analysis of individuals with a familial history of mental disorder. *Annals of Human Genetics*, *74*(6), 516-524. http://doi.org/10.1111/j.1469-1809.2010.00612.x

Seripa, D., Latina, P., Fontana, A., Gravina, C., Lattanzi, M., Savino, M., … Pilotto, A. (2015). Role of CYP2D6 polymorphisms in the outcome of postoperative pain treatment. *Pain Medicine*, *16*(10), 2012-2023. http://doi.org/10.1111/pme.12778

Sistonen, J., Fuselli, S., Palo, J. U., Chauhan, N., Padh, H., & Sajantila, A. (2009). Pharmacogenetic variation at CYP2C9, CYP2C19, and CYP2D6 at global and microgeographic scales. *Pharmacogenetics and Genomics*, *19*(2), 170-179. http://doi.org/10.1097/FPC.0b013e32831ebb30

Sosa-Macias, M., Moya, G. E., LLerena, A., Ramírez, R., Terán, E., Peñas-LLedó, E. M., … Acosta, H. (2015). Population pharmacogenetics of Ibero-Latinoamerican populations (MESTIFAR 2014). *Pharmacogenomics*, *16*(7), 673-676. http://doi.org/10.2217/pgs.15.32

Tabata, N., Hokimoto, S., Akasaka, T., Arima, Y., Sakamoto, K., Yamamoto, E., … Ogawa, H. (2015). Patients with both CYP2C19 loss-of-function allele and peripheral endothelial dysfunction are significantly correlated with adverse cardiovascular events following coronary stent implantation. *Journal of Cardiology*. http://doi.org/10.1016/j.jjcc.2015.03.010

The International HapMap Consortium. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, *467*(7311), 52-58. http://doi.org/10.1038/nature09298

Vargens, D. D., Petzl-Erler, M.-L., & Suarez-Kurtz, G. (2012). Distribution of CYP2C polymorphisms in an Amerindian population of Brazil. *Basic & Clinical Pharmacology & Toxicology*, *110*(4), 396-400. http://doi.org/10.1111/j.1742-7843.2011.00807.x

Vicente, J., González-Andrade, F., Soriano, A., Fanlo, A., Martínez-Jarreta, B., &

Sinués, B. (2014). Genetic polymorphisms of CYP2C8, CYP2C9 and CYP2C19 in Ecuadorian Mestizo and Spaniard populations: a comparative study. *Molecular Biology Reports*, *41*(3), 1267-1272. http://doi.org/10.1007/s11033-013-2971-y

Yaeger, R., Avila-Bront, A., Abdul, K., Nolan, P. C., Grann, V. R., Birchette, M. G., … Joe, A. K. (2008). Comparing genetic ancestry and self-described race in african americans born in the United States and in Africa. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *17*(6), 1329-1338. http://doi.org/10.1158/1055-9965.EPI-07-2505

Yang, J., Chen, Y., Li, X., Wei, X., Chen, X., Zhang, L., … Yin, T. (2013). Influence of CYP2C9 and VKORC1 genotypes on the risk of hemorrhagic complications in warfarin-treated patients: a systematic review and meta-analysis. *International Journal of Cardiology*, *168*(4), 4234-4243. http://doi.org/10.1016/j.ijcard.2013.07.151

Youngster, I., Zachor, D. A., Gabis, L. V, Bar-Chaim, A., Benveniste-Levkovitz, P., Britzi, M., … Berkovitch, M. (2014). CYP2D6 genotyping in paediatric patients with autism treated with risperidone: a preliminary cohort study. *Developmental Medicine and Child Neurology*, *56*(10), 990-994. http://doi.org/10.1111/dmcn.12470

[a]CEIBA Consortium of authors (*group coordinator):

- **Group 1** Authors: Graciela E. Moya*, Verónica Ferreiro. Institutions: Pontificia Universidad Católica, Buenos Aires, Argentina; Argentina & Fundación GENOS, Buenos Aires, Argentina.
- **Group 2** Author: Eduardo Tarazona-Santos*, Fernanda Rodrigues-Soares. Institution: Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.
- **Group 3** Authors: Alba P. Sarmiento*, Angélica Borbón. Institution: (previous) Pontificia Universidad Javeriana, Bogotá, Colombia.
- **Group 4** Authors: Ramiro Barrantes*, Gerardo Jiménez-Arce, Carolina Céspedes-Garro. Institution: Universidad de Costa Rica, San José, Costa Rica.
- **Group 5** Authors: Idania Rodeiro[1]*, Mayra Álvárez[2], René Delgado[3], Diadelis Remirez[4], Bárbaro Pérez[2], Luis R. Calzadilla[5]*. Institutions: [1]CEBIMAR; [2]Facultad de Medicina Calixto García; [3]CIDEM; [4]CECMED, [5]Centro Comunitario de Salud Mental La Habana Vieja, La Habana, Cuba.
- **Group 6** Authors: Enrique Terán[1]*, Santiago Terán[1], Francisco Hernández[2]. Institutions: [1]Universidad San Francisco de Quito, Quito, Ecuador, [2]Universidad Estatal de Guayaquil, Guayaquil, Ecuador.
- **Group 7** Authors: Rocío Ortiz-López*, Augusto Rojas-Martínez, Lourdes Garza-Ocañas, Yadira X. Pérez-Páramo. Institution: Universidad Autónoma de Nuevo León, Monterrey, Nuevo León, México.
- **Group 8** Authors: Marisol López-López*, Alberto Ortega-Vázquez, Nancy Monroy-Jaramillo, Helgi Jung-Cook, Ingrid Fricke-Galindo, Elisa Alonso-Vilatela, Teresa Corona-Vázquez. Institutions: Universidad Autónoma Metropolitana & Instituto Nacional de Neurología y Neurocirugía Manuel Velasco Suárez, México City, México.
- **Group 9** Authors: Martha G. Sosa-Macías*, Carlos Galaviz-Hernández, Ismael Lares-Aseff, Blanca P. Lazalde-Ramos. Institution: Instituto Politécnico Nacional-CIIDIR, Durango, México.
- **Group 10** Author: Ronald Ramírez-Roa*, Catalina Altamirano Tinoco. Institution: Universidad Nacional Autónoma de Nicaragua, León, Nicaragua.
- **Group 11** Authors: Manuela Grazina*. Institution: University of Coimbra, Coimbra, Portugal.
- **Group 12** Authors: Adrián LLerena*, Pedro Dorado, Eva M. Peñas-Lledó, Jesús Cobaleda, M. Eugenia G. Naranjo, Fernando de Andrés, Humberto Fariñas. Institution: University of Extremadura, Badajoz, Spain.
- **Group 13** Authors: Francisco E. Estévez-Carrizo*. Institution: Universidad de Montevideo, Montevideo, Uruguay.

TABLE 1

Mean of the genomic ancestry for different Costa Rican ethnic groups

Population	n	European ancestry	African ancestry	Native American ancestry
Admixed	32	0.429	0.168	0.403
Bribri	12	0.182	0.072	0.745
Chorotega	26	0.220	0.090	0.690
Guaymi	18	0.030	0.027	0.943
Afro-Caribbean	11	0.093	0.863	0.044

n: number of subjects.

TABLE 2

Frequencies (%) of *CYP2C9* alleles and phenotypes predicted from genotype in different Costa Rican ethnic groups

Self-reported Ancestry	n	**1*	**2*	**3*	**6*	gPMs
Admixed	137	88.7	7.7	3.6	0.0	0.7
Bribri	46	97.8	1.1[a]	1.1	0.0	0.0
Cabecar	27	98.1	0.0[a]	1.9	0.0	0.0
Chorotega	31	95.1	3.2	1.6	0.0	0.0
Guaymí	27	100.0	0.0[a]	0.0	0.0	0.0
Huetar	48	86.5	8.3	5.2	0.0	0.0
Maleku	15	100.0	0.0	0.0	0.0	0.0
Afro-Caribbean	45	95.6	1.1[a]	3.3	0.0	0.0

n: number of subjects; gPMs: predicted poor metabolizers from genotype. [a]$p < 0.05$ compared to the admixed and Huetar populations.

TABLE 3

Frequencies (%) of *CYP2C19* alleles and phenotypes predicted from genotype in different Costa Rican ethnic groups

Self-reported Ancestry	n	**1*	**2*	**3*	**4*	**5*	**17*	gPMs	gUMs
Admixed	141	81.9	7.1	0.0	0.7	0.0	10.3	0.0	17.7
Bribri	23	91.3	4.3	0.0	0.0	0.0	4.3	0.0	8.7
Chorotega	36	84.7	12.5	0.0	0.0	0.0	2.8	0.0	5.6
Guaymi	24	98.0	0.0	0.0	0.0	0.0	2.0	0.0	4.0
Maleku	12	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Afro-Caribbean	46	58.7	19.6[a]	0.0	0.0	0.0	21.7[b]	0.0	32.6[b]

n: number of subjects; gPMs: predicted poor metabolizers from genotype; gUMs: predicted ultra-rapid metabolizers from genotype. [a] $p < 0.05$ compared to the rest of populations with the exception of the Chorotega tribe; [b] $p < 0.05$ compared to the rest of populations.

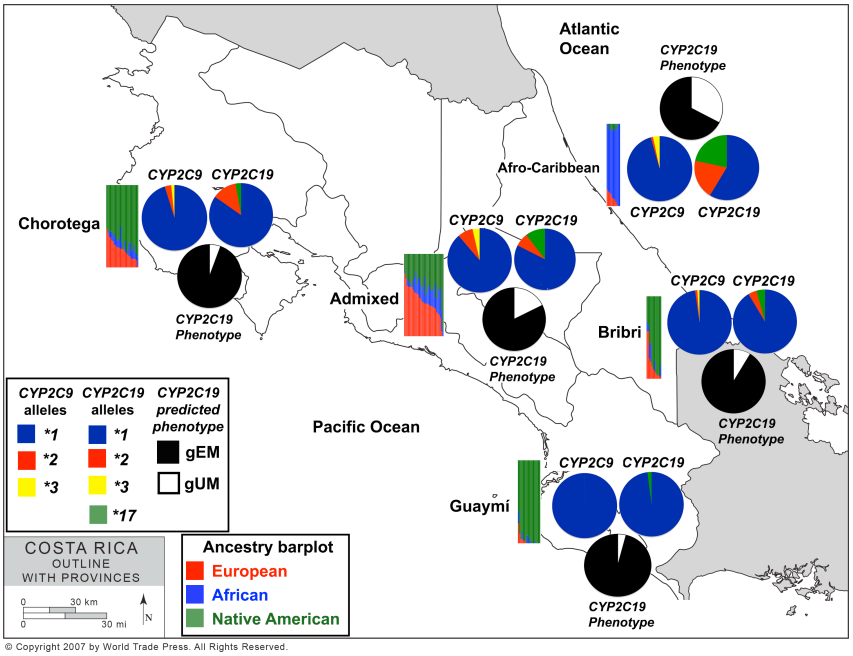

Fig. 1. Barplots of individual continental ancestry, main *CYP2C9* and *CYP2C19* alleles and CYP2C19 predicted phenotypes frequency distributions in Costa Rican populations. The approximate location of populations is shown in the chart.

PDE4B region related to acute lymphoblastic leukemia relapse is differentiated in Native Americans

A leucemia linfoide aguda (LLA) é um tipo de leucemia predominante em crianças, e os pacientes são sujeitos à poliquimioterapia. (Yang et al. 2011), estudando populações hispânicas (i.e. latino-americanos miscigenados) dos Estados Unidos, observaram que: a) existe uma maior probabilidade de recidiva da doença em indivíduos com maior componente nativo-americano e b), dois polimorfismos de ponto (SNPs): rs17039396 (no cromossomo 2, gene *MYT1L*), e rs6683977 (no gene *PDE4B*, cromossomo 1), seriam responsáveis por esta associação.

Para desenhar eficientemente estudos de farmacogenética clínica envolvendo estas questões, é importante primeiro desenvolver estudos de farmacogenética populacional, e conhecer em detalhe a diversidade nucleotídica e a estrutura haplotípica dos genes *MYT1L* e *PDE4B* (associados com maior ancestralidade nativo-americana e maior recidiva de LLA) em diferentes populações nativas e miscigenadas da América Latina, determinando o nível de diversidade dentro e entre estas populações. Por exemplo, estas informações permitirão esclarecer aspectos da associação observada entre ancestralidade nativo-americana e recidiva de LLA.

Dessa forma, o manuscrito seguinte visa determinar a diversidade nucleotídica e estrutura haplotípica em alta resolução dos genes *PDE4B* e *MYT1L* em populações nativo-americanas, normalmente negligenciadas em estudos de variabilidade genômica humana.

Neste trabalho, juntamente com Rennan Moreira, fiquei responsável pela extração de DNA das amostras de brasileiros, implementação da plataforma BeadXpress (Illumina Inc., San Diego, EUA) no ICB-UFMG, genotipagem das amostras nesta plataforma, análise de dados e escrita do manuscrito. Além disso, participei da coleta de amostras de mineiros juntamente com o doutorando Mateus Gouveia.

O trabalho está em fase de preparação. Ainda em 2016, dados de NGS serão adicionados ao estudo, visando complementar os resultados. Estes dados estão sendo gerados por Rennan Moreira e Meddly Santollala. Após a inserção e análise destes dados, o manuscrito será submetido.

PDE4B REGION RELATED TO ACUTE LYMPHOBLASTIC LEUKEMIA RELAPSE IS DIFFERENTIATED IN NATIVE AMERICANS

Fernanda Rodrigues-Soares[1*], Rennan Garcias Moreira[1*], Mateus H. Gouveia[1], Wagner Magalhães[1], Marilia O Scliar[1], Giordano Soares-Souza[1], Roxana Zamudio[1], Alexandre Pereira[2], Robert H Gilman, Eduardo Tarazona-Santos[1]

[1] Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil.
[2]Instituto do Coração, São Paulo, Brazil

* These authors equally contributed to this paper

Corresponding author:
Dr. Eduardo Tarazona-Santos
Departamento de Biologia Geral
Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais
Av. Antonio Carlos 6627, Pampulha
Belo Horizonte, MG, CEP 31270-910, Brazil
Telephone: +55 31 3409-2597
E-mail: edutars@icb.ufmg.br

Introduction

Although rare, studies with Native Americans suggest an association between Native American ancestry and complex disorders, especially cancer [1,2]. Cancer treatments are complicated due to high adverse drug reactions and relapse. It is known that individual variation in response to treatments and relapse are also due to genetic traits unique to each patient, which is the focus of pharmacogenetics. Genome wide association studies (GWAS) have played an important role in finding alleles associated with cancer predisposition [3,4], relapse [1,5] and treatment efficacy [6,7]. Notwithstanding, such studies require an extensive amount of samples in order to identify small to moderate effects due to scarcity of target alleles. An alternate and more powerful strategy is known as Admixture Mapping (AM). It relies on identifying the ancestry of each chromosome segment of the genome and associating it to disease-causing or genetic variants due to differences in allele frequency of parental populations [8]. Such approach requires fewer samples than GWAS and provides higher probabilities of finding new variants associated with diseases or traits. Thus, whereas GWAS searches for association between phenotype and genotype of a variant, AM tests whether there is association between an specific phenotype and ancestry of a small segment of a chromosome. A previous study associated two SNPs with high risk of relapse in ALL and Native American ancestry in Hispanic children from USA **(rs6683977-*PDE4B* and rs17039396-*MYT1L*)** [1]**.** Indeed, 6.0% and 18.6% of deaths among children aged from 5 to 14 years in lower- and upper middle-income countries, respectively, are caused by cancer [9].

In this sense, in order to support such findings, the main goal of this study is to evaluate genetic diversity of Native American populations as well as of Brazilian admixed populations in the genomic regions surrounding **rs6683977-*PDE4B* and rs17039396-*MYT1L***, and discuss their impact in mapping of variants related to ALL relapse in populations with Native American ancestry. According to what was already reported [1], our main hypothesis is that genetic diversity of regions surrounding **rs6683977-*PDE4B* and rs17039396-*MYT1L*** are differentiated in samples of Native Americans and also in admixed Brazilians according to their degree of native ancestry.

Methods

Tag-SNPs selection and genotyping

Following what was found by [1], regions around the admixture mapping hits (rs6683977 in *PDE4B* and rs17039396 in *MYT1L*) were selected for this study. *PDE4B* and *MYT1L* are relatively long genes, encompassing 582 kb and 542 kb respectively. Seeking a better coverage of the variation found in both genes, a search in HapMap II Project database [10] was performed in order to identify additional SNPs. We browsed HapMap II database in order to identify tag-SNPs. In October 2011, we downloaded SNPs genotypes in 1Mb regions centered in rs6683977 (*PDE4B*) and rs17039396 (*MYT1L*) in HapMap II database for the following populations: CEU (Utah residents with European ancestry, n = 174), CHB (Chinese from Beijing, n = 139), and YRI (Yoruba from Ibadan, Nigeria, n = 209), as representative of main continental groups. No Native American population was considered in HapMap 2, as well as in more recent initiatives such as 1000 Genomes. We used the package Tagzilla, implemented in the software GLU (Genotype Library and Utilities) [11] to identify tag-SNPs for each gene (minimum r^2 = 0.8). Tag-SNPs are genetic markers in linkage disequilibrium with other SNPs. Thus, genotyping such tag-SNPs provide information on other non-genotyped SNPs in linkage disequilibrium. We chose 21 tag-SNPs for *PDE4B* and 27 tag-SNPs for *MYT1L* to be genotyped, but only 24 were successfully genotyped for *MYT1L*, including rs6683977 (*PDE4B*) and rs17039396 (*MYT1L*). This number of tag-SNPs to be genotyped is a compromise between maximizing the number of tagged SNPs and the maximum limit of 48 SNPs to be genotyped by the BeadXpress platform.

We genotyped the final SNPs panel in three sets of samples (Figure 1): (1) Native Americans from Peru: Aymara (AYM, n = 91), Shimaa (SHI, n = 72) and Ashaninka (ASH, n = 92); (2) Native Americans from Brazil: Guaranis (GUA, n = 43) and Tupinikins (TUP, n = 45), (3) Admixed Brazilians from the state of Minas Gerais (AMG, n = 98), resulting in a total of 480 individuals. We inferred the genomic ancestry of each individual by genotyping 76 Ancestry Informative Markers (AIMs) for the 480 individuals [12]. We used the following Hapmap individuals as proxy of non-European continental ancestry: 176 Yoruba from Ibadan, Nigeria (YRI) and 174 European

descendants living in Utah (CEU). We estimated individual ancestry using the software Admixture v.1.2 [13].

We genotyped *PDE4B* and *MYT1L* SNPs, as well as 76 AIMs, in the 480 individuals using the *BeadXpress* platform (Illumina Inc., San Diego, USA), according to the manufacturer`s instructions. Data quality control was performed in Genome Studio software (Illumina Inc., USA) to exclude samples with Call Rate < 0.97 and SNPs with low Gene Train. For each locus we tested Hardy Weinberg equilibrium using the GLU software [3].

Pairwise F_{ST}, allelic frequencies and heterozigosity were performed with R Adegenet package [14]. Molecular variance analysis was executed using R Hierfstat package [15]. Linkage disequilibrium was measured using r^2 statistics implemented in Haploview software [16].

Results and Discussion

Definition of markers panel and quality control

HapMap II, database returned 1002 SNPs for *PDE4B* and 714 for *MYT1L* when CEU, YRI and CHB populations were considered. After analysis of linkage disequilibrium, we found 326 tag-SNPs in *PDE4B* and 198 in *MYT1L* (Table 1). We selected a subset of 21 *PDE4B* 27 *MYT1L* tag-SNPs to be included in the customized genotyping array (Table 1). After filtering samples and SNPs for such parameters, we obtained a final set of 46 SNPs, which tag a set of 478 SNPs in *PDE4B* gene and 367 SNPs in *MYT1L*.

Individual Admixture

Individual and population genomic ancestry is presented in Figure 1 and Table 2. Native Americans from Peru have less than 5% of non-Native American admixture. Native Americans from Brazil have more Old-World admixture, but the largest continental ancestry component is still Native American. Admixed Brazilians from Minas Gerais State are predominantly European.

Intra- and Inter-population diversity of genes PDE4B and MYT1L

Heterozygosity is a measure of diversity of a locus, and can be determined for a gene in a population. Intrapopulation diversity, measured by heterozygosity, was assessed for each population of the study and is reported in Table 3. The highest heterozygosity for *PDE4B* in Europeans than in Africans, contrary to expectation, suggests that the pattern of genetic diversity in this region is unusual or an ascertained bias for the HaMap II database SNPs in this genomic region. As expected Native American show lowest diversity than the other populations, particularly for *MYT1L*.

The F_{ST} statistic is a metric to measure interpopulation diversity according to its alleles frequencies [17]. The higher the F_{ST} value the higher the divergence between two populations for a specific locus or a set of loci. Pairwise F_{ST} was evaluated as a measure of genetic divergence between populations and is presented on Table 4. Considering both genomic regions, we observe a pattern and level of between-populations diversity that is common along the human genome [18]. Thus we have no evidence that the admixture mapping hit results from a genomic region that is highly differentiated along the studied 1 Mb.

Allele frequencies, AMOVA and Linkage Disequilibrium (LD)

Table 5 shows results of AMOVA, LD and allele frequencies for all SNPs genotyped for *PDE4B* and *MYT1L*. Consistently with the F_{ST} analysis for the entire *PDE4B* region, this is not a 1MB genomic region that is particularly differentiated in Native Americans, except for few SNPs represented for tag-SNPs such as rs6661245, rs6588177 and rs10889591 and rs11208774. Interestingly, the highest F_{CT} value found in *PDE4B* SNPs is for rs6683977 and considering allele frequencies and AMOVA analyses, we confirmed our expectation that the rs6683977-C-*PDE4B* allele is highly differentiated in Native Americans from Peru and Brazil, when comparing with Europeans. According to the LD analysis, rs6683977-*PDE4B* has 19 tagged-SNPs and is located in a haplotype of around 20kb, which seems to be entire differentiated in Native Americans. Differently from rs6683977-*PDE4B,* the admixture mapping hit rs17039396-*MYT1L* is not differentiated in Native Americans and there is HapMap II SNP tagged by it.

In conclusion, our analyses confirm that the LLA admixture mapping hit rs6683977 (*PDE4B*), that is associated with Native American ancestry in US Hispanic, is part of a 20Kb region that is differentiated in Native American populations from South America. We did not confirm the same pattern for the LLA admixture mapping hit rs17039396 (*MYT1L*).

References

[1] Yang JJ, Cheng C, Devidas M, Cao X, Fan Y, Campana D, et al. Ancestry and pharmacogenomics of relapse in acute lymphoblastic leukemia. Nat Genet 2011;43:237–41. doi:10.1038/ng.763.

[2] Pereira L, Zamudio R, Soares-Souza G, Herrera P, Cabrera L, Hooper CC, et al. Socioeconomic and nutritional factors account for the association of gastric cancer with Amerindian ancestry in a Latin American admixed population. PLoS One 2012;7:e41200. doi:10.1371/journal.pone.0041200.

[3] Kim HS, Minna JD, White MA. GWAS Meets TCGA to Illuminate Mechanisms of Cancer Predisposition. Cell 2013;152:387–9. doi:10.1016/j.cell.2013.01.027.

[4] Pharoah PDP, Tsai Y-Y, Ramus SJ, Phelan CM, Goode EL, Lawrenson K, et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. Nat Genet 2013;45:362–70, 370e1–2. doi:10.1038/ng.2564.

[5] Perez-Andreu V, Roberts KG, Harvey RC, Yang W, Cheng C, Pei D, et al. Inherited GATA3 variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse. Nat Genet 2013;45:1494–8. doi:10.1038/ng.2803.

[6] Low S-K, Takahashi A, Mushiroda T, Kubo M. Genome-wide association study: a useful tool to identify common genetic variants associated with drug toxicity and efficacy in cancer pharmacogenomics. Clin Cancer Res 2014;20:2541–52. doi:10.1158/1078-0432.CCR-13-2755.

[7] Li D, Wei L, Xu B, Yu D, Chang J, Yuan P, et al. Association of GWAS-identified lung cancer susceptibility loci with survival length in patients with small-cell lung

cancer treated with platinum-based chemotherapy. PLoS One 2014;9:e113574. doi:10.1371/journal.pone.0113574.

[8] Winkler CA, Nelson GW, Smith MW. Admixture mapping comes of age. Annu Rev Genomics Hum Genet 2010;11:65–89. doi:10.1146/annurev-genom-082509-141523.

[9] Gupta S, Rivera-Luna R, Ribeiro RC, Howard SC. Pediatric oncology as the next global child health priority: the need for national childhood cancer strategies in low- and middle-income countries. PLoS Med 2014;11:e1001656. doi:10.1371/journal.pmed.1001656.

[10] The International HapMap Project. Nature 2003;426:789–96. doi:10.1038/nature02168.

[11] Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a Maximally Informative Set of Single-Nucleotide Polymorphisms for Association Analyses Using Linkage Disequilibrium. October 2004:106–20.

[12] Yaeger R, Avila-Bront A, Abdul K, Nolan PC, Grann VR, Birchette MG, et al. Comparing genetic ancestry and self-described race in african americans born in the United States and in Africa. Cancer Epidemiol Biomarkers Prev 2008;17:1329–38. doi:10.1158/1055-9965.EPI-07-2505.

[13] Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res 2009;19:1655–64. doi:10.1101/gr.094052.109.

[14] Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 2008;24:1403–5. doi:10.1093/bioinformatics/btn129.

[15] de Meeûs T, Goudet J. A step-by-step tutorial to use HierFstat to analyse populations hierarchically structured at multiple levels. Infect Genet Evol 2007;7:731–5. doi:10.1016/j.meegid.2007.07.005.

[16] Barrett JC, Fry B, Maller J, Daly MJ. Haploview : analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263–5. doi:10.1093/bioinformatics/bth457.

[17] Wright S. Genetical structure of populations. Nature 1950;166:247–9.

[18] Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation. Nature 2015;526:68–74. doi:10.1038/nature15393.

[19] Otowa T, Kawamura Y, Sugaya N, Yoshida E, Shimada T, Liu X, et al. Association study of PDE4B with panic disorder in the Japanese population. Prog Neuropsychopharmacol Biol Psychiatry 2011;35:545–9. doi:10.1016/j.pnpbp.2010.12.013.

[20] Wang T, Zeng Z, Li T, Liu J, Li J, Li Y, et al. Common SNPs in Myelin Transcription Factor 1-Like (MYT1L): Association with Major Depressive Disorder in the Chinese Han Population. PLoS One 2010;5:e13662. doi:10.1371/journal.pone.0013662.

[21] Lesch K-P, Timmesfeld N, Renner TJ, Halperin R, Röser C, Nguyen TT, et al. Molecular genetics of adult ADHD: converging evidence from genome-wide association and extended pedigree linkage studies. J Neural Transm 2008;115:1573–85. doi:10.1007/s00702-008-0119-3.

[22] Zhang Y, Zhu H, Zhang X, Gu D, Zhou X, Wang M, et al. Clinical significance of MYT1L gene polymorphisms in Chinese patients with gastric cancer. PLoS One 2013;8:e71979. doi:10.1371/journal.pone.0071979.

Figures and Tables

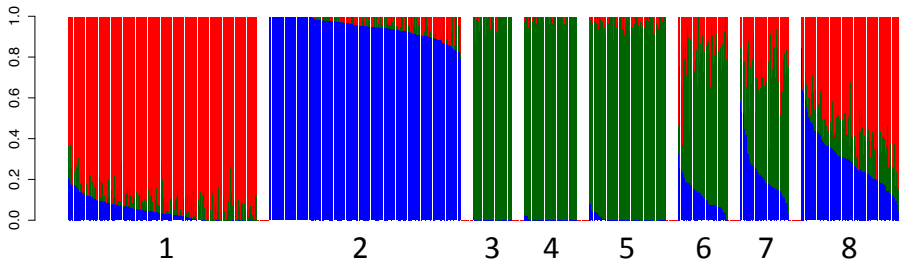

Figure 1. Individual ancestry for the following populations: 1 – European descendants from Utah (CEU); 2 – Yoruba from Nigeria (YRI); 3 – Ashaninka from Peru (ASH); 4 – Shimaa from Peru (SHI); 5 – Aymara from Peru (AYM); 6 – Guarani from Brazil (GUA); 7 – Tupinikin from Brazil (TUP); 8 – Admixed from Minas Gerais, Brazil (AMG). Each bar represents an individual, and the colors are: Green – Native American ancestry, Blue – African ancestry and Red – European ancestry.

Table 1. SNPs genotyped by gene and number of other sites covered according to HapMap II.

PDE4B	**Number of loci covered**	***MYT1L***	**Number of loci covered**
rs6683977	19	rs783036	**1**
rs12119734	24	rs17039396	**1**
rs6661245	39	rs12468168	**27**
rs10454453	32	rs11683072	**31**
rs1500951	12	rs13388663	**5**
rs12081185	10	rs10195351	**19**
rs6683604	19	rs17338491	**27**
rs11589566	9	rs17039463	**5**
rs1354060	18	rs13026825	**10**
rs11208792	13	rs6705656	**31**
rs12027416	23	rs13034457	**5**
rs2503205	23	rs6548050	**14**
rs6588177	50	rs12986730	**9**
rs10889591	14	rs17039334	**11**
rs12033425	28	rs1862114	**12**
rs12045945	19	rs6729968	**13**
rs1937453	32	rs3748988	**10**
rs1937456	14	rs2241685	**6**
rs522037	**26**	rs3748989	**6**
rs2186123	27	rs17039474	**10**
rs11208774	29	rs3924017	**22**
		rs12991351	**24**
		rs6728613	**20**
		rs12468174	**9**
		rs9284792	**14**
		rs10178240	**8**
		rs11127305	**14**

Table 2. Mean proportion of each parental ancestry across populations

	African	**Native**	**European**
CEU	0.05	0.06	0.89
YRI	0.95	0.02	0.03
ASH	0.00	0.98	0.02
SHI	0.00	0.99	0.01
AYM	0.01	0.96	0.03
GUA	0.16	0.61	0.23
TUP	0.23	0.50	0.27
AMG	0.30	0.15	0.55

Table 3. Mean heterozygosity values for each of the populations studied for *PDE4B*, *MYT1L* and both genes.

	CEU	YRI	ASH	SHI	AYM	GUA	TUP	AMG
PDE4B	0.352	0.323	0.237	0.258	0.311	0.25	0.257	0.299
MYT1L	0.298	0.311	0.159	0.095	0.204	0.215	0.272	0.299
Both genes	0.322	0.317	0.194	0.168	0.252	0.231	0.265	0.299

Table 4. Mean pairwise F_{ST} calculated using *PDE4B* and *MYT1L* SNPs data.

PDE4B	ASH	SHI	AYM	GUA	TUP	AMG	CEU
SHI	**0.011**						
AYM	0.038	**0.026**					
GUA	0.044	0.035	0.033				
TUP	0.048	0.037	0.038	**0.010**			
AMG	0.048	0.039	0.043	0.039	0.037		
CEU	**0.083**	**0.071**	0.062	0.047	0.048	0.041	
YRI	**0.105**	**0.092**	0.068	0.067	0.069	0.043	0.054

MYT1L	ASH	SHI	AYM	GUA	TUP	AMG	CEU
SHI	**0.024**						
AYM	0.018	0.037					
GUA	0.031	0.068	**0.016**				
TUP	0.059	0.111	0.034	**0.013**			
AMG	0.090	**0.131**	0.055	**0.027**	**0.010**		
CEU	0.079	**0.107**	0.052	0.033	0.025	**0.027**	
YRI	**0.151**	**0.189**	0.121	0.071	0.047	0.048	0.062

Both genes	ASH	SHI	AYM	GUA	TUP	AMG	CEU
SHI	**0.016**						
AYM	0.029	**0.030**					
GUA	0.038	0.049	**0.025**				
TUP	0.054	0.071	0.036	**0.011**			
AMG	0.070	0.086	0.049	0.033	**0.022**		
CEU	0.081	0.089	0.057	0.040	0.036	0.034	
YRI	**0.130**	**0.145**	0.097	0.069	0.057	0.046	0.058

Table 5. Allelic frequencies of *PDE4B* and *MYT1L* SNPs in each population studied

	Position v37	SNPs	Total Fct	Fct CEU x NAT	LD (r2) with hit SNP					ALLELE FREQUENCIES							
					CEU	YRI	NAT PERU	NAT BRA	AMG	CEU	YRI	ASH	SHI	AYM	GUA	TUP	AMG
PDE4B	66321193	**rs12081185.A**	**0.025**	**0.063**	0.002	0.008	0.00	nan	0.00	0.451	0.497	0.685	0.549	0.654	0.558	0.633	0.474
	66368625	**rs11589566.C**	**0.054**	**-0.019**	0.010	0.023	0.00	nan	0.00	0.250	0.153	0.380	0.278	0.308	0.209	0.233	0.148
	66410109	**rs11208774.C**	**0.548**	**0.264**	0.009	0.001	nan	nan	0.06	0.468	0.688	0.190	0.194	0.401	0.163	0.122	0.367
	66416747	**rs1937456.C**	**0.108**	**-0.074**	0.023	0.007	0.01	nan	0.01	0.621	0.889	0.772	0.653	0.445	0.337	0.467	0.724
	66435842	**rs2186123.A**	**-0.002**	**-0.160**	0.019	0.039	0.01	nan	0.00	0.632	0.628	0.777	0.681	0.560	0.430	0.300	0.679
	66439837	**rs2503205.C**	**0.083**	**NA**	nan	0.005	nan	nan	0.00	0.009	0.054	0.000	0.000	0.000	0.000	0.000	0.046
	66442338	**rs10889591.C**	**0.520**	**0.611**	0.001	0.031	nan	nan	nan	0.227	0.233	0.005	0.000	0.000	0.000	0.000	0.000
	66446672	**rs1937453.A**	**0.001**	**0.002**	0.002	0.030	0.00	nan	0.01	0.693	0.656	0.766	0.660	0.566	0.674	0.556	0.709
	66476754	**rs6588177.A**	**0.562**	**0.558**	0.014	0.058	0.10	nan	0.01	0.466	0.415	0.875	0.986	0.753	0.826	0.856	0.755
	66503107	**rs12033425.C**	**0.053**	**-0.115**	0.000	0.016	0.02	nan	0.00	0.109	0.207	0.234	0.278	0.451	0.000	0.011	0.102
	66511404	**rs1354060.A**	**0.008**	**-0.021**	0.000	0.001	0.00	nan	0.02	0.506	0.568	0.397	0.563	0.538	0.349	0.389	0.515
	66514257	**rs1500951.A**	**0.036**	**0.074**	nan	0.004	0.00	nan	0.00	0.043	0.114	0.114	0.188	0.126	0.105	0.078	0.005
	66538917	**rs12119734.C**	**0.144**	**0.346**	0.032	0.082	0.00	nan	0.01	0.784	0.889	1.000	0.993	0.995	0.907	0.900	0.791
	66541107	**rs12027416.C**	**0.052**	**-0.058**	0.002	0.015	0.06	nan	0.00	0.931	0.858	0.929	0.903	0.863	1.000	1.000	0.908
	66672532	**rs10454453.A**	**0.135**	**0.080**	0.002	0.012	0.00	nan	0.02	0.509	0.315	0.755	0.729	0.725	0.663	0.644	0.454
	66681508	**rs12045945.A**	**0.004**	**-0.005**	nan	nan	0.05	nan	nan	0.003	0.006	0.000	0.000	0.060	0.000	0.000	0.000
	66713685	**rs6661245.C**	**0.322**	**0.066**	0.046	0.006	0.00	nan	0.01	0.132	0.449	0.103	0.118	0.247	0.267	0.211	0.250
	66751901	**rs522037.C**	**0.173**	**-0.002**	**0.930**	0.013	0.01	nan	0.05	0.601	0.429	0.810	0.785	0.659	0.640	0.589	0.551
	66765778	**rs6683604.C**	**0.226**	**0.442**	**0.975**	0.061	0.00	nan	0.09	0.402	0.173	0.120	0.153	0.066	0.058	0.233	0.393
	66769100	**rs6683977.C**	**0.576**	**0.624**	**hit**	**hit**	**hit**	**nan**	**hit**	0.540	0.020	0.005	0.000	0.055	0.012	0.000	0.020
	66831370	**rs783036.A**	**0.096**	**0.013**	0.004	0.023	0.01	nan	0.00	0.506	0.824	0.609	0.660	0.500	0.558	0.578	0.587
MYT1L	1832273	**rs3924017.C**	**0.183**	**0.362**	0.036	nan	0.018	0.010	0.026	0.494	0.679	0.804	0.944	0.758	0.721	0.678	0.556
	1842774	**rs6728613.C**	**0.351**	**0.070**	0.009	nan	0.009	0.000	0.020	0.382	0.827	0.147	0.007	0.242	0.279	0.278	0.480
	1902562	**rs11127305.A**	**0.252**	**0.169**	0.114	nan	0.000	nan	0.150	0.911	0.844	0.995	1.000	0.973	1.000	1.000	0.990

1925993	**rs2241685.C**	**0.050**	**-0.002**	0.082	nan	0.020	0.000	0.048	0.888	0.756	0.793	0.979	0.934	0.953	0.900	0.878
1946914	**rs3748988.A**	**0.227**	**0.064**	0.025	nan	0.026	0.012	0.002	0.592	0.287	0.728	0.910	0.830	0.733	0.656	0.592
1946968	**rs3748989.C**	**0.015**	**-0.006**	0.092	nan	0.005	0.000	0.084	0.879	0.810	0.788	0.944	0.901	0.965	0.867	0.908
1959401	**rs12986730.A**	**0.112**	**-0.042**	0.001	nan	0.013	0.019	0.032	0.284	0.463	0.071	0.014	0.137	0.233	0.344	0.347
1962316	**rs12991351.G**	**0.019**	**0.051**	0.001	nan	nan	0.008	0.003	0.052	0.065	0.000	0.000	0.000	0.012	0.022	0.066
1978966	**rs6729968.A**	**0.326**	**-0.076**	0.014	nan	0.014	0.015	0.000	0.658	0.219	0.924	0.979	0.819	0.733	0.544	0.541
1984388	**rs6548050.C**	**0.160**	**0.147**	0.024	nan	0.013	0.001	0.029	0.155	0.233	0.000	0.000	0.077	0.047	0.078	0.189
1985248	**rs10178240.A**	**0.203**	**-0.097**	0.000	nan	0.012	0.049	0.025	0.141	0.491	0.076	0.007	0.104	0.174	0.278	0.270
1987907	**rs9284792.G**	**0.513**	**0.133**	0.004	nan	0.008	0.031	0.037	0.658	0.449	0.957	1.000	0.874	0.826	0.633	0.735
2027293	**rs12468174.A**	**0.076**	**0.042**	0.035	nan	0.010	0.001	0.000	0.095	0.145	0.000	0.000	0.093	0.047	0.056	0.107
2105842	**rs17039334.C**	**0.060**	**0.146**	0.001	nan	nan	nan	0.006	0.043	0.011	0.000	0.000	0.000	0.000	0.000	0.031
2176177	**rs11683072.G**	**0.011**	**-0.052**	0.000	nan	0.017	0.010	0.000	0.897	0.866	0.989	1.000	0.923	0.895	0.833	0.883
2190992	**rs6705656.C**	**0.581**	**0.805**	0.044	nan	0.015	0.002	0.072	0.560	0.372	0.005	0.000	0.082	0.093	0.167	0.240
2204895	**rs12468168.C**	**0.116**	**0.035**	0.071	nan	0.009	0.001	0.067	0.868	0.793	0.995	1.000	0.912	0.930	0.867	0.821
2212105	**rs17338491.A**	**0.032**	**0.079**	0.063	nan	0.020	0.000	0.134	0.115	0.097	0.000	0.000	0.088	0.058	0.089	0.117
2225144	**rs17039396.A**	**0.048**	**-0.076**	1.000	nan	1.000	1.000	1.000	0.144	0.074	0.228	0.361	0.220	0.070	0.078	0.051
2255182	**rs17039463.A**	**0.090**	**-0.032**	nan	nan	nan	0.008	0.005	0.009	0.054	0.000	0.000	0.000	0.023	0.011	0.026
2258795	**rs17039474.G**	**0.049**	**-0.002**	nan	nan	0.000	nan	0.005	0.011	0.063	0.005	0.000	0.016	0.000	0.000	0.026
2260418	**rs13034457.A**	**0.014**	**0.057**	0.046	nan	0.007	0.000	0.005	0.595	0.543	0.397	0.382	0.588	0.465	0.522	0.684
2311850	**rs13388663.A**	**0.153**	**0.156**	0.058	nan	0.019	0.055	0.012	0.431	0.449	0.533	0.556	0.516	0.744	0.700	0.561
2325234	**rs10195351.C**	**0.036**	**-0.055**	0.051	nan	0.002	0.061	0.013	0.741	0.824	1.000	1.000	0.995	0.767	0.633	0.510

Perspectiva – Breast Cancer pharmacogenetic regions are highly differentiated in Native Americans

SNPs presentes na via dos inibidores da aromatase (alvo de terapia hormonal de câncer de mama) foram descritos como diferenciados em populações com alto componente nativo americano (Soares-Souza 2010; Boone et al. 2014). Temos um trabalho em andamento, o qual sou primeira autora, com as mesmas amostras e mesma metodologia de genotipagem e análises do manuscrito da LLA visando verificar se há diferenças populacionais na estrutura haplotípica dos genes envolvidos no contexto do câncer de mama. Nossos resultados mostraram que populações com alta ancestralidade ameríndia possuem estrutura haplotípica muito diferenciada de outras populações mundiais para os genes desta via, o que pode implicar em diferenças importantes na formulação do tratamento e de dosagens de medicamentos para a terapia hormonal de câncer de mama nestes indivíduos. Um manuscrito está no início de sua preparação com perspectivas de submissão no final de 2016.

O resumo com resultados preliminares deste trabalho foi submetido e aprovado para apresentação oral no Congresso Internacional do X-meeting, 2013:

BREAST CANCER PHARMACOGENETIC REGIONS ARE HIGHLY DIFFERENTIATED IN NATIVE AMERICANS

F Rodrigues-Soares[1,3,4], RG Moreira[1,3], MH Gouveia[1], RH Gilman[2], R Zamudio[1], G Soares-Souza[1], MR Rodrigues[1], WCS Magalhães[1], E Tarazona-Santos[1]

[1] Laboratório de Diversidade Genética Humana. Departamento de Biologia Geral, Universidade Federal de Minas Gerais, Belo Horizonte, Brasil
[2] Universidade Peruana Cayetano Heredia, Peru and Johns Hopkins School of Public Health, USA.

[3] These authors contributed equally to this work
[4] fernandasoares@ufmg.br

Background: Previous studies by our group showed that some SNPs of the Aromatase Inhibitor Pathway, which is important in breast cancer treatment, might be differentiated in Native American populations. To exhaustively evaluate if there is a haplotype differentiation among Native Americans respect to other worldwide populations, we developed a bioinformatics framework to

select cosmopolitan tag-SNPs from HapMap or 1000 Genomes Project for specific genomic regions, that in this case encompasses four genes of the Aromatase Inhibitor Pathway (*CYP19A1, ESR1, ESR2* e *HSD17B1*). These SNPs were annotated in detail using the multi-agent annotation system (MASannote) developed by our group. Forty-eight tag-SNPs on these genes were genotyped using a customized kit and the BeadXpress platform (Illumina, US), and analyzed using bioinformatics QC and population genetics analysis procedure developed by our group (see the DIVERGENOMEtool platform available at http:// pggenetica.icb.ufmg.br/divergenome/pagina/dynamicpipeline/tools.php). We analyzed several worldwide populations, including original data from 384 individuals: three Peruvians Native American groups 72 Shimaa, 92 Ashaninkas and 88 Quechuas (the largest linguistic native South American group); four admixed Brazilians populations from Minas Gerais: 19 from Carmésia, 30 from Martinho Campos, 24 from Resplendor and 25 from São João das Missões. We compared these data with public database HapMap from European (CEU) ancestry and African (YRI) ancestry. Ancestry of our South American samples was estimated using a panel of 96 ancestry informative markers genotyped by the BeadXpress platform.

Results: Several population genetics analyses consistently show that Native Americans show an extremely differentiated haplotype structure when compared to Europeans, Africans and Brazilian admixed individuals.

Conclusions: The highly differentiated haplotype structure of Native Americans for genes involved deserve follow-up, to understand its clinical implication in the therapeutic efficacy of breast cancer therapy with aromatase inhibitors. The study of Native Americans is particularly important because this ethnic group is neglected in genomics and pharmacogenetics initiatives. We are currently evaluating the nucleotide diversity on these genes in detail, performing next-generation targeted sequencing of 16 Native Americans, using Sure Select Agilent technology. Funding: CNPq, CAPES, FAPEMIG.

DISCUSSÃO GERAL

Esta tese foi dividida em três capítulos que apresentam em comum o estudo genético da miscigenação populacional da América Latina nos âmbitos evolutivo (capítulo 1) e da farmacogenética (capítulos 2 e 3).

No primeiro capítulo foi feita uma caracterização genética em alta resolução da população brasileira, que apresentou história demográfica e padrões de ancestralidade complexos (Kehdy et al. 2015), que no nível continental estão correlacionados com cor/raça autodeclarada (Lima-Costa et al. 2015). Além disso, um painel mínimo de marcadores de ancestralidade foi criado visando reduzir custos de estudos de ancestralidade de alta resolução (Santos et al. 2015) e foi também verificado que a ancestralidade é uma variável a ser considerada em estudos de associação com populações miscigenadas (Zamudio et al. 2015).

Nos capítulos 2 e 3, foram realizados estudos com o intuito de caracterizar populações mundiais (através de revisões sistemáticas) e da América Latina no contexto de ancestralidade-cor/raça autodeclarada e variantes de interesse farmacogenético. As revisões sistemáticas, assim como as meta-análises, possuem amplo valor para respostas a perguntas científicas e resolução de controvérsias na bibliografia pois geralmente esgotam as fontes da literatura sobre um determinado tema, sumarizando todas as conclusões encontradas. De acordo com os resultados destes capítulos, foi encontrada uma diferenciação de frequências de variantes de interesse farmacogenético de acordo com as populações estudadas e suas ancestralidades. Em um contexto mundial, o padrão de frequências de alelos e fenótipos de CYPs é bastante diferenciado entre as populações (LLerena et al. 2014; Céspedes-Garro et al. 2015; Fricke-Galindo et al. 2015).

Os genes de interesse farmacogenético em populações da América Latina possuem padrões singulares de acordo com o padrão de ancestralidade biogeográfica estimada com base em dados genômicos e com cor/raça autodeclarados. Em indivíduos da Costa Rica, alguns alelos possuem forte correlação com a ancestralidade genômica e seguiram um padrão entre as cores/raças auto-declaradas (Céspedes-Garro et al. 2016 - *in press*). O padrão nas cores/raças autodeclaradas é encontrado porque apesar de a Costa Rica ser um país latino-americano e apresentar histórico de miscigenação, suas etnias são bem definidas, com pouca ou nenhuma miscigenação nos afro-caribenhos e ameríndios estudados; portanto, sem ancestralidades ocultas dentro das cores/raças, as correlações se tornam mais claras.

Em outros países da América Latina, como México e demais países da América Central, foi encontrada grande variabilidade interétnica nas populações estudadas utilizando os biomarcadores recomendados pela FDA e EMA (Céspedes-Garro et al. 2014; Fricke-Galindo et al. 2016). Curiosamente, o alelo *CYP2D6*4*, normalmente encontrado em populações europeias ou com grandes porcentagens de ancestralidade europeia (como o Brasil), apresentou valores de frequências significativamente altas para grupos de ameríndios da Costa Rica e do México (Céspedes-Garro et al. 2014; Fricke-Galindo et al. 2016). Como nestes estudos não foram feitas análises de ancestralidade individual, podemos supor que pode existir um componente europeu elevado nestes grupos. Em breve, novos resultados da RIBEF com dados de ancestralidade individual serão publicados e será esclarecido se o motivo desta alta frequência é proveniente de uma miscigenação europeia ou mutação independente em populações ameríndias.

Adicionalmente, os países da América Central são pouco estudados, com apenas 26 estudos de voluntários sadios em todos os países (Céspedes-Garro et al. 2014). Além disso, Belize e Honduras não possuem estudos farmacogenéticos com voluntários. Portanto, a América Central necessita de forma urgente de um maior número de estudos populacionais utilizando biomarcadores farmacogenéticos.

No Brasil, as cores/raças autodeclaradas possuem fatores não biológicos que influenciam a autodeclaração, como escolaridade, classe social e exposição ao sol (Friedrich et al. 2014). Estes fatores, juntamente com a ancestralidade complexa oculta em cada uma das cores/raças definidas pelo IBGE, tornam a população brasileira extremamente complexa com poucos padrões definidos de variantes de interesse farmacogenético (Rodrigues-Soares et al., dados não publicados).

Frequências de variantes de interesse farmacogenético variam em populações mundiais de acordo com sua ancestralidade. Atualmente, muitos estudos utilizam *arrays* (de uso geral ou especializados em biomarcadores farmacogenéticos) e *Next Generation Sequencing* (NGS) para genotipar genes relacionados à absorção, distribuição, metabolização e excreção de fármacos (ADME) - portanto de interesse farmacogenético - e relacionar estes resultados com a ancestralidade individual e cromossômica (Li et al. 2011; Ramos et al. 2014; Li, Lou, et al. 2014; Limdi et al. 2015; Li, Lao, et al. 2014). Um estudo recente em Afro-Americanos, publicado e editorializado (Price 2015) na prestigiosa revista *Blood*, recomenda que os atuais algoritmos utilizados para dosagem de varfarina devem ser estratificados por etnia, e não apenas ajustados por ela (Limdi et al. 2015), pois a ancestralidade é um fator determinante na recomendação da dose

deste medicamento. Neste estudo, os autores concluíram que mesmo realizando a genotipagem individual, a etnia autodeclarada é fundamental para a recomendação de dosagem da varfarina, pois mesmo com genótipos idênticos, descendentes de europeus e descendentes de africanos apresentaram diferenças no padrão de metabolismo, e consequente dosagem deste fármaco.

Resultados de ancestralidade cromossômica em 284 genes ADME para afro-americanos mostraram que estes indivíduos são um mosaico de haplótipos europeus e africanos, como esperado para resultados de marcadores aleatórios (Li, Lao, et al. 2014). Além disso, a análise de PCA de um estudo que levou em consideração 19 populações mundiais para 212 genes ADME mostrou que a distribuição gráfica dos indivíduos utilizando genes ADME foi similar à distribuição com marcadores aleatórios em todos os grupos continentais e étnicos (Ramos et al. 2014). Este estudo indica que existem limitações a respeito deste recente tipo de análises, como o baixo número de indivíduos por etnias mundiais e a escassez de biomarcadores farmacogenéticos validados para a maior parte dos fármacos no mercado. Além disso, a validação da maior parte destes biomarcadores e do desenvolvimento de fármacos é feito com populações de alta ancestralidade europeia (Li et al. 2011), o que poderia criar um viés nas conclusões. Por isso, é recomendado que estereótipos étnicos sejam evitados, principalmente em populações miscigenadas, pois o ideal para a prática clínica é a realização da genotipagem individualizada (Ramos et al. 2014). Estudos populacionais e de ensaios clínicos com indivíduos adicionais de múltiplas ancestralidades ainda são necessários enquanto a medicina genômica ainda começa a se incorporar na clínica. Os esforços da RIBEF, publicando resultados de ancestralidade e CYPs em toda a Ibeoamérica, estão orientados a cobrir as carências de estudos nestes locais, em especial na América Latina.

Nos Estados Unidos, os estudos de Ensaios Clínicos devem ser cadastrados no registro do *National Institutes of Health* (www.clinicaltrials.gov) para serem aprovados e regulamentados pelo governo. No Brasil, em 2012, o Ministério da Saúde criou uma base semelhante para este cadastramento, o Registro Brasileiro de Ensaios Clínicos (www.ensaiosclinicos.gov). No entanto, a base de dados dos Estados Unidos apresenta mais de 500 resultados quando se busca estudos com farmacogenética (termo "*pharmacogenetics*" no mecanismo de busca), enquanto no Brasil não se encontram Ensaios Clínicos cadastrados com esta temática. Inclusive, quando a busca é feita com o termo "genética", apenas quatro estudos brasileiros aparecem nos resultados da busca - não correspondem a farmacogenética - enquanto no site americano aparecem 11.960 estudos que levam este tema em consideração. Portanto, no nosso país há uma forte carência de estudos clínicos que

levam a genética e a farmacogenética em consideração, o que pode limitar conclusões e interpretações a respeito de predições clínicas apenas com a análise de genótipos nesta população com história demográfica complexa. O Brasil apresenta muitos estudos pequenos, com poucos indivíduos e poucas evidências clínicas ao invés de apresentar grandes estudos e ensaios clínicos, que podem ser mais valiosos para uso na prática clínica.

Recentemente, um grupo de pesquisa de Nashville, Tennessee, Estados Unidos, lançou um programa de pesquisa farmacogenômica (*Pharmacogenomic Resource for Enhanced Decisions in Care and Treatment* – PREDICT) para implementar a farmacogenética na prática clínica e melhorar o tratamento dos pacientes (Pulley et al. 2012). Interações gene-resposta a fármacos bem estabelecidas (*actionable genotypes*) são incorporadas no registro médico eletrônico e as informações destes SNPs são interrogadas para se obter um suporte de decisão clínica adequado (Van Driest et al. 2014). Cinco interações fármaco-gene (*actionable genotypes*) foram implementadas neste programa: *CYP2C19*-clopidogrel, *SLCO1B1*-simvastatina, *CYP2C9* e *VKORC1*-varfarina, *TPMT*-tiopurinas e *CYP3A5*-tacrolimus (Van Driest et al. 2014). No programa PREDICT, 10.000 pacientes utilizando um ou mais destes fármacos foram analisados e 91% apresentou pelo menos um *actionable genotype*, mostrando o benefício potencial deste tipo de estudo. Dentro do grupo dos afro-americanos, praticamente todos os indivíduos (96%) apresentaram um *actionable genotype*, enfatizando a necessidade de diversidade étnica em estudos genômicos (Van Driest et al. 2014).

Uma resposta brasileira à iniciativa do PREDICT foi publicada (Suarez-Kurtz 2014), afirmando que de acordo com os dados de brasileiros coletados pela REFARGEN, 93,5% dos indivíduos estudados possuem pelo menos um *actionable genotype*. Os autores do PREDICT enviaram uma nova resposta, afirmando que os resultados para o Brasil foram similares aos dos americanos, frisando a importância da genotipagem do maior número de populações possível para estas variantes (Van Driest & Roden 2014), corroborando sua alta prevalência independente da etnia estudada. No contexto do projeto EPIGEN, estamos quantificando os *actionable genotypes* nas maiores coortes de base populacional brasileiras.

A revisão sistemática de voluntários sadios brasileiros apresentada nesta tese realizou uma concatenação dos estudos com variantes de interesse farmacogenético no Brasil, frisando a importância de um maior número de estudos mais elaborados considerando fenótipos metabólicos. De acordo com esta revisão, o gene *GSTT1* é o mais estudado no país, tanto em número de

indivíduos como em número de estudos. Este biomarcador não é um clássico de estudos farmacogenéticos, então o motivo para ele ser o mais estudado no nosso país está relacionado com o sistema agrário brasileiro ser de forte impacto no mundo, portanto, o uso de pesticidas possui quantidades proporcionais de consumo. Consequentemente, os trabalhadores rurais expostos a estes pesticidas estão suscetíveis a contaminação por estes agentes químicos, que podem ser carcinogênicos. A enzima GSTT1 participa da detoxificação celular, portanto indivíduos que possuem a deleção de seu gene e consequentemente da enzima faz com que desenvolvam com mais facilidade processos carcinogênicos. A maior parte dos estudos encontrados são pequenos estudos de associação com trabalhadores rurais visando inferir o impacto dos pesticidas em sua saúde (Gattás et al. 2004; da Silva et al. 2008; Godoy et al. 2014). Além disso, há dois estudos relacionando consumo de peixe contaminado com mercúrio na Amazônia com processos carcinogênicos e a GSTT1 (Barcelos et al. 2013; de Oliveira et al. 2014).

Algumas limitações foram encontradas na elaboração dos trabalhos presentes nesta tese. Para as revisões, o número de trabalhos com populações da América Latina é ainda insuficiente para obter uma conclusão robusta do cenário farmacogenético desta região. Além disso, dados de fenótipos metabólicos medidos por fármacos teste é uma limitação encontrada mundialmente, portanto um maior número de estudos com este propósito são necessários. Para dados originais, o número de amostras é uma limitação a ser considerada, pois números baixos podem não ser representativos da população geral. Ainda, deve-se levar em consideração que a população brasileira estudada a partir de dados originais (dados ainda não publicados pela iniciativa RIBEF) é composta por indivíduos miscigenados de vários pequenos municípios com histórico indígena de Minas Gerais e podem não representar a população brasileira de forma abrangente.

Em geral, a nomenclatura e natureza genética das variantes de interesse farmacogenético são diversas e podem dificultar as análises dos resultados. Alguns genes apresentam simplesmente uma deleção gênica (*GSTT1, GSTM1*), outros genes apresentam haplótipos complexos compostos por vários SNPs (*CYPs, TPMT, LDLR*) e outros apresentam simplesmente SNPs. Esta é uma limitação de estudos farmacogenéticos em geral, pois as diferenças metodológicas para a identificação destas variantes podem influenciar nos resultados ou na conclusão final do estudo. Por exemplo, em genes que os alelos de interesse são haplótipos, como no caso do *CYP2D6*, em geral faz-se a genotipagem de somente um SNP representativo do haplótipo correspondente. Muitas vezes, os autores não especificam o SNP que está sendo genotipado para representar o haplótipo, e quando identificam,

muitas vezes o fazem com nomenclatura ambígua (ex: -1584C>G e -1042C>G para o haplótipo *CYP2D6*2*, que correspondem ao rs1080985). A nomenclatura de SNPs poderia ser preferivelmente identificada com o "rs" que foi genotipado para evitar ambiguidades e problemas de interpretação dos resultados. Fizemos a procura de todos os "rs" dos alelos analisados na revisão sistemática de voluntários sadios no Brasil na base de dados PharmGKB (Whirl-Carrillo et al. 2012). Os resultados estão apresentados na Tabela S2 do capítulo 2 (p. 95).

Outra importante limitação em estudos farmacogenéticos é a correlação genótipo-fenótipo. De maneira geral, o genótipo não é a predição exata do fenótipo metabólico do indivíduo, pois vários fatores contribuem para a variação na resposta a fármacos como estilo de vida, idade, níveis hormonais, funções hepáticas, entre outros (Meyer 2000). No caso de metabolizadores lentos de CYPs, a correlação é encontrada porque se o indivíduo não produz a enzima, mesmo que possua fatores que acelerem seu metabolismo, não apresentará resposta ao fármaco (LLerena et al. 2014). Este fato é válido para deleções gênicas e homozigose de alelos que não produzem a enzima. No entanto, para metabolizadores ultra-rápidos, a resposta pode ser mais variável, pois cada indivíduo terá uma quantidade e tipo de enzima distinto de outros. Além disso, há poucos estudos fenotípicos de metabolizadores ultra-rápidos no mundo, o que dificulta expressivamente a correlação (LLerena et al. 2014; Céspedes-Garro et al. 2015).

Portanto, para implementação direta na prática clínica, a farmacogenética deve ser aplicada com a genotipagem individual de cada paciente (medicina personalizada), para garantir uma previsão mais acurada da resposta ao tratamento. No contexto da farmacogenética populacional, que seria mais aplicável em estudos epidemiológicos, informações de ancestralidade e raça/cor são importantes para a predição de probabilidade de resposta a tratamentos, de acordo com a ancestralidade/raça/cor que o indivíduo possui. Como cada alelo de cada gene segue padrões mundiais populacionais, esta predição pode auxiliar em estudos epidemiológicos, quando não for possível realizar a genotipagem individual para todos os sujeitos do estudo em questão.

CONCLUSÃO

Variantes de interesse farmacogenético possuem padrões distintos entre diferentes populações mundiais, de modo que cada população possui uma identidade farmacogenética. No entanto, populações miscigenadas com história demográfica complexa, como as latino-americanas, em especial a brasileira, podem não apresentar padrões de frequências destas variantes, pois a complexidade genética herdada de populações parentais torna estes indivíduos mosaicos de seus antepassados, independente da etnia autodeclarada. Portanto, a aplicação da farmacogenética na prática clínica em populações com ancestralidade complexa como as da América Latina deveria ser feita por genotipagem individualizada pois o fenótipo metabólico não pode ser predito pela ancestralidade populacional e individual. Entretanto, em um contexto populacional e epidemiológico, as informações de cor/raça e ancestralidade genômica são informativas para inferências farmacogenéticas.

Um maior número de estudos étnicos de interesse farmacogenético, tanto a partir de genótipos quanto a partir de fenótipos metabólicos, deve ser realizado para que conclusões robustas sejam elaboradas e implementadas na prática clínica.

REFERÊNCIAS

Alexander, D.H., Novembre, J. & Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, 19(9), pp.1655–64.

Anon, 2003. The International HapMap Project. *Nature*, 426(6968), pp.789–96.

Barcelos, G.R.M. et al., 2013. Polymorphisms in glutathione-related genes modify mercury concentrations and antioxidant status in subjects environmentally exposed to methylmercury. *The Science of the total environment*, 463-464, pp.319–25.

Barreto, M.L. et al., 2006. Risk factors and immunological pathways for asthma and other allergic diseases in children: background and methodology of a longitudinal study in a large urban center in Northeastern Brazil (Salvador-SCAALA study). *BMC pulmonary medicine*, 6, p.15.

Barros, A.J.D. et al., 2006. [The 2004 Pelotas birth cohort: methods and description]. *Revista de saúde pública*, 40(3), pp.402–13.

Bertilsson, L., 1995. Geographical/interracial differences in polymorphic drug oxidation. Current state of knowledge of cytochromes P450 (CYP) 2D6 and 2C19. *Clinical pharmacokinetics*, 29(3), pp.192–209.

Boone, S.D. et al., 2014. Associations between CYP19A1 polymorphisms, Native American ancestry, and breast cancer risk and mortality: the Breast Cancer Health Disparities Study. *Cancer causes & control : CCC*, 25(11), pp.1461–71.

Cavalli-Sforza, L.L., 1998. The DNA revolution in population genetics. *Trends in genetics : TIG*, 14(2), pp.60–5.

Céspedes-Garro, C. et al., 2014. Pharmacogenetics in Central American healthy volunteers: interethnic variability. *Drug metabolism and drug interactions*, pp.1–13.

Céspedes-Garro, C. et al., 2016. Relevance of the ancestry for the variability of the Drug-Metabolizing Enzymes CYP2C9, CYP2C19 and CYP2D6 polymorphisms in a multiethnic Costa Rican population. *Revista de Biología Tropical*, in press.

Céspedes-Garro, C. et al., 2015. Worldwide interethnic variability and geographical distribution of CYP2C9 genotypes and phenotypes. *Expert opinion on drug metabolism & toxicology*, pp.1–13.

Do, C.B. & Batzoglou, S., 2008. What is the expectation maximization algorithm? *Nature biotechnology*, 26(8), pp.897–9.

Dorado, P. et al., 2006. Clinical implications of CYP2D6 genetic polymorphism during treatment with antipsychotic drugs. *Current drug targets*, 7(12), pp.1671–80.

Dorado, P. et al., 2003. CYP2C9 genotypes and diclofenac metabolism in Spanish healthy

volunteers. *European journal of clinical pharmacology*, 59(3), pp.221–5.

Dorado, P. et al., 2005. Development of a PCR-based strategy for CYP2D6 genotyping including gene multiplication of worldwide potential use. *BioTechniques*, 39(4), pp.571–4.

Van Driest, S.L. et al., 2014. Clinically actionable genotypes among 10,000 patients with preemptive pharmacogenomic testing. *Clinical pharmacology and therapeutics*, 95(4), pp.423–31.

Van Driest, S.L. & Roden, D.M., 2014. Response to "clinically actionable genotypes among Brazilians". *Clinical pharmacology and therapeutics*, 96(1), p.26. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24682028 [Accessed May 11, 2016].

Fejerman, L. et al., 2010. European ancestry is positively associated with breast cancer risk in Mexican women. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 19(4), pp.1074–82.

Flores, C. et al., 2012. African ancestry is associated with asthma risk in African Americans. *PloS one*, 7(1), p.e26807.

Fricke-Galindo, I. et al., 2016. Interethnic variability of pharmacogenetic biomarkers in Mexican healthy volunteers: a report from the RIBEF (Ibero-American Network of Pharmacogenetics and Pharmacogenomics). *Drug metabolism and personalized therapy*.

Fricke-Galindo, I. et al., 2015. Interethnic variation of CYP2C19 alleles, "predicted" phenotypes and "measured" metabolic phenotypes across world populations. *The pharmacogenomics journal*.

Friedrich, D.C. et al., 2014. Distribution of CYP2D6 alleles and phenotypes in the Brazilian population. *PloS one*, 9(10), p.e110691.

Galanter, J.M. et al., 2012. Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. *PLoS genetics*, 8(3), p.e1002554.

Gattás, G.J.F. et al., 2004. Ethnicity and glutathione S-transferase (GSTM1/GSTT1) polymorphisms in a Brazilian population. *Brazilian journal of medical and biological research = Revista brasileira de pesquisas médicas e biológicas / Sociedade Brasileira de Biofísica ... [et al.]*, 37(4), pp.451–8.

Godoy, F.R. et al., 2014. Do GSTT1 and GSTM1 polymorphisms influence intoxication events in individuals occupationally exposed to pesticides? *Environmental science and pollution research international*, 21(5), pp.3706–12.

IBGE, 2010. Censo demográfico 2010. Available at: http://www.ibge.gov.br/home/estatistica/populacao/censo2010/default.shtm [Accessed March

7, 2016].

IBGE, Indicadores Sociais Mínimos - Conceitos. Available at: http://www.ibge.gov.br/home/estatistica/populacao/condicaodevida/indicadoresminimos/conceitos.shtm.

Kehdy, F.S.G. et al., 2015. Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proceedings of the National Academy of Sciences*, p.201504447.

Khan, K.S. et al., 2003. Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine*, 96(3), pp.118–21.

Li, J., Lou, H., et al., 2014. Genetic architectures of ADME genes in five Eurasian admixed populations and implications for drug safety and efficacy. *Journal of medical genetics*, 51(9), pp.614–22.

Li, J. et al., 2011. Global patterns of genetic diversity and signals of natural selection for human ADME genes. *Human molecular genetics*, 20(3), pp.528–40.

Li, J., Lao, X., et al., 2014. Increased genetic diversity of ADME genes in African Americans compared with their putative ancestral source populations and implications for pharmacogenomics. *BMC genetics*, 15, p.52.

Lima-Costa, M.F. et al., 2015. Genomic ancestry and ethnoracial self-classification based on 5,871 community-dwelling Brazilians (The Epigen Initiative). *Scientific reports*, 5, p.9812.

Lima-Costa, M.F., Firmo, J.O.A. & Uchoa, E., 2011. Cohort profile: the Bambui (Brazil) Cohort Study of Ageing. *International journal of epidemiology*, 40(4), pp.862–7.

Limdi, N.A. et al., 2015. Race influences warfarin dose changes associated with genetic factors. *Blood*, 126(4), pp.539–45.

LLerena, A. et al., 2014. Interethnic variability of CYP2D6 alleles and of predicted and measured metabolic phenotypes across world populations. *Expert opinion on drug metabolism & toxicology*, 10(11), pp.1569–83.

Llerena, A. et al., 1993. Relationship between personality and debrisoquine hydroxylation capacity. Suggestion of an endogenous neuroactive substrate or product of the cytochrome P4502D6. *Acta psychiatrica Scandinavica*, 87(1), pp.23–8.

Meyer, U.A., 2000. Pharmacogenetics and adverse drug reactions. *Lancet (London, England)*, 356(9242), pp.1667–71.

de Oliveira, A.Á.S. et al., 2014. Genetic polymorphisms in glutathione (GSH-) related genes affect the plasmatic Hg/whole blood Hg partitioning and the distribution between inorganic and

methylmercury levels in plasma collected from a fish-eating population. *BioMed research international*, 2014, p.940952.

Pena, S.D.J. et al., 2011. The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PloS one*, 6(2), p.e17063.

Pereira, L. et al., 2012. Socioeconomic and nutritional factors account for the association of gastric cancer with Amerindian ancestry in a Latin American admixed population. *PloS one*, 7(8), p.e41200.

Pfaff, C.L. et al., 2004. Information on ancestry from genetic markers. *Genetic epidemiology*, 26(4), pp.305–15.

Price, A.L. et al., 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*, 38(8), pp.904–9.

Price, E.T., 2015. Warfarin pharmacogenomics and African ancestry. *Blood*, 126(4), pp.434–6.

Pulley, J.M. et al., 2012. Operational implementation of prospective genotyping for personalized medicine: the design of the Vanderbilt PREDICT project. *Clinical pharmacology and therapeutics*, 92(1), pp.87–95.

Ramos, E. et al., 2014. Pharmacogenomics, ancestry and clinical decision making for global populations. *The pharmacogenomics journal*, 14(3), pp.217–22.

Rodeiro, I. et al., 2012. Pharmacogenetics in Latin American populations: regulatory aspects, application to herbal medicine, cardiovascular and psychiatric disorders. *Drug metabolism and drug interactions*, 27(1), pp.57–60.

Rosenberg, N. a et al., 2002. Genetic structure of human populations. *Science (New York, N.Y.)*, 298(5602), pp.2381–5.

Salzano, F.M. & Bortolini, M.C., 2002. *The evolution and genetics of Latin American populations*, Cambridge, United Kingdom: Cambridge University Press.

Santos, H.C. et al., 2015. A minimum set of ancestry informative markers for determining admixture proportions in a mixed American population: the Brazilian set. *European journal of human genetics : EJHG*.

Shimada, T. et al., 1994. Interindividual variations in human liver cytochrome P-450 enzymes involved in the oxidation of drugs, carcinogens and toxic chemicals: studies with liver microsomes of 30 Japanese and 30 Caucasians. *The Journal of pharmacology and experimental therapeutics*, 270(1), pp.414–23.

da Silva, J. et al., 2008. Evaluation of genetic damage in a Brazilian population occupationally exposed to pesticides and its correlation with polymorphisms in metabolizing genes.

Mutagenesis, 23(5), pp.415–22.

Soares-Souza, G.B., 2010. *Identificação de genes com alta diferenciação entre populações humanas: inferências evolutivas e aplicações biomédicas*. Universidade Federal de Minas Gerais.

Suarez-Kurtz, G., 2014. Clinically actionable genotypes among Brazilians. *Clinical pharmacology and therapeutics*, 96(1), pp.25–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24598719 [Accessed May 11, 2016].

Suarez-Kurtz, G., 2004. Pharmacogenomics in admixed populations: the Brazilian pharmacogenetics/pharmacogenomics network--REFARGEN. *The pharmacogenomics journal*, 4(6), pp.347–8.

Tarazona-Santos, E., Raimondi, S. & Fuselli, S., 2007. Chapter 2: Controlling the effects of population stratification by admixture in pharmacogenetics. In G. Suarez-Kurtz, ed. *Pharmacogenomics in Admixed populations*. Austin, TX: Landes Bioscience, pp. 12–27.

Via, M., Ziv, E. & Burchard, E.G., 2009. Recent advances of genetic ancestry testing in biomedical research and direct to consumer testing. *Clinical genetics*, 76(3), pp.225–35.

Whirl-Carrillo, M. et al., 2012. Pharmacogenomics knowledge for personalized medicine. *Clinical pharmacology and therapeutics*, 92(4), pp.414–7.

Wollstein, A. & Lao, O., 2015. Detecting individual ancestry in the human genome. *Investigative genetics*, 6, p.7.

Yaeger, R., Avila-bront, A., et al., 2008. Comparing Genetic Ancestry and Self-Described Race in African Americans Born in the United States and in Africa Comparing Genetic Ancestry and Self-Described Race in African Americans Born in the United States and in Africa. , pp.1329–1338.

Yaeger, R., Avila-Bront, A., et al., 2008. Comparing genetic ancestry and self-described race in african americans born in the United States and in Africa. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, 17(6), pp.1329–38.

Yang, J.J. et al., 2011. Ancestry and pharmacogenomics of relapse in acute lymphoblastic leukemia. *Nature genetics*, 43(3), pp.237–41.

Zamudio, R. et al., 2015. Population, Epidemiological, and Functional Genetics of Gastric Cancer Candidate Genes in Peruvians with Predominant Amerindian Ancestry. *Digestive Diseases and Sciences*.

ANEXOS

Anexo A – A minimum set of ancestry informative markers for determining admixture proportions in a mixed American population: the Brazilian set

European Journal of Human Genetics (2015), 1–7

www.nature.com/ejhg

ARTICLE

A minimum set of ancestry informative markers for determining admixture proportions in a mixed American population: the Brazilian set

Hadassa C Santos[1], Andréa VR Horimoto[1], Eduardo Tarazona-Santos[2], Fernanda Rodrigues-Soares[2], Mauricio L Barreto[3], Bernardo L Horta[4], Maria F Lima- Costa[5], Mateus H Gouveia[2], Moara Machado[2], Thiago M Silva[3], José M Sanches[1], Nubia Esteban[1], Wagner CS Magalhaes[2], Maíra R Rodrigues[2], Fernanda SG Kehdy[2] and Alexandre C Pereira*[,1] The Brazilian EPIGEN Project Consortium

The Brazilian population is considered to be highly admixed. The main contributing ancestral populations were European and African, with Amerindians contributing to a lesser extent. The aims of this study were to provide a resource for determining and quantifying individual continental ancestry using the smallest number of SNPs possible, thus allowing for a cost- and time-efficient strategy for genomic ancestry determination. We identified and validated a minimum set of 192 ancestry informative markers (AIMs) for the genetic ancestry determination of Brazilian populations. These markers were selected on the basis of their distribution throughout the human genome, and their capacity of being genotyped on widely available commercial platforms. We analyzed genotyping data from 6487 individuals belonging to three Brazilian cohorts. Estimates of individual admixture using this 192 AIM panels were highly correlated with estimates using ~370 000 genome-wide SNPs: 91%, 92%, and 74% of, respectively, African, European, and Native American ancestry components. Besides that, 192 AIMs are well distributed among populations from these ancestral continents, allowing greater freedom in future studies with this panel regarding the choice of reference populations. We also observed that genetic ancestry inferred by AIMs provides similar association results to the one obtained using ancestry inferred by genomic data (370 K SNPs) in a simple regression model with rs1426654, related to skin pigmentation, genotypes as dependent variable. In conclusion, these markers can be used to identify and accurately quantify ancestry of Latin Americans or US Hispanics/Latino individuals, in particular in the context of fine-mapping strategies that require the quantification of continental ancestry in thousands of individuals.
European Journal of Human Genetics advance online publication, 23 September 2015; doi:10.1038/ejhg.2015.187

Anexo B – Population, Epidemiological, and Functional Genetics of Gastric Cancer Candidate Genes in Peruvians with Predominant Amerindian Ancestry

Dig Dis Sci
DOI 10.1007/s10620-015-3859-6

ORIGINAL ARTICLE

Population, Epidemiological, and Functional Genetics of Gastric Cancer Candidate Genes in Peruvians with Predominant Amerindian Ancestry

Roxana Zamudio[1,2] · Latife Pereira[1] · Carolina D. Rocha[4] · Douglas E. Berg[5] · Thaís Muniz-Queiroz[1] · Hanaisa P. Sant Anna[1] · Lilia Cabrera[2,3] · Juan M. Combe[6] · Phabiola Herrera[2,3] · Martha H. Jahuira[2,3] · Felipe B. Leão[4] · Fernanda Lyon[1] · William A. Prado[7] · Maíra R. Rodrigues[1] · Fernanda Rodrigues-Soares[1] · Meddly L. Santolalla[1] · Camila Zolini[1] · Aristóbolo M. Silva[4] · Robert H. Gilman[2,3,8] · Eduardo Tarazona-Santos[1] · Fernanda S. G. Kehdy[1,9]

Received: 15 March 2015 / Accepted: 25 August 2015

Abstract
Background Gastric adenocarcinoma is associated with chronic infection by *Helicobacter pylori* and with the host inflammatory response triggered by it, with substantial inter-person variation in the immune response profile due to host genetic factors.
Aim To investigate the diversity of the proinflammatory genes *IL8*, its receptors and *PTGS2* in Amerindians; to test whether candidate SNPs in these genes are associated with gastric cancer in an admixed population with high Amerindian ancestry from Lima, Peru; and to assess whether an *IL8RB* promoter-derived haplotype affects gene expression.
Methods We performed a Sanger-resequencing population survey, a candidate-gene association study (220 cases, 288 controls) and meta-analyses. We also performed an in vitro validation by a reporter gene assay of *IL8RB* promoter.
Results The diversity of the promoter of studied genes in Native Americans is similar to Europeans. Although an association between candidate SNPs and gastric cancer was not found in Peruvians, trend in our data is consistent with meta-analyses results that suggest *PTGS2*-rs689466-A is associated with *H. pylori*-associated gastric cancer in East Asia. *IL8RB* promoter-derived haplotype (rs3890158-A/rs4674258-T), common in Peruvians, was up-regulated by TNF-α unlike the ancestral haplotype (rs3890158-G/rs4674258-C). Bioinformatics analysis suggests that this effect stemmed from creation of a binding site for the FOXO3 transcription factor by rs3890158G>A.
Conclusions Our updated meta-analysis reinforces the role of *PTGS2*-rs689466-A in gastric cancer in Asians, although more studies that control for ancestry are necessary to clarify its role in Latin Americans. Finally, we suggest that *IL8RB*-rs3890158G>A is a *cis*-regulatory SNP.

Keywords Amerindians · Ancestry · Association studies · Gastric cancer · Meta-analyses · Proinflammatory genes

Roxana Zamudio and Latife Pereira have equally contributed to this paper.

Electronic supplementary material The online version of this article (doi:10.1007/s10620-015-3859-6) contains supplementary material, which is available to authorized users.

Review

EXPERT OPINION

Taylor & Francis
Taylor & Francis Group

Worldwide interethnic variability and geographical distribution of CYP2C9 genotypes and phenotypes

Carolina Céspedes-Garro[*], Ingrid Fricke-Galindo[*], María Eugenia G Naranjo, Fernanda Rodrigues-Soares, Humberto Fariñas, Fernando de Andrés, Marisol López-López, Eva M Peñas-Lledó & Adrián LLerena[†]
[†]*CICAB Clinical Research Centre, Extremadura University Hospital and Medical School, Badajoz 06080, Spain*

Introduction: Notably differences in *CYP2C9* allele frequencies among worldwide populations have been reported, with an interesting low frequency of the *CYP2C9*2* allele in Amerindians compared with Admixed and European populations.
Areas covered: Literature was searched using the PubMed database and was focused on worldwide original research papers on *CYP2C9* alleles and CYP2C9 phenotypes ("predicted" from *CYP2C9* genotypes and "measured" metabolic phenotype with a probe drug) among healthy volunteers according to their ethnicity and geographical distribution. Seventy-eight original research articles including a total of 31,978 subjects were identified.
Expert opinion: *CYP2C9*2* allele is the most frequent in Caucasian populations (average 14%), with the lowest frequencies for Africans (0.46%), East Asians (0.56%) and Native Americans (1.25%), which is in agreement with the hypothesis about the low prevalence in Amerindians. *CYP2C9*3* shows the highest frequency among South Asians (11.7%), while *CYP2C9*5* (1.56%) and **8* (4.70%) in African Americans. The predicted poor metabolizers (gPMs) were found overall in a low frequency, with the highest frequency detected for South Asians, in accordance with the *CYP2C9*3* frequency in these populations. This study shows the worldwide variability in the *CYP2C9* allele frequencies across different ethnic and geographic groups. Data about CYP2C9 "measured" metabolic phenotypes is still limited.

Keywords: CYP2C9, ethnicity, genotype, phenotype

Expert Opin. Drug Metab. Toxicol. [Early Online]

1. Introduction

Inter-individual variation in the drug metabolism is greatly explained by genetic polymorphisms [1] and other factors, such as age, gender and tobacco, caffeine or ethanol consumption.[2–4] Moreover, this variability influences in the drug response and the occurrence of adverse drug reactions.[1,5] CYP2C9 is one of the major cytochrome P450 enzymes, and it is the responsible for the phase I metabolism of several commonly used drugs such as oral anticoagulants, non-steroidal anti-inflammatory drugs, hypoglycemic agents, fluoxetine and antiepileptic drugs,[6] some of them with narrow therapeutic index.[7] *CYP2C9* gene is polymorphic, with more than 60 allelic variants described up to date. Some *CYP2C9* alleles are related to a null hydroxylation capacity (i.e. *CYP2C9*6, *15, *25* and **35*), whereas most of the other variants described so far are

*These authors contributed equally to this work.

10.1517/17425255.2015.1111871 © 2015 Taylor & Francis. ISSN 1742-5255, e-ISSN 1744-7607

Article highlights.

- African and Oceanian populations are under-represented in CYP2C9 studies across worldwide.
- *CYP2C9*2* frequency was higher in Caucasians than in the other groups.
- Lower frequencies of *CYP2C9*2* were found among Native Americans (1.25%) when compared with American admixed populations (7.12%) and the frequency in American admixed populations was also lower than those reported for Europeans.
- *CYP2C9*3* presented the highest frequency in South Asians, while *CYP2C9*5, *6* was mostly found among African groups.
- A low frequency of "predicted" poor metabolizers from genotype (gPMs) was found in all groups. The maximum frequencies have been found in South Asians and Central Europeans.
- *CYP2C9* "measured" metabolic phenotype has not been very extensively studied.

This box summarizes key points contained in the article.

associated with reduced enzyme capacity (e.g., *CYP2C9*2, *3, *5* and **11*).[8] Furthermore, a single nucleotide polymorphism (SNP) recently described (*CYP2C9* IVS8-109A>T) has been related to lower CYP2C9 hydroxylation capacity in an Ecuadorian population,[9] although to higher hydroxylation capacity in a Swedish population.[10]

To determine the actual CYP2C9 activity, two approaches are utilized: the "measured" metabolic phenotype by administering specific probe drugs and most frequently, the "predicted" phenotype from *CYP2C9* genotypes. The "measured" CYP2C9 metabolic phenotype has been determined by administering in healthy volunteers a single dose of probe drugs like tolbutamide, phenytoin, losartan, diclofenac, and warfarin.[11–18] Thus, those individuals with an impaired enzyme activity are identified as poor metabolizers (mPMs). But also, individuals who carry two null alleles or the reduced *CYP2C9*3* allele are identified as genotypically "predicted" poor metabolizers (gPMs).

Some *CYP2C9* alleles and the gPM and mPM groups have been related to important clinical implications. For instance, individuals carriers of *CYP2C9*3* and/or *CYP2C9*2* alleles present an increased risk of over-anticoagulation after warfarin administration,[19,20] and a higher relative risk of acute upper gastrointestinal bleeding has been reported in patients under non-steroidal anti-inflammatory drugs treatment.[21] Consequently, the Food and Drug Administration (FDA) has included pharmacogenetic information on *CYP2C9* genotype to be considered when patients are treated with the drugs celecoxib, flurbiprofen and/or warfarin.[22] Therefore, severe cases of neurotoxicity have been described in patients treated with phenytoin who exhibited higher drug plasma levels and were shown to carry genetic variants that encoded for an enzyme with reduced hydroxylation capacity.[23–25]

It has also been reported that CYP2C9 plays an important role on depression. CYP2C9 has been related to the metabolism of endogenous substrates like melatonin, which is involved in several processes such as circadian rhythms, mood, reproduction, ageing, etc. [26]; and arachidonic acid, that has been related to depression.[27] Interestingly, a higher frequency of *CYP2C9*3* allele was found among patients diagnosed with major depression when compared with schizophrenic patients and healthy volunteers,[28] and this very same allele was related to the lack of adherence to fluoxetine treatment in Mexican–American depressive patients.[29]

Finally, it is well known the interethnic variability in the frequencies of *CYP* alleles or metabolic groups.[30–33] With regard to *CYP2C9* allelic variants, *CYP2C9*2* is quite frequent in Caucasians, whereas in Asian populations the *CYP2C9*3* is the most frequent allele found.[34–36] In Africans, other *CYP2C9* alleles have been reported to be present at a high prevalence.[24,37] We have previously reported a significant lower *CYP2C9*2* frequency in Mexican Amerindian populations than in Mexican admixed and Spaniards, hypothesizing a negative correlation between the *CYP2C9*2* frequency and the degree of Asian ancestry.[38–40]

The aim of this study was to evaluate the worldwide distribution of the *CYP2C9* alleles, "predicted" phenotypes from genotypes (gPMs) and "measured" metabolic phenotypes (mPMs) among healthy volunteers from different ethnic groups and geographic areas; as well as to determine the relationship between the "predicted" and "measured" CYP2C9 metabolic phenotypes.

2. Review methods

The bibliographic search was conducted using the PubMed database by including the terms 'CYP2C9' and one by one, the names of world countries. Studies on "measured" metabolic phenotypes were obtained by searching the terms 'CYP2C9' and the probe drugs utilized for *in vivo* CYP2C9 phenotyping studies, that is, losartan, tolbutamide, phenytoin, diclofenac and warfarin.

Studies were selected according to the following inclusion criteria: *CYP2C9* allelic frequencies and/or measured metabolic phenotypes were reported, articles in which one or more *CYP2C9* alleles were studied, participants selected were healthy volunteers and at least 50 individuals were studied, except for those reports from Native populations from America, Oceania and the Circumpolar region, according to previous methods [32,41] (Figure 1).

Subjects included in this study were each different and not overlapping in the samples as stated in the original research article. Furthermore, individuals were assumed to be unrelated as the original research articles included in this study were aimed at calculating populations' frequencies. Studies

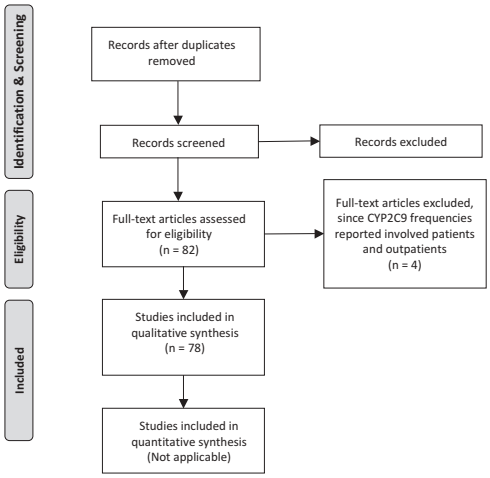

Figure 1. Flow diagram of studies included in analyses.

including subjects selected as controls or patients from case–control studies were excluded.

As a result, 78 original research papers on CYP2C9 were selected out of 82 preselected studies; four were excluded since *CYP2C9* allele and phenotype frequencies were calculated from healthy volunteers, patients and outpatients. [18,42–44] Selected articles were then classified by either geographic region or ethnicity. Additionally, the bibliographic search has covered papers published from January 1999 to January 2015.

Six major geographic regions were considered: Africa, America, Asia, Europe, Oceania and the Middle East. The ethnic groups were classified following previous methods. [32,41] However, Asians were divided into South Asians (involving people from India and Romani population living in Hungary), East Asians (people living in China, Japan, Korea, Malaysia, Vietnam, and Thailand) and North Asians (non-Caucasian Russian populations) (Table 1).

The frequencies of *CYP2C9* alleles, the "predicted" phenotype from genotype and "measured" metabolic phenotypes were calculated for each geographic region and ethnic group. To determine the frequencies of "predicted" phenotypes from *CYP2C9* genotype, only those individuals homozygous for *CYP2C9*3* allele or for null activity alleles, such as *CYP2C9*6*, were classified as poor metabolizers (gPMs). For "measured" metabolic phenotype analysis, subjects reported as poor metabolizers according to the CYP2C9 metabolic activity determined by administering a specific probe drug, were also presented in this review as poor metabolizers (mPMs). CYP2C9 frequencies calculated for geographic regions and ethnic groups could vary according to the populations included for each analysis.

The frequencies calculated for each geographic region and ethnic group were compared by a chi-squared test with Yate's correction, using GraphPad Software (GraphPad Software Inc., La Jolla, CA).[114] The allelic frequencies were compared versus a weighted average of the allelic frequencies of the remaining groups and to calculate the weighted average, the total number of subjects studied was taken into account. SPSS Base 19.0 (SPSS Inc., Chicago, IL) was used for the simple linear regression correlation analysis between CYP2C9 gPMs and mPMs.

Nevertheless, the methodological approach utilized in this study has some limitations: the single nucleotide polymorphisms (SNPs) studied for the *CYP2C9* alleles might differ from one study to another. However, the *CYP2C9* alleles from the original studies are used independently of the genotyped SNPs. Furthermore, the *CYP2C9*1* frequency is not calculated since this allele is determined by the absence of other *CYP2C9* SNPs and consequently, it depends on the number of the polymorphisms studied in the original article.

3. Worldwide CYP2C9 studies

Reports on CYP2C9 included in the present study comprised a total of 31,978 healthy volunteers (4.55×10^{-4}% of the world population), of whom 40.58% were Asians, 27.25% Americans, 17.91% Europeans, 7.91% were Africans, 6.16% Middle Easterns and 0.19% were Oceanians. More specifically, *CYP2C9* alleles were studied in 31,871 subjects, whereas the CYP2C9 "predicted" phenotypes from genotype were determined in 27,143 subjects. However, the CYP2C9 mPM was calculated only in 301 subjects (0.94% of the total number of subjects that studied CYP2C9).

According to the search results, frequencies of 41 *CYP2C9* allelic variants were reported, although 60 allelic variants are included in the Human Cytochrome P450 (CYP) Allele Nomenclature Database.[8] Within the analyzed alleles, *CYP2C9*2* and *CYP2C9*3* have been reported for most of the populations (Table 1), whereas other alleles like *CYP2C9*9, CYP2C9*10* and *CYP2C9*12* have only been assessed in Black Africans and East Asians. Moreover, *CYP2C9*2* and **3* alleles have not been extensively studied in Circumpolar and Native Oceanian populations. Minor *CYP2C9* alleles, such as *CYP2C9*14, *16, *19, *23, *27, *29, *33* or **34*, have been reported only for East Asians, for which a low frequency has been found [19] (data not shown).

Up to date, only a few number of studies involving 1040 healthy volunteers have assessed the CYP2C9 "measured" metabolic phenotype [9,18,45,67,77,84,115,116] and according to the previously defined inclusion criteria, only two of them, involving 301 subjects, were selected to include in this study. More precisely, the "measured" metabolic phenotypes have been carried out for American admixed and South Asian populations.[9,116]

Table 1. *CYP2C9* allele frequencies in different ethnic groups across the world.

Ethnic groups (N)	*CYP2C9* allele frequencies[1]											Country/zone	References
	**2*	**3*	**4*	**5*	**6*	**8*	**9*	**10*	**11*	**12*	**13*		
Black													
Black Africans (2284)	0.46	0.72	0.00	0.72	2.70	8.60	15.70	0.00	2.36	0.00	-	Ethiopia, Tanzania, Benin, Ghana, Mozambique, Cabinda, São Tome and Principe, Central Africa, Nigeria, South Africa, Mandenka (ethnic group)	[45–53]
	(2046)	(963)	(424)	(757)	(109)	(109)	(109)	(109)	(424)	(109)	-		
	[0.0 – 4.3]	[0.0 – 2.3]	-	[0.0 – 1.8]	-	-	-	-	[2.0 – 2.7]	-	-		
North Africans (247)	12.00	6.00	-	-	-	-	-	-	-	-	-	Egypt	[54]
	(247)	(247)	-	-	-	-	-	-	-	-	-		
African Americans (700)	3.11	1.79	0.00	1.56	1.00	4.70	-	-	1.30	-	0.00	USA	[37,51,55]
	(700)	(420)	(300)	(420)	(300)	(300)	-	-	(300)	-	(300)		
	[1.0 – 5.0]	[1.3 – 2.0]	-	[1.5 – 1.7]	-	-	-	-	-	-	-		
Middle Eastern													
Middle Easterns (1969)	10.88	7.76	0.00	0.10	0.00	0.30	-	-	-	-	-	Iran, Israel, Oman, Syria, Saudi Arabia, Turkey, USA[2]	[51,56–62]
	(1969)	(1789)	(502)	(502)	(502)	(722)	-	-	-	-	-		
	[5.0 – 15.2]	[0.0 – 19.1]	-	-	-	[0.0 – 1.0]	-	-	-	-	-		
Natives													
Native Americans (1882)	1.25	2.38	-	0.00	0.00	2.20	-	-	0.00	-	-	Brazil, Mexico, Peru, USA	[40,51,63–66]
	(1882)	(1484)	-	(180)	(685)	(180)	-	-	(180)	-	-		
	[0.0 – 5.2]	[0.0 – 10.4]	-	-	-	[0.0 – 4.4]	-	-	-	-	-		
Native Oceanians (60)	0.00	-	-	-	-	-	-	-	-	-	-	Papua New Guinea, Micronesia	[51]
	(60)	-	-	-	-	-	-	-	-	-	-		
Circumpolar populations (49)	-	8.20	-	-	-	-	-	-	-	-	-	Finland	[53]
	-	(49)	-	-	-	-	-	-	-	-	-		
Admixed[3]													
American admixed (5545)	7.12	3.94	0.00	0.37	0.17	1.75	-	-	0.73	-	0.00	Brazil, Chile, Ecuador, Mexico, USA, Cuba	[9,38,39,51,59, 64,65,67–75]
	(5545)	(5102)	(908)	(1942)	(908)	(229)	-	-	(1135)	-	(101)		
	[0.5 – 10.5]	[1.5 – 9.0]	-	[0.0 – 1.5]	[0.0 – 0.5]	[1.5 – 2.0]	-	-	[0.7 – 1.0]	-	-		
Caucasian[3]													
American Caucasians (587)	14.68	6.49	0.00	0.00	0.00	0.21	-	-	0.50	-	0.00	USA, Cuba	[37,53,59,67,76]
	(521)	(587)	(238)	(378)	(238)	(238)	-	-	(106)	-	(106)		
	[13.2 – 17.0]	[4.3 – 11.0]	-	-	-	[0.0 – 0.4]	-	-	-	-	-		

(*continued*)

Table 1. *CYP2C9* allele frequencies in different ethnic groups across the world. (continued).

Ethnic groups (N)	*CYP2C9* allele frequencies[1]											Country/zone	References
	**2*	**3*	**4*	**5*	**6*	**8*	**9*	**10*	**11*	**12*	**13*		
Scandinavians (990)	14.32 (919) [10.7 – 18.8]	6.23 (939) [0.0 – 16.8]	- - -	0.00 (430) - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	Denmark, Finland, Sweden	[51,53,77–79]
Central Europeans (1395)	12.13 (1231) [5.0 – 16.0]	7.95 (1193) [5.0 – 9.5]	- -	- -	- -	- -	- -	- -	- -	- -	- -	England, Germany, Hungary, Ireland, Netherlands, Scotland	[51,53,80–83]
Mediterranean- South Europeans (2282)	13.83 (1945) [11.2 – 16.2]	8.37 (2282) [3.7 – 11.2]	0.00 (327) -	0.00 (327) -	0.00 (327) -	- - -	- - -	- - -	- - -	- - -	- - -	France, Greece, Italy, Portugal, Spain	[38,46,53,71, 84–89]
Slavs (1012)	12.31 (1012) [6.0 – 16.5]	7.57 (958) [5.0 – 9.5]	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	Croatia, Romania, Russia (ethnic group: Adygei)	[51,90–93]
Asian													
South Asians (1085)	7.26 (1023) [0.6 – 11.8]	11.66 (1085) [3.9 – 15.5]	0.00 (165) -	0.00 (165) -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	India and Romani population living in Hungary	[53,82,94–96]
East Asians (11332)	0.56 (9319) [0.0 – 9.6]	3.36 (8583) [2.8 – 5.1]	0.00 (1209) -	0.00 (1185) -	0.00 (302) -	1.64 (502) [1.0 – 1.8]	0.00 (200) -	0.00 (200) -	0.03 (3187) [0.0 – 0.1]	0.00 (200) -	0.32 (6189) [0.0 – 0.5]	China, Japan, Korea, Malaysia, Thailand, Vietnam	[19,47,51,57, 59,76,94,97–113]
North Asians (452)	2.20 (452) [0.0 – 5.7]	4.14 (351) [0.6 – 9.2]	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	Russian ethnic groups: Tuvinians, Buryats, Altaians, Yakuts, Khanty	[51,91]

[1]Frequencies expressed as %. In parentheses, the number of subjects studied for each allele. In square brackets, the maximum and minimum range of *CYP2C9* allele frequencies.
[2]Ashkenazi population living in USA.
[3]Studies for African admixed and Australian Caucasians were not found according to inclusion criteria.
N: Total number of subjects from a specific ethnic group studied for a *CYP2C9* allele.

4. *CYP2C9* allele variability

The frequencies of the main allelic variants studied, according to the ethnic groups and geographic regions, are shown in Table 1 and Figure 2, respectively. With regard to ethnicity, the frequency of *CYP2C9*2* was higher in Caucasians than among the other groups ($p < 0.001$) and within Caucasians, the highest frequency of this allele was observed in American Caucasians (14.68%). On the contrary, this allele shows the lowest frequency among Black Africans (0.46%) and East Asians (0.56%). Interestingly, a very low frequency is also observed among Native Americans (1.25%) when compared with American admixed populations (7.12%) ($p < 0.001$). In turn, the frequency in American admixed populations was also lower than those reported for American Caucasians, Central Europeans and Mediterranean–South Europeans ($p < 0.001$, in all cases). No studies of *CYP2C9* allele frequencies were found for African admixed and Australian Caucasians and only one report was found in Native Oceanians, in which the 60 studied individuals were considered as homozygous for the wild type allele.

*CYP2C9*3* is present at the highest frequency in South Asians (11.66%, $p < 0.001$). However, *CYP2C9*5* was more frequently found among African Americans (1.56%), and *CYP2C9*6, CYP2C9*8* and *CYP2C9*11* in Black African populations (2.70%, 8.60%, and 3.82%, respectively). *CYP2C9*9* allele has only been studied once among Black Africans, in whom a high frequency was found (15.70%). *CYP2C9*13* was only observed in East Asians (0.32%), whereas *CYP2C9*4, *10* and *12* were not detected in any population.

Regarding the geographic distribution, the frequency of *CYP2C9*2* was higher in Europe than in the rest of the regions (13.21%, $p < 0.001$) and *CYP2C9*3* was more frequent in the Middle East and Europe when compared with the other regions (7.76%, in both cases; $p < 0.001$). As in the ethnicity analysis, *CYP2C9*5, *6, *8* and *11* alleles were more frequent in Africa than in the rest of the regions ($p < 0.01$ in all cases).

5. Variability of the CYP2C9 "predicted" phenotype from genotype

The frequencies of CYP2C9 gPMs and mPMs according to ethnic groups are presented in Table 2. A low frequency of gPMs was found in all groups, being the maximum frequencies those described in South Asians (1.38%, $p < 0.001$) and Central Europeans (0.77%, $p = 0.005$). No gPMs were observed in Black African, African American and Circumpolar populations.

Considering the geographic distribution, the highest frequency of gPMs was found in Europe (0.66%, $p < 0.001$), whereas they were quite uncommon in Africa and America (0.10% and 0.13%, respectively) (Table 3).

6. Variability of the CYP2C9 "measured" metabolic phenotype

The CYP2C9 "measured" metabolic phenotype has not been very extensively studied and consequently, the number of analyzed individuals is still low. As shown in Table 2, mPMs are frequent in a population from South Asia (14.02%), but completely absent in an Ecuadorian admixed population. As the total number of individuals studied is quite reduced it was not possible to assess the potential correlation among "predicted" phenotypes from genotypes and "measured" metabolic phenotypes.

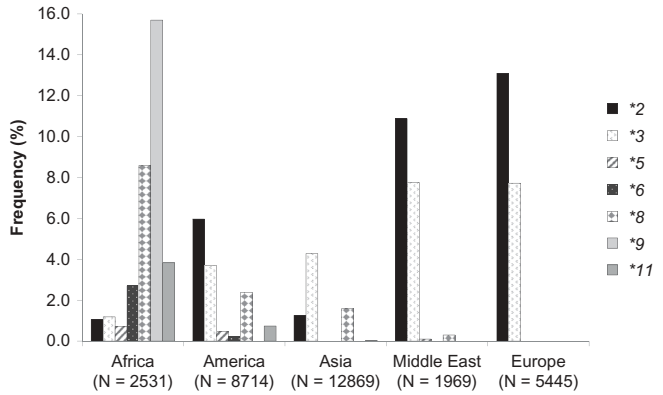

Figure 2. Distribution of *CYP2C9* allele across five geographic regions. N: Total number of subjects studied within each geographic region. Oceania was not included in the chart (n = 60, wt = 100%).

Table 2. Frequencies of CYP2C9 "predicted" phenotypes from genotype and "measured" metabolic phenotypes found in different studied ethnic groups.

Ethnic group		"Predicted" phenotype			"Measured" metabolic phenotype			References
		N	n gPMs	% gPMs	N	n mPMs	% mPMs	
Black	Black Africans	759	0	0.00	-	-	-	[46–53]
	North Africans	247	1	0.40	-	-	-	[54]
	African Americans	420	0	0.00	-	-	-	[37,55]
Middle Eastern	Middle Easterns	3103	16	0.52	-	-	-	[56–62,77,117]
Natives	Native Americans	1423	2	0.14	-	-	-	[40,63–66,118]
	Circumpolar	49	0	0.00	-	-	-	[53]
Admixed	American admixed	5263	6	0.11	194	0	0.00	[38,39,59,64,65,67–73,75] mPMs [9]
Caucasian	American Caucasian	587	2	0.34	-	-	-	[37,53,59,67,76]
	Scandinavians	939	6	0.64	-	-	-	[53,78,79]
	Central Europeans	1293	10	0.77	-	-	-	[53,80–83]
	Mediterranean–South Europeans	2282	14	0.61	-	-	-	[38,46,53,71,84,85,87–89]
	Slavs	758	5	0.66	-	-	-	[91–93]
Asian	South Asians	1085	15	1.38	107	15	14.02	[53,82,94–96] mPMs [116]
	East Asians	8584	12	0.14	-	-	-	[19,57,59,76,94,97–106,109–111]
	North Asians	351	1	0.28	-	-	-	[91]

N: Total number of subjects studied; n: Number of subjects considered as poor metabolizers either from the genotype (gPMs) or from the "measured" metabolic phenotype (mPMs); PMs: Poor metabolizers.

Table 3. Frequencies of CYP2C9 "predicted" phenotypes from genotype and "measured" metabolic phenotypes across different geographic regions.

Continent	"Predicted" phenotype			"Measured" metabolic phenotype			References
	N	n gPMs	% gPMs	N	n mPMs	% mPMs	
Africa	1006	1	0.10	-	-	-	[45–54]
Middle East	3103	16	0.52	-	-	-	[56–62,77,117]
Europe	5321	35	0.66	-	-	-	[38,46,53,71,78–89,91–93]
America	7693	10	0.13	194	0	0.00	[37–40,53,55,59,63–73,75,76,118] mPMs [9]
Asia	10,020	28	0.28	107	15	14.28	[19,53,57,59,76,82,91,94–106,109–111] mPMs [116]
Oceania	-	-	-	-	-	-	-

N: Total number of subjects studied; n: Number of subjects considered as poor metabolizers, either from the genotype (gPMs) or from the "measured" metabolic phenotype (mPMs); PMs: Poor metabolizers.

7. Conclusion

Interethnic and geographical variability of *CYP2C9* alleles and CYP2C9 metabolic phenotypes frequencies was observed. *CYP2C9*2* was more frequent among Caucasians, conversely to the low frequency of this allele observed among Native American and American admixed populations, which can be explained by the potential influence of Asian populations on Native Americans. In South Asians, *CYP2C9*3* was the predominant allele, while in Black Africans a high prevalence of the *CYP2C9*9* allele was observed. Nevertheless, there is still a lack of knowledge on *CYP2C9* genetic polymorphisms and CYP2C9 metabolic phenotype worldwide. Consequently, additional research studies on *CYP2C9* genetic polymorphisms and metabolic phenotypes are required, especially for African and Oceanian populations.

Moreover, studies determining whether a correlation between genotypes and "measured" metabolic phenotypes exists are definitely required, as of the scarce results found up to date.

8. Expert opinion

This study shows the worldwide variability in the *CYP2C9* allele frequencies across different ethnic and geographic groups, with *CYP2C9*2* being the most frequent in Caucasians and *CYP2C9*3* in South Asians. The gPMs were found overall in a low frequency, with the highest frequency detected for South Asians, accordingly to the *CYP2C9*3* frequency in these populations. Moreover, data about CYP2C9 "measured" metabolic phenotypes are currently limited.

Most *CYP2C9* allele frequencies studies have been developed in Europe and America (Table 1), so further research is needed in the least studied populations, particularly in Africans and Oceanians. Thus, an increase of CYP2C9 research on these under-represented populations is highly recommended, even more considering the differences observed in allele frequencies within the African populations studied, and the scarce information on *CYP2C9* genotypes for Oceanians.

Furthermore, the variability in *CYP2C9* allele frequencies has been assumed to be owing to the studied subject's ethnical background and not to population selection bias. However, for Africans especially for North Africans, further studies are needed to verify the implication of the ethnicity in the frequency of *CYP2C9* alleles, as only one study was found for this ethnic group.

The *CYP2C9*2* allele frequency was found to be lower in Native Americans than in the American admixed and European populations, which is in accordance to previous reports supporting the hypothesis of Asian genetic background in Amerindian populations.[38–40] It is also well recognized that populations from America are characterized by an heterogeneous genetic background due to the initial migration of Asians, more specifically, of North-Eastern Asian across Beringia and, afterwards, the arrival of Europeans and Africans.[119,120] Accordingly, Asians' genetic background is predominant among Amerindians and, conversely, it was diluted in admixed populations due to the European population and other populations' influence, as it is observed in the variable frequencies of *CYP2C9*2* among Native American, American admixed and Caucasian populations presented in this review.

The high frequency of *CYP2C9*3* allele found for South Asians in the analysis by ethnic groups was different from that found in the study of *CYP2C9* alleles by geographic region, in which the top frequencies were observed for Middle East and Europe. This could be due to the low *CYP2C9*3* frequency in East and North Asians (3.36% and 4.14%, respectively) in comparison to that reported for South Asians (11.66%).

Seventeen *CYP2C9* alleles with decreased enzyme activity *in vitro* [8] have been reported among selected population studies, though many of them have been studied in almost none of the populations. For instance, the decreased activity *CYP2C9*8* allele, described among Black Africans and African Americans (8.60% and 4.70%, respectively) and to a minor extent, in Native Americans, American admixed and East Asians (2.20%, 1.75% and 1.64%, respectively), has not been yet analyzed in other ethnic groups and due to its aforementioned prevalence, its analysis should be included.

Additionally, the CYP2C9 enzyme activity related to some allelic variants, such as *CYP2C9*9*, has not been fully described.[8] Thus, in spite of *CYP2C9*9* is the most frequent allele among Black Africans (15.70%) and it has been additionally studied in East Asians (0.00%), only one study analyzed the correlation between the CYP2C9 enzyme activity and the presence of *CYP2C9*9* in a Black population phenotyped with phenytoin.[45] Hence, further studies focused on the *CYP2C9* genotype-phenotype correlation are needed, especially for these non-characterized alleles.

The frequency of gPMs was <1% in all different ethnic groups studied, but for South Asians (1.38%). Although gPMs theoretically include individuals homozygous for *CYP2C9*3* (with an extreme reduced activity enzyme) or for the null activity *CYP2C9* alleles as *CYP2C9*6*, in practice only individuals *CYP2C9*3/*3* were classified as gPMs, since *CYP2C9*6* and other null activity alleles were not frequent across worldwide populations and no homozygous individual for these alleles was observed among selected studies. Moreover, the reduced number of studies on CYP2C9 "measured" metabolic phenotype limited the execution of a correlation analysis between frequencies of gPMs and mPMs as it has been described in other studies.[32,33] The frequencies of gPMs and mPMs were both reported only in the American admixed (0.11% and 0.00%, respectively) and South Asian (1.38% and 14.02%, respectively) populations (Table 2). The discordance of these frequencies among South Asians, with mPMs 10-fold higher than gPMs, could not seen as a complete lack of correlation between "predicted" and "measured" phenotypes as there is only one report of "measured" metabolic phenotype in that population and the number of individuals studied for phenotype in that ethnic group is 10-fold different (1085 for "predicted" phenotype from genotype versus 107 individuals studied for "measured" metabolic phenotype). This highlights the necessity of further studies on CYP2C9 "measured" metabolic phenotype that could lead to more accurate conclusions.

In addition, there are more studies that evaluated the correlation between *CYP2C9* genotype and "measured" metabolic phenotype using phenytoin and diclofenac as probe drugs.[18,45,77,84,115] In these studies significant differences in the corresponding drug metabolic ratios, after stratifying the individuals according to their genotype, were found. However, the frequency of mPMs was not reported and hence, the studies were not included in the analysis. Second, another study assessed the CYP2C9 enzyme activity in Cubans and Spaniards using diclofenac as a probe drug, [67] but the resulted distribution of the metabolic ratio was unimodal and the precise number of mPMs was not determined. Furthermore, this study evaluated the influence of environmental factors in the CYP2C9 enzyme activity but did not find an influence of tobacco smoking, caffeine or alcohol intake or gender on "measured" metabolic

phenotype. Thus, additional research on CYP2C9 "measured" metabolic activity and genetic polymorphisms is required to identify not only the genetic, but also the environmental and/or epigenetic factors that could explain the observed variability in drug metabolism by CYP2C9.

Declaration of interest

This study was supported by Junta de Extremadura, AEXCID 13IA001 (to SIFF) and coordinated by the network Red Iberoamericana de Farmacogenética y Farmacogenómica (www.ribef.com). Consejo Nacional de Ciencia y Tecnología de México (CONACyT) (#167261). Some researchers were supported by fellowships from the University of Costa Rica (PhD fellowship to Carolina Céspedes-Garro in Spain), Instituto de Salud Carlos III-Sara Borrell Postdoc Program (CD13/00348 to Fernando de Andrés), Brazilian Government CAPES grant (BEX 12350/2013-03 fellowship to Fernanda Rodrigues-Soares and 110/2012) and from CONACYT (#369708 to Ingrid Fricke-Galindo). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Bibliography

Papers of special note have been highlighted as either of interest (*) or of considerable interest (**) to readers.

1. Evans WE, Mcleod HL. Pharmacogenomics – drug disposition, drug targets, and side effects. N Engl J Med. 2003;348(6):538–549.
2. Llerena A, Cobaleda J, Martínez C, et al. Interethnic differences in drug metabolism: influence of genetic and environmental factors on debrisoquine hydroxylation phenotype. Eur J Drug Metab Pharmacokinet. 1996;21(2):129–138.
3. Vesell ES. Genetic and environmental factors affecting drug disposition in man. Clin Pharmacol Ther. 1977;22(5 Pt 2):659–679.
4. Wilson K. Sex-related differences in drug disposition in man. Clin Pharmacokinet. 1984;9(3):189–202.
5. Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into rational therapeutics. Science. 1999;286(5439):487–491.
6. He S-M, Zhou Z-W, Li X-T, et al. Clinical drugs undergoing polymorphic metabolism by human cytochrome P450 2C9 and the implication in drug development. Curr Med Chem. 2011;18(5):667–713.
7. Higashi MK, Veenstra DL, Kondo LM, et al. Association between CYP2C9 genetic variants and anticoagulation-related outcomes during warfarin therapy. JAMA. 2002;287(13):1690–1698.
8. CYP2C9 allele nomenclature [Internet]. The human cytochrome P450 (CYP) allele nomenclature database; 2015 [cited 2014 Dec 3]. Available from: http://www.cypalleles.ki.se/cyp2c9.htm.
9. Dorado P, Beltrán LJ, Machín E, et al. Losartan hydroxylation phenotype in an Ecuadorian population: influence of CYP2C9 genetic polymorphism, habits and gender. Pharmacogenomics. 2012;13(15):1711–1717.

* **First study reporting the CYP2C9 hydroxylation phenotypes and genotypes in a Hispanic population.**

10. Hatta FHM, Teh LK, Helldén A, et al. Search for the molecular basis of ultra-rapid CYP2C9-catalysed metabolism: relationship between SNP IVS8-109A>T and the losartan metabolism phenotype in Swedes. Eur J Clin Pharmacol. 2012;68(7):1033–1042.
11. Dorado P, Berecz R, Cáceres MC, et al. Analysis of diclofenac and its metabolites by high-performance liquid chromatography: relevance of CYP2C9 genotypes in diclofenac urinary metabolic ratios. J Chromatogr B Analyt Technol Biomed Life Sci. 2003;789(2):437–442.
12. Dorado P, Berecz R, Cáceres MC, et al. Reproducibility over time of the urinary diclofenac/4′-OH diclofenac ratio among different CYP2C9 genotypes. Eur J Drug Metab Pharmacokinet. 2003;28(3):213–215.
13. Turpault S, Brian W, Van Horn R, et al. Pharmacokinetic assessment of a five-probe cocktail for CYPs 1A2, 2C9, 2C19, 2D6 and 3A. Br J Clin Pharmacol. 2009;68(6):928–935.
14. Dorado P, Machín E, De Andrés F, et al. Development of a HPLC method for the determination of losartan urinary metabolic ratio to be used for the determination of CYP2C9 hydroxylation phenotypes. Drug Metabol Drug Interact. 2012;27(4):217–223.
15. De Andrés F, Sosa-Macías M, Llerena A. A rapid and simple LC-MS/MS method for the simultaneous evaluation of CYP1A2, CYP2C9, CYP2C19, CYP2D6 and CYP3A4 hydroxylation capacity. Bioanalysis. 2014;6(5):683–696.

* **Novel method that allows evaluating several enzymes hydroxylation capacity simultaneously, including CYP2C9.**

16. Yasar U, Forslund-Bergengren C, Tybring G, et al. Pharmacokinetics of losartan and its metabolite E-3174 in relation to the CYP2C9 genotype. Clin Pharmacol Ther. 2002;71(1):89–98.

* **It was evaluated, for the first time, the pharmacokinetics of losartan considering all known CYP2C9 genotypes.**

17. Jetter A, Kinzig-Schippers M, Skott A, et al. Cytochrome P450 2C9 phenotyping using low-dose tolbutamide. Eur J Clin Pharmacol. 2004;60(3):165–171.
18. Aynacioglu AS, Brockmöller J, Bauer S, et al. Frequency of cytochrome P450 CYP2C9 variants in a Turkish population and functional relevance for phenytoin. Br J Clin Pharmacol. 1999;48(3):409–415.
19. Dai D-P, Xu R-A, Hu L-M, et al. CYP2C9 polymorphism analysis in Han Chinese populations: building the largest allele frequency database. Pharmacogenomics J. 2014;14(1):85–92.
20. Yang J, Chen Y, Li X, et al. Influence of CYP2C9 and VKORC1 genotypes on the risk of hemorrhagic complications in warfarin-treated patients: a systematic review and meta-analysis. Int J Cardiol. 2013;168(4):4234–4243.
21. Martin JH, Begg EJ, Kennedy MA, et al. Is cytochrome P450 2C9 genotype associated with NSAID gastric ulceration? Br J Clin Pharmacol. 2001;51(6):627–630.
22. FDA. Genomics – table of pharmacogenomic biomarkers in drug labeling [Internet]. Silver Spring (MD): Center for Drug Evaluation and Research, Food and Drug Administration; 2015 [cited 2015

May 5]. Available from: http://www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics/ucm083378.htm

23. Dorado P, López-Torres E, Peñas-Lledó EM, et al. Neurological toxicity after phenytoin infusion in a pediatric patient with epilepsy: influence of CYP2C9, CYP2C19 and ABCB1 genetic polymorphisms. Pharmacogenomics J. 2013;13(4):359–361.
24. Kidd RS, Curry TB, Gallagher S, et al. Identification of a null allele of CYP2C9 in an African-American exhibiting toxicity to phenytoin. Pharmacogenetics. 2001;11(9):803–808.
25. Ninomiya H, Mamiya K, Matsuo S, et al. Genetic polymorphism of the CYP2C subfamily and excessive serum phenytoin concentration with central nervous system intoxication. Ther Drug Monit. 2000;22(2):230–232.
26. Facciolá G, Hidestrand M, von Bahr C, et al. Cytochrome P450 isoforms involved in melatonin metabolism in human liver microsomes. Eur J Clin Pharmacol. 2001;56(12):881–888.
27. Adams PB, Lawson S, Sanigorski A, et al. Arachidonic acid to eicosapentaenoic acid ratio in blood correlates positively with clinical symptoms of depression. Lipids. 1996;31(Suppl):S157–S161.
28. LLerena A, Berecz R, Dorado P, et al. CYP2C9 gene and susceptibility to major depressive disorder. Pharmacogenomics J. 2003;3(5):300–302.

* **Important study of the CYP2C9 role in relationship with mental disorders.**

29. LLerena A, Dorado P, Caceres M, et al. CYP2C9 and clinical response to antidepressant drugs in Mexican-Americans. Clin Pharmacol Ther. 2005;77(2):P24.
30. Xie HG, Kim RB, Wood AJ, et al. Molecular basis of ethnic differences in drug disposition and response. Annu Rev Pharmacol Toxicol. 2001;41:815–850.
31. Xie H-G, Prasad HC, Kim RB, et al. CYP2C9 allelic variants: ethnic distribution and functional significance. Adv Drug Deliv Rev. 2002;54(10):1257–1270.
32. LLerena A, Naranjo MEG, Rodrigues-Soares F, et al. Interethnic variability of CYP2D6 alleles and of predicted and measured metabolic phenotypes across world populations. Expert Opin Drug Metab Toxicol. 2014;10(11):1569–1583.
33. Fricke-Galindo I, Céspedes-Garro C, Rodrigues-Soares F, et al. Interethnic variation of CYP2C19 alleles, "predicted" phenotypes and "measured" metabolic phenotypes across world populations. Pharmacogenomics J. 2015.doi:10.1038/tpj.2015.70. [Epub ahead of print].
34. Schwarz UI. Clinical relevance of genetic polymorphisms in the human CYP2C9 gene. Eur J Clin Invest. 2003;33(Suppl 2):23–30.
35. Kirchheiner J, Brockmöller J. Clinical consequences of cytochrome P450 2C9 polymorphisms. Clin Pharmacol Ther. 2005;77(1):1–16.
36. Sistonen J, Fuselli S, Palo JU, et al. Pharmacogenetic variation at CYP2C9, CYP2C19, and CYP2D6 at global and microgeographic scales. Pharmacogenet Genomics. 2009;19(2):170–179.

** **Precedent study of *CYP2C9* allele frequencies variation in worldwide populations.**

37. Dickmann LJ, Rettie AE, Kneller MB, et al. Identification and functional characterization of a new CYP2C9 variant (CYP2C9*5) expressed among African Americans. Mol Pharmacol. 2001;60(2):382–387.
38. Dorado P, Sosa-Macias MG, Peñas-Lledó EM, et al. CYP2C9 allele frequency differences between populations of Mexican-Mestizo, Mexican-Tepehuano, and Spaniards. Pharmacogenomics J. 2011;11(2):108–112.
39. LLerena A, Dorado P, O'Kirwan F, et al. Lower frequency of CYP2C9*2 in Mexican-Americans compared to Spaniards. Pharmacogenomics J. 2004;4(6):403–406.

** **First study to report *CYP2C9* allele frequencies in Hispanic and to state the lower frequency of *CYP2C9*2* allele in admixed populations than in Spaniards.**

40. Sosa-Macías M, Lazalde-Ramos BP, Galaviz-Hernández C, et al. Influence of admixture components on CYP2C9*2 allele frequency in eight indigenous populations from Northwest Mexico. Pharmacogenomics J. 2013;13(6):567–572.
41. Crews KR, Gaedigk A, Dunnenberger HM, et al. Clinical Pharmacogenetics Implementation Consortium guidelines for cytochrome P450 2D6 genotype and codeine therapy: 2014 update. Clin Pharmacol Ther. 2014;95(4):376–382.
42. Gaikovitch EA, Cascorbi I, Mrozikiewicz PM, et al. Polymorphisms of drug-metabolizing enzymes CYP2C9, CYP2C19, CYP2D6, CYP1A1, NAT2 and of P-glycoprotein in a Russian population. Eur J Clin Pharmacol. 2003;59(4):303–312.
43. Pchelina SN, Sirotkina OV, Taraskina AE, et al. The frequency of cytochrome P450 2C9 genetic variants in the Russian population and their associations with individual sensitivity to warfarin therapy. Thromb Res. 2005;115(3):199–203.
44. Stubbins MJ, Harries LW, Smith G, et al. Genetic analysis of the human cytochrome P450 CYP2C9 locus. Pharmacogenetics. 1996;6(5):429–439.
45. Allabi AC, Gala J-L, Horsmans Y. CYP2C9, CYP2C19, ABCB1 (MDR1) genetic polymorphisms and phenytoin metabolism in a Black Beninese population. Pharmacogenet Genomics. 2005;15(11):779–786.
46. Scordo MG, Aklillu E, Yasar U, et al. Genetic polymorphism of cytochrome P450 2C9 in a Caucasian and a black African population. Br J Clin Pharmacol. 2001;52(4):447–450.
47. Yasar U, Aklillu E, Canaparo R, et al. Analysis of CYP2C9*5 in Caucasian, Oriental and black-African populations. Eur J Clin Pharmacol. 2002;58(8):555–558.
48. Allabi AC, Gala J-L, Desager J-P, et al. Genetic polymorphisms of CYP2C9 and CYP2C19 in the Beninese and Belgian populations. Br J Clin Pharmacol. 2003;56(6):653–657.
49. Kudzi W, Dodoo AN, Mills JJ. Characterisation of CYP2C8, CYP2C9 and CYP2C19 polymorphisms in a Ghanaian population. BMC Med Genet. 2009;10:124.
50. Oliveira E, Pereira R, Amorim A, et al. Patterns of pharmacogenetic diversity in African populations: role of ancient and recent history. Pharmacogenomics. 2009;10(9):1413–1422.
51. Speed WC, Kang SP, Tuck DP, et al. Global variation in CYP2C8-CYP2C9 functional haplotypes. Pharmacogenomics J. 2009;9(4):283–290.

* **First *CYP2C9* study involving Native American populations.**

52. Dandara C, Lombard Z, Du Plooy I, et al. Genetic variants in CYP (–1A2, –2C9, –2C19, –3A4 and –3A5), VKORC1 and ABCB1 genes in a black South African population: a window into diversity. Pharmacogenomics. 2011;12(12):1663–1670.

53. Pimenoff VN, Laval G, Comas D, et al. Similarity in recombination rate and linkage disequilibrium at CYP2C and CYP2D cytochrome P450 gene regions among Europeans indicates signs of selection and no advantage of using tagSNPs in population isolates. Pharmacogenet Genomics. 2012;22(12):846–857.

54. Hamdy SI, Hiratsuka M, Narahara K, et al. Allele and genotype frequencies of polymorphic cytochromes P450 (CYP2C9, CYP2C19, CYP2E1) and dihydropyrimidine dehydrogenase (DPYD) in the Egyptian population. Br J Clin Pharmacol. 2002;53(6):596–603.

55. Scott SA, Jaremko M, Lubitz SA, et al. CYP2C9*8 is prevalent among African-Americans: implications for pharmacogenetic dosing. Pharmacogenomics. 2009;10 (8):1243–1255.

56. Peyvandi F, Spreafico M, Karimi M, et al. Allele frequency of CYP2C9 gene polymorphisms in Iran. Thromb Haemost. 2002;88(5):874–875.

57. Nakai K, Habano W, Nakai K, et al. Ethnic differences in CYP2C9*2 (Arg144Cys) and CYP2C9*3 (Ile359Leu) genotypes in Japanese and Israeli populations. Life Sci. 2005;78(1):107–111.

58. Zand N, Tajik N, Moghaddam AS, et al. Genetic polymorphisms of cytochrome P450 enzymes 2C9 and 2C19 in a healthy Iranian population. Clin Exp Pharmacol Physiol. 2007;34(1–2):102–105.

59. Scott SA, Khasawneh R, Peter I, et al. Combined CYP2C9, VKORC1 and CYP4F2 frequencies among racial and ethnic groups. Pharmacogenomics. 2010;11(6):781–791.

60. Hashemi-Soteh SMB, Shahabi-Majd N, Gholizadeh A-R, et al. Allele and genotype frequencies of CYP2C9 within an Iranian population (Mazandaran). Genet Test Mol Biomarkers. 2012;16(7):817–821.

61. Pathare AV, Al Zadjali S, Misquith R, et al. Warfarin pharmacogenetics: polymorphisms of the CYP2C9, CYP4F2, and VKORC1 loci in a genetically admixed Omani population. Hum Biol. 2012;84 (1):67–77.

62. Alzahrani AM, Ragia G, Hanieh H, et al. Genotyping of CYP2C9 and VKORC1 in the Arabic population of Al-Ahsa, Saudi Arabia. Biomed Res Int. 2013;2013:315980.

63. Vargens DD, Petzl-Erler M-L, Suarez-Kurtz G. Distribution of CYP2C polymorphisms in an Amerindian population of Brazil. Basic Clin Pharmacol Toxicol. 2012;110(4):396–400.

64. Soares RAG, Santos PCJL, Machado-Coelho GLL, et al. CYP2C9 and VKORC1 polymorphisms are differently distributed in the Brazilian population according to self-declared ethnicity or genetic ancestry. Genet Test Mol Biomarkers. 2012;16(8):957–963.

65. Castelán-Martínez OD, Hoyo-Vadillo C, Sandoval-García E, et al. Allele frequency distribution of CYP2C9 2 and CYP2C9 3 polymorphisms in six Mexican populations. Gene. 2013;523(2):167.

66. Fohner A, Muzquiz LI, Austin MA, et al. Pharmacogenetics in American Indian populations: analysis of CYP2D6, CYP3A4, CYP3A5, and CYP2C9 in the Confederated Salish and Kootenai Tribes. Pharmacogenet Genomics. 2013;23 (8):403–414.

67. Llerena A, Alvarez M, Dorado P, et al. Interethnic differences in the relevance of CYP2C9 genotype and environmental factors for diclofenac metabolism in Hispanics from Cuba and Spain. Pharmacogenomics J. 2014;14(3):229–234.

68. Bravo-Villalta HV, Yamamoto K, Nakamura K, et al. Genetic polymorphism of CYP2C9 and CYP2C19 in a Bolivian population: an investigative and comparative study. Eur J Clin Pharmacol. 2005;61(3):179–184.

69. Aguilar B, Rojas JC, Collados MT. Prevalence of CYP2C9 variants in the Mexican population. Arch Med Res. 2008;39(4):463.

70. Roco A, Quiñones L, Agúndez JAG, et al. Frequencies of 23 functionally significant variant alleles related with metabolism of antineoplastic drugs in the chilean population: comparison with caucasian and asian populations. Front Genet. 2012;3:229.

71. Vicente J, González-Andrade F, Soriano A, et al. Genetic polymorphisms of CYP2C8, CYP2C9 and CYP2C19 in Ecuadorian Mestizo and Spaniard populations: a comparative study. Mol Biol Rep. 2014;41 (3):1267–1272.

72. Suarez-Kurtz G, Genro JP, de Moraes MO, et al. Global pharmacogenomics: Impact of population diversity on the distribution of polymorphisms in the CYP2C cluster among Brazilians. Pharmacogenomics J. 2012;12(3):267–276.

73. Kramer MA, Rettie AE, Rieder MJ, et al. Novel CYP2C9 promoter variants and assessment of their impact on gene expression. Mol Pharmacol. 2008;73 (6):1751–1760.

74. Orengo-Mercado C, Nieves B, López L, et al. Frequencies of functional polymorphisms in three pharmacokinetic genes of clinical interest within the admixed Puerto Rican population. J Pharmacogenomics Pharmacoproteomics. 2013;4:113–125.

75. Duconge J, Cadilla CL, Windemuth A, et al. Prevalence of combinatorial CYP2C9 and VKORC1 genotypes in Puerto Ricans: implications for warfarin management in Hispanics. Ethn Dis. 2009;19(4):390–395.

76. Myrand SP, Sekiguchi K, Man MZ, et al. Pharmacokinetics/genotype associations for major cytochrome P450 enzymes in native and first- and third-generation Japanese populations: comparison with Korean, Chinese, and Caucasian populations. Clin Pharmacol Ther. 2008;84 (3):347–361.

77. Babaoglu MO, Yasar U, Sandberg M, et al. CYP2C9 genetic variants and losartan oxidation in a Turkish population. Eur J Clin Pharmacol. 2004;60(5):337–342.

78. Yasar U, Eliasson E, Dahl ML, et al. Validation of methods for CYP2C9 genotyping: frequencies of mutant alleles in a Swedish population. Biochem Biophys Res Commun. 1999;254(3):628–631.

79. Halling J, Petersen MS, Damkier P, et al. Polymorphism of CYP2D6, CYP2C19, CYP2C9 and CYP2C8 in the Faroese population. Eur J Clin Pharmacol. 2005;61(7):491–497.

80. Burian M, Grösch S, Tegeder I, et al. Validation of a new fluorogenic real-time PCR assay for detection of CYP2C9 allelic variants and CYP2C9 allelic distribution in a German population. Br J Clin Pharmacol. 2002;54(5):518–521.

81. Bosch TM, Doodeman VD, Smits PHM, et al. Pharmacogenetic screening for polymorphisms in drug-metabolizing enzymes and drug transporters in a Dutch population. Mol Diagn Ther. 2006;10(3):175–185.

82. Sipeky C, Lakner L, Szabo M, et al. Interethnic differences of CYP2C9 alleles in healthy Hungarian and Roma population samples: relationship to worldwide allelic frequencies. Blood Cells Mol Dis. 2009;43(3):239–242.

83. Vogl S, Lutz RW, Schönfelder G, et al. CYP2C9 genotype vs. metabolic phenotype for individual drug dosing–a correlation analysis using flurbiprofen as probe drug. PLoS One. 2015;10(3):e0120403.

** **Study that evaluates the contribution of *CYP2C9* alleles to the hydroxylation of flurbiprofen and questions the individualized drug dosing regimen based only on genotype.**

84. Dorado P, Berecz R, Norberto M-J, et al. CYP2C9 genotypes and diclofenac metabolism in Spanish healthy volunteers. Eur J Clin Pharmacol. 2003;59(3):221–225.

85. Scordo MG, Caputi AP, D'Arrigo C, et al. Allele and genotype frequencies of CYP2C9, CYP2C19 and CYP2D6 in an Italian population. Pharmacol Res. 2004;50(2):195–200.

86. Yang JQ, Morin S, Verstuyft C, et al. Frequency of cytochrome P450 2C9 allelic variants in the Chinese and French populations. Fundam Clin Pharmacol. 2003;17(3):373–376.

87. Arvanitidis K, Ragia G, Iordanidou M, et al. Genetic polymorphisms of drug-metabolizing enzymes CYP2D6, CYP2C9, CYP2C19 and CYP3A5 in the Greek population. Fundam Clin Pharmacol. 2007;21(4):419–426.

88. Oliveira E, Marsh S, van Booven DJ, et al. Pharmacogenetically relevant polymorphisms in Portugal. Pharmacogenomics. 2007;8(7):703–712.

89. Hadjipanagi D, Chrysanthou S, Voskarides K, et al. Genetic polymorphisms in warfarin and tacrolimus-related genes VKORC1, CYP2C9 and CYP3A5 in the Greek-Cypriot population. BMC Res Notes. 2014;7:123.

90. Bozina N, Granić P, Lalić Z, et al. Genetic polymorphisms of cytochromes P450: CYP2C9, CYP2C19, and CYP2D6 in Croatian population. Croat Med J. 2003;44(4):425–428.

91. Makeeva O, Stepanov V, Puzyrev V, et al. Global pharmacogenetics: genetic substructure of Eurasian populations and its effect on variants of drug-metabolizing enzymes. Pharmacogenomics. 2008;9(7):847–868.

92. Gra O, Mityaeva O, Berdichevets I, et al. Microarray-based detection of CYP1A1, CYP2C9, CYP2C19, CYP2D6, GSTT1, GSTM1, MTHFR, MTRR, NQO1, NAT2, HLA-DQA1, and AB0 allele frequencies in native Russians. Genet Test Mol Biomarkers. 2010;14(3):329–342.

93. Buzoianu AD, Trifa AP, Mureşanu DF, et al. Analysis of CYP2C9*2, CYP2C9*3 and VKORC1 –1639 G>A polymorphisms in a population from South-Eastern Europe. J Cell Mol Med. 2012;16(12):2919–2924.

94. Zainuddin Z, Teh LK, Suhaimi AWM, et al. Malaysian Indians are genetically similar to Caucasians: CYP2C9 polymorphism. J Clin Pharm Ther. 2006;31(2):187–191.

95. Rathore SS, Agarwal SK, Pande S, et al. Frequencies of VKORC1 –1639 G>A, CYP2C9*2 and CYP2C9*3 genetic variants in the Northern Indian population. Biosci Trends. 2010;4(6):333–337.

96. Nahar R, Deb R, Saxena R, et al. Variability in CYP2C9 allele frequency: a pilot study of its predicted impact on warfarin response among healthy South and North Indians. Pharmacol Rep. 2013;65(1):187–194.

97. Nasu K, Kubota T, Ishizaki T. Genetic analysis of CYP2C9 polymorphism in a Japanese population. Pharmacogenetics. 1997;7(5):405–409.

98. Kimura M, Ieiri I, Mamiya K, et al. Genetic polymorphism of cytochrome P450s, CYP2C19, and CYP2C9 in a Japanese population. Ther Drug Monit. 1998;20(3):243–247.

99. Yoon YR, Shon JH, Kim MK, et al. Frequency of cytochrome P450 2C9 mutant alleles in a Korean population. Br J Clin Pharmacol. 2001;51(3):277–280.

100. Bae J-W, Kim H-K, Kim J-H, et al. Allele and genotype frequencies of CYP2C9 in a Korean population. Br J Clin Pharmacol. 2005;60(4):418–422.

101. Bae J, Choi C, Kim M, et al. Frequency of CYP2C9 alleles in Koreans and their effects on losartan pharmacokinetics. Acta Pharmacol Sin. 2011;32(10):1303–1308.

102. Lee SS, Kim K-M, Thi-Le H, et al. Genetic polymorphism of CYP2C9 in a Vietnamese Kinh population. Ther Drug Monit. 2005;27(2):208–210.

103. Honda M, Ogura Y, Toyoda W, et al. Multiple regression analysis of pharmacogenetic variability of carvedilol disposition in 54 healthy Japanese volunteers. Biol Pharm Bull. 2006;29(4):772–778.

104. Ngow HA, Wan Khairina WM, Teh LK, et al. CYP2C9 polymorphism: prevalence in healthy and warfarin-treated Malay and Chinese in Malaysia. Singapore Med J. 2009;50(5):490–493.

105. Perini JA, Vargens DD, Santana ISC, et al. Pharmacogenetic polymorphisms in Brazilian-born, first-generation Japanese descendants. Braz J Med Biol Res. 2009;42(12):1179–1184.

106. Yang ZF, Cui HW, Hasi T, et al. Genetic polymorphisms of cytochrome P450 enzymes 2C9 and 2C19 in a healthy Mongolian population in China. Genet Mol Res. 2010;9(3):1844–1851.

107. Shi Y, Xiang P, Li L, et al. Analysis of 50 SNPs in CYP2D6, CYP2C19, CYP2C9, CYP3A4 and CYP1A2 by MALDI-TOF mass spectrometry in Chinese Han population. Forensic Sci Int. 2011;207(1–3):183–187.

108. Si D, Wang J, Zhang Y, et al. Distribution of CYP2C9*13 allele in the Chinese Han and the long-range haplotype containing CYP2C9*13 and CYP2C19*2. Biopharm Drug Dispos. 2012;33(6):342–345.

109. Xiong Y, Wang M, Fang K, et al. A systematic genetic polymorphism analysis of the CYP2C9 gene in four different geographical Han populations in mainland China. Genomics. 2011;97(5):277–281.

110. Zeng WT, Zheng QS, Huang M, et al. Genetic polymorphisms of VKORC1, CYP2C9, CYP4F2 in Bai, Tibetan Chinese. Pharmazie. 2012;67(1):69–73.

111. Lee HW, Lim M, Lee J, et al. Frequency of CYP2C9 variant alleles, including CYP2C9*13 in a Korean population and effect on glimepiride pharmacokinetics. J Clin Pharm Ther. 2012;37(1):105–111.

112. Zuo J, Xia D, Jia L, et al. Genetic polymorphisms of drug-metabolizing phase I enzymes CYP3A4, CYP2C9, CYP2C19 and CYP2D6 in Han, Uighur, Hui and Mongolian Chinese populations. Pharmazie. 2012;67(7):639–644.

113. Wu Z, Zhang X, Shen L, et al. A systematically combined genotype and functional combination analysis of CYP2E1, CYP2D6, CYP2C9, CYP2C19 in different geographic areas of mainland China–a basis for personalized therapy. PLoS One. 2013;8(10):e71934.

114. Graphpad Software. QuickCalcs [Internet]. La Jolla (CA): GraphPad Software, Inc.; 2015 [cited 2015 Apr 1]. Available from: http://www.graphpad.com/quickcalcs/contingency1/

115. Sandberg M, Johansson I, Christensen M, et al. The impact of CYP2C9 genetics and oral contraceptives on cytochrome P450 2C9 phenotype. Drug Metab Dispos. 2004;32(5):484–489.

116. Varshney E, Saha N, Tandon M, et al. Genotype-phenotype correlation of cytochrome P450 2C9 polymorphism in Indian National Capital Region. Eur J Drug Metab Pharmacokinet. 2013;38 (4):275–282.

117. Efrati E, Elkin H, Sprecher E, et al. Distribution of CYP2C9 and VKORC1 risk alleles for warfarin sensitivity and resistance in the Israeli population. Curr Drug Saf. 2010;5 (3):190–193.

118. Vargens DD, Damasceno A, Petzl-Erler M-L, et al. Combined CYP2C9, VKORC1 and CYP4F2 frequencies among Amerindians, Mozambicans and Brazilians. Pharmacogenomics. 2011;12 (6):769–772.

119. Schurr TG. The peopling of the new world: perspectives from molecular anthropology. Annu Rev Anthropol. 2004;33(1):551–583.

120. Salzano FM, Sans M. Interethnic admixture and the evolution of Latin American populations. Genet Mol Biol. 2014;37(1 Suppl):151–170.

Affiliation

Carolina Céspedes-Garro[1,2], Ingrid Fricke-Galindo[3], María Eugenia G Naranjo[1], Fernanda Rodrigues-Soares[1,4], Humberto Fariñas[1], Fernando De Andrés[1], Marisol López-López[5], Eva M Peñas-Lledó[1] & Adrián LLerena[†1]

[†]Author for correspondence
[1]CICAB Clinical Research Centre, Extremadura University Hospital and Medical School, Badajoz 06080, Spain
Tel: +34 924 218 040
Fax: +34 924 219 881
E-mail: allerena@unex.es
[2]Genetics Section, School of Biology, University of Costa Rica, 11501-2060 San José, Costa Rica
[3]Doctorate in Biological and Health Sciences, Metropolitan Autonomous University, 04960 Mexico City, Mexico
[4]Department of General Biology, Federal University of Minas Gerais, 31270-901 Belo Horizonte, Brazil
[5]Department of Biological Systems, Metropolitan Autonomous University, 04960 Mexico City, Mexico

Anexo D – Interethnic variation of CYP2C19 alleles, 'predicted' phenotypes and 'measured' metabolic phenotypes across world populations

The Pharmacogenomics Journal (2015), 1–11

www.nature.com/tpj

REVIEW

Interethnic variation of *CYP2C19* alleles, 'predicted' phenotypes and 'measured' metabolic phenotypes across world populations

I Fricke-Galindo[1,2], C Céspedes-Garro[1,3], F Rodrigues-Soares[1,4], MEG Naranjo[1], Á Delgado[1], F de Andrés[1], M López-López[5], E Peñas-Lledó[1] and A LLerena[1]

The present study evaluates the worldwide frequency distribution of *CYP2C19* alleles and CYP2C19 metabolic phenotypes ('predicted' from genotypes and 'measured' with a probe drug) among healthy volunteers from different ethnic groups and geographic regions, as well as the relationship between the 'predicted' and 'measured' CYP2C19 metabolic phenotypes. A total of 52 181 healthy volunteers were studied within 138 selected original research papers. *CYP2C19*17* was 42- and 24-fold more frequent in Mediterranean-South Europeans and Middle Easterns than in East Asians ($P < 0.001$, in both cases). Contrarily, *CYP2C19*2* and *CYP2C19*3* alleles were more frequent in East Asians (30.26% and 6.89%, respectively), and even a twofold higher frequency of these alleles was found in Native populations from Oceania (61.30% and 14.42%, respectively; $P < 0.001$, in all cases), which may be a consequence of genetic drift process in the Pacific Islands. Regarding CYP2C19 metabolic phenotype, poor metabolizers (PMs) were more frequent among Asians than in Europeans, contrarily to the phenomenon reported for CYP2D6. A correlation has been found between the frequencies of CYP2C19 poor metabolism 'predicted' from *CYP2C19* genotypes (gPMs) and the poor metabolic phenotype 'measured' with a probe drug (mPMs) when subjects are either classified by ethnicity ($r = 0.94$, $P < 0.001$) or geographic region ($r = 0.99$, $P = 0.002$). Nevertheless, further research is needed in African and Asian populations, which are under-represented, and additional *CYP2C19* variants and the 'measured' phenotype should be studied.

The Pharmacogenomics Journal advance online publication, 27 October 2015; doi:10.1038/tpj.2015.70

INTRODUCTION

Cytochrome P450 enzymes are responsible for the phase I metabolism of several endogenous and exogenous substrates. Within the CYP2C subfamily, CYP2C19 is relevant because it is highly polymorphic[1] and is involved in the metabolism of drugs widely used in clinical practice such as several antidepressants, proton pump inhibitors and warfarin and has a key role in the activation of the prodrug clopidogrel.[2] Therefore, *CYP2C19* polymorphisms account for inter-individual variability in the metabolism of abovementioned drugs.

The CYP2C19 hydroxylation capacity can be determined by two approaches: 'measured' metabolic phenotype with a probe drug (for example, mephenytoin, omeprazole and so on)[3] or 'predicted' phenotype from *CYP2C19* genotype. The CYP2C19 'measured' metabolic phenotype is variable, and a group of subjects with impaired activity, the so-called poor metabolizers (mPMs), have been described. The mPM phenotype is inherited as an autosomic recessive character.[4]

The gene that encodes CYP2C19 enzyme belongs to the *CYP2C* gene cluster located on 10q24 and for which >48 *CYP2C19* allelic variants have been described, ranging from null to increased enzymatic activity.[5] Individuals carrying two inactive *CYP2C19* alleles (that is, *CYP2C19*2*, *CYP2C19*3*) have been classified as having a 'predicted' poor metabolism (gPMs), and those carrying the increased activity allele, **17*, that is *wt/*17* or **17/*17* genotypes, are considered as ultrarapid metabolizers (gUMs). Individuals considered as gPMs or gUMs have shown variation in their clinical response, including well-documented adverse drug reactions.[6–10] For instance, a high incidence of side effects has been observed in patients with inactive *CYP2C19* alleles treated with antidepressants such as amitriptyline,[6] and neurological toxicity induced by the antiepileptic phenytoin has been reported for patients who carry *CYP2C9* and *CYP2C19* null alleles.[7] Moreover, in psychiatric patients using mephenytoin as a probe drug, the lack of inhibitory effect of CYP2C19 compared with CYP2D6 was demonstrated[8] as well as its lack of implication in clozapine metabolism.[9] On the contrary, patients who possess the *CYP2C19*17* allele exhibit a higher clearance of warfarin enantiomers than those individuals homozygous for the wild-type allele, which could impact the international normalized ratio value in the anticoagulant therapy.[10] However, the relation of this variant on an increased CYP2C19 activity for other drug therapies still remains controversial.[11,12] These and other reports have led regulatory agencies (that is, Food and Drug Administration, European Medicine Agency) to recognize *CYP2C19* as a pharmacogenetic biomarker for 16 drugs,[13,14] and CPIC (Clinical Pharmacogenetics Implementation Consortium) has developed

[1]CICAB Clinical Research Centre, Extremadura University Hospital and Medical School, Badajoz, Spain; [2]Doctorate in Biological and Health Sciences, Metropolitan Autonomous University, Mexico City, Mexico; [3]Genetics Section, School of Biology, University of Costa Rica, San José, Costa Rica; [4]Department of General Biology, Federal University of Minas Gerais, Belo Horizonte, Brazil and [5]Department of Biological Systems, Metropolitan Autonomous University, Mexico City, Mexico. Correspondence: Professor A LLerena, CICAB, Clinical Research Centre, Extremadura University Hospital and Medical School, Avda de Elvas s/n, Badajoz 06080, Spain.
E-mail: allerena@unex.es
Received 1 June 2015; revised 15 July 2015; accepted 19 August 2015

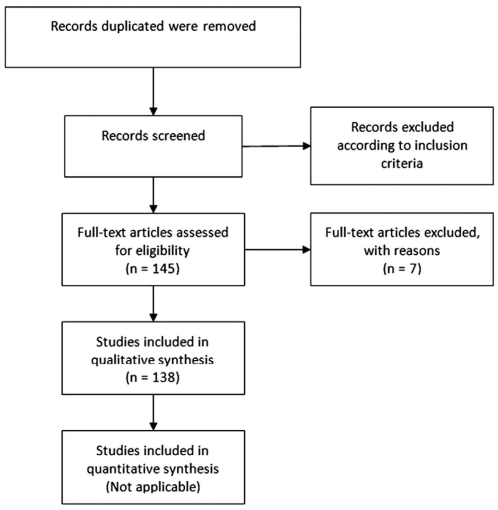

Figure 1. Flow diagram of the studies included in the analyses.

guidelines with therapeutic recommendations for clopidogrel and amitriptyline according to *CYP2C19* genotype.[2]

Nevertheless, benefits of a CYP2C19 impaired metabolism have also been reported. For instance, mPMs showed an increased healing rate in the treatment of gastric ulcers with proton pump inhibitors as a result of increased drug plasma levels;[15] and a synergic effect of clopidogrel and calcium-channel blockers on the reduction of blood pressure has been reported, which might be explained by the metabolic interaction between these two drugs and the influence of *CYP2C19*2* in the regulation of aldosterone.[16]

There is also an implication of CYP2C19 activity in the metabolism of endogenous compounds as substrates of the central nervous system.[17] It has been reported that individuals with an increased CYP2C19 activity show more depressive symptoms,[18] and recently, in a transgenic mouse model it has been shown that it influences the brain development and affective behavior.[19] Furthermore, a high combined CYP2D6–CYP2C19 metabolic capacity show an increased risk for a severe suicide attempt.[20] Therefore, *CYP2C19* genetic polymorphisms not only could impact the metabolism drugs but also could affect the equilibrium of endogenous compounds.

Finally, *CYP2C19* allele frequencies show interethnic variability. Thus *CYP2C19*3* is mainly found in Asians,[21] whereas in European and African populations *CYP2C19*2* is the most frequently found allelic variant. Similarly, *CYP2C19*17* is common among Europeans.[22] Regarding 'measured' metabolic phenotype, the frequency of CYP2C19 mPMs in Caucasians ranges from 1% to 5%,[23–27] and a high variability has also been observed across populations, with a high prevalence among individuals from Oriental countries (10–23%).[28–30] Ultimately, pharmacogenetic knowledge about CYP2C19 geographic and interethnic variability is expected to be useful to improve the quality of health care and to prevent adverse drug reactions or therapeutic failures during the treatment with CYP2C19 substrates within specific populations.

Hence, the aim of this study was to evaluate the worldwide frequency distribution of the *CYP2C19* alleles, 'predicted' phenotypes from genotypes and 'measured' metabolic phenotypes among healthy volunteers from different ethnic groups and geographic regions, as well as to determine the relationship between the 'predicted' phenotypes from genotypes and the 'measured' metabolic phenotypes for CYP2C19.

METHODS

The present study was conducted by search within PubMed database using the terms 'CYP2C19' and one by one the names of the different countries. The 'measured' metabolic phenotypes search was carried out by typing the terms 'CYP2C19' and the probe drugs known to have been utilized for the *in vivo* CYP2C19 phenotyping studies (mephenytoin, omeprazole and proguanil).

The following inclusion criteria were used for study selection: (i) one or more *CYP2C19* alleles were studied; (ii) *CYP2C19* allelic frequencies and/or metabolic phenotypes were reported; (iii) individuals selected randomized as healthy volunteers; and (iv) at least 50 individuals were studied, except for those studies on Native populations from America, Oceania and the Circumpolar region, as previously stated[31,32] (Figure 1). Studies included within this article were each distinct and not overlapping in the samples according to the available information in the original research papers. Moreover, it was assumed that subjects were unrelated as the studies included in this manuscript were aimed at calculating population frequencies. Methodologies utilized in the studies comprised PCR, reverse transcriptase–PCR and sequencing approaches. Studies including individuals selected as controls or patients from case–control studies were excluded, as well as those reports that measured actual metabolic phenotype but did not report the number of PMs.

As a result, 138 original research papers were selected and classified either by geographic region or by ethnic group. Six major geographic regions were considered: Africa, America, Asia, Europe, Oceania, and the Middle East. The ethnic groups were classified following the previous methods,[31,32] with a modification (Table 1): Asians were divided into South Asians (involving people from India, Sri Lanka and Romani population living in Hungary), East Asians (Asians and Asian descendants living in Canada, China, Japan, Korea, Malaysia, Thailand, USA and Vietnam), and North Asians (Russians).

The frequencies of CYP2C19 alleles, 'predicted' phenotypes and 'measured' metabolic phenotypes at each geographic region and for every ethnic group were calculated. To determine the 'predicted' phenotype from genotype frequencies, individuals carriers of *CYP2C19*2/*2*, **2/*3* and **3/*3* genotypes were classified as gPMs, whereas those individuals homozygous or heterozygous (*wt/*17*) for *CYP2C19*17* allele were considered as gUMs.[2] Regarding 'measured' metabolic phenotypes, individuals reported as PMs according to their actual metabolic phenotype were identified as mPMs. Frequencies calculated for geographic regions and ethnic groups could vary according to the populations included for each analysis.

The frequencies calculated for each geographic region and ethnic group were compared by a chi-squared test with Yate's correction using the GraphPad Software (GraphPad Software, Inc., La Jolla, CA, USA).[33] Furthermore, the comparisons between the allelic, 'predicted' phenotypes and 'measured' metabolic phenotypes frequencies for each geographic region and ethnic group were made vs a weighted average of the frequencies of the remaining groups.[13] Thus, to calculate the weighted average, the number of studied subjects was taken into account. Moreover, the simple linear regression analysis of correlation between CYP2C19 gPMs and mPMs was carried out using SPSS Base 19.0 (SPSS Inc., Chicago, IL, USA).

Still, this methodological approach has some limitations as, in some cases, the single-nucleotide polymorphism studied for the *CYP2C19* alleles differed from one study to another. However, the *CYP2C19* alleles from the 138 original papers were used independently of the genotyped single-nucleotide polymorphisms. Moreover, *CYP2C19*1* frequency is not calculated as this

Table 1. Frequencies of *CYP2C19* alleles analyzed in different ethnic groups

Ethnic groups			*CYP2C19 allele frequencies*[a]							*Country/zone*	*References*
		N	**2*	**3*	**4*	**5*	**17*	**27*	**28*		
Black	Black Africans	2830	15.48 (2604)	0.27 (2520)			17.71 (326)	33.00 (100)	5.00 (100)	Benin, Ethiopia, Gambia, Ghana, Nigeria, South Africa, Tanzania, Uganda, Zimbabwe	51–64
	North Africans	367	8.65 (367)	0.20 (247)						Egypt	65,66
	African Americans	336	22.31 (336)	0.80 (236)						USA	67,68
Middle East	Middle Easterns	2154	13.46 (2154)	1.88 (1703)	0.47 (1047)	0.00 (404)	24.87 (573)			USA[b], Turkey, Saudi Arabia, Iran	69–79
Natives	Native Americans	638	10.84 (638)	0.00 (455)	0.00 (365)	0.00 (365)	10.95 (273)			Brazil, Mexico	80–82
	Native Oceanians	6766	61.30 (6766)	14.42 (6766)						Australia, Melanesia, New Zealand, Papua New Guinea, Vanuatu	36,38–42,83
	Circumpolar populations	311	14.86 (311)	0.00 (159)						Canada	84,85
Admixed	African admixed	75	17.00 (75)				14.00 (75)	8.00 (75)	4.50 (75)	South Africa	61
	American admixed	4980	12.25 (4980)	0.17 (3772)	0.00 (334)	0.00 (334)	20.57 (2360)			Bolivia, Brazil, Chile, Colombia, Ecuador, Mexico, Puerto Rico, USA, Venezuela	68,80,81,86–95
Caucasians	American Caucasians	416	13.01 (416)	0.59 (416)	3.00 (143)	0.00 (143)	20.10 (143)			USA	68,96
	Australian Caucasians	99	14.60 (99)	0.00 (99)						Australia	97
	Scandinavians	1304	16.20 (1304)	0.00 (94)	0.00 (94)		20.80 (770)			Sweden, Norway, Finland, Denmark	98–101
	Central Europeans	5044	13.07 (743)	0.17 (603)						Belgium, Germany, Hungary, Netherlands	59,102–105
	Mediterranean-South Europeans	1204	12.51 (1204)	0.08 (1078)			42.00 (282)			Greece, Italy, Portugal, Spain	94,106–109
	Slavs	724	12.27 (724)	0.23 (372)						Russia	110–112
Asians	South Asians	2117	30.34 (1947)	0.48 (1947)			13.74 (189)			India, Romani population, Sri Lanka	105,113–120
	East Asians	12 525	30.26 (11857)	6.89 (11788)	0.19 (790)	0.10 (793)	0.96 (3777)			Canada, China, Japan, Korea, Malaysia, Thailand, USA, Vietnam	47,68,85,96,101,121–152
	North Asians	351	18.51 (351)	4.43 (351)						Russia	111

N: Total number of subjects from a specific ethnic group that have been studied for a *CYP2C19* allele. References [36,53,63,66,71,77,94,105,114,117,118,124,132,136,139,140,144,146,149]: studies that determined *CYP2C19* alleles through sequencing methods. [a]Frequencies are expressed in percentage (%) and in brackets is the number of subjects that have been studied for the specific allele. [b]Ashkenazi population living in USA (United States of America).

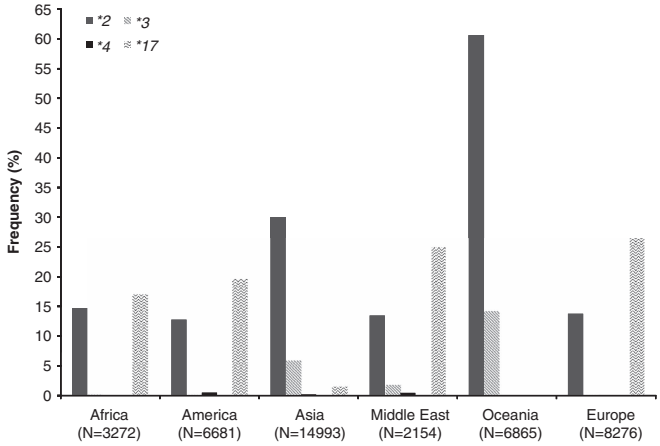

Figure 2. *CYP2C19* major alleles' frequencies across different geographic regions. *N*: total number of subjects studied for the *CYP2C19* allele at each geographic region.

allele is determined by the absence of other *CYP2C19* polymorphisms and thus can vary depending on the number of studied alleles.

RESULTS

CYP2C19 studies across the world population

The studies on CYP2C19 included a total of 52 181 healthy volunteers (7.43×10^{-4}% of the world population) distributed as follows: 31.71% were Asians, 14.37% Americans, 28.72% Europeans, 6.68% Africans, 5.09% Middle Easterns, and 13.43% individuals were from Oceania. Specifically, *CYP2C19* alleles were studied in 42 241 subjects, while the CYP2C19 'predicted' phenotype from genotype was determined for 31 834 subjects and the CYP2C19 metabolic phenotype was measured using a probe drug in 13 802 subjects (81.0, 61.0 and 26.5% of the studied population, respectively). Mephenytoin was used as a probe drug in 11 403 subjects, proguanil in 588 subjects and omeprazole in 1811 subjects. The antimodes reported ranged 0.8–1.7, 0.9–1.2 and 0.6–1.7 for mephenytoin, proguanil and omeprazole studies, respectively.

Up to date, 48 *CYP2C19* allelic variants have been reported in the Human Cytochrome P450 (CYP) Allele Nomenclature Database,[5] though only 17 alleles have been studied in different populations. Among them, *CYP2C19*2* has been studied in all ethnic groups and geographic regions, and *CYP2C19*3* has been analyzed in every population included in this article, except for the Africans admixed. *CYP2C19*17* allele has also been widely studied though its evaluation in more populations is highly advisable (Table 1). Other minor *CYP2C19* alleles analyzed such as **6*, **7*, **8*, **9*, **15*, **18*, **23*, **24*, **25* and **26* have only been studied in Chinese, Colombian, Arabian Saudi, Egyptian and Korean populations, but their frequencies are quite lower or even not found (data not shown).

CYP2C19 allele variability across the ethnic and geographic groups

The allele frequencies of *CYP2C19* by ethnic group and geographic regions are shown in Table 1 and Figure 2, respectively. When classified by ethnicity, the *CYP2C19*2* allele was more frequent in Native Oceanians (61.30%) than in the rest of the populations ($P < 0.001$), followed by East and South Asians (30.3%). The *CYP2C19*3* allele was also more frequent in Native Oceanians and East Asians (14.42% and 6.89%, respectively; $P < 0.001$ in both cases), whereas it was rarely found in the rest of the ethnic groups. On the other hand, *CYP2C19*17* was more common in Mediterranean and South-Europeans (42.00%, $P < 0.001$) and in Middle East populations (24.87%, $P < 0.001$).

Geographically, *CYP2C19*2* and *CYP2C19*3* alleles are more frequent in Oceania than in the rest of the geographic regions (60.63% and 14.22%, respectively; $P < 0.001$ in both cases), while *CYP2C19*17*, associated with an increased enzyme activity, predominates in Europe and Middle East regions (26.48% and 24.87%, respectively; $P < 0.001$). Additionally, this allele has been found in a very low frequency in Asia when compared with other regions (1.57%, $P < 0.001$).

Variability of the *CYP2C19* 'predicted' phenotypes from genotype across the ethnic and geographic groups

The frequencies of gPMs and gUMs according to ethnicity are shown in Table 2. The highest frequency of gPMs was found in Native Oceanians (58.96%, $P < 0.001$), followed by East Asians (12.82%, $P < 0.001$), whereas gPMs were quite less frequent among North Africans (0.82%, $P < 0.001$). Furthermore, gUMs presented a high frequency in most of the ethnic groups, with the highest frequency reported for Middle Easterns (35.96%, $P < 0.001$), American Caucasians (32.87%. $P < 0.001$) and American admixed (30.09%, $P < 0.001$). However, individuals identified as gUMs were the 1.26% of East Asians ($P < 0.001$).

Considering the geographic regions, gPMs were more prevalent in Oceania than in the rest of the world (58.15%, $P < 0.001$), whereas, contrarily, the lowest frequency was found in populations from Europe (1.76%, $P < 0.001$). As in the analysis by ethnic groups, gUMs were also more frequently found in Middle East region (35.96%, $P < 0.001$), as well as were quite frequent in America (29.30%) and Europe (28.72%), but they were not as prevalent in countries from Asia (3.47%, $P < 0.001$; Table 3).

Variability of the *CYP2C19* 'measured' metabolic phenotype across the ethnic and geographic groups

Considering ethnicity, CYP2C19 mPMs were more frequent between East and South Asians (14.99% and 12.78%, respectively, $P < 0.001$ in both cases), and less frequent in Circumpolar populations (1.97%). Additionally, no studies on actual metabolic phenotype were found for North Asians, North Africans, African admixed or Australian Caucasians populations (Table 2).

Table 2. Frequencies of *CYP2C19* 'predicted' phenotypes from genotype and 'measured' metabolic phenotypes in different ethnic groups

Ethnic group		*'Predicted' phenotype from genotype*						*'Measured' metabolic phenotype*			*References*
		N	n *gPMs*	% *gPMs*	N	n *gUMs*	% *gUMs*	N	n *mPMs*	% *mPMs*	
Black	Black Africans	2259	75	3.32	199	33	16.58	881	41	4.65	51–55,57–62,64,153
	North Africans	367	3	0.82	—	—	—	—	—	—	65,66
	African Americans	336	12	3.57	—	—	—	229	9	3.93	67,68
Middle East	Middle Easterns	2149	60	2.79	381	137	35.96	821	33	4.02	48,69–79,154–157
Natives	Native Americans	548	16	2.92	183	38	20.77	302	21	6.95	80,81,158,159
	Native Oceanians	6706	3954	58.96	—	—	—	157	11	7.01	36,38–44
	Circumpolar	311	12	3.86	—	—	—	152	3	1.97	84,85
Admixed	African admixed	75	6	8.00	75	13	17.33	—	—	—	61
	American admixed	3946	94	2.38	1326	399	30.09	347	15	4.32	68,80,81,86–89,91–95,160
Caucasians	American Caucasian	416	15	3.61	143	47	32.87	546	36	6.59	25,30,68,96,161
	Australian Caucasian	99	3	3.03	—	—	—	—	—	—	97
	Scandinavians	679	15	2.21	585	173	29.57	1317	37	2.81	24,27,28,98,100,162
	Central Europeans	603	8	1.33	—	—	—	—	96	2.04	59,102–105,163
	Mediterranean-South Europeans	925	17	1.84	282	76	26.95	505	13	2.57	23,26,94,107,109
	Slavs	642	10	1.56	—	—	—	366	8	2.19	110,112,164
Asians	South Asians	1889	174	9.21	189	38	20.11	681	87	12.78	105,113–116,118–120,165–169
	East Asians	9884	1267	12.82	1426	18	1.26	2802	420	14.99	28,30,47,68,85,96,121–124,126,128–135,138,140–142,144,146–151,155,169–174
	North Asians	—	—	—	—	—	—	—	—	—	

Abbreviations: PM, poor metabolizers; UM, ultrarapid metabolizers. Individuals were considered as gUMs either when they were heterozygous or homozygous for *CYP2C19*17* allele; therefore, the frequency of gUMs reported in this article should be interpreted cautiously (see Discussion section). *N*, total number of subjects studied. The number of studied subjects differs in gPMs, gUMs and mPMs; therefore, *n* is the number of subjects considered as PMs or UMs either from the genotype (gPMs or gUMs) or phenotype (mPMs) evaluation.

Table 3. Frequencies of *CYP2C19* 'predicted' phenotypes from genotype and 'measured' metabolic phenotypes in the different geographic regions

Continent	*'Predicted' phenotype from genotype*						*'Measured' metabolic phenotype*			*References*
	N	n *gPMs*	% *gPMs*	N	n *gUMs*	% *gUMs*	N	n *mPMs*	% *mPMs*	
Africa	2701	84	3.11	274	46	16.79	881	41	4.65	51–55,57–62,64–66,142,153
Middle East	2149	60	2.79	381	137	35.96	821	33	4.02	48,69–79,154–157
Europe	2849	50	1.76	867	249	28.72	6884	154	2.24	23,24,26–28,59,94,98,100,102–105,107,109,110,112,162–164
America	5557	149	2.68	1652	484	29.30	1576	84	5.33	25,30,67,68,80,81,84–89,91–96,158–161
Asia	11 773	1441	12.24	1615	56	3.47	3483	507	14.56	28,30,47,68,80,85,96,105,113–116,118–124,126,128–135,138,140–142,144,146–151,155,165–174
Oceania	6805	3957	58.15	—	—	—	157	11	7.01	36,38–44,97

Abbreviations: PM, poor metabolizers; UM, ultrarapid metabolizers. Individuals were considered as gUMs either when they were heterozygous or homozygous for *CYP2C19*17* allele; therefore, the frequency of gUMs reported in this article should be interpreted cautiously (see Discussion section). *N*, total number of subjects studied. The number of studied subjects differs in gPMs, gUMs and mPMs, and therefore, *n* indicates the number of subjects considered as PMs or UMs either from the genotype (gPMs or gUMs) or phenotype (mPMs) evaluation.

Geographically, the mPMs were found in a high frequency in Asia (14.56%, $P < 0.001$), while the lowest frequency was observed in populations from Europe (2.24%, $P < 0.001$), as indicated in Table 3.

Correlation between the 'predicted' phenotype from genotype and 'measured' metabolic phenotype

High correlation coefficients were found between the frequencies of CYP2C19 gPMs and mPMs when either ethnicity ($r = 0.935$, $P < 0.001$) or geographic region ($r = 0.988$, $P = 0.002$) were selected as classification criteria (Figure 3). Oceania and Native Oceanian populations were excluded for these analyses as they are heterogenous populations and there are no studies of gPM and mPM frequencies in the same populations, resulting in a remarkable difference between the gPM and mPM frequency (Tables 2 and 3).

DISCUSSION

The frequencies of *CYP2C19* alleles present interethnic variability and show different geographic pattern of distribution from other *CYPs*.[31] *CYP2C19*2* and *CYP2C19*3* alleles are frequent among Asians and Oceanians, as previously reported.[1] Regarding the metabolic phenotype, the PMs are common among Asians, while the ultrarapid metabolism 'predicted' from genotype is frequent in

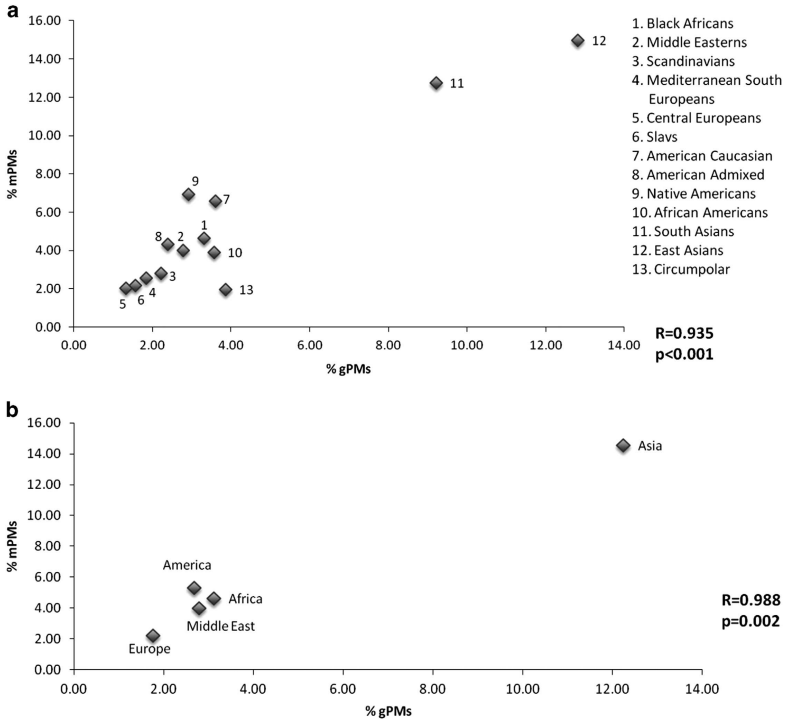

Figure 3. Relationship between the 'predicted' CYP2C19 poor metabolizer (PM) phenotype from genotype (gPMs) and CYP2C19 PM 'measured' metabolic phenotype (mPMs) in different ethnic groups (**a**) and geographic regions (**b**)*. *North Africans, African Admixed and North Asians were not included in the correlation analysis by ethnic groups as no data for 'measured' metabolic phenotype were available for these populations. Moreover, Native Oceanian (**a**) and populations from Oceania (**b**) were excluded for these analyses as these are heterogeneous populations and there are no studies of gPM and mPM frequencies in the same populations (see Discussion section).

populations from Africa, Middle East, America and Europe. To the best of our knowledge, this study is the first to determine *CYP2C19* alleles' frequencies by ethnic groups and geographic regions across worldwide populations, as well as the variability of 'predicted' phenotypes from genotypes and 'measured' metabolic phenotypes, and their potential correlation.

According to the studied population size of each geographic region,[34] the current knowledge about CYP2C19 across worldwide populations seems to be founded on an overrepresentation of European and Oceanian subjects. The reports for Africans and Asians are under-represented as the percentage of subjects studied for CYP2C19 is below the world average for the African population (6.68% vs 14.72%) and for Asians (31.71% vs 55.67%).

Currently, some of the better known *CYP2C19* alleles have been widely studied worldwide; however, there is a gap regarding many other allelic variants not extensively studied, which could be of interest. Many of the reported *CYP2C19* alleles are non-synonymous polymorphisms and could affect the protein product. Consequently, analyzing a higher number of alleles would not only increase the knowledge on *CYP2C19* variants but also their potential importance for predicting the responses to drug therapy while administering CYP2C19 substrates, particularly in those geographic regions where a high prevalence of PMs or UMs has been detected. With regard to the ethnicity analyses, *CYP2C19* allele frequencies showed an interethnic variability throughout the global population that could be influenced by ancestral migrations and the human population history.[35] In this sense, the unusual high frequencies of *CYP2C19*2* and *CYP2C19*3* alleles in Oceania may be due to several factors: first, >99% of the studied population in Oceania were Native Oceanians from Papua New Guinea, Vanuatu and Melanesia, which comprise geographically isolated populations ethnically diverse; and, second, the colonization of the Pacific Islands is believed to have started from Southeast Asia, where *CYP2C19*2* and *CYP2C19*3* alleles are highly prevalent, and thus a genetic drift process may be involved.[36] For the rest of the geographic regions, *CYP2C19*2* and *CYP2C19*3*, which are associated with reduced activity, are frequently found in Asian populations, whereas only *CYP2C19*2* has been observed in European, American and African populations. *CYP2C19*17* allele is common in Africa, America, Middle East and Europe, whereas it is rarely found in Asia. The variability on the distribution of *CYP2C19* alleles throughout the world may have clinical implications for drug development processes and clinical trials. Therefore, ethnic differences and migration phenomena should alert physicians about the importance of underlying genetic determinants on drug response when drugs are used in populations or subjects from different genetic backgrounds.[37]

The discrepancy between the CYP2C19 'predicted' phenotype from genotype and the CYP2C19 'measured' metabolic phenotype (58.15% vs 7.01%, respectively) observed for Oceania could be due to different reasons. Studies were carried out in different populations within Oceania, and while genotype analyses were developed in populations from Vanuatu, Papua New Guinea, Australia and Melanesia,[36,38–41] the metabolic phenotype studies

were just carried out in New Zealand.[42–44] Also, the ethnic diversity and geographical isolation of these populations hinders the correlation studies, and even more, none of these studies included both genotypic and phenotypic measures within the same population. Nevertheless, in the rest of the geographic regions and ethnic groups, the frequency of gPMs was strongly correlated with the mPM frequency, which could indicate that *CYP2C19* genotype could adequately predict the poor metabolic phenotype.

Both CYP2C19 gPMs and mPMs were more frequently found in Asians than in Caucasians, which is opposite to the phenomenon observed for the CYP2D6 geographic distribution, for which gPMs are more common in Europe and other Caucasian populations than across Asians.[31] Regarding the so-called ultrarapid metabolism, individuals were considered as gUMs either when they were heterozygous or homozygous for *CYP2C19*17* allele, as previously stated in CPIC guidelines,[2] as well as in other studies.[45,46] However, some other reports have considered individuals with *wt/*17* genotype as extensive metabolizers.[47,48] Therefore, the frequency of gUMs reported in this study should be cautiously considered, and further studies are required to confirm the enhanced activity of this variant across all populations.

Interethnic differences on gPMs and mPMs frequencies could have important clinical implications, and genotype–clinical outcome association studies are warranted to evaluate the clinical implications of these differences. For instance, there is a high prevalence of malaria in Pacific Islands and proguanil is highly prescribed, therefore, dose-dependent adverse effects could be expected, even though this has not been fully confirmed.[49] Regarding the pro-drug clopidogrel, the clinical consequences in CYP2C19 PMs are more severe, as they could present a lack of efficacy in antiplatelet therapy. Recently, a fatal case was reported in a Hawaiian individual under treatment with clopidogrel who was postmortem genotyped as *CYP2C19*2/*3*. The authors highlighted the importance of *CYP2C19* genotyping prior to antiplatelet prescription in Pacific Island and populations with a high prevalence of *CYP2C19* null alleles.[50] Moreover, there is no enough information about the correlation of CYP2C19 phenotype and genotype in African populations. However, owing to much genetic heterogeneity in this region, it is probable that this relation will be less strong. Thus Regulatory Agencies should not generalize their CYP2C19-based PGx dosing recommendations in all populations but rather be specific to certain geographical regions or ethnic groups.

To sum up, most of the studies included in this article just reported 'predicted' phenotype from genotype or 'measured' metabolic phenotype, and only few studies performed both analyses simultaneously. More precisely, for 20.5% of the individuals in whom metabolic phenotype was measured using a probe drug, *CYP2C19* genotype was also determined. Furthermore, when the correlation analysis between these phenotypes' frequencies was carried out we did not include the same individuals, which represents a limitation of this study. Consequently, additional research including both 'predicted' phenotypes from genotype and 'measured' metabolic phenotypes is still necessary to confirm whether *CYP2C19* genotype can predict accurately the actual metabolic phenotype for all the individuals.

CONCLUSION

Worldwide genetic variability of the *CYP2C19* gene is a widely recognized phenomenon. *CYP2C19*2* and *CYP2C19*3* alleles are more frequent in East Asians and in Native Oceanians than in the rest of the world, whereas *CYP2C19*17* was more frequent in Mediterranean-South Europeans and Middle Easterns ($P < 0.001$, for all cases). Regarding the 'predicted' metabolic phenotype from genotype, gPMs have been found to be frequent in Native Oceanians and East Asians, while gUMs are common in Middle Easterns and the admixed and Caucasian populations from America and Europe. Although the 'measured' metabolic phenotype has not been studied across all populations, a high frequency of mPMs was found in South and East Asians.

The knowledge on *CYP2C19* allele frequencies is therefore necessary to take better decisions to improve drug therapies when CYP2C19 substrates are administered. However, more studies are required, especially in those less analyzed populations such as Africans and Asians, as well as additional knowledge on those allelic variants scarcely studied. According to the data presented here, *CYP2C19* genotyping strategies are quite useful to identify CYP2C19 PMs, with the exception of specific populations from Oceania for which simultaneous studies of genotype and metabolic phenotype are required. Additional *CYP2C19* variants in all populations as well as the 'measured' metabolic phenotype should be still analyzed in larger populations from different ethnic groups. Further research is warranted to characterize the ultrarapid metabolic phenotype.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

This study was supported by Gobierno de Extremadura, AEXCID 13IA001 (to SIFF) and coordinated by the network Red Iberoamericana de Farmacogenética y Farmacogenómica (http://www.ribef.com). Consejo Nacional de Ciencia y Tecnología de México (CONACyT; no. 167261). Some researchers were supported by fellowships from the University of Costa Rica (PhD fellowship to CCG in Spain), Gobierno de Extremadura, Consejería de Empleo, Empresa e Innovación and Fondo Social Europeo (FSE) (TE14002 to AD), Instituto de Salud Carlos III-Sara Borrell Postdoc Program (CD13/00348 to FdA), Brazilian Government CAPES grant (BEX 12350/2013-03 fellowship to FR-S and 110/2012) and from CONACYT (no. 369708 to IF-G).

REFERENCES

1 Sistonen J, Fuselli S, Palo JU, Chauhan N, Padh H, Sajantila A. Pharmacogenetic variation at CYP2C9, CYP2C19, and CYP2D6 at global and microgeographic scales. *Pharmacogenet Genomics* 2009; **19**: 170–179.
2 Clinical Pharmacogenetics Implementation Consortium (CPIC). CYP2C19. Pharmacogenomics Knowledge Database. Available at: https://www.pharmgkb.org/cpic/alleles (accessed 20 March 2014).
3 De Andrés F, Sosa-Macías M, Llerena A. A rapid and simple LC-MS/MS method for the simultaneous evaluation of CYP1A2, CYP2C9, CYP2C19, CYP2D6 and CYP3A4 hydroxylation capacity. *Bioanalysis* 2014; **6**: 683–696.
4 Llerena A, Valdivielso MJ, Benítez J, Bertilsson L. Reproducibility over time of mephenytoin and debrisoquine hydroxylation phenotypes. *Pharmacol Toxicol* 1993; **73**: 46–48.
5 CYP2C19 allele nomenclature. Hum. Cytochrome P450 Allele Nomencl. Database. Available at: http://www.cypalleles.ki.se/cyp2c19.htm (accessed 15 May 2014).
6 Steimer W, Zöpf K, von Amelunxen S, Pfeiffer H, Bachofer J, Popp J *et al.* Amitriptyline or not, that is the question: pharmacogenetic testing of CYP2D6 and CYP2C19 identifies patients with low or high risk for side effects in amitriptyline therapy. *Clin Chem* 2005; **51**: 376–385.
7 Dorado P, López-Torres E, Peñas-Lledó EM, 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* 2013; **13**: 359–361.
8 LLerena A, Herraíz AG, Cobaleda J, Johansson I, Dahl ML. Debrisoquin and mephenytoin hydroxylation phenotypes and CYP2D6 genotype in patients treated with neuroleptic and antidepressant agents. *Clin Pharmacol Ther* 1993; **54**: 606–611.
9 Dahl ML, Llerena A, Bondesson U, Lindström L, Bertilsson L. Disposition of clozapine in man: lack of association with debrisoquine and S-mephenytoin hydroxylation polymorphisms. *Br J Clin Pharmacol* 1994; **37**: 71–74.
10 Chang M, Söderberg MM, Scordo MG, Tybring G, Dahl M-L. CYP2C19*17 affects R-warfarin plasma clearance and warfarin INR/dose ratio in patients on stable warfarin maintenance therapy. *Eur J Clin Pharmacol* 2015; **71**: 433–439.
11 Larsen P, Johnston L, Holley A, La Flamme A, Smyth L, Chua E *et al.* Prevalence and significance of CYP2C19*2 and CYP2C19*17 alleles in a New Zealand acute coronary syndrome population. *Intern Med J* 2015; **45**: 537–545.

12 Román M, Ochoa D, Sánchez-Rojas SD, Talegón M, Prieto-Pérez R, Rivas Â *et al.* Evaluation of the relationship between polymorphisms in CYP2C19 and the pharmacokinetics of omeprazole, pantoprazole and rabeprazole. *Pharmacogenomics* 2014; **15**: 1893–1901.
13 FDA. Genomics - Table of Pharmacogenomic Biomarkers in Drug Labeling. Available at: http://www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics/ucm083378.htm (accessed 16 June 2015).
14 Heads of Medicines Agencies: CMDh. Available at: http://www.hma.eu/cmdh.html (accessed 15 May 2015).
15 Klotz U. Clinical impact of CYP2C19 polymorphism on the action of proton pump inhibitors: a review of a special problem. *Int J Clin Pharmacol Ther* 2006; **44**: 297–302.
16 Stathopoulou MG, Monteiro P, Shahabi P, Peñas-Lledó E, El Shamieh S, Silva Santos L *et al.* Newly identified synergy between clopidogrel and calcium-channel blockers for blood pressure regulation possibly involves CYP2C19 rs4244285. *Int J Cardiol* 2013; **168**: 3057–3058.
17 Sim SC, Kacevska M, Ingelman-Sundberg M. Pharmacogenomics of drug-metabolizing enzymes: a recent update on clinical implications and endogenous effects. *Pharmacogenomics J* 2013; **13**: 1–11.
18 Sim SC, Nordin L, Andersson TM-L, Virding S, Olsson M, Pedersen NL *et al.* Association between CYP2C19 polymorphism and depressive symptoms. *Am J Med Genet B Neuropsychiatr Genet* 2010; **153B**: 1160–1166.
19 Persson A, Sim SC, Virding S, Onishchenko N, Schulte G, Ingelman-Sundberg M. Decreased hippocampal volume and increased anxiety in a transgenic mouse model expressing the human CYP2C19 gene. *Mol Psychiatry* 2014; **19**: 733–741.
20 Peñas-Lledó E, Guillaume S, Naranjo MEG, Delgado A, Jaussent I, Blasco-Fontecilla H *et al.* A combined high CYP2D6-CYP2C19 metabolic capacity is associated with the severity of suicide attempt as measured by objective circumstances. *Pharmacogenomics J* 2015; **15**: 172–176.
21 Xie HG, Kim RB, Wood AJ, Stein CM. Molecular basis of ethnic differences in drug disposition and response. *Annu Rev Pharmacol Toxicol* 2001; **41**: 815–850.
22 Kurose K, Sugiyama E, Saito Y. Population differences in major functional polymorphisms of pharmacokinetics/pharmacodynamics-related genes in Eastern Asians and Europeans: implications in the clinical trials for novel drug development. *Drug Metab Pharmacokinet* 2012; **27**: 9–54.
23 Reviriego J, Bertilsson L, Carrillo JA, Llerena A, Valdivielso MJ, Benítez J. Frequency of S-mephenytoin hydroxylation deficiency in 373 Spanish subjects compared to other Caucasian populations. *Eur J Clin Pharmacol* 1993; **44**: 593–595.
24 Drøhse A, Bathum L, Brøsen K, Gram LF. Mephenytoin and sparteine oxidation: genetic polymorphisms in Denmark. *Br J Clin Pharmacol* 1989; **27**: 620–625.
25 Guttendorf RJ, Britto M, Blouin RA, Foster TS, John W, Pittman KA *et al.* Rapid screening for polymorphisms in dextromethorphan and mephenytoin metabolism. *Br J Clin Pharmacol* 1990; **29**: 373–380.
26 Jacqz E, Dulac H, Mathieu H. Phenotyping polymorphic drug metabolism in the French Caucasian population. *Eur J Clin Pharmacol* 1988; **35**: 167–171.
27 Sanz EJ, Villén T, Alm C, Bertilsson L. S-mephenytoin hydroxylation phenotypes in a Swedish population determined after coadministration with debrisoquin. *Clin Pharmacol Ther* 1989; **45**: 495–499.
28 Bertilsson L, Lou YQ, Du YL, Liu Y, Kuang TY, Liao XM *et al.* Pronounced differences between native Chinese and Swedish populations in the polymorphic hydroxylations of debrisoquin and S-mephenytoin. *Clin Pharmacol Ther* 1992; **51**: 388–397.
29 Jurima M, Inaba T, Kadar D, Kalow W. Genetic polymorphism of mephenytoin p(4′)-hydroxylation: difference between Orientals and Caucasians. *Br J Clin Pharmacol* 1985; **19**: 483–487.
30 Nakamura K, Goto F, Ray WA, McAllister CB, Jacqz E, Wilkinson GR *et al.* Interethnic differences in genetic polymorphism of debrisoquin and mephenytoin hydroxylation between Japanese and Caucasian populations. *Clin Pharmacol Ther* 1985; **38**: 402–408.
31 LLerena A, Naranjo MEG, Rodrigues-Soares F, Penas-LLedó EM, Fariñas H, Tarazona-Santos E. Interethnic variability of CYP2D6 alleles and of predicted and measured metabolic phenotypes across world populations. *Expert Opin Drug Metab Toxicol* 2014; **10**: 1569–1583.
32 Crews KR, Gaedigk A, Dunnenberger HM, Leeder JS, Klein TE, Caudle KE *et al.* Clinical Pharmacogenetics Implementation Consortium guidelines for cytochrome P450 2D6 genotype and codeine therapy: 2014 update. *Clin Pharmacol Ther* 2014; **95**: 376–382.
33 Graphpad Software. Available at: http://www.graphpad.com/quickcalcs/contingency1/ (accessed 1 April 2015).
34 World population statistics. Available at: www.worldpopulationstatistics.com (accessed 1 May 2014).
35 Henn BM, Feldman MW. The great human expansion. *Proc Natl Acad Sci USA* 2012; **109**: 17758–17764.
36 Hsu H-L, Woad KJ, Woodfield DG, Helsby NA. A high incidence of polymorphic CYP2C19 variants in archival blood samples from Papua New Guinea. *Hum Genomics* 2008; **3**: 17–23.
37 Yusuf IRAWAN, Djojosubroto MW, RIKA Wan, Lum K, Kaneko A, Marzukt S. Ethnic and geographical distributions of CYP2C19 alleles in the populations of Southeast Asia. In: Marzuki S, Verhoef J, Snippe H (eds). *Tropical Diseases*. Springer: New York, USA, 2003, pp 37–46.
38 Kaneko A, Kaneko O, Taleo G, Björkman A, Kobayakawa T. High frequencies of CYP2C19 mutations and poor metabolism of proguanil in Vanuatu. *Lancet* 1997; **349**: 921–922.
39 Kaneko A, Lum JK, Yaviong L, Takahashi N, Ishizaki T, Bertilsson L *et al.* High and variable frequencies of CYP2C19 mutations: medical consequences of poor drug metabolism in Vanuatu and other Pacific islands. *Pharmacogenetics* 1999; **9**: 581–590.
40 Griese EU, Ilett KF, Kitteringham NR, Eichelbaum M, Powell H, Spargo RM *et al.* Allele and genotype frequencies of polymorphic cytochromes P4502D6, 2C19 and 2E1 in aborigines from western Australia. *Pharmacogenetics* 2001; **11**: 69–76.
41 Masta A, Lum JK, Tsukahara T, Hwaihwanje I, Kaneko A, Paniu MM *et al.* Analysis of Sepik populations of Papua New Guinea suggests an increase of CYP2C19 null allele frequencies during the colonization of Melanesia. *Pharmacogenetics* 2003; **13**: 697–700.
42 Wanwimolruk S, Bhawan S, Coville PF, Chalcroft SC. Genetic polymorphism of debrisoquine (CYP2D6) and proguanil (CYP2C19) in South Pacific Polynesian populations. *Eur J Clin Pharmacol* 1998; **54**: 431–435.
43 Wanwimolruk S, Pratt EL, Denton JR, Chalcroft SC, Barron PA, Broughton JR. Evidence for the polymorphic oxidation of debrisoquine and proguanil in a New Zealand Maori population. *Pharmacogenetics* 1995; **5**: 193–198.
44 Wanwimolruk S, Thou MR, Woods DJ. Evidence for the polymorphic oxidation of debrisoquine and proguanil in a Khmer (Cambodian) population. *Br J Clin Pharmacol* 1995; **40**: 166–169.
45 Rudberg I, Mohebi B, Hermann M, Refsum H, Molden E. Impact of the ultrarapid CYP2C19*17 allele on serum concentration of escitalopram in psychiatric patients. *Clin Pharmacol Ther* 2008; **83**: 322–327.
46 Bauer T, Bouman HJ, van Werkum JW, Ford NF, ten Berg JM, Taubert D. Impact of CYP2C19 variant genotypes on clinical efficacy of antiplatelet treatment with clopidogrel: systematic review and meta-analysis. *BMJ* 2011; **343**: d4588.
47 Sugimoto K, Uno T, Yamazaki H, Tateishi T. Limited frequency of the CYP2C19*17 allele and its minor role in a Japanese population. *Br J Clin Pharmacol* 2008; **65**: 437–439.
48 Payan M, Rouini MR, Tajik N, Ghahremani MH, Tahvilian R. Hydroxylation index of omeprazole in relation to CYP2C19 polymorphism and sex in a healthy Iranian population. *Daru* 2014; **22**: 81.
49 Kaneko A, Bergqvist Y, Taleo G, Kobayakawa T, Ishizaki T, Björkman A. Proguanil disposition and toxicity in malaria patients from Vanuatu with high frequencies of CYP2C19 mutations. *Pharmacogenetics* 1999; **9**: 317–326.
50 Wu AH, WhiteSam MJ, Burchard E. The Hawaii clopidogrel lawsuit : the possible effect on clinical laboratory testing. *Per Med* 2015; **12**: 179–181.
51 Masimirembwa C, Bertilsson L, Johansson I, Hasler JA, Ingelman-Sundberg M. Phenotyping and genotyping of S-mephenytoin hydroxylase (cytochrome P450 2C19) in a Shona population of Zimbabwe. *Clin Pharmacol Ther* 1995; **57**: 656–661.
52 Persson I, Aklillu E, Rodrigues F, Bertilsson L, Ingelman-Sundberg M. S-mephenytoin hydroxylation phenotype and CYP2C19 genotype among Ethiopians. *Pharmacogenetics* 1996; **6**: 521–526.
53 Herrlin K, Massele AY, Jande M, Alm C, Tybring G, Abdi YA *et al.* Bantu Tanzanians have a decreased capacity to metabolize omeprazole and mephenytoin in relation to their CYP2C19 genotype. *Clin Pharmacol Ther* 1998; **64**: 391–401.
54 Bathum L, Skjelbo E, Mutabingwa TK, Madsen H, Hørder M, Brøsen K. Phenotypes and genotypes for CYP2D6 and CYP2C19 in a black Tanzanian population. *Br J Clin Pharmacol* 1999; **48**: 395–401.
55 Dandara C, Masimirembwa CM, Magimba A, Sayi J, Kaaya S, Sommers DK *et al.* Genetic polymorphism of CYP2D6 and CYP2C19 in east- and southern African populations including psychiatric patients. *Eur J Clin Pharmacol* 2001; **57**: 11–17.
56 Dandara C, Lombard Z, Du Plooy I, McLellan T, Norris SA, Ramsay M. Genetic variants in CYP (-1A2, -2C9, -2C19, -3A4 and -3A5), VKORC1 and ABCB1 genes in a black South African population: a window into diversity. *Pharmacogenomics* 2011; **12**: 1663–1670.
57 Aklillu E, Herrlin K, Gustafsson LL, Bertilsson L, Ingelman-Sundberg M. Evidence for environmental influence on CYP2D6-catalysed debrisoquine hydroxylation as demonstrated by phenotyping and genotyping of Ethiopians living in Ethiopia or in Sweden. *Pharmacogenetics* 2002; **12**: 375–383.
58 Bolaji OO, Sadare IO, Babalola CP, Ogunbona FA. Polymorphic oxidative metabolism of proguanil in a Nigerian population. *Eur J Clin Pharmacol* 2002; **58**: 543–545.

59 Allabi AC, Gala J-L, Desager J-P, Heusterspreute M, Horsmans Y. Genetic polymorphisms of CYP2C9 and CYP2C19 in the Beninese and Belgian populations. *Br J Clin Pharmacol* 2003; **56**: 653–657.
60 Miura J, Obua C, Abbo C, Kaneko S, Tateishi T. Cytochrome P450 2C19 genetic polymorphisms in Ugandans. *Eur J Clin Pharmacol* 2009; **65**: 319–320.
61 Drögemöller BI, Wright GEB, Niehaus DJH, Koen L, Malan S, Da Silva DM *et al.* Characterization of the genetic profile of CYP2C19 in two South African populations. *Pharmacogenomics* 2010; **11**: 1095–1103.
62 Dodgen TM, Hochfeld WE, Fickl H, Asfaha SM, Durandt C, Rheeder P *et al.* Introduction of the AmpliChip CYP450 Test to a South African cohort: a platform comparative prospective cohort study. *BMC Med Genet* 2013; **14**: 20.
63 Janha RE, Worwui A, Linton KJ, Shaheen SO, Sisay-Joof F, Walton RT. Inactive alleles of cytochrome P450 2C19 may be positively selected in human evolution. *BMC Evol Biol* 2014; **14**: 71.
64 Kudzi W, Dodoo AN, Mills JJ. Characterisation of CYP2C8, CYP2C9 and CYP2C19 polymorphisms in a Ghanaian population. *BMC Med Genet* 2009; **10**: 124.
65 Hamdy SI, Hiratsuka M, Narahara K, El-Enany M, Moursi N, Ahmed MS-E *et al.* Allele and genotype frequencies of polymorphic cytochromes P450 (CYP2C9, CYP2C19, CYP2E1) and dihydropyrimidine dehydrogenase (DPYD) in the Egyptian population. *Br J Clin Pharmacol* 2002; **53**: 596–603.
66 Ellison CA, Abou El-Ella SS, Tawfik M, Lein PJ, Olson JR. Allele and genotype frequencies of CYP2B6 and CYP2C19 polymorphisms in Egyptian agricultural workers. *J Toxicol Environ Health A* 2012; **75**: 232–241.
67 Marinac JS, Balian JD, Foxworth JW, Willsie SK, Daus JC, Owen R *et al.* Determination of CYP2C19 phenotype in black Americans with omeprazole: correlation with genotype. *Clin Pharmacol Ther* 1996; **60**: 138–144.
68 Luo H-R, Poland RE, Lin K-M, Wan Y-JY. Genetic polymorphism of cytochrome P450 2C19 in Mexican Americans: a cross-ethnic comparative study. *Clin Pharmacol Ther* 2006; **80**: 33–40.
69 Aynacioglu AS, Sachse C, Bozkurt A, Kortunay S, Nacak M, Schröder T *et al.* Low frequency of defective alleles of cytochrome P450 enzymes 2C19 and 2D6 in the Turkish population. *Clin Pharmacol Ther* 1999; **66**: 185–192.
70 Sviri S, Shpizen S, Leitersdorf E, Levy M, Caraco Y. Phenotypic-genotypic analysis of CYP2C19 in the Jewish Israeli population. *Clin Pharmacol Ther* 1999; **65**: 275–282.
71 Luo H-R, Aloumanis V, Lin K-M, Gurwitz D, Wan Y-JY. Polymorphisms of CYP2C19 and CYP2D6 in Israeli ethnic groups. *Am J Pharmacogenomics* 2004; **4**: 395–401.
72 Scott SA, Edelmann L, Kornreich R, Erazo M, Desnick RJ. CYP2C9, CYP2C19 and CYP2D6 allele frequencies in the Ashkenazi Jewish population. *Pharmacogenomics* 2007; **8**: 721–730.
73 Zand N, Tajik N, Moghaddam AS, Milanian I. Genetic polymorphisms of cytochrome P450 enzymes 2C9 and 2C19 in a healthy Iranian population. *Clin Exp Pharmacol Physiol* 2007; **34**: 102–105.
74 Djaffar Jureidini I, Chamseddine N, Keleshian S, Naoufal R, Zahed L, Hakime N. Prevalence of CYP2C19 polymorphisms in the Lebanese population. *Mol Biol Rep* 2011; **38**: 5449–5452.
75 Yousef A-M, Bulatova NR, Newman W, Hakooz N, Ismail S, Qusa H *et al.* Allele and genotype frequencies of the polymorphic cytochrome P450 genes (CYP1A1, CYP3A4, CYP3A5, CYP2C9 and CYP2C19) in the Jordanian population. *Mol Biol Rep* 2012; **39**: 9423–9433.
76 Zalloum I, Hakooz N, Arafat T. Genetic polymorphism of CYP2C19 in a Jordanian population: influence of allele frequencies of CYP2C19*1 and CYP2C19*2 on the pharmacokinetic profile of lansoprazole. *Mol Biol Rep* 2012; **39**: 4195–4200.
77 Al-Jenoobi FI, Alkharfy KM, Alghamdi AM, Bagulb KM, Al-Mohizea AM, Al-Muhsen S *et al.* CYP2C19 genetic polymorphism in Saudi Arabians. *Basic Clin Pharmacol Toxicol* 2013; **112**: 50–54.
78 Tabari RG, Marjani A, Ataby OA, Mansourian AR, Samai NM. Genetic polymorphism of cytochrome p450 (2C19) enzyme in Iranian Turkman Ethnic group. *Oman Med J* 2013; **28**: 237–244.
79 Saeed LH, Mayet AY. Genotype-phenotype analysis of CYP2C19 in healthy saudi individuals and its potential clinical implication in drug therapy. *Int J Med Sci* 2013; **10**: 1497–1502.
80 Santos PCJL, Soares RAG, Santos DBG, Nascimento RM, Coelho GLLM, Nicolau JC *et al.* CYP2C19 and ABCB1 gene polymorphisms are differently distributed according to ethnicity in the Brazilian general population. *BMC Med Genet* 2011; **12**: 13.
81 Salazar-Flores J, Torres-Reyes LA, Martínez-Cortés G, Rubi-Castellanos R, Sosa-Macías M, Muñoz-Valle JF *et al.* Distribution of CYP2D6 and CYP2C19 polymorphisms associated with poor metabolizer phenotype in five Amerindian groups and western Mestizos from Mexico. *Genet Test Mol Biomarkers* 2012; **16**: 1098–1104.
82 Vargens DD, Petzl-Erler M-L, Suarez-Kurtz G. Distribution of CYP2C polymorphisms in an Amerindian population of Brazil. *Basic Clin Pharmacol Toxicol* 2012; **110**: 396–400.
83 Lea RA, Roberts RL, Green MR, Kennedy MA, Chambers GK. Allele frequency differences of cytochrome P450 polymorphisms in a sample of New Zealand Māori. *N Z Med J* 2008; **121**: 33–37.
84 Jurima-Romet M, Goldstein JA, LeBelle M, Aubin RA, Foster BC, Walop W *et al.* CYP2C19 genotyping and associated mephenytoin hydroxylation polymorphism in a Canadian Inuit population. *Pharmacogenetics* 1996; **6**: 329–339.
85 Nowak MP, Sellers EM, Tyndale RF. Canadian Native Indians exhibit unique CYP2A6 and CYP2C19 mutant allele frequencies. *Clin Pharmacol Ther* 1998; **64**: 378–383.
86 Bravo-Villalta HV, Yamamoto K, Nakamura K, Bayá A, Okada Y, Horiuchi R. Genetic polymorphism of CYP2C9 and CYP2C19 in a Bolivian population: an investigative and comparative study. *Eur J Clin Pharmacol* 2005; **61**: 179–184.
87 Isaza C, Henao J, Martínez JHI, Sepúlveda Arias JC, Beltrán L. Phenotype-genotype analysis of CYP2C19 in Colombian mestizo individuals. *BMC Clin Pharmacol* 2007; **7**: 6.
88 Duconge J, Cadilla CL, Renta JY, Silén-Rivera P, Piovanetti P, García-Berdecía R *et al.* Prevalence of CYP2C19 gene polymorphisms in the Puerto Rican population: a preliminary report. *P R Health Sci J* 2008; **27**: 357–358.
89 Roco A, Quiñones L, Agúndez JAG, García-Martín E, Squicciarini V, Miranda C *et al.* Frequencies of 23 functionally significant variant alleles related with metabolism of antineoplastic drugs in the Chilean population: comparison with Caucasian and Asian populations. *Front Genet* 2012; **3**: 229.
90 Suarez-Kurtz G, Genro JP, de Moraes MO, Ojopi EB, Pena SDJ, Perini JA *et al.* Global pharmacogenomics: Impact of population diversity on the distribution of polymorphisms in the CYP2C cluster among Brazilians. *Pharmacogenomics J* 2012; **12**: 267–276.
91 Castro de Guerra D, Flores S, Izaguirre MH. Distribution of CYP2C19*2 and CYP2C19*3 polymorphisms in Venezuelan populations with different admixture. *Ann Hum Biol* 2013; **40**: 197–200.
92 Orengo-Mercado C, Nieves B, López L, Vallés-Ortiz N, Renta JY, Santiago-Borrero PJ *et al.* Frequencies of functional polymorphisms in three pharmacokinetic genes of clinical interest within the admixed Puerto Rican population. *J Pharmacogenomics Pharmacoproteomics* 2013; **4**: 113.
93 Kohlrausch FB, Carracedo Á, Hutz MH. Characterization of CYP1A2, CYP2C19, CYP3A4 and CYP3A5 polymorphisms in South Brazilians. *Mol Biol Rep* 2014; **41**: 1453–1460.
94 Vicente J, González-Andrade F, Soriano A, Fanlo A, Martínez-Jarreta B, Sinués B. Genetic polymorphisms of CYP2C8, CYP2C9 and CYP2C19 in Ecuadorian Mestizo and Spaniard populations: a comparative study. *Mol Biol Rep* 2014; **41**: 1267–1272.
95 Vargas-Alarcón G, Ramírez-Bello J, de la Peña A, Calderón-Cruz B, Peña-Duque MA, Martínez-Ríos MA *et al.* Distribution of ABCB1, CYP3A5, CYP2C19, and P2RY12 gene polymorphisms in a Mexican Mestizos population. *Mol Biol Rep* 2014; **41**: 7023–7029.
96 Myrand SP, Sekiguchi K, Man MZ, Lin X, Tzeng R-Y, Teng C-H *et al.* Pharmacokinetics/genotype associations for major cytochrome P450 enzymes in native and first- and third-generation Japanese populations: comparison with Korean, Chinese, and Caucasian populations. *Clin Pharmacol Ther* 2008; **84**: 347–361.
97 Hoskins JM, Shenfield GM, Gross AS. Relationship between proguanil metabolic ratio and CYP2C19 genotype in a Caucasian population. *Br J Clin Pharmacol* 1998; **46**: 499–504.
98 Bramness JG, Skurtveit S, Fauske L, Grung M, Molven A, Mørland J *et al.* Association between blood carisoprodol:meprobamate concentration ratios and CYP2C19 genotype in carisoprodol-drugged drivers: decreased metabolic capacity in heterozygous CYP2C19*1/CYP2C19*2 subjects? *Pharmacogenetics* 2003; **13**: 383–388.
99 Hilli J, Rane A, Lundgren S, Bertilsson L, Laine K. Genetic polymorphism of cytochrome P450s and P-glycoprotein in the Finnish population. *Fundam Clin Pharmacol* 2007; **21**: 379–386.
100 Pedersen RS, Brasch-Andersen C, Sim SC, Bergmann TK, Halling J, Petersen MS *et al.* Linkage disequilibrium between the CYP2C19*17 allele and wildtype CYP2C8 and CYP2C9 alleles: identification of CYP2C haplotypes in healthy Nordic populations. *Eur J Clin Pharmacol* 2010; **66**: 1199–1205.
101 Ramsjö M, Aklillu E, Bohman L, Ingelman-Sundberg M, Roh H-K, Bertilsson L. CYP2C19 activity comparison between Swedes and Koreans: effect of genotype, sex, oral contraceptive use, and smoking. *Eur J Clin Pharmacol* 2010; **66**: 871–877.
102 Brockmöller J, Rost KL, Gross D, Schenkel A, Roots I. Phenotyping of CYP2C19 with enantiospecific HPLC-quantification of R- and S-mephenytoin and comparison with the intron4/exon5 G–>A-splice site mutation. *Pharmacogenetics* 1995; **5**: 80–88.
103 Tamminga WJ, Wemer J, Oosterhuis B, Weiling J, Wilffert B, de Leij LF *et al.* CYP2D6 and CYP2C19 activity in a large population of Dutch healthy volunteers: indications for oral contraceptive-related gender differences. *Eur J Clin Pharmacol* 1999; **55**: 177–184.

104 Rideg O, Háber A, Botz L, Szücs F, Várnai R, Miseta A *et al.* Pilot study for the characterization of pharmacogenetically relevant CYP2D6, CYP2C19 and ABCB1 gene polymorphisms in the Hungarian population. *Cell Biochem Funct* 2011; **29**: 562–568.
105 Sipeky C, Weber A, Szabo M, Melegh BI, Janicsek I, Tarlos G *et al.* High prevalence of CYP2C19*2 allele in Roma samples: study on Roma and Hungarian population samples with review of the literature. *Mol Biol Rep* 2013; **40**: 4727–4735.
106 Ruas JL, Lechner MC. Allele frequency of CYP2C19 in a Portuguese population. *Pharmacogenetics* 1997; **7**: 333–335.
107 Scordo MG, Caputi AP, D'Arrigo C, Fava G, Spina E. Allele and genotype frequencies of CYP2C9, CYP2C19 and CYP2D6 in an Italian population. *Pharmacol Res* 2004; **50**: 195–200.
108 Oliveira E, Marsh S, van Booven DJ, Amorim A, Prata MJ, McLeod HL. Pharmacogenetically relevant polymorphisms in Portugal. *Pharmacogenomics* 2007; **8**: 703–712.
109 Arvanitidis K, Ragia G, Iordanidou M, Kyriaki S, Xanthi A, Tavridou A *et al.* Genetic polymorphisms of drug-metabolizing enzymes CYP2D6, CYP2C9, CYP2C19 and CYP3A5 in the Greek population. *Fundam Clin Pharmacol* 2007; **21**: 419–426.
110 Gaikovitch EA, Cascorbi I, Mrozikiewicz PM, Brockmöller J, Frötschl R, Köpke K *et al.* Polymorphisms of drug-metabolizing enzymes CYP2C9, CYP2C19, CYP2D6, CYP1A1, NAT2 and of P-glycoprotein in a Russian population. *Eur J Clin Pharmacol* 2003; **59**: 303–312.
111 Makeeva O, Stepanov V, Puzyrev V, Goldstein DB, Grossman I. Global pharmacogenetics: genetic substructure of Eurasian populations and its effect on variants of drug-metabolizing enzymes. *Pharmacogenomics* 2008; **9**: 847–868.
112 Gra O, Mityaeva O, Berdichevets I, Kozhekbaeva Z, Fesenko D, Kurbatova O *et al.* Microarray-based detection of CYP1A1, CYP2C9, CYP2C19, CYP2D6, GSTT1, GSTM1, MTHFR, MTRR, NQO1, NAT2, HLA-DQA1, and AB0 allele frequencies in native Russians. *Genet Test Mol Biomarkers* 2010; **14**: 329–342.
113 Adithan C, Gerard N, Vasu S, Rosemary J, Shashindran CH, Krishnamoorthy R. Allele and genotype frequency of CYP2C19 in a Tamilian population. *Br J Clin Pharmacol* 2003; **56**: 331–333.
114 Jose R, Chandrasekaran A, Sam SS, Gerard N, Chanolean S, Abraham BK *et al.* CYP2C9 and CYP2C19 genetic polymorphisms: frequencies in the south Indian population. *Fundam Clin Pharmacol* 2005; **19**: 101–105.
115 Panchabhai TS, Noronha SF, Davis S, Shinde VM, Kshirsagar NA, Gogtay NJ. Evaluation of the activity of CYP2C19 in Gujrati and Marwadi subjects living in Mumbai (Bombay). *BMC Clin Pharmacol* 2006; **6**: 8.
116 Ghodke Y, Joshi K, Arya Y, Radkar A, Chiplunkar A, Shintre P *et al.* Genetic polymorphism of CYP2C19 in Maharashtrian population. *Eur J Epidemiol* 2007; **22**: 907–915.
117 Satyanarayana CRU, Devendran A, Sundaram R, Gopal SD, Rajagopal K, Chandrasekaran A. Genetic variations and haplotypes of the 5′ regulatory region of CYP2C19 in South Indian population. *Drug Metab Pharmacokinet* 2009; **24**: 185–193.
118 Anichavezhi D, Chakradhara Rao US, Shewade DG, Krishnamoorthy R, Adithan C. Distribution of CYP2C19*17 allele and genotypes in an Indian population. *J Clin Pharm Ther* 2012; **37**: 313–318.
119 Shalia KK, Shah VK, Pawar P, Divekar SS, Payannavar S. Polymorphisms of MDR1, CYP2C19 and P2Y12 genes in Indian population: effects on clopidogrel response. *Indian Heart J* 2013; **65**: 158–167.
120 Gulati S, Yadav A, Kumar N, Kanupriya, Kumar G, Aggarwal N *et al.* Frequency distribution of high risk alleles of CYP2C19, CYP2E1, CYP3A4 genes in Haryana population. *Environ Toxicol Pharmacol* 2014; **37**: 1186–1193.
121 De Morais SM, Goldstein JA, Xie HG, Huang SL, Lu YQ, Xia H *et al.* Genetic analysis of the S-mephenytoin polymorphism in a Chinese population. *Clin Pharmacol Ther* 1995; **58**: 404–411.
122 Kubota T, Chiba K, Ishizaki T. Genotyping of S-mephenytoin 4′-hydroxylation in an extended Japanese population. *Clin Pharmacol Ther* 1996; **60**: 661–666.
123 Takakubo F, Kuwano A, Kondo I. Evidence that poor metabolizers of (S)-mephenytoin could be identified by haplotypes of CYP2C19 in Japanese. *Pharmacogenetics* 1996; **6**: 265–267.
124 Xiao ZS, Goldstein JA, Xie HG, Blaisdell J, Wang W, Jiang CH *et al.* Differences in the incidence of the CYP2C19 polymorphism affecting the S-mephenytoin phenotype in Chinese Han and Bai populations and identification of a new rare CYP2C19 mutant allele. *J Pharmacol Exp Ther* 1997; **281**: 604–609.
125 Kimura M, Ieiri I, Mamiya K, Urae A, Higuchi S. Genetic polymorphism of cytochrome P450s, CYP2C19, and CYP2C9 in a Japanese population. *Ther Drug Monit* 1998; **20**: 243–247.
126 Kimura M, Ieiri I, Wada Y, Mamiya K, Urae A, Iimori E *et al.* Reliability of the omeprazole hydroxylation index for CYP2C19 phenotyping: possible effect of age, liver disease and length of therapy. *Br J Clin Pharmacol* 1999; **47**: 115–119.
127 Kubota T, Hibi N, Chiba K. Linkage of mutant alleles of CYP2C18 and CYP2C19 in a Japanese population. *Biochem Pharmacol* 1998; **55**: 2039–2042.
128 Yao TW, Zeng S, Wang TW, Chen SQ. Phenotype analysis of cytochrome P450 2C19 in Chinese subjects with mephenytoin S/R enantiomeric ratio in urine measured by chiral GC. *Biomed Chromatogr* 2001; **15**: 9–13.
129 Tassaneeyakul W, Tawalee A, Tassaneeyakul W, Kukongviriyapan V, Blaisdell J, Goldstein JA *et al.* Analysis of the CYP2C19 polymorphism in a North-eastern Thai population. *Pharmacogenetics* 2002; **12**: 221–225.
130 Tassaneeyakul W, Mahatthanatrakul W, Niwatananun K, Na-Bangchang K, Tawalee A, Krikreangsak N *et al.* CYP2C19 genetic polymorphism in Thai, Burmese and Karen populations. *Drug Metab Pharmacokinet* 2006; **21**: 286–290.
131 He N, Yan F-X, Huang S-L, Wang W, Xiao Z-S, Liu Z-Q *et al.* CYP2C19 genotype and S-mephenytoin 4′-hydroxylation phenotype in a Chinese Dai population. *Eur J Clin Pharmacol* 2002; **58**: 15–18.
132 Niu CY, Luo JY, Hao ZM. Genetic polymorphism analysis of cytochrome P4502C19 in Chinese Uigur and Han populations. *Chin J Dig Dis* 2004; **5**: 76–80.
133 Yang ZF, Cui HW, Hasi T, Jia SQ, Gong ML, Su XL. Genetic polymorphisms of cytochrome P450 enzymes 2C9 and 2C19 in a healthy Mongolian population in China. *Genet Mol Res* 2010; **9**: 1844–1851.
134 Ishiguro A, Kubota T, Soya Y, Sasaki H, Yagyu O, Takarada Y *et al.* High-throughput detection of multiple genetic polymorphisms influencing drug metabolism with mismatch primers in allele-specific polymerase chain reaction. *Anal Biochem* 2005; **337**: 256–261.
135 Lee SS, Lee S-J, Gwak J, Jung H-J, Thi-Le H, Song I-S *et al.* Comparisons of CYP2C19 genetic polymorphisms between Korean and Vietnamese populations. *Ther Drug Monit* 2007; **29**: 455–459.
136 Lee S-J, Kim W-Y, Kim H, Shon J-H, Lee SS, Shin J-G. Identification of new CYP2C19 variants exhibiting decreased enzyme activity in the metabolism of S-mephenytoin and omeprazole. *Drug Metab Dispos* 2009; **37**: 2262–2269.
137 Matsumoto N, Kakihara F, Kimura S, Kurebayashi Y, Hirai M, Yohda M *et al.* Single nucleotide polymorphism genotyping of CYP2C19 using a new automated system. *Anal Biochem* 2007; **370**: 121–123.
138 Wang J-H, Li P-Q, Fu Q-Y, Li Q-X, Cai W-W. Cyp2c19 genotype and omeprazole hydroxylation phenotype in Chinese Li population. *Clin Exp Pharmacol Physiol* **34**: 421–424.
139 Chen L, Qin S, Xie J, Tang J, Yang L, Shen W *et al.* Genetic polymorphism analysis of CYP2C19 in Chinese Han populations from different geographic areas of mainland China. *Pharmacogenomics* 2008; **9**: 691–702.
140 Jin SK, Kang TS, Eom SO, Kim J-I, Lee HJ, Roh J. CYP2C19 haplotypes in Koreans as a marker of enzyme activity evaluated with omeprazole. *J Clin Pharm Ther* 2009; **34**: 437–446.
141 Perini JA, Vargens DD, Santana ISC, Moriguchi EH, Ribeiro-Dos-Santos AKC, Tsutsumi M *et al.* Pharmacogenetic polymorphisms in Brazilian-born, first-generation Japanese descendants. *Braz J Med Biol Res* 2009; **42**: 1179–1184.
142 Veiga MI, Asimus S, Ferreira PE, Martins JP, Cavaco I, Ribeiro V *et al.* Pharmacogenomics of CYP2A6, CYP2B6, CYP2C19, CYP2D6, CYP3A4, CYP3A5 and MDR1 in Vietnam. *Eur J Clin Pharmacol* 2009; **65**: 355–363.
143 Zhou Q, Yu XM, Lin HB, Wang L, Yun QZ, Hu SN *et al.* Genetic polymorphism, linkage disequilibrium, haplotype structure and novel allele analysis of CYP2C19 and CYP2D6 in Han Chinese. *Pharmacogenomics J* 2009; **9**: 380–394.
144 Kim K-A, Song W-K, Kim K-R, Park J-Y. Assessment of CYP2C19 genetic polymorphisms in a Korean population using a simultaneous multiplex pyrosequencing method to simultaneously detect the CYP2C19*2, CYP2C19*3, and CYP2C19*17 alleles. *J Clin Pharm Ther* 2010; **35**: 697–703.
145 Shi Y, Xiang P, Li L, Shen M. Analysis of 50 SNPs in CYP2D6, CYP2C19, CYP2C9, CYP3A4 and CYP1A2 by MALDI-TOF mass spectrometry in Chinese Han population. *Forensic Sci Int* 2011; **207**: 183–187.
146 Wei W, Fang L, Wang N, Zhang T, Zeng J, Lin M. Prevalence of CYP2C19 polymorphisms involved in clopidogrel metabolism in Fujian Han population. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* 2012; **29**: 420–425.
147 Hu L-M, Dai D-P, Hu G-X, Yang J-F, Xu R-A, Yang L-P *et al.* Genetic polymorphisms and novel allelic variants of CYP2C19 in the Chinese Han population. *Pharmacogenomics* 2012; **13**: 1571–1581.
148 Yin S-J, Ni Y-B, Wang S-M, Wang X, Lou Y-Q, Zhang G-L. Differences in genotype and allele frequency distributions of polymorphic drug metabolizing enzymes CYP2C19 and CYP2D6 in mainland Chinese Mongolian, Hui and Han populations. *J Clin Pharm Ther* 2012; **37**: 364–369.
149 Wu Z, Zhang X, Shen L, Xiong Y, Wu X, Huo R *et al.* A systematically combined genotype and functional combination analysis of CYP2E1, CYP2D6, CYP2C9, CYP2C19 in different geographic areas of mainland China--a basis for personalized therapy. *PLoS One* 2013; **8**: e71934.
150 Honda M, Ogura Y, Toyoda W, Taguchi M, Nozawa T, Inoue H *et al.* Multiple regression analysis of pharmacogenetic variability of carvedilol disposition in 54 healthy Japanese volunteers. *Biol Pharm Bull* 2006; **29**: 772–778.
151 Roh HK, Dahl ML, Tybring G, Yamada H, Cha YN, Bertilsson L. CYP2C19 genotype and phenotype determined by omeprazole in a Korean population. *Pharmacogenetics* 1996; **6**: 547–551.

152 Zuo LJ, Guo T, Xia DY, Jia LH. Allele and genotype frequencies of CYP3A4, CYP2C19, and CYP2D6 in Han, Uighur, Hui, and Mongolian Chinese populations. *Genet Test Mol Biomarkers* 2012; **16**: 102–108.
153 Skjelbo E, Mutabingwa TK, Bygbjerg Ib, Nielsen KK, Gram LF, Brøosen K. Chloroguanide metabolism in relation to the efficacy in malaria prophylaxis and the S-mephenytoin oxidation in Tanzanians. *Clin Pharmacol Ther* 1996; **59**: 304–311.
154 Basci NE, Bozkurt A, Kortunay S, Isimer A, Sayal A, Kayaalp SO. Proguanil metabolism in relation to S-mephenytoin oxidation in a Turkish population. *Br J Clin Pharmacol* 1996; **42**: 771–773.
155 Evans DA, Krahn P, Narayanan N. The mephenytoin (cytochrome P450 2C 19) and dextromethorphan (cytochrome P450 2D6) polymorphisms in Saudi Arabians and Filipinos. *Pharmacogenetics* 1995; **5**: 64–71.
156 Hadidi HF, Irshaid YM, Woosley RL, Idle JR, Flockhart DA. S-mephenytoin hydroxylation phenotypes in a Jordanian population. *Clin Pharmacol Ther* 1995; **58**: 542–547.
157 Kortunay S, Basci NE, Bozkurt A, Isimer A, Sayal A, Kayaalp SO. The hydroxylation of omeprazole correlates with S-mephenytoin and proguanil metabolism. *Eur J Clin Pharmacol* 1997; **53**: 261–264.
158 Arias TD, Jorge LF. Population pharmacology and ethnopharmacokinetics in 3 Amerindian groups from Panama: Cuna, Ngawbe Guaymi and Teribe. *Rev Med Panama* 1990; **15**: 20–26.
159 Inaba T, Jorge LF, Arias TD. Mephenytoin hydroxylation in the Cuna Amerindians of Panama. *Br J Clin Pharmacol* 1988; **25**: 75–79.
160 Gonzalez HM, Romero EM, Peregrina AA, de J Chávez T, Escobar-Islas E, Lozano F *et al.* CYP2C19- and CYP3A4-dependent omeprazole metabolism in West Mexicans. *J Clin Pharmacol* 2003; **43**: 1211–1215.
161 Wedlund PJ, Aslanian WS, McAllister CB, Wilkinson GR, Branch RA. Mephenytoin hydroxylation deficiency in Caucasians: frequency of a new oxidative drug metabolism polymorphism. *Clin Pharmacol Ther* 1984; **36**: 773–780.
162 Marandi T, Dahl ML, Rägo L, Kiivet R, Sjöqvist F. Debrisoquine and S-mephenytoin hydroxylation polymorphisms in a Russian population living in Estonia. *Eur J Clin Pharmacol* 1997; **53**: 257–260.
163 Küpfer A, Preisig R. Pharmacogenetics of mephenytoin: a new drug hydroxylation polymorphism in man. *Eur J Clin Pharmacol* 1984; **26**: 753–759.
164 Marandi T, Dahl ML, Kiivet RA, Rägo L, Sjöqvist F. Debrisoquin and S-mephenytoin hydroxylation phenotypes and CYP2D6 genotypes in an Estonian population. *Pharmacol Toxicol* 1996; **78**: 303–307.
165 Doshi BS, Kulkarni RD, Chauhan BL, Wilkinson GR. Frequency of impaired mephenytoin 4′-hydroxylation in an Indian population. *Br J Clin Pharmacol* 1990; **30**: 779–780.
166 Lamba JK, Dhiman RK, Kohli KK. Genetic polymorphism of the hepatic cytochrome P450 2C19 in north Indian subjects. *Clin Pharmacol Ther* 1998; **63**: 422–427.
167 Rosemary J, Adithan C, Padmaja N, Shashindran CH, Gerard N, Krishnamoorthy R. The effect of the CYP2C19 genotype on the hydroxylation index of omeprazole in South Indians. *Eur J Clin Pharmacol* 2005; **61**: 19–23.
168 Weerasuriya K, Jayakody RL, Smith CA, Wolf CR, Tucker GT, Lennard MS. Debrisoquine and mephenytoin oxidation in Sinhalese: a population study. *Br J Clin Pharmacol* 1994; **38**: 466–470.
169 Yang YS, Wong LP, Lee TC, Mustafa AM, Mohamed Z, Lang CC. Genetic polymorphism of cytochrome P450 2C19 in healthy Malaysian subjects. *Br J Clin Pharmacol* 2004; **58**: 332–335.
170 Horai Y, Nakano M, Ishizaki T, Ishikawa K, Zhou HH, Zhou BI *et al.* Metoprolol and mephenytoin oxidation polymorphisms in Far Eastern Oriental subjects: Japanese versus mainland Chinese. *Clin Pharmacol Ther* 1989; **46**: 198–207.
171 Roh HK, Dahl ML, Johansson I, Ingelman-Sundberg M, Cha YN, Bertilsson L. Debrisoquine and S-mephenytoin hydroxylation phenotypes and genotypes in a Korean population. *Pharmacogenetics* 1996; **6**: 441–447.
172 Setiabudy R, Kusaka M, Chiba K, Darmansjah I, Ishizaki T. Dapsone N-acetylation, metoprolol alpha-hydroxylation, and S-mephenytoin 4-hydroxylation polymorphisms in an Indonesian population: a cocktail and extended phenotyping assessment trial. *Clin Pharmacol Ther* 1994; **56**: 142–153.
173 Sohn DR, Kusaka M, Ishizaki T, Shin SG, Jang IJ, Shin JG *et al.* Incidence of S-mephenytoin hydroxylation deficiency in a Korean population and the interphenotypic differences in diazepam pharmacokinetics. *Clin Pharmacol Ther* 1992; **52**: 160–169.
174 Xie HG, Huang SL, Xu ZH, Xiao ZS, He N, Zhou HH. Evidence for the effect of gender on activity of (S)-mephenytoin 4′-hydroxylase (CYP2C19) in a Chinese population. *Pharmacogenetics* 1997; **7**: 115–119.

DE GRUYTER Drug Metab Drug Interact 2014; aop

Review

Carolina Céspedes-Garro*, María-Eugenia G. Naranjo, Ronald Ramírez, Víctor Serrano, Humberto Fariñas, Ramiro Barrantes, Adrián LLerena and CEIBA Consortium of the Ibero-American Network of Pharmacogenetics and Pharmacogenomics RIBEF

Pharmacogenetics in Central American healthy volunteers: interethnic variability

Abstract: Ethnicity is one of the major factors involved in interindividual variability to drug response. This study aims to describe the frequency of the most relevant pharmacogenetic biomarkers and metabolic phenotypes in Central American healthy volunteers and to determine its interethnic variability. Twenty-six original research articles on allelic, genotypes or metabolic phenotype frequencies were analyzed, in which a total number of 7611 Central American healthy volunteers were included (6118 were analyzed for genotype and 1799 for metabolic phenotype). No reports were available for population from Belize and Honduras. The *CYP2D6*4* and **5* frequencies in Amerindian populations from Costa Rica have shown to be among the highest frequencies so far reported in the world. Furthermore, *NAT2*5* and **6* presented higher frequencies in admixed populations than in Amerindians, but, inversely, the *NAT2*7* was more frequent in Amerindians compared to an admixed population. Likewise, different patterns of distribution have been shown in *HLA-A*02*, **03* and *HLA-B*07* among Native populations from Latin America. Reports on Central American populations were also found for the CYP2C19, *LDLR*, *CYP2E1*, *MDR1*, *G6PD*, *TP53*, CYP1A2, *CYP3A4* and *CYP3A5* biomarkers, but no data were available for the other 91 pharmacogenetic biomarkers revised in Central American populations. Differences in the frequency of some pharmacogenetic biomarkers and metabolic phenotypes were found, showing interethnic variability within Central American and with other Latin American populations.

Keywords: admixed; Afro-Caribbean; Amerindian; biomarkers; Central America; CYP1A2; CYP2C19; CYP2D6; *CYP2E1*; *G6PD*; *HLA*; *LDLR*; *MDR1*; NAT2; pharmacogenetics; *TP53*.

DOI 10.1515/dmdi-2014-0025
Received July 9, 2014; accepted November 5, 2014

***Corresponding author: Carolina Céspedes-Garro,** CICAB, Clinical Research Centre, Extremadura University Hospital and Medical School, Badajoz 06080, Spain, Phone: +34 924218040, Fax: +34 924219881, E-mail: ccg004@gmail.com
Carolina Céspedes-Garro and Ramiro Barrantes: Genetics Section, School of Biology, University of Costa Rica, San José, Costa Rica
María-Eugenia G. Naranjo, Humberto Fariñas and Adrián LLerena: CICAB Clinical Research Centre, Extremadura University Hospital and Medical School, Badajoz, Spain
Ronald Ramírez: Facultad de Medicina, UNAN Universidad Autónoma Nacional de Nicaragua, León, Nicaragua
Víctor Serrano: CIIMET Centro de Investigación e Información de Medicamentos y Tóxicos, Facultad de Medicina, Universidad de Panamá, Panamá, Panamá

Introduction

Central America is inhabited by more than 44.5 million people and is one of the richest bio-cultural regions of the world having a complex ancestry background, involving different degrees of ethnic admixture including diverse Amerindian cultural influences. Two Native groups are distinguished in the region: (a) the Mesoamerican tribes located in the northern part of the isthmus (Mexico, Guatemala, Belize, El Salvador and Honduras), and (b) tribes of the lower Central America (Costa Rica and Panama) [1]. Additionally, during the 16th century, the arrival of European population mainly from the Iberian Peninsula and African people traded as slaves boosted the genetic diversity of the Central American (CA) region [2].

The interindividual variability in the response to pharmacological treatment has been attributed to factors such as environment, epigenetics, genetics, physiology and pathology. Additionally, ethnicity has been associated also with variability on drug response among individuals, due to the varying prevalence of genetic polymorphisms on drug-metabolizing enzymes (DMEs) among ethnic groups [3].

Most of the pharmacogenetic studies involved the phase I and II enzymes, transporters and receptors.

CYP2D6, *CYP2C19*, *NAT2*, *HLA* and *LDLR* are some of the most relevant pharmacogenetic biomarkers, which are further characterized below.

CYP2D6 plays an important role in the metabolism of about 25% of clinically important drugs, including antidepressants, antipsychotics, antiarrythmics, antihistamines, etc. About 100 allelic variants have been described for the *CYP2D6* gene [4], some of which are associated with null activity (e.g., *CYP2D6*3, *4, *5* and **6*), decreased activity (e.g., *CYP2D6*10, *17* and **29*), normal (e.g., *CYP2D6*1*, **2* and **35*) or increased activity (e.g., *CYP2D6*1xN*, **2xN* and **35xN*). A high frequency of *CYP2D6*17* and **29* is found in African populations, while *CYP2D6*10* is more frequent in Asians than in any other populations [5]. Recently, a new nonfunctional variant (*CYP2D6*31*) associated with poor metabolism has been described in Hispanic populations [6].

The presence of two non-active alleles in an individual classifies a subject into the predicted poor metabolizer (PM) group. Individuals with one or two active alleles are classified in the predicted extensive metabolizers (EMs) group; among them, subjects with more than two active alleles are classified as predicted ultra-rapid metabolizers (UMs). These metabolic groups are associated with drug efficacy or adverse drug reactions during pharmacological therapy [7].

Cytochrome CYP2C19 is responsible for the metabolism of drugs such as antiepileptics, antiretroviral, antidepressants and proton-pump inhibitors. To date, 34 alleles for *CYP2C19* have been described. Certain polymorphisms have been associated with lack of enzyme activity, such as *CYP2C19*2* through **8*, but only *CYP2C19*17* has been associated with increased hydroxylation activity [8]. *CYP2C19*2* and **3* frequencies are higher among Asian populations than in European and African populations [9].

Depending on the *CYP2C19* allelic combinations, individuals can be classified into three groups: PMs (carriers of two inactive variants), EMs (with one or two active alleles) and UMs (carriers of the *CYP2C19*17*) [10]. These predicted phenotypes have shown clinical implications during pharmacological treatment [11].

Regarding phase II enzymes, arylamine *N*-acetyltransferase 2 (NAT2) has been one of the first enzymes shown to be variable [12]. This enzyme is responsible for the metabolism of drugs such as isoniazid and dapsone. Thirty-six allele variants have been described, some of which are related to slow metabolism (*NAT2*5, *6* and **7*), normal and increased metabolism (*NAT2*11*, **12* and **13*). In Europeans and Africans, the frequencies of *NAT2*5* are high and those of allele *NAT2*7* are low, whereas the opposite is observed in Asian populations. Slow acetylators (SAs) are determined by the homozygosis of alleles with diminished activity; furthermore, these subjects have increased risk of adverse effects during drug therapy [13].

Human leukocyte antigens (HLAs) are the proteins responsible for presenting processed peptides to T cells. Nowadays, 2884 alleles have been described for *HLA-A*, some of which are related to a null expression (e.g., *HLA-A*01:01:01:02N, A*01:04N* and *A*01:11N*) and others are related to an alternative expression (e.g., *A*01:01:38L, A*01:147Q* and *A*02:01:01:02L*). *HLA-B* has been reported to have approximately 3590 alleles, 13 of which having an alternative expression [14]. Furthermore, the *HLA* system is considered to be the most polymorphic region in the human genome. *HLA* allelic variants have been related to cutaneous adverse reactions in the treatment with antiepileptic drugs [15].

Low-density lipoprotein receptor (LDLR) binds to the LDL, the major cholesterol-carrying lipoprotein of plasma, and transports it into cells by endocytosis. More than 1000 allele variants have been reported for the *LDLR* gene [16], and some of these polymorphisms have been found in specific ethnic groups [17]. Variants causing altered activity of the receptor have been related to hypercholesterolemia in pharmacological therapy with statins [18].

Despite the growing number of pharmacogenetic studies describing biomarkers involving populations of the world, there is a gap in the information regarding Hispanics in general and CA populations in particular. Therefore, this review intends to describe the frequency of the most relevant pharmacogenetic biomarkers and metabolic phenotypes in healthy volunteers from CA populations and to determine its interethnic variability.

Materials and methods

A literature search was conducted in June 2014 using PubMed, Scopus and Scielo databases.

The search terms were “country name” and “biomarker”, looked up one by one. The countries included in the search were Guatemala, Belize, Honduras, El Salvador, Nicaragua, Costa Rica and Panama (Figure 1). A panel of clinical relevant pharmacogenetic biomarkers was selected using different resources: (1) Table of Pharmacogenomic Biomarkers in Drug Labeling (from the US Food and Drug Administration); (2) European Public Assessment Reports (from the European Medicines Agency); and, additionally, a pharmacogenomics knowledge database (3) Table of CPIC Pairs (from the Pharmacogenomics Knowledgebase) (see Table 1). Moreover, a search was carried out to include metabolic phenotype studies using

Figure 1 Number of pharmacogenetic studies and healthy volunteers reported for each country from the Central American region. Values in parenthesis refer to the number of studies and number of healthy volunteers recruited. #Both studies reported allele frequencies for the same populations from El Salvador and Nicaragua.

Table 1 Panel of pharmacogenetic biomarkers based on the clinical implication according to the drug regulatory agencies from Europe and America (European Medicines Agency and Food and Drug Administration, respectively), and the Pharmacogenomics Knowledgebase CPIC gene/drug pairs [19–21].

ABCC4	*COQ2*	*ERBB2*	*HPRT1*	*SCN1A*
ABCG2	*CRHR1*	*ERCC1*	*HTR2A*	*SLC22A1*
ADD1	*CRHR2*	*ESR1/PGR*	*HTR2C*	*SLC22A2*
ADORA2A	*CYP1A1*	*F5*	*IFNL3*	*SLC22A6*
ADRB1	*CYP1A2*	*FCGR3A*	*IL2RA*	*SLC6A4*
ADRB2	*CYP2A7P1*	*FDPS*	*ITPA*	*SLCO1B1*
ALK	*CYP2B6*	*FIP1L1/PDGFRA*	*KIT*	*SLCO2B1*
ANKK1	*CYP2C19*	*FLOT1*	*KRAS*	*SOD2*
ATIC	*CYP2C8*	*G6PD*	*LDLR*	*SULT1A1*
BCR-ABL	*CYP2C9*	*GGCX*	*LTC4S*	*TMEM43*
BRAF	*CYP2D6*	*GNB3*	*MDR1/ABCB1*	*TNFRSF8*
BTG3	*CYP2E1*	*GRIK4*	*MRP1/ABCC1*	*TP53*
C11orf65	*CYP3A4*	*GSTA1*	*MRP2/ABCC2*	*TPMT*
CACNB2	*CYP3A5*	*GSTM1*	*MS4A1*	*UGT1A1*
CALU	*CYP4F2*	*GSTP1*	*MTHFR*	*UGT1A4*
CBR3	*del (5q)*	*GSTT1*	*MTRR*	*UGT2B15*
CCR5	*DPYD*	*HER1*	*NAT2*	*UMPS*
CES1	*DRD1*	*HERG1*	*OATP1B1*	*VKORC1*
CFTR	*DRD2*	*HLA-A*	*OPRM1*	*XRCC1*
COL22A1	*EGFR*	*HLA-B*	*PDGFRA*	*YEATS4*
COMT	*EPHX1*	*HMGCR*	Philadelphia chromosome	

the terms “country name” and “probe drugs”. The probe drugs searched were debrisoquine, sparteine, metoprolol, dextromethorphan, mephenytoin, omeprazole, tolbutamide, losartan, diclofenac, isoniazid and caffeine.

The inclusion criteria were as follows (1) studies reporting genotype, allele and/or metabolic phenotype frequencies for any of the biomarkers revised; (2) healthy volunteers (control groups from case-control studies were not taken into consideration); and (3) subjects living in the CA region (Figure 1).

Additionally, unpublished data on *CYP2D6* allelic and genotype frequencies for six Amerindian groups from Costa Rica [22] were included.

Predicted phenotype was determined for the *CYP2D6* and *NAT2* genes. Individual carriers of two inactive alleles for *CYP2D6* were classified as PMs, and those with more than two active *CYP2D6* genes were classified as predicted UMs, following previously published methodology [23, 24]. And, finally, those subjects with two *NAT2* alleles related to a decreased *in vivo* activity were classified as SAs [25].

To study the biomarker’s variability within populations from the CA region, the allele, metabolic phenotype and/or predicted phenotype frequencies were compared to those in other CA populations. The differences in allele, predicted phenotype or metabolic phenotype frequencies were compared using the χ^2 test followed by the Marascuilo procedure. p-Values <0.05 were regarded as statistically significant.

Results

A total of 7611 healthy volunteers from Central America were included in 26 studies published between 1986 and 2014. Healthy volunteers from five of seven countries in the CA region were studied. No reports were available for population from Belize and Honduras.

Populations involved in this review were (a) Amerindian populations from Guatemala (Maya), Panama (Kuna, Embera, Ngawbe Guaymi and Teribe) and Costa Rica (Bribri, Cabecar, Chorotega, Guatuso, Guaymi, Huetar and Teribe); (b) admixed populations from Nicaragua, Costa Rica, El Salvador and Panama; and (c) an Afro-Caribbean population from Costa Rica. Demographic data were not specified for most of these populations.

Over 77% of the subjects were analyzed for genotype (n=6118), and only 23% of the subjects were recruited for metabolic phenotype analysis (n=1799), most of them from Panama. With the available data, analysis of the variability was possible for CYP2D6, CYP2C19, NAT2, *HLA*, *LDLR*, *CYP2E1*, *MDR1*, *G6PD*, *TP53*, CYP1A2, *CYP3A4* and *CYP3A5*.

CYP2D6

CYP2D6 polymorphism was studied in volunteers from Nicaragua, Costa Rica and Panama (Tables 2 and 3), including populations with different ethnic backgrounds such as admixed, Amerindian and African ancestries (see Tables 2 and 3).

CYP2D6 metabolic phenotype, genotype and/or allelic frequencies were reported in nine studies. Allele frequencies by country and populations from Central America are shown in Table 2.

The frequency of the inactive allele *CYP2D6*4* remained similar in most of the admixed and Amerindian populations assessed in the region (10.4%–19.2%), except for the Cabecar and Bribri Amerindian populations, which had the highest frequencies (26.8% and 31.9%, respectively).

The frequency of *CYP2D6*10* was different among the Amerindian groups of the CA region. The highest frequency of *CYP2D6*10* was found in the Ngawbe Guaymi tribe (17.5%), more than in any other Amerindian populations studied ($p<0.05$), except for the Embera tribe. The admixed and Afro-Caribbean populations studied showed low frequencies that ranged between 1.1% and 3.1% (Table 2).

The frequency of *CYP2D6*17* was higher among the African descendant population from Costa Rica (18.4%, $p<0.05$) than in any other admixed population of Central America and some Amerindians (Table 2).

The presence of genotypes associated with null activity enzyme was higher in two Costa Rica Amerindian groups (Table 2), due to the presence of *CYP2D6*4* and *CYP2D6*5*. Furthermore, high percentages of predicted UMs were found in the Costa Rican Guatuso (18.8%) and admixed (10.1%) populations (Table 2).

The assessment of CYP2D6 metabolic phenotype was carried out in different Amerindian populations from Panama, and the PM frequency ranged from 0.04% in the Kuna group to 5.9% in the Ngawbe Guaymi group. The prevalence of PMs in the admixed population from Nicaragua was 6%, with no UMs reported (Table 3).

CYP2C19

The frequency of CYP2C19 PM metabolic phenotype varied from 0% in the Kuna Amerindians [32, 33] to 15.8% and 18.6% in the Teribe and Ngawbe populations from Panama [32], respectively. No data were available for the *CYP2C19* genotypes.

Table 2 *CYP2D6* allele and predicted phenotype frequencies in populations from the Central American region with different ethnic backgrounds.

Country	Ancestry	n	*1+*2	*3	*4	*4xN	*5	*6	*9	*10	*17	*29	*35	*41	Active alleles xN	% predicted PMs	% predicted UMs	References
Nicaragua	Admixed	98	76.1	0	14.2	0	4.6	0	–	3.1	0	–	–	–	2.0	4.1	3.0	[24]
Panama-Colombia	Embera[a]	136	–	0	14.0	–	0	1.1	0	6.9[b]	–	–	–	–	–	–	–	[26]
Panama-Colombia	Ngawbe[a]	105	–	0	17.1	–	0	0.5	0	17.5[c]	–	–	–	–	–	–	–	[26]
Costa Rica	Admixed	139	66.2	1.4	10.4	1.4	3.2	0.7	–	1.1	2.2	1.4	0.4	6.1	5.4	1.4	10.1	[22]
Costa Rica	Afro-Caribbean	49	43.8	1.0	4.1[d]	2.0	4.1	0	–	3.1	18.4[e]	11.2	1.0	1.0	7.2	2.0	8.2	[22]
Costa Rica	Chorotega[a]	32	71.9	0	17.2	0	0	0	–	0	4.7	3.1	1.6	0	1.6	6.3	3.1	[22]
Costa Rica	Guatuso[a]	16	68.8	0	15.6	0	0	0	–	0	6.3	0	0	0	9.4	6.3	18.8	[22]
Costa Rica	Cabecar[a]	28	58.9	0	26.8	0	14.3	0	–	0	0	0	0	0	0	10.7	0	[22]
Costa Rica	Bribri[a]	47	52.2	0	31.9	0	11.7	0	–	0	3.2	0	0	1.1	0	21.3	0	[22]
Costa Rica	Guaymi[a]	26	78.9	0	19.2	0	0	0	–	0	0	0	1.9	0	0	7.7	0	[22]
Costa Rica	Huetar[a]	48	68.8	0	18.8	0	0	0	–	1.0	0	0	1.0	7.3	3.1	4.2	6.3	[22]

[a]Amerindian population. [b]$p<0.05$ compared with the other Amerindian populations from Costa Rica studied. [c]$p<0.05$ compared with the other populations from Nicaragua and Costa Rica studied. [d]$p<0.05$ compared with the Bribri from Costa Rica studied. [e]$p<0.05$ compared with the admixed from Nicaragua and Costa Rica and the Cabecar, Guaymi and Huetar populations studied.

Table 3 CYP2D6 metabolic phenotype frequencies for populations from Central America.

Country	Population	n	Probe drug	PMs, %	UMs, %	References
Nicaragua	Admixed	133	DB	6.0	0	[24]
Panama	Kuna[a]	51	SP	0	–	[27]
Panama	Kuna[a]	170	SP	0	–	[28]
Panama	Ngawbe Guaymi[a]	121	SP	5.2	–	[29]
Panama	Ngawbe Guaymi[a]	84	SP, DB	5.9	–	[30]
Panama	Kuna[a]	89	DB	0	–	[31]
Panama	Kuna[a]	89	DB	0	–	[32]
Panama	Kuna[a]	250	SP	0.04	–	[32]
Panama	Ngawbe Guaymi[a]	170	DB	4.7	–	[32]
Panama	Ngawbe Guaymi[a]	284	SP	4.9	–	[32]
Panama	Teribe[a]	19	DB, SP	5.3	–	[32]
Panama	Ngawbe Guaymi[a]	344	SP	4.4	–	[26]
Panama	Embera[a]	153	SP	2.2	–	[26]

[a]Amerindian populations. DB, debrisoquine; SP, sparteine.

NAT2

Three reports with *NAT2* allele frequencies were included (Table 4). *NAT2*5D* was the most frequent allele in the admixed population from Nicaragua (35.8%), whereas the frequency of this allele in two Amerindian populations from Panama ranged from 2.4% to 9.9%. The frequency of *NAT2*6* was higher in the admixed population from Nicaragua than in the Amerindian populations from Panama ($p<0.05$). The *NAT2*7* allele was not detected in the Nicaraguan admixed population, and its frequency varied from 20.8% to 23.9% in Amerindians ($p<0.05$).

Predicted phenotype frequencies are given in Table 4, and data were available for two Amerindian populations from Panama and for an admixed population from Nicaragua. The frequency of SAs was higher in the admixed population from Nicaragua than in any other Amerindian population from Panama (Ngawbe and Embera) ($p<0.05$).

The NAT2 metabolic phenotype frequencies are given in Table 4. The metabolic phenotype determination was assessed only in Amerindian populations from Panama, specifically in the Kuna, Teribe and Embera tribes, using mostly isoniazid and, in just one case, caffeine as probe drugs. The percentage of SAs varied from 18.1% to 29% (Table 4).

HLA

The *HLA* allelic frequencies for populations from Guatemala, Costa Rica and Panama are shown in Table 5.

Table 4 Allele, metabolic phenotype and predicted phenotype frequencies for the biomarker NAT2 in Central American populations.

Country	Population	Allele frequency						Metabolic phenotype and predicted phenotype			References
		n	*5D (341T>C)	*6B (590G>A)	*7A (857G>A)	*11A (481C>T)	*12A (803A>G)	*13 (282C>T)	n	Probe drug	SAs, %	
Nicaragua	Admixed	137	35.8	17.5[a]	0[a]	–	–	–	137	X	49.6[a]	[34]
Panama-Colombia	Ngawbe[b]	71	–	0	23.9	2.1	–	–	71	X	7.0	[35]
Panama-Colombia	Embera[b]	101	–	3.5	20.8	10.4	–	–	101	X	12.0	[35]
Panama	Embera[b]	136	9.9	3.7	22.8	9.2	9.9	29.0	72	CF	18.1	[36]
Panama	Ngawbe[b]	105	2.4	0	23.3	2.4	1.9	25.2	105	X	7.7	[36]
Panama	Kuna[b]	–	–	–	–	–	–	–	45	IZ	24.4	[32, 37]
Panama	Teribe[b]	–	–	–	–	–	–	–	62	IZ	29.0	[32, 37]
Panama	Embera[b]	–	–	–	–	–	–	–	136	X	14.7	[36]

[a]$p<0.05$ compared with the other populations studied using the χ^2 test followed by the Marascuilo procedure. [b]Amerindian populations. Probe drugs: CF, caffeine; IZ, isoniazid; X, studies reporting predicted phenotype; Sas, slow acetylators for NAT2.

Table 5 *HLA-A*, *HLA-B* and *LDLR* allele frequencies of healthy volunteers from Central America.

Country	Population	n								Alleles	References
HLA-A			****1***	****02***	****3***	****11***	****23***	****24***	****29***	****30***	
Guatemala	Amerindian	132	1.5	36.3	1.1	1.5	–	21.2	2.3	–	[38]
Costa Rica	Admixed	130	8.1	21.2	8.9	2.7	1.2	15.4	5.4	5.0	[39]
Panama	Admixed	965	–	17.6	7.7	–	10.7	15.8	–	13.4	[40]
HLA-B			****07***	****1522***	****35***		****3512***		****4002***	****44***	
Guatemala	Amerindian	132	1.5	4.2	40.1		1.9		1.1	3.0	[38]
Costa Rica	Admixed	130	11.5	7.7	16.9		–		10.0	8.5	[39]
Panama	Admixed	965	7.7	–	19.5		–		–	9.0	[40]
Panama	Amerindian	8	–	18.8	–		6.3		6.3	–	[41]
LDLR				***A***				***B***			
Nicaragua	Admixed	100		54.0				46.0			[42]
Costa Rica	Admixed	1204		53.1				46.9			[43]
Costa Rica	Afro-Caribbean	72		36.1				63.9			[43]
Costa Rica	Bribri-Cabecar	100		72.9				27.1			[43]

[a]$p<0.05$ compared with the other populations studied using the χ^2 test followed by the Marascuilo procedure.

The frequency of *HLA-A*02* was lower in the admixed populations than in an Amerindian population from Guatemala (36.3%; $p<0.05$). In contrast, the frequency of *HLA-A*03* was higher in the admixed populations than in the Mayan Amerindian population from Guatemala (1.1%; $p<0.05$).

A variability on the frequencies of *HLA-B*07* and **35* was observed between the admixed populations from Costa Rica (11.5% and 16.9%, respectively) and Panama (7.7% and 19.5%, respectively) in comparison with the Amerindian population from Guatemala (1.5% and 40.1%, respectively; $p<0.05$).

LDLR

The *LDLR* allele frequencies for populations with different ethnic background from Nicaragua and Costa Rica are given in Table 5. The lowest frequency of *LDLR-A* was found in the Afro-Caribbean ethnic group (36.1%; $p<0.05$), while the Amerindian population (72.9%; $p<0.05$) showed the highest frequency among the CA populations studied.

Besides the aforementioned information, reports were found for the *CYP2E1*, *MDR1*, *G6PD*, *TP53*, *CYP3A4* and *CYP3A5* genotypes. Allelic frequencies are reported in Table 6. Comparison with other CA populations was not possible due to the lack of scientific reports.

The CYP1A2 metabolic phenotype was determined using caffeine as a probe drug for the Costa Rican admixed population, obtaining a unimodal distribution [48].

Discussion

This is the first review that addresses the available population pharmacogenetic data in CA populations, including a unique sample of diverse Amerindian populations.

Variability on the pharmacogenetic biomarkers and DME phenotypes is described below for CYP2D6, CYP2C19, NAT2, *HLA*, *LDLR*, *CYP2E1*, *MDR1*, *G6PD*, *TP53*, CYP1A2, *CYP3A4* and *CYP3A5*.

CYP2D6

Data regarding CYP2D6 are abundant in populations of the region.

Among Central Americans, the Bribri and Cabecar Amerindian populations from Costa Rica have the highest *CYP2D6*4* frequencies (31.9% and 26.8%, respectively). Interestingly, the Amerindian Bari and Chorote populations from South America show the highest worldwide *CYP2D6*4* frequencies (42.5% and 37.5%, respectively) [49, 50]. The frequencies in these four populations are higher than in any Caucasian population, with the exception of the Faroese people (33.4%) [51], despite the fact that *CYP2D6*4* has been considered to be a predominantly Caucasian allele.

Furthermore, and contrary to other Native and admixed populations from Latin America having low *CYP2D6*5* frequencies, the Amerindian Cabecar and Bribri populations from Costa Rica have the highest frequencies for *CYP2D6*5* hitherto detected in the Amerindian

Table 6 Allele frequencies of some pharmacogenetic biomarkers studied in healthy volunteers from Central America.

Country	Population	n	Alleles					References
CYP2E1					****5***			
Nicaragua	Admixed	137			16.4			[34]
MDR1					**C3435**			
El Salvador	Admixed	112			51.0			[44]
Nicaragua	Admixed	117			54.0			[44]
G6PD			**A**		**B**			
Costa Rica	Bribri (Atlantic)	123	0.9		99.1			[45]
	Cabecar (Atlantic)	99	–		100			[45]
	Bribri (Pacific)	99	–		100			[45]
	Cabecar (Pacific)	60	–		100			[45]
	Guatuso	80	–		100			[45]
	Teribe	63	–		100			[45]
	Guaymi	770	–		100			[45]
TP53			**141**	**136**	**131**	**126**	**121**	
Costa Rica	Chibcha	576	0.3	38.7	8.9	50.6	1.1	[46]
	Chorotega (Matambú)	75	–	48.7	12.7	35.3	3.3	[46]
				CYP3A4			***CYP3A5***	
Other biomarkers			****1A***	****1B***		****1***	****3***	
El Salvador	Admixed	112	87.5	12.5		23.7	76.3	[47]
Nicaragua	Admixed	120						

populations [52]. These native populations are the third and fourth highest frequencies reported worldwide so far, only surpassed by two African populations (Bantu 18.8% and South Africa 17.7%) [53, 54]. Previous studies analyzing ABO, adenosin deaminase and *G6PD* loci inferred an admixture with non-Amerindians in the Bribri and Cabecar populations [45]. However, a later study analyzing 48 loci of enzymatic systems, blood groups and serum proteins revealed a low admixture with European and African populations in these Amerindian populations (<0.01%) [55]. Further ancestry-based studies are required to clarify this point.

*CYP2D6*10* is related to Asian ancestry [5], which might explain the frequency of this allele found in the Ngawbe Amerindian population from Panama (17.5%), the highest *CYP2D6*10* frequency reported in Amerindians so far [26, 52]. However, further ancestral analysis of Native populations is needed, taking into account that low *CYP2D6*10* frequencies are commonly found within Amerindian populations from North and South America, mostly ranging from 0 to 3.3% [22, 49, 56–61] and rising up to 7.1% in some specific cases [50].

In agreement with the hypothesis of the African ancestry [62, 63], the Costa Rican Afro-Caribbean population showed the highest *CYP2D6*17* frequency in Central America (18.4%) [22]. Similar frequencies of this allele in America were only found in Afro-American populations from the USA (19.1% and 20.9%) [64, 65].

The frequency of predicted PMs and UMs varied in the CA populations, regardless of the ethnicity, revealing a wide diversity despite having such a reduced territory. Furthermore, an unclear pattern for the predicted phenotypes was found in the populations from the CA region.

It is remarkable that the predicted PM frequency for the Bribri population of Costa Rica, which was the highest in the region and the second highest reported so far for the American continent and worldwide, was surpassed only by the Bari Amerindian population from Venezuela (25%) [49, 52]. Furthermore, the frequency of the predicted UMs of the Guatuso Amerindian population from Costa Rica is ranked third in the Americas, the first two places being found in Amerindian populations from Mexico (Huicholes 20.6% and Guarijíos 20%) [61]. In the global ranking, the CA population shares the fourth place with an Arabian Saudi population (18.8%) [66], and the first place is led by an Ethiopian population (29%) [52, 67].

Interestingly, most of the CYP2D6 metabolic phenotype frequencies worldwide reported for Amerindians were from the CA region, with the only exception of the Mexican Tepehuano population [56]. The lowest PM frequencies in the region are reported for the Kuna Amerindian population, which was characterized by the authors to have a low racial admixture (<1%) [28, 55]. Furthermore, the Tepehuano population from Mexico also reported the lack of this CYP2D6 metabolic phenotype [56]. The other Panamanian Amerindian populations (Ngawbe Guaymi,

Embera and Teribe) reported higher PM frequencies, which might be related to the genetic divergence from the Kuna population [28], as these populations are also known as low admixed tribes (<2%) [26, 55].

More CYP2D6 metabolic phenotype studies in admixed populations are available. However, the only one available from Central America is for a Nicaraguan admixed population (PMs: 6%), which presented similar PM frequencies to those of Mexican populations (PMs reported: 6%–6.8%) [56, 68].

CYP2C19

No genetic studies on *CYP2C19* in populations from Central America are available, although this gene is the second most investigated, being associated with different therapeutic areas [69]. Further studies on Central Americans are needed in order to identify the genetic diversity of these populations and their possible relevance during pharmacological treatment.

Regarding the CYP2C19 metabolic phenotype, the only available report for Amerindians in Latin America is for three Panamanian tribes. Like those reported for CYP2D6, CYP2C19 PMs are also absent among individuals belonging to the Kuna tribe [32, 33]. And the PM frequencies for the Teribe and Ngawbe Amerindian groups (15.8% and 18.6%, respectively) [32] from Panama are higher than those for other Hispanic populations from America [70, 71] but similar to those of Asian populations (in whom the PMs vary from 12% to 23%) [72], which is in agreement with the Asian ancestry of these populations [52, 73].

NAT2

NAT2 allelic variants, *NAT2*5*, of abolished activity, and *NAT2*6* and **7*, both of decreased activity, were mainly reported for CA populations. NAT2 frequencies varied according to the ethnic background of the population. As expected, the prevalence of *NAT2*5* is higher in the Nicaraguan admixed population (35.8%), due to the European influence [9, 74], than in the Amerindian populations of Ngawbe and Embera (2.4% and 9.9%). The finding of the low *NAT2*5* frequency within CA Native populations is controversial due to the high *NAT2*5* frequencies present in other Amerindian populations [74], except in the San Martín tribe from Peru and the Cayapa tribe from Ecuador (9.1% and 13.3%, respectively) [74].

Likewise, the decreased *NAT2*6* frequencies of the Amerindian populations from Panama are in agreement with the low frequencies reported for other populations from North and South America with the same ethnicity (Pima, Piapoco and Curripaco, San Martín, Karitana, Surui, Lengua, Ayoreo and Wichi) [50, 74]. Furthermore, the high *NAT2*6* frequency in admixed Nicaraguans is in agreement with the hypothesis for this allele as being an indicator of recent admixture from Europe or Africa [74].

In contrast, the *NAT2*7* frequencies of the Panamanian Amerindian populations (Ngawbe and Embera) are among the highest reported in America [74], which might be explained by the proposed Asian ancestry [25, 75]. Moreover, in admixed populations, lower *NAT2*7* frequencies are commonly found like those from Nicaragua (0%) and from previously reported populations [74–77].

As expected, Spanish descendant populations such as the Nicaraguans reported high frequencies of predicted SAs (49.6%) [25, 75].

A pattern in the distribution of the frequency of predicted SAs in Amerindians was observed, showing that groups from North and South America [74] presented higher values (with an average frequency of predicted SAs of 25%), whereas the value in Native populations from the lower CA region seemed to decrease.

Studies reporting the NAT2 metabolic phenotype in Latin American populations are scant. The available data from Central America, specifically in Amerindian tribes from Panama, are the only report for populations with this ethnic background. A slight variation within these populations was observed, showing higher NAT2 SA frequencies in tribes from the northern part (Teribe 29% and Kuna 24.4%) than in a tribe from the southern part (Embera 19%) of the country.

HLA

The *HLA-A*02* frequency of the Maya tribe from Guatemala (36.3%) seemed to be similar to that of other Amerindian populations from Mexico (Teenek 39.8% and Mixe 31.1%) [78, 79] and Colombia (Wayu 35.8%) [80]. Nonetheless, in Native populations from South America, the frequency of this allele rose to more than 50% in the Quechua (Bolivia), Guarani (Paraguay) and Lamas (Peru) populations (50.7%, 54% and 63.4%, respectively) [81–83]. In contrast, *HLA-A*02* frequency in admixed populations from the CA region and Mexico was slightly lower [84].

*HLA-A*03* frequencies seemed to vary independently from the location of the Amerindians; populations such as the Maya [38], Zapotec, Mixtec, Mixe [79], Mayos [85], Quechua [81] and Lamas [83] (1.1%, 0.7%, 0%, 0%, 1.7%, 2.2% and 1.2%, respectively) have lower frequencies

in comparison to Native Americans from the USA [86], Nahuas from Mexico [85] and Mapuche from Chile [87], in whom higher frequencies of the allele are predominant (6.7%, 8.2% and 7.7%).

The frequency of *HLA-B* alleles seemed to vary among the Amerindian and admixed populations from the CA region [38, 41]. *HLA-B*07* is present at low frequencies in the Mayan population from Guatemala (1.5%). However, as higher frequencies of the allele are found among Asian populations than in any other worldwide group [88], which is considered as an indicator of Asian ancestry, high frequencies were also expected within Amerindian populations from Central America. Previous reports agree with the findings that, in Amerindian tribes, the *HLA-B*07* frequency tends to be low [38, 78, 81, 85, 89].

At present, studies considering *HLA* allele frequencies in Amerindian populations [90] have not included lower CA groups. Admixed populations from the CA region present slightly higher *HLA-B*07* frequencies.

LDLR

The variation in the frequency of *LDLR-A* is documented for populations with different ethnic backgrounds from Nicaragua and Costa Rica. However, studies that further characterize other CA populations, including different SNPs for this gene, are needed.

Other biomarkers

Regarding the frequency of *CYP2E1*5*, the values for the Nicaraguan population (16.4%) were similar to those for other Latin American admixed populations from Mexico (14% and 16%) [91, 92] and Chile (16%) [93].

The frequency of *MDR1* 3534C in the CA populations from El Salvador and Nicaragua (51% and 54%, respectively) is similar to those in the Mexican (52.2%) [94] and Ecuadorian (52%) [44] populations and in other Amerindian (48.6%) [95] and white (55%, 56.8% and 57.1%) [95–97] populations from Brazil, whereas in most of the South American populations from Chile [98] and Brazil [95–97, 99], the frequency of this allele is higher (>61%).

Data on *G6PD* and *TP53* are only available for Amerindian populations from Costa Rica. Moreover, there are no reports involving healthy volunteers for other Hispanic populations. Further studies are needed in order to determine the variation within Central Americans, as well as with other Latin American populations.

The frequency of the CYP1A2 metabolic phenotype has been determined only in the Costa Rican population in Latin America. No deficiency of the enzyme was found in the population [48].

The frequency of *CYP3A4*1B* in the CA population (12.5%) [47] seemed to be different from those in the Mexican population [100] and Hispanics living in the US [101] (6% and 9%, respectively). In contrast, the frequency of *CYP3A5*3* in the population from El Salvador and Nicaragua (78% and 74%, respectively) was similar to those reported for Brazilian populations [102].

In order to address the limited data available in Latin American populations, the CEIBA consortium of RIBEF (Red Ibero-Latino Americana de Farmacogenética y Farmacogenómica; http://www.ribef.com) was created. RIBEF is a collaboration network that brings together more than 40 research groups with the aim to increase the pharmacogenetic knowledge in this multiethnic and multicultural region. The network has published several original research papers and reviews [3, 22, 24, 103–107]. Although further studies are required in the area, the actions of RIBEF have increased the pharmacogenetic knowledge on these populations.

In conclusion, differences in the allelic, genotype or metabolic phenotype frequency of some relevant pharmacogenetic biomarkers (CYP2D6, CYP2C19, NAT2, *HLA* and *LDLR*) were found, showing the interethnic variability within the CA populations, mainly among Natives. Amerindian populations have shown to be unique, with some of the highest frequencies worldwide for some alleles, such as *CYP2D6*4* and *CYP2D6*5*. No data are available for most of the biomarkers, so further studies are warranted.

Acknowledgments: This study was supported by the University of Costa Rica (PhD fellowship to CCG in Spain), Gobierno de Extremadura (AEXCID 13IA001 to SIFF) and Consejería de Empleo, Empresa e Innovación and Fondo Social Europeo (FSE) (grant PD10199 to MEGN), and was coordinated by the network RIBEF Red Iberoamericana de Farmacogenética y Farmacogenómica (http://www.ribef.com).

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved the submission.

Research funding: None declared.

Employment or leadership: None declared.

Honorarium: None declared.

Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

CEIBA Consortium of authors (*group coordinator):
- Group 1: Graciela E. Moya*, Verónica Ferreiro. Institutions: Pontificia Universidad Católica, Buenos Aires, Argentina; Argentina & Fundación GENOS, Buenos Aires, Argentina.
- Group 2: Eduardo Tarazona-Santos*, Fernanda Rodrigues-Soares. Institution: Universidade Federal de Minas Gerais, Belo Horizonte, Brasil.
- Group 3: Alba P. Sarmiento*, Angélica Borbón. Institution: (previous) Pontificia Universidad Javeriana, Bogotá, Colombia.
- Group 4: Ramiro Barrantes*, Gerardo Jiménez-Arce, Carolina Céspedes-Garro. Institution: Universidad de Costa Rica, San José, Costa Rica.
- Group 5: Idania Rodeiro[1]*, Mayra Álvarez[2], René Delgado[3], Diadelis Remirez[4], Bárbaro Pérez[2], Luis R. Calzadilla[5]*. Institutions: [1]CEBIMAR; [2]Facultad de Medicina Calixto García; [3]CIDEM; [4]CECMED; [5]Centro Comunitario de Salud Mental La Habana Vieja, La Habana, Cuba.
- Group 6: Enrique Terán[1]*, Santiago Terán[1], Francisco Hernández[2]. Institutions: [1]Universidad San Francisco de Quito, Quito, Ecuador; [2]Universidad Estatal de Guayaquil, Guayaquil, Ecuador.
- Group 7: Rocío Ortiz-López*, Augusto Rojas-Martínez, Lourdes Garza-Ocañas, Yadira X. Pérez-Páramo. Institution: Universidad Autónoma de Nuevo León, Monterrey, Nuevo León, México.
- Group 8: Marisol López-López*, Alberto Ortega-Vázquez, Nancy Monroy-Jaramillo, Helgi Jung-Cook, Ingrid Fricke-Galindo, Elisa Alonso-Vilatela, Teresa Corona-Vázquez. Institutions: Universidad Autónoma Metropolitana & Instituto Nacional de Neurología y Neurocirugía Manuel Velasco Suárez, México City, México.
- Group 9: Martha G. Sosa-Macías*, Carlos Galaviz-Hernández, Ismael Lares-Aseff, Blanca P. Lazalde-Ramos. Institution: Instituto Politécnico Nacional-CIIDIR, Durango, México.
- Group 10 Author: Ronald Ramírez-Roa*, Catalina Altamirano Tinoco. Institution: Universidad Nacional Autónoma de Nicaragua, León, Nicaragua.
- Group 11: Manuela Grazina*. Institution: University of Coimbra, Coimbra, Portugal.
- Group 12: Adrián LLerena*, Pedro Dorado, Eva M. Peñas-Lledó, Jesús Cobaleda, M. Eugenia G. Naranjo, Fernando de Andrés, Humberto Fariñas. Institution: University of Extremadura, Badajoz, Spain.
- Group 13: Francisco E. Estévez-Carrizo*, Nicolás González-Vacarezza. Institution: Universidad de Montevideo, Montevideo, Uruguay.

References

1. Barrantes R, Smouse PE, Mohrenweiser HW, Gershowitz H, Azofeifa J, Arias TD, et al. Microevolution in lower Central America: genetic characterization of the Chibcha-speaking groups of Costa Rica and Panama, and a consensus taxonomy based on genetic and linguistic affinity. Am J Hum Genet 1990;46:63–84.
2. Barrantes R. Evolución en el trópico: Los amerindios de Costa Rica y Panamá. San José, Costa Rica: Editorial de la Universidad de Costa Rica, 1993.
3. Sosa-Macías M, Llerena A. Cytochrome P450 genetic polymorphisms of Mexican indigenous populations. Drug Metabol Drug Interact 2013;28:193–208.
4. CYP2D6 allele nomenclature [Internet]. The Human Cytochrome P450 (CYP) Allele Nomenclature Database. 2014. Available at: http://www.cypalleles.ki.se/cyp2d6.htm (accessed 30 Apr 2013).
5. Bradford LD. CYP2D6 allele frequency in European Caucasians, Asians, Africans and their descendants. Pharmacogenomics 2002;3:229–43.
6. Gaedigk A, Isidoro-García M, Pearce RE, Sánchez S, García-Solaesa V, Lorenzo-Romo C, et al. Discovery of the nonfunctional CYP2D6 31 allele in Spanish, Puerto Rican, and US Hispanic populations. Eur J Clin Pharmacol 2010;66:859–64.
7. Dorado P, Berecz R, Peñas-Lledó EM, Llerena A. Antipsychotic drugs and QTc prolongation: the potential role of CYP2D6 genetic polymorphism. Expert Opin Drug Metab Toxicol 2007;3:9–19.
8. CYP2C19 allele nomenclature [Internet]. The Human Cytochrome P450 (CYP) Allele Nomenclature Database. 2014. Available at: http://www.cypalleles.ki.se/cyp2c19.htm (accessed 12 November 2014).
9. Umamaheswaran G, Kumar DK, Adithan C. Distribution of genetic polymorphisms of genes encoding drug metabolizing enzymes & drug transporters – a review with Indian perspective. Indian J Med Res 2014;139:27–65.
10. Peñas-Lledó E, Guillaume S, Naranjo M, Delgado A, Jaussent I, Blasco-Fontecilla H, et al. A combined high CYP2D6-CYP2C19 metabolic capacity is associated with the severity of suicide attempt as measured by objective circumstances. Pharmacogenomics J 2014; doi:10.1038/tpj.2014.42. [Epub ahead of print].
11. Sim SC, Kacevska M, Ingelman-Sundberg M. Pharmacogenomics of drug-metabolizing enzymes: a recent update on clinical implications and endogenous effects. Pharmacogenomics J 2013;13:1–11.
12. Crettol S, Petrovic N, Murray M. Pharmacogenetics of phase I and phase II drug metabolism. Curr Pharm Des 2010;16:204–19.
13. Spielberg SP. N-acetyltransferases: pharmacogenetics and clinical consequences of polymorphic drug metabolism. J Pharmacokinet Biopharm 1996;24:509–19.
14. Robinson J, Mistry K, McWilliam H, Lopez R, Parham P, Marsh SG. The IMGT/HLA database. Nucleic Acids Res 2011;39:D1171–6.
15. Kaniwa N, Saito Y. The risk of cutaneous adverse reactions among patients with the HLA-A* 31:01 allele who are given carbamazepine, oxcarbazepine or eslicarbazepine: a perspective review. Ther Adv Drug Saf 2013;4:246–53.
16. Leigh SE, Foster AH, Whittall RA, Hubbart CS, Humphries SE. Update and analysis of the University College London low density lipoprotein receptor familial hypercholesterolemia database. Ann Hum Genet 2008;72:485–98.

17. Fokkema IF, den Dunnen JT, Taschner PE. LOVD: easy creation of a locus-specific sequence variation database using an "LSDB-in-a-box" approach. Hum Mutat 2005;26:63–8.
18. Blakey JD, Hall IP. Current progress in pharmacogenetics. Br J Clin Pharmacol 2011;71:824–31.
19. Center for Drug Evaluation and Research. Genomics – Table of Pharmacogenomic Biomarkers in Drug Labeling [Internet]. Available at: http://www.fda.gov/drugs/scienceresearch/researchareas/pharmacogenetics/ucm083378.htm (accessed 21 Feb 2014).
20. Table of CPIC Gene/Drug Pairs [Internet]. The Pharmacogenomics Knowledgebase. Available at: http://www.pharmgkb.org/cpic/pairs (accessed 23 June 2014).
21. European Medicines Agency. European public assessment reports [Internet]. 2014. Available at: http://www.ema.europa.eu/ema/index.jsp?curl=/pages/medicines/landing/epar_search.jsp&mid=WC0b01ac058001d124 (accessed 10 June 2014).
22. Céspedes-Garro C, Jiménez-Arce G, Naranjo M, Barrantes R, Llerena A. Ethnic background and CYP2D6 genetic polymorphisms in Costa Ricans. Rev Biol Trop 2014;62:1659–71.
23. Gaedigk A, Simon SD, Pearce RE, Bradford LD, Kennedy MJ, Leeder JS. The CYP2D6 activity score: translating genotype information into a qualitative measure of phenotype. Clin Pharmacol Ther 2008;83:234–42.
24. Llerena A, Dorado P, Ramírez R, González I, Alvarez M, Peñas-Lledó EM, et al. CYP2D6 genotype and debrisoquine hydroxylation phenotype in Cubans and Nicaraguans. Pharmacogenomics J 2012;12:176–83.
25. Sabbagh A, Langaney A, Darlu P, Gérard N, Krishnamoorthy R, Poloni ES. Worldwide distribution of NAT2 diversity: implications for NAT2 evolutionary history. BMC Genet 2008;9:21.
26. Jorge LF, Eichelbaum M, Griese EU, Inaba T, Arias TD. Comparative evolutionary pharmacogenetics of CYP2D6 in Ngawbe and Embera Amerindians of Panama and Colombia: role of selection versus drift in world populations. Pharmacogenetics 1999;9:217–28.
27. Arias TD, Jorge L, Lee D, Guerrero de Solís N, Escobar J, Barrantes R, et al. Biotransformation of sparteine in the Cuna American Indians of Panama. Rev Méd Panamá 1986;11:199–209.
28. Arias TD, Jorge LF, Lee D, Barrantes R, Inaba T. The oxidative metabolism of sparteine in the Cuna Amerindians of Panama: absence of evidence for deficient metabolizers. Clin Pharmacol Ther 1988;43:456–65.
29. Arias TD, Inaba T, Cooke RG, Jorge LF. A preliminary note on the transient polymorphic oxidation of sparteine in the Ngawbé Guaymí Amerindians: a case of genetic divergence with tentative phylogenetic time frame for the pathway. Clin Pharmacol Ther 1988;44:343–52.
30. Arias TD, Jorge LF. An observation on the ethnic uniqueness of the debrisoquine and sparteine antimodes: a study in the Ngawbé Guaymí Amerindians of Panamá. Br J Clin Pharmacol 1989;28:493–4.
31. Jorge LF, Arias TD, Inaba T, Jackson PR. Unimodal distribution of the metabolic ratio for debrisoquine in Cuna Amerindians of Panama. Br J Clin Pharmacol 1990;30:281–5.
32. Arias TD, Jorge LF. Population pharmacology and ethnopharmacokinetics in 3 Amerindian groups from Panama: Cuna, Ngawbe Guaymi and Teribe. Rev Méd Panamá 1990;15:20–6.
33. Inaba T, Jorge LF, Arias TD. Mephenytoin hydroxylation in the Cuna Amerindians of Panama. Br J Clin Pharmacol 1988;25:75–9.
34. Martínez C, Agúndez JA, Olivera M, Llerena A, Ramirez R, Hernández M, et al. Influence of genetic admixture on polymorphisms of drug-metabolizing enzymes: analyses of mutations on NAT2 and C gamma P2E1 genes in a mixed Hispanic population. Clin Pharmacol Ther 1998;63:623–8.
35. Arias TD, Jorge LF, Griese EU, Inaba T, Eichelbaum M. Polymorphic N-acetyltransferase (NAT2) in Amerindian populations of Panama and Colombia: high frequencies of point mutation 857A, as found in allele S3/M3. Pharmacogenetics 1993;3:328–31.
36. Jorge-Nebert LF, Eichelbaum M, Griese E-U, Inaba T, Arias TD. Analysis of six SNPs of NAT2 in Ngawbe and Embera Amerindians of Panama and determination of the Embera acetylation phenotype using caffeine. Pharmacogenetics 2002;12:39–48.
37. Inaba T, Arias TD. On phenotyping with isoniazid: the use of urinary acetylation ratio and the uniqueness of antimodes. Study of two Amerindian populations. Clin Pharmacol Ther 1987;42:493–7.
38. Gómez-Casado E, Martínez-Laso J, Moscoso J, Zamora J, Martin-Villa M, Perez-Blas M, et al. Origin of Mayans according to HLA genes and the uniqueness of Amerindians. Tissue Antigens 2003;61:425–36.
39. Arrieta-Bolaños E, Maldonado-Torres H, Dimitriu O, Hoddinott MA, Fowles F, Shah A, et al. HLA-A, -B, -C, -DQB1, and -DRB1,3,4,5 allele and haplotype frequencies in the Costa Rica Central Valley population and its relationship to worldwide populations. Hum Immunol 2011;72:80–6.
40. Vernaza-Kwiers AA, de Góómez IJ, Díaz-Isaacs M, Cuero CJ, Pérez Guardia E, Moreno Saavedra M. Gene frequency and haplotypes of the HLA system in the Panamanian population. Rev Méd Panamá 1995;20:116–23.
41. Iwanaga KK, Eberle M, Kolman CJ, Bermingham E, Watkins DI. Further diversification of the HLA-B locus in Central American Amerindians: new B*39 and B*51 alleles in the Kuna of Panama. Tissue Antigens 1997;50:251–7.
42. Morera B, Sánchez-Rivera G, Jiménez-Arce G, Calafell F, Morales-Cordero AI. Nicaraguan population data on LDLR, GYPA, D7S8, HBGG, GC and HLA-DQA1 loci. Rev Biol Trop 2001;49:1253–60.
43. Morales AI, Morera B, Jiménez-Arce G, Sánchez-Rivera G, Calafell F, Barrantes R. Allele frequencies of markers LDLR, GYPA, D7S8, HBGG, GC, HLA-DQA1 and D1S80 in the general and minority populations of Costa Rica. Forensic Sci Int 2001;124:1–4.
44. Vicente J, Sinues B, Fanlo A, Vasquez P, Medina JC, Martinez-Jarreta B. Polymorphism C3435T of the MDR1 gene in Central Americans and Spaniards. Mol Biol Rep 2008;35:473–8.
45. Azofeifa J, Barrantes R. Genetic variation in the Bribri and Cabecar Amerindians from Talamanca, Costa Rica. Rev Biol Trop 1991;39:249–53.
46. Azofeifa J, Hahn M, Ruiz E, Hummerich L, Morales AI, Jiménez G, et al. The STR polymorphism (AAAAT)n within the intron 1 of the tumor protein 53 (TP53) locus in 17 populations of different ethnic groups of Africa, America, Asia and Europe. Rev Biol Trop 2004;52:645–57.
47. Sinues B, Vicente J, Fanlo A, Vasquez P, Medina JC, Mayayo E, et al. CYP3A5*3 and CYP3A4*1B allele distribution and genotype combinations: differences between Spaniards and Central Americans. Ther Drug Monit 2007;29:412–6.

48. González Argüello R. El citocromo 1A2 (CYP 1A2) en una población universitaria de Costa Rica. Rev Costarric Cienc Médicas 2002;23:25–31.
49. Griman P, Moran Y, Valero G, Loreto M, Borjas L, Chiurillo MA. CYP2D6 gene variants in urban/admixed and Amerindian populations of Venezuela: pharmacogenetics and anthropological implications. Ann Hum Biol 2012;39:137–42.
50. Bailliet G, Santos MR, Alfaro EL, Dipierri JE, Demarchi DA, Carnese FR, et al. Allele and genotype frequencies of metabolic genes in Native Americans from Argentina and Paraguay. Mutat Res 2007;627:171–7.
51. Halling J, Petersen MS, Damkier P, Nielsen F, Grandjean P, Weihe P, et al. Polymorphism of CYP2D6, CYP2C19, CYP2C9 and CYP2C8 in the Faroese population. Eur J Clin Pharmacol 2005;61:491–7.
52. Llerena A, Naranjo M, Rodrigues-Soares F, Peñas-Lledó EM, Fariñas H, Tarazona-Santos E. Interethnic variability of CYP2D6 polymorphisms in the world populations. Expert Opin Drug Metab Toxicol 2014;10:1569–83.
53. Gaedigk A, Coetsee C. The CYP2D6 gene locus in South African Coloureds: unique allele distributions, novel alleles and gene arrangements. Eur J Clin Pharmacol 2008;64:465–75.
54. Wright GE, Niehaus DJ, Drögemöller BI, Koen L, Gaedigk A, Warnich L. Elucidation of CYP2D6 genetic diversity in a unique African population: implications for the future application of pharmacogenetics in the Xhosa population. Ann Hum Genet 2010;74:340–50.
55. Barrantes R. Diversidad genética y mezcla racial en los amerindios de Costa Rica y Panamá. Rev Biol Trop 1993;41:379–84.
56. Sosa-Macías M, Elizondo G, Flores-Pérez C, Flores-Pérez J, Bradley-Alvarez F, Alanis-Bañuelos RE, et al. CYP2D6 genotype and phenotype in Amerindians of Tepehuano origin and Mestizos of Durango, Mexico. J Clin Pharmacol 2006;46:527–36.
57. Jurima-Romet M, Goldstein JA, LeBelle M, Aubin RA, Foster BC, Walop W, et al. CYP2C19 genotyping and associated mephenytoin hydroxylation polymorphism in a Canadian Inuit population. Pharmacogenetics 1996;6:329–39.
58. Muñoz S, Vollrath V, Vallejos MP, Miquel JF, Covarrubias C, Raddatz A, et al. Genetic polymorphisms of CYP2D6, CYP1A1 and CYP2E1 in the South-Amerindian population of Chile. Pharmacogenetics 1998;8:343–51.
59. Sistonen J, Sajantila A, Lao O, Corander J, Barbujani G, Fuselli S. CYP2D6 worldwide genetic variation shows high frequency of altered activity variants and no continental structure. Pharmacogenet Genomics 2007;17:93–101.
60. Salazar-Flores J, Torres-Reyes LA, Martínez-Cortés G, Rubi-Castellanos R, Sosa-Macías M, Muñoz-Valle JF, et al. Distribution of CYP2D6 and CYP2C19 polymorphisms associated with poor metabolizer phenotype in five Amerindian groups and western Mestizos from Mexico. Genet Test Mol Biomark 2012;16:1098–104.
61. Lazalde-Ramos BP, Martínez-Fierro Mde L, Galaviz-Hernández C, Garza-Veloz I, Naranjo M, Sosa-Macías M, et al. CYP2D6 gene polymorphisms and predicted phenotypes in eight indigenous groups from northwestern Mexico. Pharmacogenomics 2014;15:339–48.
62. Fuselli S, de Filippo C, Mona S, Sistonen J, Fariselli P, Destro-Bisol G, et al. Evolution of detoxifying systems: the role of environment and population history in shaping genetic diversity at human CYP2D6 locus. Pharmacogenet Genomics 2010;20:485–99.
63. Gaedigk A, Bhathena A, Ndjountché L, Pearce RE, Abdel-Rahman SM, Alander SW, et al. Identification and characterization of novel sequence variations in the cytochrome P4502D6 (CYP2D6) gene in African Americans. Pharmacogenomics J 2005;5:173–82.
64. Gaedigk A, Bradford LD, Marcucci KA, Leeder JS. Unique CYP2D6 activity distribution and genotype-phenotype discordance in black Americans. Clin Pharmacol Ther 2002;72:76–89.
65. Gaedigk A, Ndjountché L, Divakaran K, Dianne Bradford L, Zineh I, Oberlander TF, et al. Cytochrome P4502D6 (CYP2D6) gene locus heterogeneity: characterization of gene duplication events. Clin Pharmacol Ther 2007;81:242–51.
66. McLellan RA, Oscarson M, Seidegård J, Evans DA, Ingelman-Sundberg M. Frequent occurrence of CYP2D6 gene duplication in Saudi Arabians. Pharmacogenetics 1997;7:187–91.
67. Aklillu E, Persson I, Bertilsson L, Johansson I, Rodrigues F, Ingelman-Sundberg M. Frequent distribution of ultrarapid metabolizers of debrisoquine in an ethiopian population carrying duplicated and multiduplicated functional CYP2D6 alleles. J Pharmacol Exp Ther 1996;278:441–6.
68. Casner PR. The effect of CYP2D6 polymorphisms on dextromethorphan metabolism in Mexican Americans. J Clin Pharmacol 2005;45:1230–5.
69. Holmes MV, Shah T, Vickery C, Smeeth L, Hingorani AD, Casas JP. Fulfilling the promise of personalized medicine? Systematic review and field synopsis of pharmacogenetic studies. PlOS ONE 2009;4:e7960.
70. Gonzalez HM, Romero EM, Peregrina AA, de J Chávez T, Escobar-Islas E, Lozano F, et al. CYP2C19- and CYP3A4-dependent omeprazole metabolism in West Mexicans. J Clin Pharmacol 2003;43:1211–5.
71. Luo H-R, Poland RE, Lin K-M, Wan Y-J. Genetic polymorphism of cytochrome P450 2C19 in Mexican Americans: a cross-ethnic comparative study. Clin Pharmacol Ther 2006;80:33–40.
72. Goldstein JA. Clinical relevance of genetic polymorphisms in the human CYP2C subfamily. Br J Clin Pharmacol 2001;52:349–55.
73. Henn BM, Cavalli-Sforza LL, Feldman MW. The great human expansion. Proc Natl Acad Sci USA 2012;109:17758–64.
74. Fuselli S, Gilman RH, Chanock SJ, Bonatto SL, De Stefano G, Evans CA, et al. Analysis of nucleotide diversity of NAT2 coding region reveals homogeneity across Native American populations and high intra-population diversity. Pharmacogenomics J 2007;7:144–52.
75. Patin E, Barreiro LB, Sabeti PC, Austerlitz F, Luca F, Sajantila A, et al. Deciphering the ancient and complex evolutionary history of human arylamine N-acetyltransferase genes. Am J Hum Genet 2006;78:423–36.
76. García-Martín E, Martínez C, Ladero JM, Agúndez JA. Interethnic and intraethnic variability of CYP2C8 and CYP2C9 polymorphisms in healthy individuals. Mol Diagn Ther 2006;10:29–40.
77. Johnson N, Bell P, Jonovska V, Budge M, Sim E. NAT gene polymorphisms and susceptibility to Alzheimer's disease: identification of a novel NAT1 allelic variant. BMC Med Genet 2004;5:6.
78. Vargas-Alarcón G, Hernández-Pacheco G, Moscoso J, Pérez-Hernández N, Murguía LE, Moreno A, et al. HLA genes in Mexican Teeneks: HLA genetic relationship with other worldwide populations. Mol Immunol 2006;43:790–9.

79. Hollenbach JA, Thomson G, Cao K, Fernandez-Vina M, Erlich HA, Bugawan TL, et al. HLA diversity, differentiation, and haplotype evolution in Mesoamerican Natives. Hum Immunol 2001;62:378–90.
80. Silvera C, Vargas-Alarcon G, Areces C, Rey D, Parga-Lozano C, Gomez-Prieto P, et al. HLA genes in Wayu Amerindians from Colombia. Immunol Invest 2011;40:92–100.
81. Martinez-Laso J, Siles N, Moscoso J, Zamora J, Serrano-Vela JI, R-A-Cachafeiro JI, et al. Origin of Bolivian Quechua Amerindians: their relationship with other American Indians and Asians according to HLA genes. Eur J Med Genet 2006;49:169–85.
82. Benitez O, Busson M, Charron D, Loiseau P. HLA polymorphism in a Guarani-Indian population from Paraguay and its usefulness for the Hispano-Indian admixture study in Paraguay. Int J Immunogenet 2011;38:7–11.
83. Moscoso J, Seclen S, Serrano-Vela JI, Villena A, Martinez-Laso J, Zamora J, et al. HLA genes in Lamas Peruvian-Amazonian Amerindians. Mol Immunol 2006;43:1881–9.
84. Klitz W, Gragert L, Maiers M, Tu B, Lazaro A, Yang R, et al. Four-locus high-resolution HLA typing in a sample of Mexican Americans. Tissue Antigens 2009;74:508–13.
85. Arnaiz-Villena A, Moscoso J, Granados J, Serrano-Vela JI, de la Peña A, Reguera R, et al. HLA Genes in Mayos Population from Northeast Mexico. Curr Genomics 2007;8:466–75.
86. Williams R, Chen Y-F, Endres R, Middleton D, Trucco M, Williams JD, et al. Molecular variation at the HLA-A, B, C, DRB1, DQA1, and DQB1 loci in full heritage American Indians in Arizona: private haplotypes and their evolution. Tissue Antigens 2009;74:520–33.
87. Rey D, Parga-Lozano C, Moscoso J, Areces C, Enriquez-de-Salamanca M, Fernández-Honrado M, et al. HLA genetic profile of Mapuche (Araucanian) Amerindians from Chile. Mol Biol Rep 2013;40:4257–67.
88. Gonzalez-Galarza FF, Christmas S, Middleton D, Jones AR. Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations. Nucleic Acids Res 2011;39:D913–9.
89. Silveira Vda S, Canalle R, Scrideli CA, Queiroz RG, Tone LG. Polymorphisms in genes encoding drugs and xenobiotic metabolizing enzymes in a Brazilian population. Biomarkers 2009;14:111–7.
90. Arnaiz-Villena A, Parga-Lozano C, Moreno E, Areces C, Rey D, Gomez-Prieto P. The origin of Amerindians and the peopling of the Americas according to HLA genes: admixture with Asian and Pacific people. Curr Genomics 2010;11:103–14.
91. Konishi T, Smith JL, Lin K-M, Wan Y-J. Influence of genetic admixture on polymorphisms of alcohol-metabolizing enzymes: analyses of mutations on the CYP2E1, ADH2, ADH3 and ALDH2 genes in a Mexican-American population living in the Los Angeles area. Alcohol Alcohol 2003;38:93–4.
92. Gordillo-Bastidas E, Panduro A, Gordillo-Bastidas D, Zepeda-Carrillo EA, García-Bañuelos JJ, Muñoz-Valle JF, et al. Polymorphisms of alcohol metabolizing enzymes in indigenous Mexican population: unusual high frequency of CYP2E1*c2 allele. Alcohol Clin Exp Res 2010;34:142–9.
93. Quiñones L, Lucas D, Godoy J, Cáceres D, Berthou F, Varela N, et al. CYP1A1, CYP2E1 and GSTM1 genetic polymorphisms. The effect of single and combined genotypes on lung cancer susceptibility in Chilean people. Cancer Lett 2001;174:35–44.
94. Vargas-Alarcon G, Ramírez-Bello J, de la Peña A, Calderón-Cruz B, Peña-Duque MA, Martínez-Ríos MA, et al. Distribution of ABCB1, CYP3A5, CYP2C19, and P2RY12 gene polymorphisms in a Mexican Mestizos population. Mol Biol Rep 2014;41:7023–9.
95. Santos PC, Soares RA, Santos DB, Nascimento RM, Coelho GL, Nicolau JC, et al. CYP2C19 and ABCB1 gene polymorphisms are differently distributed according to ethnicity in the Brazilian general population. BMC Med Genet 2011;12:13.
96. Estrela RC, Ribeiro FS, Carvalho RS, Gregório SP, Dias-Neto E, Struchiner CJ, et al. Distribution of ABCB1 polymorphisms among Brazilians: impact of population admixture. Pharmacogenomics 2008;9:267–76.
97. Sortica Vde A, Ojopi EB, Genro JP, Callegari-Jacques S, Ribeiro-Dos-Santos A, de Moraes MO, et al. Influence of genomic ancestry on the distribution of SLCO1B1, SLCO1B3 and ABCB1 gene polymorphisms among Brazilians. Basic Clin Pharmacol Toxicol 2012;110:460–8.
98. Wielandt AM, Vollrath V, Chianale J. Polymorphisms of the multiple drug resistance gene (MDR1) in Mapuche, Mestizo and Maori populations in Chile. Rev Méd Chile 2004;132:1061–8.
99. Scheiner MA, Damasceno AM, Maia RC. ABCB1 single nucleotide polymorphisms in the Brazilian population. Mol Biol Rep 2010;37:111–8.
100. Reyes-Hernández OD, Arteaga-Illán G, Elizondo G. Detection of CYP3A4*1B and CYP3A4*2 polymorphisms by RFLP. Distribution frequencies in a Mexican population. Clin Genet 2004;66:166–8.
101. Ball SE, Scatina J, Kao J, Ferron GM, Fruncillo R, Mayer P, et al. Population distribution and effects on drug metabolism of a genetic variant in the 5′ promoter region of CYP3A4. Clin Pharmacol Ther 1999;66:288–94.
102. Suarez-Kurtz G, Perini JA, Bastos-Rodrigues L, Pena SD, Struchiner C. Impact of population admixture on the distribution of the CYP3A5*3 polymorphism. Pharmacogenomics 2007;8:1299–306.
103. Dorado P, Heras N, Machín E, Hernández F, Teran E, Llerena A. CYP2D6 genotype and dextromethorphan hydroxylation phenotype in an Ecuadorian population. Eur J Clin Pharmacol 2012;68:637–44.
104. Dorado P, Beltrán LJ, Machín E, Peñas-Lledó EM, Terán E, Llerena A, et al. Losartan hydroxylation phenotype in an Ecuadorian population: influence of CYP2C9 genetic polymorphism, habits and gender. Pharmacogenomics 2012;13:1711–7.
105. Dorado P, Sosa-Macias MG, Peñas-Lledó EM, Alanis-Bañuelos RE, Wong M-L, Licinio J, et al. CYP2C9 allele frequency differences between populations of Mexican-Mestizo, Mexican-Tepehuano, and Spaniards. Pharmacogenomics J 2011;11:108–12.
106. Sosa-Macías M, Dorado P, Alanis-Bañuelos RE, Llerena A, Lares-Asseff I. Influence of CYP2D6 deletion, multiplication, -1584C-->G, 31G-->A and 2988G-->a gene polymorphisms on dextromethorphan metabolism among Mexican Tepehuanos and mestizos. Pharmacology 2010;86:30–6.
107. López M, Dorado P, Ortega A, Peñas-Lledó E, Monroy N, Silva-Zolezzi I, et al. Interethnic differences in UGT1A4 genetic polymorphisms between Mexican Mestizo and Spanish populations. Mol Biol Rep 2013;40:3187–92.