UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE BIOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA

RONALD RODRIGUES DE MOURA

INFERÊNCIA DA MISCIGENAÇÃO GENÉTICA NA POPULAÇÃO PERNAMBUCANA E SUA APLICAÇÃO EM ESTUDOS DE ASSOCIAÇÃO

Recife
2018

RONALD RODRIGUES DE MOURA

INFERÊNCIA DA MISCIGENAÇÃO GENÉTICA NA POPULAÇÃO PERNAMBUCANA E SUA APLICAÇÃO EM ESTUDOS DE ASSOCIAÇÃO

Tese apresentada ao Programa de Pós-Graduação em Genética da, Universidade Federal de Pernambuco, como parte dos requisitos exigidos para obtenção do título de Doutor em Genética.

Orientador: Lucas André Cavalcanti Brandão
Co-orientadores: Sergio Crovella e Valdir de Queiroz Balbino

Recife
2018

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD

Moura, Ronald Rodrigues de

Inferência da miscigenação genética na população pernambucana e sua aplicação em estudos de associação/ Ronald Rodrigues de Moura- 2018.

92 folhas: il., fig., tab.
Orientador: Lucas André Cavalcanti Brandão
Coorientadores: Sérgio Crovella e Valdir de Queiroz Balbino
Tese (doutorado) – Universidade Federal de Pernambuco. Centro de Biociências. Programa de Pós-Graduação em Genética. Recife, 2018.
Inclui referências e anexos

1. Genética de populações 2. Miscigenação 3. Pernambuco I. Brandão, Lucas André Cavalcanti (orient.) II. Crovella, Sérgio (coorient) III. Balbino, Valdir de Queiroz (coorient.) IV. Título

576.58 CDD (22.ed.) UFPE/CB-2018-143

Elaborado por Elaine C. Barroso CRB4/1728

RONALD RODRIGUES DE MOURA

INFERÊNCIA DA MISCIGENAÇÃO GENÉTICA NA POPULAÇÃO PERNAMBUCANA E SUA APLICAÇÃO EM ESTUDOS DE ASSOCIAÇÃO

Tese apresentada ao Programa de Pós-Graduação em Genética, Área de Concentração Genética, da Universidade Federal de Pernambuco, como parte dos requisitos exigidos para obtenção do título de Doutor em Genética.

Aprovada em 15/02/2018

COMISSÃO EXAMINADORA

_____

Dr. Lucas André Cavalcanti Brandão
Universidade Federal de Pernambuco

_____

Dr. Marcos André Cavalcanti Bezerra
Universidade Federal de Pernambuco

_____

Dr. Antônio Roberto Lucena de Araújo
Universidade Federal de Pernambuco

_____

Dr. Rodrigo Augusto Torres
Universidade Federal de Pernambuco

_____

Dr. Paulo Roberto Eleutério de Souza
Universidade Federal Rural de Pernambuco

AGRADECIMENTOS

Ao meu orientador, Prof. Lucas André Cavalcanti Brandão, e meus co-orientadores, Prof. Sergio Crovella e Prof. Valdir Queiroz Balbino, pelos ensinamentos e conselhos ao longo desse tempo.

Aos meus colegas de trabalho, dos quais não citarei em nomes para não cometer injustiças (vocês sabem quem são), por tornar o exercício da ciência ainda mais prazeroso.

Aos professores e funcionários do Programa de Pós-graduação em Genética, pelo suporte, que já vem desde a época do mestrado.

Aos membros da banca, que avaliarão meu trabalho de maneira justa e competente.

Aos órgãos de fomento (CNPq, CAPES e FACEPE), pela confiança no trabalho e pelo empenho financeiro que contribuiu decisivamente para a realização desse projeto.

A vós, agradeço!

“Todo o brasileiro, mesmo o alvo, de cabelo louro, traz na alma, quando não na alma e no corpo [...] a sombra, ou pelo menos a pinta, do indígena ou do negro. [...] Na ternura, na mímica excessiva, no catolicismo em que se deliciam nossos sentidos, na música, no andar, na fala, no canto de ninar menino pequeno, em tudo que é expressão sincera de vida, trazemos quase todos a marca da influência negra”.

(Gilberto Freyre)

RESUMO

A natureza tri-híbrida da população brasileira de um modo geral, faz com que seja possível tecer um diálogo entre os acontecimentos históricos e evolutivos que se sucederam nesses últimos cinco séculos. Os Pernambucanos são bons representantes disso, porém poucos trabalhos avaliaram esse fenômeno em nosso estado. A presente tese pretende ajudar a preencher essa lacuna, além de reforçar aplicações importantes desse tipo de análise, como nos estudos de associação genética. Utilizando levantamentos sistemáticos da literatura, foi possível verificar que a contribuição europeia (EUR) predomina em todo o território brasileiro, com presença em 62% do nosso genoma, enquanto a africana (AFR) contribui com 21% e a nativa americana (AMR) com 17%. Em particular, a população pernambucana segue esse padrão com 60%, 23% e 17% de proporções EUR, AFR e AMR, respectivamente. Por fim, vale mencionar a importância que a estimativa da miscigenação populacional tem em estudos de associação caso/controle no sentido desse fenômeno produzir resultados espúrios por inflacionar a frequência de um alelo em um dos grupos de estudo, não por influência desse no fenótipo estudado, mas por ser mais presente em indivíduos com certa composição genética. Assim, verificamos se há influência da miscigenação na presença de alelos em genes relacionados a recuperação imunológica de pacientes HIV+, sob tratamento antirretroviral. Como resultado, as estimativas não apontaram diferenças significativas na contribuição EUR, AFR e AMR entre os grupos avaliados (p-value > 0.05).

Palavras-chave: Marcadores Informativos de Ancestralidade. Estruturação Populacional. Fluxo Gênico. SNP.

ABSTRACT

The tri-hybrid aspect of the general Brazilian population makes possible to establish a dialogue between Historical and Evolutionary events that toke place in the past five centuries. The ‘Pernambucanos’ are a good example of it; however, a few studies evaluated this phenomenon in Pernambuco. This thesis helps to fulfill this gap and reinforce the importance of applying this analysis in genetic studies, such as those with case/control design. From a systematic screening of the bibliography, we verified that the European genetic contribution (EUR) is predominant throughout all Brazilian territory, composing 62% of our genome, whereas African (AFR) contributes with 21% and Native American (AMR) with 17%. Pernambuco’s population follows this pattern with 60%, 23% and 17% for EUR, AFR and AMR proportions, respectively. Lastly, it is worth to note the importance that estimating genetic admixture has on genetic association studies, once this evolutionary force can produce spurious results by inflate the frequency of an allele in one of the groups. This may happen not because the allele is associated to the phenotype, but because it is more frequent in individuals with a given genetic background. Thus, we verified the influence of genetic admixture in the presence of alleles from genes related to immunologic recovery of HIV^{+} patients, under antiretroviral treatment. As result, the estimates did not pointed to significant differences of EUR, AFR or AMR contributions (p-value > 0.05).

Key-words: Ancestry Informative Markers. Genetic Substructure. Gene Flow. SNP.

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

IBGE - Instituto Brasileiro de Geografia e Estatística

AIM - *Ancestry Informative Marker*

SNP - *Single Nucleotide Polymorphism*

INDEL - Inserção/Deleção

PCA - *Principal Component Analysis*

PCR - *Polymerase Chain Reaction*

SUMÁRIO

1 INTRODUÇÃO

Os processos sociais que ocorreram desde as primeiras interações entre Portugueses, tribos indígenas nativas do Brasil e africanos trazidos forçadamente a partir do século XVI contribuíram para a composição atual do genoma da população brasileira. Hoje em dia está claro que o genoma do brasileiro é, de um modo geral, trí-hibrido, com diferentes proporções de herança genética de um dos grupos citados acima de modo que cada indivíduo tenha praticamente uma proporção única de ancestralidade africana, europeia e indígena.

Porém, essa diversidade entre indivíduos não se percebe ao nível populacional, onde populações de regiões geográfica distintas (sudeste e nordeste, por exemplo) podem apresentar contribuições genéticas bastante semelhantes. Essa afirmação pode ser constatada após mais de três décadas de estudos genéticos em populações brasileiras de diferentes regiões que buscam caracterizar essa diversidade. A população pernambucana também emergiu de forma similar que a brasileira como um todo. No entanto, ao longo dessas décadas de estudo, poucos trabalhos pesquisaram sobre como as contribuições africanas, ameríndias e europeias aparecem no genoma do povo pernambucano.

Outro aspecto discutido ao longo dos estudos da miscigenação genética brasileira (e também no artigo do segundo capítulo) foi a relação da miscigenação genética com a “raça” ou “etnia” dos indivíduos. Apesar do ponto de vista biológico raças humanas não existirem, as pessoas que compartilham certos traços fisionômicos e culturais tendem a se categorizar em grupos. Na sociedade brasileira atual, existe um debate intenso sobre políticas afirmativas e racismo. Apesar de nesses casos a ciência não tenha um caráter normativo, a Genética pode informar e talvez enriquecer o debate.

Não só em Pernambuco, mas em todo o país, injúrias raciais são praticadas e propagadas cotidianamente. O tipo racismo no Brasil é o de “marca”, onde os principais alvos do preconceito são traços físicos sendo cor da pele o mais explorado (escárnios em relação ao tipo de cabelo e formato do nariz também podem ser vistos). Assim, ainda no segundo capítulo estudamos a relação da miscigenação genética com a cor da pele autodeclarada, de acordo com as categorias adotadas pelo Instituto Brasileiro de Geografia e Estatística (IBGE).

Além de poder esclarecer a dinamica populacional do ponto de vista genético, entender a miscigenação genética dos indivíduos em um grupo pode ajudar em estudos de associação genética, onde se objetiva estabelecer uma relação de causa e efeito entre variações genéticas e um fenótipo estudado (características físicas, doenças, etc.). A ocorrência de determinado alelo estudado em uma amostra da população pode ser influenciada pelo grau de miscigenação que aquela amostra possui. Por exemplo, um alelo mais frequente em indivíduos nativos americanos é mais visto em populações miscigenadas, quanto maior foi a proporção de contribuição nativo americana que ela possua. Assim, duas amostras de uma mesma população miscigenada podem ter diferenças significativas entre as frequências de um mesmo alelo, caso haja um viés com relação a miscigenação que essas amostras apresentam.

Portanto, estudos que buscam caracterizar a estruturação genética continuam sendo relevantes atualmente por ajudar a compreender a história de uma população e auxiliar na compreensão da influência genética em fenótipos de interesse.

2 REVISÃO DA LITERATURA

Miscigenação, do inglês *admixture*, é uma forma de fluxo gênico que refere-se ao processo no qual duas ou mais populações com diferentes frequências alélicas para mesmos loci se intercruzam, formando uma população híbrida (Mielke,et al., 2006). Consideráveis níveis de miscigenação genética vem ocorrendo na América Latina durante os últimos 500 anos entre populações africanas, europeias e indígenas.

2.1 MISCIGENAÇÃO GENÉTICA NAS AMÉRICAS E CARIBE

Considerando a matriz genética tri-híbrida observada nas populações ao longo dos continentes americanos, podemos destacar a grande variação de contribuições Africanas, Ameríndias e Europeias. A contribuição genética de populações ancestrais africanas pode variar de 1%, nos Estados Unidos, a 96%, no Haiti; a proporção de genoma nativo americano pode variar entre 0% e 92% (Haiti e Peru, respectivamente); e a contribuição europeia oscila entre 4%, nas Bahamas e 98%, nos Estados Unidos (Halder et al., 20009; Simms et al., 2010; Moura et al., 2015). Devido a falta de estudos sistematizados produzimos um relato, descrito no Capítulo I desse manuscrito.

2.2 MISCIGENAÇÃO GENÉTICA NO BRASIL

Um recente estudo aponta que os momentos-chave de fluxo gênico que contribuíram para a composição atual do genoma dos brasileiros ocorreram em três pulsos de miscigenação ocorridos há 18-16 (no século XVII), 12-10 (século XVIII) e 6-4 gerações (século XIX). O Genoma brasileiro passou de uma maior contribuição africana até os dois últimos pulsos, onde imigraram aproximadamente quatro milhões de europeus, em um processo conhecido como “branqueamento do Brasil”,

aumentando a contribuição europeia que passou a predominar até os dias de hoje (KHEDY ET AL., 2015).

No Brasil, as proporções de miscigenação genética também variam tanto dentro de um Estado, como em São Paulo (CARDENA ET AL., 2013; MANTA ET AL., 2013), quanto entre regiões (PENA, ET AL., 2011; MANTA ET AL., 2013). Uma das explicações para esses dados são as diferentes rotas de colonização que existiram no país, bem como as proporções de africanos, europeus e indígenas que coexistiram nas regiões e estados, como por exemplo em Pernambuco, que recebeu populações de Angola durante a escravidão (IBGE, 2000).

Em Pernambuco, três distintos estudos foram conduzidos: Alves-Silva, et al. (2000), usando marcadores no DNA mitocondrial, encontrou 44%, 34% e 22% de proporções africana, europeia e indígena, respectivamente; Carvalho-Silva et al. (2001), utilizando Y-DNA observou 4.1%, 95.9 e 0%, na mesma ordem que o estudo anterior; e Manta, et al. (2013), utilizando padrões autossômicos de inserções/deleções, encontrou valores de 28%, 57% e 15% das respectivas populações ancestrais. Ao nosso conhecimento, não há nenhum estudo que investigue a extensão das contribuições genéticas ancestrais em várias regiões de Pernambuco, como as do Agreste e Sertão usando os mesmos marcadores. No o artigo presente no Capítulo I desse manuscrito encontram-se mais informações sobre estudos no Brasil, enquanto que no Capítulo II estão dispostas mais informações sobre estudos em Pernambuco.

2.3 MARCADORES INFORMATIVOS DE ANCESTRALIDADE

A esse ponto, pode-se questionar como foi possível estimar a contribuição de cada população parental para uma população miscigenada. Atualmente a resposta inicial é utilizando o DNA. De fato, marcadores no DNA são os mais usados. Porém, no início desses estudos, os marcadores utilizados eram especialmente marcadores para sistemas sanguíneos (SCHULLER ET AL., 1982; SANTOS ET AL., 1987; RIBEIRO-DOS-SANTOS ET AL., 1995). Além de sequências de DNA mitocondrial, microssatélites e sítios *Alu* também foram utilizados em um segundo momento (FERREIRA ET AL., 2005; SCLIAR ET AL., 2009). Desde 2010, os marcadores que

vem predominando são os SNPs (do inglês, *single nucleotide polymorphism*) e INDELs (inserções/deleções) (SANTOS ET AL., 2010; PENA ET AL., 2011; GIOLO ET AL., 2012; MANTA ET AL., 2013). Com a alta disponibilidade de dados de genotipagem de SNPs, esse tipo de marcador vem tendo preferência superior aos INDELs, especialmente porque apesar do grau de informatividade de microssatélites serem de oito a dez vezes maior que a de SNPs, 2 a 12% dos SNPs conhecidos tem informatividade superior que a mediana da informatividade de microssatélites, além de atualmente serem mais viáveis técnica e economicamente (ROSENBERG ET AL.,2003).

Independentemente do tipo, esse marcadores devem ter em comum o fato de serem marcadores informativos de ancestralidades [do inglês, *ancestry informative marker* (AIM)]. Um AIM ideal tem um alelo fixado em uma população ancestral e outro ausente nas outras populações ancestrais. Entretanto, a maioria dos alelos são compartilhados entre as populações. Portanto, é importante identificar os AIMs que conseguem melhor discriminar as proporções de cada população ancestral em uma população miscigenada. Várias medidas para verificar o grau de informatividade dos marcadores tem sido desenvolvidas com o crescente aumento de informação disponível nos bancos de dados públicos. Dentre elas destacam-se: Diferença absoluta entre as frequências alélicas (δ), Conteúdo informativo de Shannon (SIC), conteúdo informativo de Fisher (FIC), estatísticas F (FST) e informatividade por medição atribuída (In) (DING, ET AL., 2011).

O algoritmo In informa o logaritmo esperado da taxa de verossimilhança que um alelo é atribuído a uma das populações ancestrais comparado com uma população “media hipotética” cujas as frequências alélicas são iguais a média das frequências alélicas entre as K populações. Quanto maior a diferença nas frequências alélicas entre as populações, maior será o valor de In (ROSENBERG ET AL.,2003). A partir da comparação dos cinco métodos citados acima, Ding, et al. (2011) concluíram que AIMs escolhidos baseados no algoritmo In produzem uma estimativa de ancestralidade genética com um menor viés estatístico e uma menor variância com uma menor quantidade de marcadores (DING, ET AL., 2011).

2.4 ESTRATIFICAÇÃO POPULACIONAL

Estratificação populacional é uma das fontes de associações espúrias em estudos caso-controle, envolvendo doenças ou resposta à fármacos, por exemplo, podendo levar tanto à resultados falsos-positivos quanto falso-negativos (BALDING, 2006; LEWIS; KNIGHT, 2012). Esse fator de confundimento se torna evidente quando casos e controles possuem diferentes proporções de ancestralidade genética; e também quando o fenótipo que está sendo estudado possui uma variação em relação com a base genética ancestral das populações (THOMAS; WITTE, 2002).

Três métodos para correção da estratificação genética em estudos caso-controle vem sendo implementados: Associação Estruturada (PRITCHARD ET AL., 2000), Controle Genômico (DEVLIN ET AL., 2001) e Análise de Componentes Principais – PCA, do inglês, *Principal Component Analysis* – (PATTERSON ET AL., 2006). Esses métodos são capazes de corrigir os resultados de associação através da análise de um conjunto de loci espalhados pelo genoma que não estão ligados ao *locus* candidato (ENOCH, ET AL., 2006).

Portanto, para estudos de associação genética em populações miscigenadas, como a brasileira, é crucial a inferência da contribuição das populações parentais no genoma dos indivíduos estudados a fim de evitar resultados falso-positivos ou falso-negativos.

3 OBJETIVOS

3.1 OBJETIVO GERAL

Inferir a contribuição africana, ameríndia e europeia no genoma de populações pernambucanas e avaliar o efeito dessa estruturação populacional em estudos de associação genética.

3.2 OBJETIVOS ESPECÍFICOS

1. Realizar o levantamento do estado da arte dos estudos de miscigenação genética no Brasil;
2. Inferir a ancestralidade da população geral pernambucana;
3. Realizar estudos de associação genética com doenças e outros fenótipos de interesse aplicando os marcadores de ancestralidade em ambos casos e controles.

4 META-ANALYSIS OF BRAZILIAN GENETIC ADMIXTURE AND COMPARISON WITH OTHER LATIN AMERICA COUNTRIES

Autores: Ronald Rodrigues de Moura, Antonio Victor Campos Coelho, Valdir de Queiroz Balbino, Sergio Crovella, Lucas André Cavalcanti Brandão
Revista: American Journal Of Human Biology (2015)
Fator de Impacto (QUALIS): 1.78 (B2)

O texto referente a esse Capítulo pode ser encontrado no Apêndice A.

5 A RAPID SCREENING OF ANCESTRY FOR GENETIC ASSOCIATION STUDIES IN AN ADMIXED POPULATION FROM PERNAMBUCO, BRAZIL

Autores: Antonio Victor Campos Coelho, Ronald Rodrigues de Moura, Catarina Addobbati Jordão Cavalcanti, Rafael Lima Guimarães, Paula Sandrin Garcia, Sergio Crovella, Lucas André Cavalcanti Brandão

Revista: Genetics and Molecular Research (2015)

Fator de Impacto (QUALIS): 1.013 (B4)

O texto referente a esse Capítulo pode ser encontrado no Apêndice B.

6 ON THE USE OF CHINESE POPULATION AS A PROXY OF AMERINDIAN ANCESTORS IN GENETIC ADMIXTURE STUDIES WITH LATIN AMERICAN POPULATIONS

Autores: Ronald Rodrigues de Moura, Valdir de Queiroz Balbino, Sergio Crovella, Lucas André Cavalcanti Brandão

Revista: European Journal of Human Genetics (2015)

Fator de Impacto (QUALIS): 4.287 (A2)

O texto referente a esse Capítulo pode ser encontrado no Apêndice C.

7 ANTIRETROVIRAL THERAPY IMMUNOLOGIC NON-RESPONSE IN A BRAZILIAN POPULATION: ASSOCIATION STUDY USING PHARMACO- AND IMMUNOGENETIC MARKERS

Autores: Antônio Victor Campos Coelho, Ronald Rodrigues de Moura, Rafael Lima Guimarães, Lucas André Cavalcanti Brandão. Sergio Crovella
Revista: Pharmacogenetics and Genomics
Fator de Impacto (QUALIS): 2.184 (B1)

O texto referente a esse Capítulo pode ser encontrado no Apêndice D.

8 DISCUSSÃO GERAL

Com base nos resultados encontrados até agora, podemos observar que há miscigenação genética em todas as América e Caribe, respeitando uma matriz composta de populações parentais africanas, ameríndias e europeias. As proporções de ancestralidade de cada uma dessas populações variam dramaticamente, muito em parte dos eventos históricos e sociais ocorridos ao longo dos períodos de povoamento e migração. No Brasil, a situação é semelhante, destacando a contribuição europeia predominante (acima de 50%) ao longo de todo o território.

Em Pernambuco, encontramos um exemplo do que é observado para todo o Brasil. Além disso, ao avaliar uma relação entre miscigenação genética e cor da pele autodeclarada, percebemos apesar de haver uma relação, essa correlação é fraca de modo que há uma alta sobreposição das proporções de ancestralidade europeia e africana entre indivíduos brancos, pardos e pretos. Sendo assim, cor da pele não é visto como um bom indicador de ancestralidade genética.

Um aspecto técnico das inferências de miscigenação genética explorado foi o uso correto das populações parentais nos testes, onde a recomendação é a de que sejam utilizadas populações que representem os grupos biogeográficos de onde as populações parentais vieram. Portanto, para populações latino-americanas, deve-se usar populações de referência europeia, africana e nativo-americana.

Outro ponto a ser considerado é a possível estruturação genética que pode ser observada em estudos de associação em populações miscigenadas, como a população de Pernambuco. Apesar dessa estruturação não ter sido detectada no estudo de associação apresentação do capítulo IV, outros estudos com desenho experimental similar podem sofrer desse efeito durante o processo de amostragem.

9 CONCLUSÕES

Como conclusão dessa tese, pudemos observar que: a população brasileira em geral, possui natureza tri-hibrida, com proporções das populações parentais variando ligeiramente de acordo com as regiões geográficas do pais; a população pernambucana segue a tendência da população brasileira, que é uma contribuição genética europeia predominante; e ressaltou a importância de avaliar a miscigenação presente em amostragens para estudos de associação genética.

REFERÊNCIAS

ALVES-SILVA J, ET AL. The Ancestry of Brazilian mtDNA Lineages. **Am J Hum Genet**, v. 67, p. 444-461, 2000.

BALDING D. A tutorial on statistical methods for population association studies. **Nature Rev Genet**, v. 7, p. 781-791, 2006.

CARDENA M, ET AL. Assessment of the Relationship between Self-Declared Ethnicity, Mitochondrial Haplogroups and Genomic Ancestry in Brazilian Individuals. **PloS One**, v. 8, n. 4, p. e62005, 2013.

CARVALHO-SILVA D, ET AL. The Phylogeography of Brazilian Y-Chromosome Lineages. **Am J Hum Genet**, v. 68, p. 281-286, 2001

DEVLIN B, ET AL. Genomic control, a new approach to genetic-based association studies. **Theoretical Population Biology**, v. 60, p. 155-166, 2001.

DING L, ET AL. Comparison of measures of marker informativeness for ancestry and admixture mapping. **BMC Genomics**, v. 12, p. 622-640, 2011.

ENOCH M, ET AL. Using ancestry-informative markers to define populations and detect population stratification. **J Psychopharmacol**, v. 20, n. 4, p. 19-26, 2006.

FERREIRA F, ET AL. Genetic characterization of the population of São Luís, MA, Brazil. **Genet Mol Biol**, v. 28, n. 1, p. 22-31, 2005.

GIOLO S, ET AL. Brazilian urban population genetic structure reveals a high degree of admixture. **Eur J Hum Genet**, v. 20, p. 111-116, 2012.

HALDER I, ET AL. Measurement of admixture proportions and description of admixture structure in different US populations. **Hum Mutat**, v. 30, n. 9, p. 1299-1309, 2009.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Brasil: 500 anos de povoamento**. IBGE, Rio de Janeiro 232 p., 2000

KEHDY, F, ET AL. Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. **PNAS**, v. 112, n. 28, p. 8696–8701, 2015.

LEWIS C, KNIGHT J. Introduction to Genetic Association Studies. **Cold Spring Harb Protoc**, v. 3, p. 297-306, 2012.

MANTA F, ET AL. Revisiting the genetic ancestry of brazilians using autosomal AIM-Indels. **PLoS one**, v. 8, n. 9, p. e75145, 2013.

MIELKE J, ET AL. **Human biological variation**. Oxford University Press, New York, 417 p., 2006.

MOURA R, ET AL. Meta-Analysis of Brazilian Genetic Admixture and Comparison with Other Latin America Countries. **Am J Hum Biol**, v. 27, p. 674-680, 2015.

PATTERSON N, ET AL. Population Structure and Eigenanalysis. **PLoS Genet**, v. 2, n. 12, p. e190, 2006.

PENA S, ET AL. The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. **PLoS One**, v. 6, n. 2, p. e17063, 2011.

PRITCHARD J, ET AL. Association mapping in structured populations. **Am J Hum Genet**, v. 67, p. 170-181, 2000.

RIBEIRO-DOS-SANTOS Â, ET AL. Demographic and Genetic Structure of the population of Castanhal, in the Amazon Region of Brazil. **Braz J of Genet**, v. 18, n. 3, p. 469-474, 1995.

ROSENBERG N, ET AL. Informativeness of genetic markers for inference of ancestry. **Am J Hum Genet**, v. 73, p. 1402-1422, 2003.

SANTOS N, ET AL. Assessing Individual Interethnic Admixture and Population Substructure Using a 48–Insertion-Deletion (INSEL) Ancestry-Informative Marker (AIM) Panel. **Hum Mutat**, v. 31, n. 2, p. 184-190, 2010.

SANTOS S, ET AL. Mobility, blood genetic traits and race mixture in the Amazonian population of Oriximiná. **Braz J Genet**, v. 10, n. 4, p. 745-759, 1987.

SCHULER L, ET AL. Demographic and Blood Genetic Characteristics in an Amazon Population. **J Hum Evol**, v. 11, p. 549-558, 1982.

SCLIAR M, ET AL. 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**, v. 139, p. 591-595, 2009.

SIMMS T, ET AL. The genetic structure of populations from Haiti and Jamaica reflect divergent demographic histories. **Am J Phys Anthropol**, v. 142, p. 49-66, 2010.

THOMAS D, WITTE S. Population stratification: a problem for case-control studies of candidate-gene associations? **Cancer Epidemiol Biomarkers Prev**, v. 11, n. 6, p. 505-512, 2002.

APÊNDICE A - META-ANALYSIS OF BRAZILIAN GENETIC ADMIXTURE AND COMPARISON WITH OTHER LATIN AMERICA COUNTRIES

AMERICAN JOURNAL OF HUMAN BIOLOGY 27:674–680 (2015)

Original Research Article

Meta-Analysis of Brazilian Genetic Admixture and Comparison with Other Latin America Countries

RONALD RODRIGUES DE MOURA,[1,2]* ANTONIO VICTOR CAMPOS COELHO,[1,2] VALDIR DE QUEIROZ BALBINO,[1]
SERGIO CROVELLA,[1,2] AND LUCAS ANDRÈCAVALCANTI BRANDÃO[2,3]
[1]Department of Genetics, Federal University of Pernambuco (UFPE), Recife, Pernambuco, Brazil
[2]Laboratory of Immunopatology Keizo Asami (LIKA/UFPE), Recife, Pernambuco, Brazil
[3]Department of Pathology, Federal University of Pernambuco (UFPE), Recife, Pernambuco, Brazil

Objectives: This study aims at performing a systematic review and meta-analysis with the studies of genetic admixture inference of Brazilian population and to compare these results with the genetic admixture levels in other Latin American countries.

Methods: We searched for articles regarding the estimation of Brazilian genetic admixture published between 1980 and 2014 that used autosomal markers. Then, conducted meta-analyses at the whole-country and regional level. Finally, we compared the results of Brazil with other estimates from other South, Central and North American countries.

Results: We analyzed data from 25 studies in 38 different Brazilian populations. European (EUR) ancestry is the major contributor to the genetic background of Brazilians, followed by African (AFR), and Amerindian (AMR) ancestries. The pooled ancestry contributions were 0.62 EUR, 0.21 AFR, and 0.17AMR. The Southern region had a greater EUR contribution (0.77) than other regions. Individuals from the Northeast (NE) region had the highest AFR contribution (0.27) whereas individuals from the North regions had more AMR contribution (0.32). In the Latin America context, Brazil has the 5th high EUR contribution, the 12th for the AFR component and the 10th for the AMR ancestry.

Conclusions: Admixture proportions vary greatly among Brazilian populations and also through Latin America. More studies in the Center-West, North and NE regions are needed to capture a more complete picture of the genomic ancestry of Brazil. Am. J. Hum. Biol. 27:674–680, 2015.

Latin America populations exhibit varying degrees of genetic admixture due to different historical processes that have occurred since the end of the 15th century, leaving genetic traces of European (EUR), African (AFR), and Native American populations in the genomes of these individuals. Brazilian populations are not an exception to this general pattern (Salzano; Sans, 2014).

The colonization history of Brazil began in the 16th century, when the first Portuguese settlers (about a half million) started to mix with the indigenous populations (about 2.5 million) and then with AFR slaves (about 4 million) (IBGE, 2007). Moreover, after the establishment of the Republic of Brazil, in the 19th century, individuals from other nations migrated to Brazil (including Italians, Germans, and Japanese) (IBGE, 2007).

According to the last national census, Brazil has a population of about 200 million (IBGE, 2013). Genetic admixture has been directly influenced by this colonization process resulting in Brazil becoming a genetically trihybrid population. The genomic inheritance of EUR, AFR, and Amerindian (AMR) groups can be traced through the analysis of autosomal (Manta et al., 2013; Pena et al., 2011), sex chromosomes, and mitochondrial genetic information (Alves-silva et al., 2000; Palha et al., 2012).

Despite early insights about the Brazilian genetic ancestry emerging in the 1960s (Krieger et al., 1965), researchers have only extensively investigated the genetic contribution of EUR, AFR, and AMR ancestors to the genetic background of Brazil population from the 1980s onwards (Callegari-Jacques et al., 2003; Pena et al., 2009; Santos and Guerreiro, 1995, Schneider and Salzano, 1979).

Genomic admixture studies support the idea that it is not possible to use externally visible characteristics, such as hair, eye and skin color or even facial morphology, to infer the genetic ancestry of Brazilian individuals. These observations have clinical and social implications for affirmative policies implementation, case-control studies design, disease association studies and pharmacogenetics studies (Lins et al., 2011; Pena, 2005; Pena and Birchal, 2006; Suarez-kurtz et al., 2012).

To assess the ancestral proportion in Brazilian individuals, different parts of the genome have been analyzed, such as mitochondrial DNA (mtDNA) (Alves-silva et al. 2000), short tandem repeats (STR) (Callegari-Jacques et al., 2003), insertion/deletions (INDELS) (Santos et al., 2010) and single nucleotide polymorphisms (SNP) (Giolo et al., 2012). Criteria for ancestry informative markers (AIM) selection also varies among the studies, such as the absolute difference between the allele frequencies (d), F statistics and the Informativeness for assignment (In) measure (Ding et al., 2011). Different methods to select the AIMs, the number of markers adopted in the study

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: Fundação de Amparo à Pesquisa do Estado de Pernambuco (FACEPE); Contract grant number: APQ-1530-2.02/12; Contract grant sponsor: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); Contract grant number: 475350/2011-1.

*Correspondence to: Ronald Rodrigues de Moura, Department of Genetics, Federal University of Pernambuco (UFPE), Recife, Brazil.
E-mail: ronaldmoura1989@gmail.com

Received 6 May 2014; Revision received 22 January 2015; Accepted 11 February 2015

DOI: 10.1002/ajhb.22714
Published online 26 March 2015 in Wiley Online Library (wileyonlinelibrary.com).

TABLE 1. Characterization of the studies included in the meta-analysis

Year	Author	State	Region	Type of marker	N° of markers	N° individuals	EUR[a]	AFR[a]	AMR[a]
1982	Schuller et al.	Amazonas	North	Blood systems	7	595	0.3600	0.1300	0.5100
1983	Santos et al.	Amazonas	North	Blood systems	3	954	0.6100	0.1200	0.2700
1984	Rosa et al.	Amazonas	North	Blood systems	8	811	0.4300	0.1400	0.4300
1987	Santos et al.	Para ́	North	Blood systems	10	206	0.5700	0.1500	0.2800
1993	Guerreiro et al.	Para ́	North	Blood systems	11	250	0.3800	0.1000	0.5200
1995	Ribeiro-dos-Santos et al.	Para ́	North	Blood systems	13	500	0.4200	0.3300	0.2500
2005	Ferreira et al.	Maranha ̃o	Northeast	STR/VNTR	5	177	0.4200	0.1900	0.3900
2006	Ferreira et al.	Sa ̃o Paulo	Southeast	STR	8	400	0.7900	0.1400	0.0700
2009	Scliar et al.	Minas Gerais	Southeast	STR/VNTR	13	234	0.6600	0.3200	0.0200
2010	Santos et al.	Para ́	North	INDEL	48	196	0.6140	0.1170	0.2690
2010	Felix et al.	Bahia	Northeast	Alu/INDEL/ Restiction sites	7	289	0.4400	0.4900	0.0700
2010	Silva et al.	Minas Gerais	Southeast	SNP	14	24	0.5200	0.3900	0.0900
2010	Silva et al.	Minas Gerais	Southeast	SNP	14	30	0.7300	0.1900	0.0800
2011	Francez et al.	Amapa ́	North	STR	12	307	0.4600	0.1900	0.3500
2011	Martins et al.	Sa ̃o Paulo	Southeast	STR	15	403	0.7600	0.1800	0.0600
2011	Pena et al.	Para ́	North	INDEL	40	203	0.7820	0.0770	0.1410
2011	Pena et al.	Ceara ́	Northeast	INDEL	40	82	0.7580	0.1330	0.1090
2011	Pena et al.	Bahia	Northeast	INDEL	40	147	0.6680	0.2440	0.0880
2011	Pena et al.	Rio de Janeiro	Southeast	INDEL	40	264	0.8610	0.0740	0.0650
2011	Pena et al.	Rio Grande do Sul	South	INDEL	40	189	0.8600	0.0500	0.0900
2011	Lins et al.	Distrito Federal	Centre-West	SNP/INDEL	13	189	0.6290	0.2540	0.1170
2011	Leite et al.	Distrito Federal	Centre-West	SNP	21	172	0.6900	0.2100	0.1000
2012	Francez et al.	Amapa ́	North	INDEL	48	130	0.5000	0.2900	0.2100
2012	Giolo et al.	Sa ̃o Paulo	Southeast	SNP	100	138	0.6100	0.2400	0.1500
2012	Pereira et al.	Para ́	North	INDEL	46	226	0.5370	0.1680	0.2950
2012	Manta et al.	Rio de Janeiro	Southeast	INDEL	46	280	0.5520	0.3110	0.1370
2013	Cardena et al.	Sa ̃o Paulo	Southeast	INDEL	48	492	0.5740	0.2830	0.1430
2013	Manta et al.	Amazonas	North	INDEL	46	42	0.4590	0.1630	0.3780
2013	Manta et al.	Pernambuco	Northeast	INDEL	46	133	0.5680	0.2790	0.1530
2013	Manta et al.	Alagoas	Northeast	INDEL	46	104	0.5470	0.2660	0.1870
2013	Manta et al.	Mato Grosso do Sul	Centre-West	INDEL	46	84	0.5880	0.2590	0.1530
2013	Manta et al.	Minas Gerais	Southeast	INDEL	46	88	0.5920	0.2890	0.1190
2013	Manta et al.	Esp ́rito Santo	Southeast	INDEL	46	92	0.7410	0.1340	0.1250
2013	Manta et al.	Sa ̃o Paulo	Southeast	INDEL	46	49	0.6290	0.2550	0.1160
2013	Manta et al.	Parana ́	South	INDEL	46	21	0.7100	0.1750	0.1150
2013	Manta et al.	Santa Catarina	South	INDEL	46	20	0.7970	0.1140	0.0890
2013	Manta et al.	Rio Grande do Sul	South	INDEL	46	23	0.7290	0.1400	0.1300
2013	Queiroz et al.	Minas Gerais	Southeast	SNP	15	189	0.5030	0.3330	0.1640

[a]EUR, AFR, and AMR ancestries. STR: short tandem repeat; VNTR: variable number tandem repeat; INDEL: insertion/deletion; SNP: Single nucleotide polymorphism.

and different sampling strategies may contribute to divergent results among the studies (Manta et al., 2013).

In this context, we performed a systematic review with meta-analysis of the genetic admixture studies of the Brazilian population to estimate the pooled proportions of the EUR, AFR, and AMR contributions and compare these findings with the estimates of other countries in Latin America using the same approach.

MATERIAL AND METHODS

Literature search

The literature search was performed using the PubMed (www.ncbi.nlm.nih.gov/pubmed), Web of knowledge (www.webofknowledge.com), Scielo (www.scielo.org) and Google scholar (www.scholar.google.com) databases for studies published from 1980 to 2014 and also through the references cited in the articles. The key terms for the literature search were: "Brazilian Human Genetic admixture," "Brazilian Human ancestry," "Brazilian Human substructure," "Brazilian Human admixture" and "ancestry informative markers Brazilian population."

Inclusion/exclusion criteria

The inclusion criteria for the meta-analysis were (1) studies that provided the number of individuals studied per population (see the Supporting Information 2 for the Brazilian geographic distribution); (2) the use of nuclear markers to infer the genetic ancestry; and (3) explicitly information concerning the genetic proportions of the EUR, AFR, and AMR ancestries for each population (city) studied. In order to avoid possible bias, we also excluded from the meta-analysis studies focusing on completely or partially isolated populations, such as Asian and EUR colonies, indigenous tribes or AFR communities (Quilombos) remaining from the extensive period of slavery in the country (from 16th to late 18th century).

Data extraction

After excluding the articles that did not comply with the inclusion criteria, we retrieved the name of the first author, year of publication, type of genetic marker used (e.g., INDELs, SNP), number of genetic markers, geographic location of the population (e.g., Sa ̃o Paulo, Rio de Janeiro, Recife), sample size and the proportions of EUR, AFR, and AMR ancestries.

Statistical analysis

Standardized raw proportions of ancestry. The number of AIMs varied across the selected studies. To correct for any possible bias due to this variation, we developed two simple formulae that yielded raw proportions of EUR, AFR, and AMR ancestry:

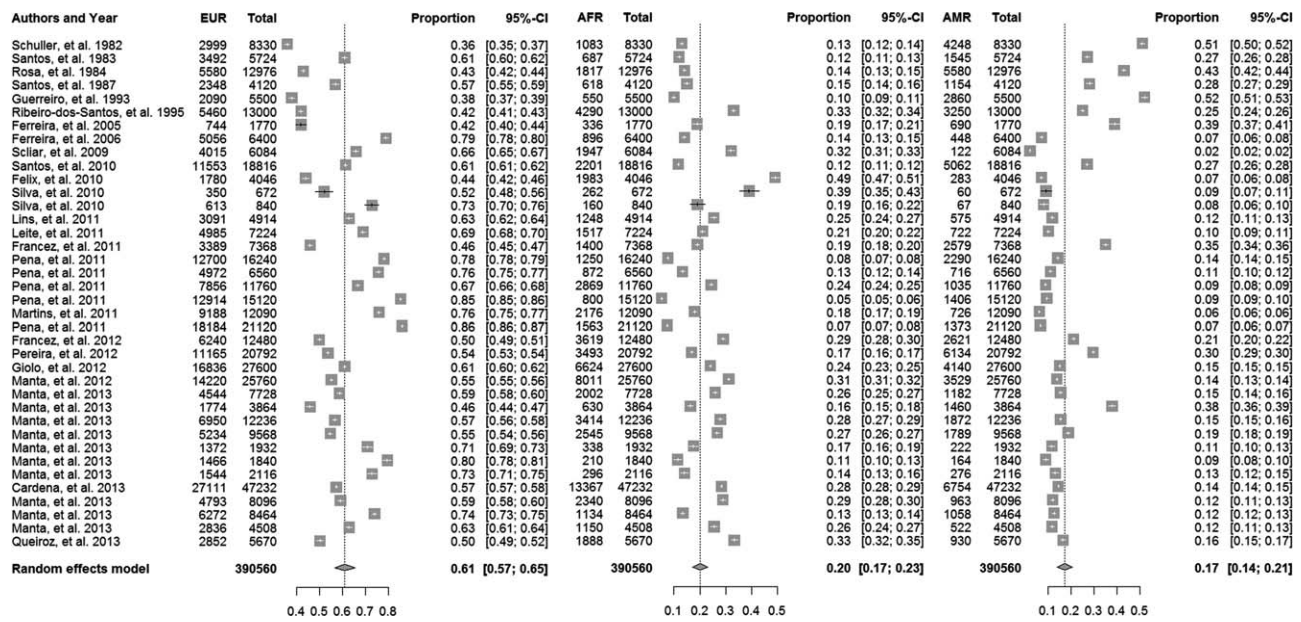


Fig. 1. Forest plot of the whole-Brazil EUR, AFR, and AMR ancestry proportions meta-analysis. The count of alleles, raw proportion, the exact confidence intervals (95% CI) and the weight (W) of each study included in the meta-analysis are shown. In the lower part of the plot, the pooled proportions with its 95% CI, and the values of the between-studies heterogeneity (s^2 and I^2) can also be seen.

$$r = a_i x b_i x c_i x d \quad (1)$$

$$n = b_i x c_i x d \quad (2)$$

In the equations above, r is the probable number of alleles of a given ancestry; n is the total number of alleles sampled; a denotes a value between 0 and 1 which correspond to the inferred ancestry proportion for the study; b, represents the sample size; c, the number of AIMs used; and d represents the number of chromosome sets of the organism (humans are diploid). A raw proportion is simply a division between r and n (r/n).

As a demonstration, in a study with 196 individuals from Pará state, using 48 INDELs as AIMs, Santos et al. (2010) found values of 0.6140, 0.1170, and 0.2690 for the contributions of EUR, AFR, and AMR ancestries, respectively. Using the Eqs. (1) and (2) for the EUR estimate, we found $r_{eur} = 0.6140 \times 196 \times 48 \times 2$, which is 11,553; $n = 196 \times 48 \times 2$, which is 18,816. These numbers can represent the number of "European alleles" (r_{eur}) among all alleles (n). The same procedures can be done to estimate the AFR and AMR contribution.

With these formulae, we were able to obtain an ancestry estimate corrected by the number of AIMs used by each authors in their studies.

Meta-analysis

With the corrected ancestry estimates, we performed a meta-analysis for single proportions using the R package "meta" (R Core Team, 2013) for each ancestry separately. This package is able to calculate an overall proportion and its variance from datasets reporting a single proportion.

The I^2 and s^2 measures were adopted to inform heterogeneity between-studies. Since high values of heterogeneity were found ($I^2 > 90\%$; $P < 0.1000$), we applied random effect model for all meta-analyses (DerSimonian and Laird, 1986). We calculated the transformed and pooled proportions using the Freeman-Tukey Double arcsine transformation (Freeman and Tukey, 1950), with its Clopper-Pearson confidence intervals, using both fixed and random effect models as well as the weight of the individual studies in the meta-analysis. All these results were displayed in the form of forest plots.

RESULTS

Characteristics of the elected studies

After a global literature search including papers published from 1980 to 2014, we selected 25 studies to be considered in our meta-analysis. In these studies, the nuclear genetic ancestry (EUR, AFR, and AMR) in 38 different populations from 17 different Brazilian states was assessed. Pará was the state with most populations studied (six) followed by São Paulo and Minas Gerais (both with five populations). Table 1 shows a summary of all studies included in the meta-analysis.

Regarding the type of AIMs used in these analyses, we subdivided the studies into four main categories: those using INDELs (Cardena et al., 2013; Kimura et al., 2013; Manta et al., 2012, 2013; Pena et al., 2011; Pereira et al., 2012; Santos et al., 2010), SNPs (Leite et al., 2011; Lins et al., 2011; Giolo et al., 2012; Queiroz et al., 2013; Silva et al., 2010), STR/VNTRs (Ferreira et al. 2005, 2006; Francez et al., 2011, 2012; Martins et al., 2011; Scliar et al., 2009) or blood group markers (Santos et al., 1983, 1987; Schuler et al., 1982; Rosa et al. 1984; Guerreiro et al. 1993; Ribeiro-dos-Santos et al., 1995). The only exception was the study of Felix et al. (2010), which employed a set of Alu elements, INDELs and restriction sites.

TABLE 2. Pooled ancestry proportions by Brazilian's geopolitical region

Region	Ancestry proportions (95% CI)		
	EUR[a]	AFR[a]	AMR[a]
North	0.51 (0.44–0.59)	0.16 (0.12–0.21)	0.32 (0.26–0.39)
Northeast	0.58 (0.48–0.66)	0.27 (0.19–0.34)	0.15 (0.10–0.21)
Centre-West	0.64 (0.58–0.69)	0.24 (0.21–0.27)	0.12 (0.09–0.16)
South	0.77 (0.69–0.85)	0.12 (0.06–0.19)	0.11 (0.09–0.12)
Southeast	0.67 (0.59–0.72)	0.23 (0.18–0.29)	0.10 (0.08–0.12)

[a]EUR, AFR, and AMR ancestries.

TABLE 3. Genetic admixture proportions of EUR, AFR, and AMR parental populations in American countries

Country	EUR	AFR	AMR	Reference
Peru	0.06	0.02	0.92	This study[a]
Ecuador	0.19	0.08	0.73	González-Andrade et al. (2007)
Mexico	0.31	0.06	0.62	This study[a]
Chile	0.42	0.02	0.56	Wang et al. (2008)
Guatemala	0.40	0.07	0.53	Wang et al. (2008)
Colombia	0.42	0.11	0.44	This study[a]
Argentina	0.54	0.03	0.42	This study[a]
Costa Rica	0.58	0.04	0.38	Ruiz-Narváez et al. (2010)
WC-USA[b]	0.56	0.08	0.36	Halder et al. (2009)
Venezuela	0.60	0.14	0.25	This study[a]
Brazil	0.62	0.21	0.17	This study[a]
EC-USA[b]	0.65	0.18	0.17	Halder et al. (2009)
Dominica	0.28	0.56	0.16	Torres et al. (2013)
Puerto Rico	0.65	0.20	0.14	This study[a]
Nicaragua	0.69	0.20	0.11	Nuñez et al. (2010)
Uruguay	0.84	0.06	0.10	Hidalgo et al. (2005)
Trinidad and Tobago	0.16	0.75	0.09	Torres et al. (2013)
Jamaica	0.10	0.82	0.08	Torres et al. (2013)
St. Lucia	0.18	0.75	0.07	Torres et al. (2013)
Grenada	0.12	0.81	0.07	Torres et al. (2013)
St. Thomas	0.17	0.77	0.06	Torres et al. (2013)
St. Vincent	0.13	0.81	0.06	Torres et al. (2013)
St. Kitts and Nevis	0.08	0.86	0.06	Torres et al. (2013)
AA-USA[b]	0.16	0.81	0.04	Halder et al. (2009)
EA-USA[b]	0.98	0.01	0.01	Halder et al. (2009)
Cuba	0.73	0.26	0.01	Diaz-Horta et al. (2010)
Bahamas	0.04	0.96	0.00	Simms et al. (2010)
Haiti	0.04	0.96	0.00	Simms et al. (2010)

The data are listed in descendent order of AMR contribution.
[a]With the exception of the references used in the Brazilian meta-analyses, the complete references for this table can be found in the Supporting Information 3.
[b]AA-USA stands for AFR American from United States of America (USA); EA is European American; EC is East Coast Hispanics; WC is West Coast Hispanics.

The first group employed sets of INDELs ranging from 40 to 48 mostly based on a larger dataset (Weber et al., 2002), sometimes sharing selected markers (Pena et al., 2011; Pereira et al., 2012; Santos et al., 2010) or the whole set in other cases (Manta et al., 2012, 2013; Pereira et al., 2012). The SNP group chose the AIMs from more sources (Bonilla et al., 2004; Fernández et al., 2003; Packer et al., 2006; Smith et al., 2004; Tian et al., 2006). They also share selected markers (Leite et al., 2011; Lins et al., 2011) or the whole set (Leite et al., 2011; Queiroz et al., 2013). Although the studies shared some variants in their sets, the source of the AIMs in the STR/VNTR group were not detailed with the exception of the study of Martins et al. (2011), which used a commercial kit. Finally, the blood group markers used in the six studies are detailed in the paper from Guerreiro et al. (1993).

The INDEL group elaborated its datasets using d and F_{ST} measures. The SNP group was more divergent regarding the criteria of selection using combinations of d, F_{ST}, In, and even principal component analysis (Giolo et al., 2012). The authors who used STR/VNTR and blood group markers did not specify the criteria used to select their AIMs.

Some studies used autosomal markers, but analyzed the data according to region and not according to state (Callegari-Jacques et al., 2003; Godinho et al., 2008; Lins et al., 2010) or they evaluated the ancestral proportions in semi or completely isolated populations (Callegari-jacques and Salzano, 1999; Maciel et al., 2011; Salzano et al., 1997). These studies were not included in the meta-analysis, although they served for comparison with our results.

Brazilian genetic admixture

We conducted the meta-analysis including the 25 studies that aggregated 3,90,560 alleles in 8,733 Brazilian individuals. The pooled proportions of European, AFR, and AMR ancestry with their respective confidence intervals are summarized in the forest plots (Fig. 1). The major contribution came from Europeans (0.61), followed by AFRs (0.20) and AMRs (0.17). It should be noted that the sum of the proportions is not equal to 1 (indeed, it is 0.98). This may happen because of rounding processes that occurred during the calculations. To circumvent this problem, we simply standardized the values dividing each pooled contribution by the sum of them. Therefore, for the EUR contribution we have 0.62 (i.e., 0.61/0.98), for AFR 0.21 (0.20/0.98) and 0.17 (0.17/0.98) for AMR ancestry.

We performed separated tests for the five Brazilian's geopolitical regions: North (N), NE, Centre-West (CW), South (S), Southeast (SE). The number of populations (and the total number of individuals in parenthesis) per region was as follows: N = 12 (4,420), NE = 6 (932), CW = 3 (445), S = 4 (253), and SE = 13 (2,683). Table 2 summarizes the pooled proportions of the EUR, AFR, and AMR ancestries per region.

The Southern region of Brazil had a greater EUR contribution (0.77) than other regions. The NE with 0.27 and the North with 0.32 were the regions with greater AFR and AMR contributions, respectively. The complete analysis of Brazilian genomic ancestry according to geographic region is described in the Supporting Information 1.

Admixture proportions in Latin America countries

Recently, Salzano and Sans (2014) published a review which discussed the genetic admixture in Latin American populations, although they did not explore the data using a meta-analytical approach. Therefore, based on their article, we expanded our analysis in order to verify the proportions of parental populations (AFR, AMR, and EUR) in other countries from Latin America using meta-analyses of genetic admixture studies carried out in populations from Argentina (20 populations), Colombia (25), Mexico (23), Peru (25), Puerto Rico (8), and Venezuela (5) using the same inclusion criteria applied for Brazilian meta-analysis. These results are summarized in Table 3. In this table, we also listed admixture proportions for other American countries' populations (including USA), which did not have sufficient published data to allow meta-analysis to better represent genetic admixture in Latin America.

After conducting the meta-analyses, the Mexican population has a pooled 0.31, 0.06, and 0.62 of EUR, AFR, and

AMR ancestry proportions, respectively. In Central America, the Nicaraguan population had the highest EUR and AFR contributions (0.69 and 0.20, respectively), whilst the highest AMR contribution was described in Guatemala (0.53). Among the South American countries, Peruvians showed 0.92 of AMR contribution, whereas Brazilians had 0.21 AFR contribution. Uruguayans showed the highest EUR contribution (0.84).

Among Caribbean Islands, Haiti, and the Bahamas had almost total AFR contributions (0.96), while Cuba had 0.73 of EUR ancestry and Dominica 0.16 AMR ancestry.

DISCUSSION

In the last two decades, genetic admixture in Brazilian populations has been a matter of concerted investigation. These investigations attempted to delineate the composition of Brazilians' genetic background in uniparental (Alves-Silva et al., 2000; Carvalho-silva et al., 2001) and biparental contexts (Santos and Guerreiro, 1995; Schneider and Salzano, 1979; Salzano et al., 1997; Callegari-jacques et al., 2003; Parra et al., 2003).

Although several studies have demonstrated the genetic admixture in various populations through all Brazil and other Latin America countries, there is no systematic review compiling these studies in order to establish overall results based on available data. This was the main objective of the present work.

Throughout our search, we found a considerable number of articles that used only uniparental markers, such as Marrero et al. (2005), Guerreiro-Junior et al. (2009), and Bernardo et al. (2014) who used mtDNA, and Silva et al. (2006) and Carvalho et al. (2010) who used Y-chromosome markers. However, the majority of data came from Southeastern and Southern regions, which will bias the meta-analysis results. Therefore, we decided to concentrate on only autosomal markers for this article.

Considering all Brazilian populations studied, we found that the pooled EUR, AFR and AMR ancestry proportions were 0.62, 0.21, and 0.17, respectively. These results are in agreement with other findings describing the highest contribution of EURs, followed by AFRs and AMRs, to Brazilians (Godinho et al., 2008; Lins et al., 2010). We observed high values of between-study heterogeneity. This heterogeneity might be due to the differences in sample sizes and different number and sets of markers used. For example, Manta and et al. (2013) and Scliar et al. (2009), using INDELs and microsatellite data, respectively, found different ancestry proportions for Brazilians (see Table 1). We can also hypothesize that different contexts of admixture processes occurred in the study populations as well as social and cultural influences that may account for this heterogeneity.

In the Latin American context, Brazil has the 5th highest EUR genetic ancestry (0.62) after Uruguay (0.84), Cuba (0.73), Nicaragua (0.69) and Puerto Rico (0.65). In general, EUR ancestors made a larger genetic contribution in the Atlantic side of the American continent, whereas the AMR contribution occurred predominantly at the Pacific side. Peru, Ecuador and Mexico are the three countries with the highest American Native contributions (0.92, 0.73 and 0.62, respectively). These countries were formerly the core of Inca (Ecuador and Peru) and Aztec (Mexico) Empires. Despite the massive depopulation due to epidemics, exploitation and war during the Spanish invasion, natives from these populations were more involved in the admixture process than other indigenous populations (Salzano and Callegari-Jacques, 1988). Brazil is the 10th population with respect to AMR proportion (0.17).

Although Brazil has the highest AFR contribution among South American countries (0.21), it is the 12th for AFR contribution if we consider Latin America as a whole, since Caribbean countries have the highest estimates for AFR ancestry, such as Bahamas and Haiti with 96% of AFR contribution. During the colonization of the Caribbean Islands by British, France, Spanish and other EUR countries, almost all indigenous people were killed or deported, requiring them to import a large number of AFR Slaves to work on sugar-cane and coffee plantations (Knight, 1997). These slaves played a vital role in the historical and genetic composition of Caribbean countries. In Haiti, for example, after the Haitian revolution in the early 19th century, much of the French colonizers left the island (Pamphile, 2001).

The same scenario evidenced at national scale has also been observed at regional scale in Brazil. The EUR contribution is the highest in all five regions of Brazil. With the exception of Lins et al. (2010), who found more AMR contribution in Central-West than in North Brazil, our results agree with other authors who reported an increased value for the AMR ancestry in the North in comparison with the other regions and an higher EUR contribution in the South (Callegari-jacques et al., 2003; Godinho et al., 2008; Manta et al., 2013; Pena et al., 2011).

Although the Southern region of Brazil has the lowest AFR proportion (0.12), this value is high compared to the bordering countries such as Argentina (0.03) and Uruguay (0.06). On the other hand, in the Northern region the AMR proportion is the highest (0.32) in the country, also higher when compared to Venezuela (0.25) but lower if compared to Colombia (0.44) and Peru (0.92).

The regional genomic distribution found in Brazilian is linked with the different colonization history of each region. For example, the South region received successive migration cycles of EURs during the 16th (Portuguese) and 19th (Germans and Italians) centuries. Moreover, the majority of AFR slaves arrived and settled in Brazil's NE and SE before moving to other Brazilian regions (Conrad, 1973; Levy, 1974; IBGE, 2007). These observations help to explain why, according to our meta-analysis, the South region has 77% of EUR genomic ancestry, whereas NE has 58%. The opposite can be seen when we look at the AFR contribution: 12% in the South and 27% in NE.

Using F-statistics, two studies reported significant genetic distance between populations from different regions from Brazil. Lins et al. (2010) obtained significant values when comparing the South region population with those from other regions. In the other study, Manta et al. (2013) also verified divergence between urban populations from South and populations from North, NE and CW regions. It is important to bear in mind that, although these authors reported significant distance between Brazilian populations, the magnitude of these differences did not reach more than 10%, which represents a low to moderate genetic divergence according to Wright's qualitative guidelines (Wright, 1978). If compared with proxies of parental populations (e.g. CEU and YRI populations from HapMap Project), the F_{ST} can reach values of 44%

depending on the proportion of a given ancestry present in the Brazilian population that is been compared (Manta et al., 2013).

Some articles not included in our meta-analysis evaluated admixture in AFR (Quilombos) and AMR communities. These studies found AFR contributions ranging from 0.32 to 0.92 (Kimura et al., 2013; Maciel et al., 2011; Scliar et al., 2009) and AMR contributions ranging from 0.25 to 0.97 (Callegari-jacques and Salzano, 1999; Manta et al., 2013; Salzano et al., 1997). These values are superior to the pooled AFR and AMR contribution of our meta-analysis for the whole of Brazil (0.21 and 0.17, respectively).

Some of the studies included in our meta-analysis investigated the correlation between self-reported skin color and genomic ancestry in Brazilian population (Leite et al., 2011; Lins et al., 2011; Pena et al., 2011; Queiroz et al., 2013). From those papers, Leite et al. (2011) used quantitative measures to evaluate the associations between melanin index, self-reported skin color and genomic ancestry among pairs of siblings.

Although they found statistical differences between self-reported skin color and melanin index, there was considerable overlap between groups. Furthermore, the correlations between self-reported skin color versus genomic ancestry and melanin index versus genomic ancestry also had considerable overlap between skin color categories and continental ancestry. These results are in agreement with other studies in Latin America that pointed to similar findings (Parra et al., 2004; Ruiz-Linares et al., 2014).

Moreover, self-perception of ancestry is biased not only by skin color but also by other phenotypic traits, such as iris and hair color, which tend to overestimate EUR ancestry, and hair type and some facial characteristics that may overestimate AFR ancestry. Apart from phenotypic traits, socioeconomic factors also contribute to this self-perception, with wealth and education also tending to overestimate EUR ancestry (Ruiz-Linares et al., 2014).

Considering the geographic distribution of the study populations (Supporting Information 2), 10 states had no information about the nuclear genetic ancestry of their populations; most of them are from the North and NE regions. Moreover, only three studies were conducted in populations from CW. Therefore, further studies are required to better elucidate and correctly describe the genomic ancestry of CW, North and NE Brazilian regions. It is well worth noting that the study of the Brazilian ancestry is also making an important contribution to clinical research, providing fundamental information about the development of genetic association studies searching for disease-related markers or clinical epidemiological analyses (Lins et al., 2011; Pinto et al., 2012; Suarez-kurtz et al., 2012): Moreover, ancestry studies are subsidizing others issues, such as sociological debates (Pena, 2005; Pena and Birchal, 2006).

CONCLUSION

In the present work, we used a meta-analytic approach to compile various studies on the genetic ancestry of Brazilian populations, based on nuclear markers, and compared the results with the admixture data for other countries in Latin America. We concluded that the pooled proportions of EUR, AFR, and AMR ancestries in Brazil (globally considered) are 0.62, 0.21, and 0.17, respectively. These values, mainly for AFR and AMR contributions, are intermediate when compared to other Latin American countries. At the regional level, as expected, the highest AMR contribution occurred in the Northern region, the highest AFR contribution in the Northeastern region and the highest EUR contribution is in the Southern region. More studies in the CW, North and NE regions are needed to capture the whole landscape of the genomic ancestry of Brazil.

LITERATURE CITED

Alves-silva J, Santos MS, Guimarães PEM, Ferreira ACS, Bandelt HJ, Pena SDJ, Prado VF. 2000. The ancestry of Brazilian mtDNA lineages. Am J Hum Genet 67:444–461.

Bernardo S, Hermida R, Desidério M, Silva DA, Carvalho EF. 2014. mtDNA ancestry of Rio de Janeiro population, Brazil. Mol Biol Rep 41: 1945–1950.

Bonilla C, Shriver MD, Parra EJ, Jones A, Fernández JR. 2004. Ancestral proportions and their association with skin pigmentation and bone mineral density in Puerto Rican women from New York City. Hum Genet 115:57–68.

Callegari-jacques SM, Grattapaglia D, Salzano FM, Salamoni SP, Grossetti SG, Ferreira ME, Hutz MH. 2003. Historical genetics: spatio-temporal analysis of the formation of the Brazilian population. Am J Hum Biol 15:824–834.

Callegari-jacques SM, Salzano FM. 1999. Brazilian Indian/non- Indian interactions and their effects. Cienc Cult 51:166–174.

Cardena MMSG, Ribeiro-dos-Santos AKC, Santos S, Mansur AJ, Pereira AC, Fridman C. 2013. Assessment of the relationship between self-declared ethnicity, mitochondrial haplogroups and genomic ancestry in Brazilian individuals. PloS One 8:e62005.

Carvalho M, Brito P, Lopes V, Andrade L, Anjos MJ, Real FC, Gusmão L. 2010. Analysis of paternal lineages in Brazilian and African populations. Genet Mol Biol 33:422–427.

Carvalho-silva DR, Santos FR, Rocha J, Pena SDJ. 2001. The Phylogeography of Brazilian Y-Chromosome Lineages. Am J Hum Genet 68:281–286.

Conrad R. 1973. The destruction of Brazilian slavery 1850-1888. Oakland, CA: University of California. p 368.

DerSimonian R, Laird N. 1986. Meta-analysis in clinical trials. Contr Clin Trials 7:177–188.

Ding L, Wiener H, Abebe T, Altaye M, Go RCP, Kercsmar C, Grabowski G, Martin LJ, Hershey GKK, Chakorborty R, Baye TM. 2011. Comparison of measures of marker informativeness for ancestry and admixture mapping. BMC Genomics 12:622–640.

Felix GES, Abe-sandes K, Bonfim TM, Bendicho MT, Guedes R, Brandão CJF, Torres AJL, Brites C, Netto EM, Meyer R, Freire SM. 2010. Ancestry informative markers and complete blood count parameters in Brazilian blood donors. Rev Bras Hematol Hemoter 32:282–285.

Fernández JR, Shriver MD, Beasley TM, Rafla-demetrious N, Parra E, Albu J, Nicklas B, Ryan AS, McKeigue PM, Hoggart CL, Weinster RL, Allison DB. 2003. Association of African genetic admixture with resting metabolic rate and obesity among women. Obesity Res 11:904–911.

Ferreira FL, Leal-mesquita ER, Santos SEB, Ribeiro-dos-Santos AKC. 2005. Genetic characterization of the population of São Luís, MA, Brazil. Genet Mol Biol 28:22–31.

Ferreira LB, Mendes-junior CT, Wiezel CEV, Luizon MR, Simões AL. 2006. Genomic Ancestry of a Sample Population from the State of São Paulo, Brazil. Am J Hum Biol 18:702–705.

Francez PAC, Ribeiro-Rodrigues EM, Frazão GF, Borges NDR, Santos SEB. 2011. Allelic frequencies and statistical data obtained from 12 codis STR loci in an admixed population of the Brazilian Amazon. Genet Mol Biol 34:35–39.

Francez PAC, Ribeiro-Rodrigues EM, Santos SEB. 2012. Allelic frequencies and statistical data obtained from 48 AIM INDEL loci in an admixed population from the Brazilian Amazon. Forensic Sci Int 6:132–135.

Freeman MF, Tukey JW. 1950. Transformations Related to the Angular and the Square Root. Ann Math Stat 21:607–611.

Giolo SR, Soler JMP, Greenway SC, Almeida MAA, Andrade M, Seidman JG, Seidman CE, Krieger JE, Pereira AC. 2012. Brazilian urban population genetic structure reveals a high degree of admixture. Eur J Hum Genet 20:111–116.

Godinho NMO, Contijo CC, Diniz MECG, Falcão-Alencar G, Dalton GC, Amorim CEG, Barcelos RSS, Klautau-Guimarães MN, Oliveira, SF. 2008. Regional patterns of genetic admixture in South America. Forensic Sci Int 1:329–330.

Gravlee CC, Dressler WW, Bernard HR. 2005. Skin color, social classification, and blood pressure in southeastern Puerto Rico. Am J Public Health 95:2191–2197.

Guerreiro JF, Ribeiro-dos-Santos AK, Santos EJM, Cayres IMV, Santos SEB. 1993. Genetic Structure and Demography of the Human Population of Óbidos, in the Brazilian Amazon. Braz J Genet 16:1075–1084.

Guerreiro-Junior V, Bisso-Machado R, Marrero A, Hünemeier T, Salzano FM, Bortolini MC. 2009. Genetic signatures of parental contribution in black and white populations in Brazil. Genet Mol Biol 32:1–11.
Knight FW. 1997. General history of the Caribbean—volume III: the slave societies of the Caribbean. London, UK: UNESCO Publishing. p 380.
Instituto Brasileiro de Estatística (IBGE). 2007. Brasil: 500 anos de povoamento. Rio de Janeiro: IBGE. p 237.
Instituto Brasileiro de Estatística (IBGE). 2013. Atlas do Censo Demográfico 2010. Rio de Janeiro: IBGE. p 216.
Kimura L, Ribeiro-rodrigues EM, Auricchio MTBM, Vicente JP, Santos SEB, Mingroni-Netto RC. 2013. Genomic Ancestry of Rural African-Derived Populations from Southeastern Brazil. Am J Hum Biol 25:35–41.
Krieger H, Morton NE, Azevedo E, Freire-Maia N, Yasuda N. 1965. Racial admixture in Northeastern Brazil. Ann Hum Genet 29:113–125.
Leite TKM, Fonseca RMC, França NM, Parra EJ, Pereira RW. 2011. Genomic Ancestry, Self-Reported "Color" and Quantitative Measures of Skin Pigmentation in Brazilian Admixed Siblings. PloS One 6:e27162.
Levy MSF. 1974. The role of international migration on the evolution of the Brazilian population (1872 to 1972). Rev saude publ 8:49–90.
Lins TC, Vieira RG, Abreu BS, Gentil P, Moreno-lima R, Oliveira RJ, Pereira RW. 2011. Genetic Heterogeneity of Self-Reported Ancestry Groups in an Admixed Brazilian Population. J Epidemiol 21:240–245.
Lins TC, Vieira RG, Abreu BS, Grattapaglia D, Pereira RW. 2010. Genetic Composition of Brazilian Population Samples Based on a Set of Twenty Eight Ancestry Informative SNPs. Am J Hum Biol 22:187–192.
Lins TC, Vieira RG, Grattapaglia D, Pereira RW. 2011. Population analysis of vitamin D receptor polymorphisms and the role of genetic ancestry in an admixed population. Genet Mol Biol 34:377–385.
Maciel LGL, Ribeiro-rodrigues EM, Santos NPC, Ribeiro-dos-Santos AKC, Guerreiro JF, Santos S. 2011. Afro-Derived Amazonian Populations: inferring Continental Ancestry and Population Substructure. Hum Biol 83:627–636.
Manta FSN, Pereira R, Vianna R, Araújo ARB, Gitaí GDL, Silva DA, Wolfgramm EV, Pontes IMP, Aguiar JI, Moraes MO, Carvalho EF, Gusmão L. 2013. Revisiting the genetic ancestry of Brazilians using autossomal AIM-INDELs. PloS One 8:e75145.
Manta FSN, Pereira R, Caiafa A, Silva DA, Gusmão L, Carvalho EF. 2012. Analysis of genetic ancestry in the admixed Brazilian population from Rio de Janeiro using 46 autosomal ancestry-informative indel markers. Ann Hum Biol 40:1–5.
Marrero AR, Leite FP, Carvalho BA, Peres LM, Kommers TC, Cruz IM, Bortolini MC. 2005. Heterogeneity of the Genome Ancestry of Individuals Classified as White in the State of Rio Grande do Sul, Brazil. Am J Hum Biol 17:496–506.
Martins JA, Figueiredo RF, Yoshizaki CS, Paneto GG, Cicarelli RMB. 2011. Genetic data of 15 autosomal STR loci: an analysis of the Araraquara population colonization (São Paulo, Brazil). Mol Biol Rep 38: 5397–5403.
Packer BR, Yeager M, Burdett L, Welch R, Beerman M, Qi L, Sicotte H, Staats B, Acharya M, Crenshaw A, Eckert A, Puri V, Gerhard DS, Chanock SJ. 2006. SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes. Nucleic Acids Res 34:D617–D621.
Palha T, Ribeiro-rodrigues E, Ribeiro-dos-Santos AKC, Santos S. 2012. Fourteen short tandem repeat loci Y chromosome haplotypes: genetic analysis in populations from northern Brazil. Forensic Sci Int 6:413–418.
Pamphile LD. 2001. Haitians and African Americans. Gainseville, FL: University Press of Florida.
Parra FC, Amado RC, Lambertucci R, Rocha J, Antunes CM, Pena SDJ. 2003. Color and genomic ancestry in Brazilians. PNAS 100:177–182.
Parra EJ, Kittles RA, Shriver MD. 2004. Implications of correlations between skin color and genetic ancestry for biomedical research. Nat Genet 36:S54–S60.
Pena SDJ. 2005. Razões para banir o conceito de raça da medicina brasileira. Hist Cienc Saude 12:319–346.
Pena SDJ, Bastos-rodrigues L, Pimenta JR, Bydlowski SP. 2009. DNA tests probe the genomic ancestry of Brazilians. Braz J Med Biol Res 42: 870–876.
Pena SDJ, Birchal TS. 2006. A inexistência biológica vesus a existência social de raças humanas: pode a ciência instruir o etos social? Rev USP 68:10–21.
Pena SDJ, Pietro GD, Fuchshuber-moraes M, Genro JP, Hutz MH, Kehdy FSG, Kohlrausch F, Magno LAV, Montenegro RC, Moraes MO, Moraes MEA, Moraes MR, Ojopi EB, Perini JA, Racciopi C, Ribeiro-dos-Santos AKC, Rios-Santos F, Romano-Silva MA, Sortica VA, Suarez-kurtz, G. 2011. The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. PloS One 6:e17063.
Pereira R, Phillips C, Santos C, Santos SEB, Carracedo A, Gusmão L. 2012. Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. PloS One 7:e29684.
Pinto P, Salgado CG, Santos N, Alencar DO, Santos S, Hutz MH, Ribeiro-dos-Santos AKC. 2012. Polymorphisms in the CYP2E1 and GSTM1 genes as possible protection factors for leprosy patients. PloS One 7: e47498.
Queiroz EM, Santos AM, Castro IM, Machado-Coelho GL, Cândido APC, Leite TM, Pereira RW, Freitas RN. 2013. Genetic composition of a Brazilian population: the footprint of the gold cycle. Genet Mol Res 12: 5124–5133.
R Core Team. 2013. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/
Ribeiro-dos-Santos ÂKC, Santos EJM, Guerreiro JF, Santos SEB. 1995. Demographic and genetic structure of the population of Castanhal, in the Amazon region of Brazil. Braz J Genet 18:469–474.
Rosa VL, Salzano FM, Franco MH, Freitas MJ. 1984. Blood genetic studies in five Amazonian populations. Braz J Genet 7:569–582.
Ruiz-Linares A, Adhikari K, Acuña-Alonzo V, Quinto-Sanchez M, Jaramillo C, Arias W, Gonzalez-José R. 2014. Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals. PLoS Genet 10: e1004572.
Salzano FM and Callegari-Jacques SM. 1988. South American Indians: a case study in evolution. Oxford, UK: Clarendon Press. p 259.
Salzano FM, Callegari-jacques SM, Weimer TA, Franco MHLP, Hutz MH, Petzl-erler ML. 1997. Electrophoretic protein polymorphisms in Kaingang and Guarani Indians of Southern Brazil. Am J Hum Biol 9:502–512.
Salzano FM, Sans M. 2014. Interethnic admixture and the evolution of Latin American populations. Genet Mol Biol 37:151–170.
Santos SEB, Salzano FM, Helena M, Franco LP, Freitas MJM. 1983. Mobility, genetic markers, susceptibility to malaria and race mixture in Manaus, Brazil. J Hum Evol 12:373–381.
Santos S, Guerreiro JF, Salzano FM, Weimer TA, Hutz MH, Franco MHLP. 1987. Mobility, blood genetic traits and race mixture in the Amazonian population of Oriximiná Braz J Genet 10:745–759.
Santos S, Guerreiro J. 1995. The indigenous contribution to the formation of the population of the Brazilian Amazon Region. Rev Bras Genet 18: 311–315.
Santos NPC, Ribeiro-rodrigues EM, Ribeiro-dos-Santos AKC, Pereira R, Gusmão L, Amorim A, Guerreiro JF, Zago MA, Matte C, Hutz MH, Santos SEB. 2010. Assessing individual interethnic admixture and population substructure using a 48–insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. Hum Mutat 31:184–190.
Schneider H, Salzano FM. 1979. Gm allotypes and racial admixture in two Brazilian populations. Hum Genet 53:101–105.
Schüler L, Salzano FM, Franco MHLP, Freitas MJM, Mestriner MA, Simões AL. 1982. Demographic and blood genetic characteristics in an Amazon population. J Hum Evol 11:549–558.
Scliar MO, Vaintraub MT, Vaintraub PMV, Fonseca CG. 2009. 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 139:591–595.
Silva DA, Carvalho EF, Costa G, Tavares L, Amorim A, Gusmão L. 2006. Y-chromosome genetic variation in Rio de Janeiro population. Am J Hum Biol 18:829–837.
Silva MCF, Zuccherato LW, Soares-Souza GB, Vieira ZM, Cabrera L, Herrera P, Balqui J, Romero C, Jahuira H, Gilman RH, Martins ML, Tarazona-santos E. 2010. Development of two multiplex mini-sequencing panels of ancestry informative SNPs for studies in Latin Americans: an application to populations of the State of Minas Gerais (Brazil). Genet Mol Res 9:2069–2085.
Smith MW, Patterson N, Lautenberger JA, Truelove AL, Mcdonald GJ, Waliszewska A, Kessing BD, Malasky MJ, Scafe C, Le E, De Jager PL, Mignault AA, Yi Z, The G, Essex M, Sankale J, Moore JH, Poku K, Phair JP, Goedert JJ, Vlahov D, Williams SM, Tishkoff SA, Winkler CA, De la Vega FM, Woodgate T, Sninsky JJ, Hafler DA, Altshuler D, Gilbert DA, O'Brien SJ, Reich, D. 2004. A high-density admixture map for disease gene discovery in African Americans. Am J Hum Genet 74:1001–1013.
Suarez-kurtz G, Pena SDJ, Struchiner CJ, Hutz MH. 2012. Pharmacogenomic diversity among Brazilians: influence of ancestry, self-reported color, and geographical origin. Front Pharmaco 3:1–7.
Tian C, Hinds DA, Shigeta R, Kittles R, Ballinger DG, Seldin MF. 2006. Genome-wide single-nucleotide-polymorphism panel with high ancestry information for African American admixture mapping. Am J Hum Genet 79:640–649.
Weber JL, David D, Heil J, Fan Y, Zhao C, Marth G. 2002. Human diallelic insertion/deletion polymorphisms. Am J Hum Genet 71:854–862.
Wright S. 1978. Variability within and among natural populations. Chicago, IL: University of Chicago, v. 4, 590 p.

APÊNDICE B - A RAPID SCREENING OF ANCESTRY FOR GENETIC ASSOCIATION STUDIES IN AN ADMIXED POPULATION FROM PERNAMBUCO, BRAZIL

GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

A rapid screening of ancestry for genetic association studies in an admixed population from Pernambuco, Brazil

A.V.C. Coelho[1,2], R.R. Moura[1,2], C.A.J. Cavalcanti[1,2], R.L. Guimarães[1,2], P. Sandrin-Garcia[1,2], S. Crovella[3] and L.A.C. Brandão[2,4]

[1]Departamento de Genética, Universidade Federal de Pernambuco, Recife, PE, Brasil
[2]Laboratório de Imunopatologia Keizo Asami, Universidade Federal de Pernambuco, Recife, PE, Brasil
[3]IRCCS Burlo Garofolo, Trieste, Itália
[4]Departamento de Patologia, Universidade Federal de Pernambuco, Recife, PE, Brasil

Corresponding author: R.R. Moura
E-mail: ronaldmoura1989@gmail.com

Genet. Mol. Res. 14 (1): 2876-2884 (2015)
Received March 21, 2014
Accepted November 13, 2014
Published March 31, 2015
DOI http://dx.doi.org/10.4238/2015.March.31.18

ABSTRACT. Genetic association studies determine how genes influence traits. However, non-detected population substructure may bias the analysis, resulting in spurious results. One method to detect substructure is to genotype ancestry informative markers (AIMs) besides the candidate variants, quantifying how much ancestral populations contribute to the samples' genetic background. The present study aimed to use a minimum quantity of markers, while retaining full potential to estimate ancestries. We tested the feasibility of a subset of the 12 most informative markers from a previously established study to estimate influence from three ancestral populations: European, African and Amerindian. The results showed that in a sample with a diverse ethnicity (N = 822) derived

from 1000 Genomes database, the 12 AIMs had the same capacity to estimate ancestries when compared to the original set of 128 AIMs, since estimates from the two panels were closely correlated. Thus, these 12 SNPs were used to estimate ancestry in a new sample (N = 192) from an admixed population in Recife, Northeast Brazil. The ancestry estimates from Recife subjects were in accordance with previous studies, showing that Northeastern Brazilian populations show great influence from European ancestry (59.7%), followed by African (23.0%) and Amerindian (17.3%) ancestries. Ethnicity self-classification according to skin-color was confirmed to be a poor indicator of population substructure in Brazilians, since ancestry estimates overlapped between classifications. Thus, our streamlined panel of 12 markers may substitute panels with more markers, while retaining the capacity to control for population substructure and admixture, thereby reducing sample processing time.

Key words: Brazilian genetic admixture; Population structure; Ethnicity; Ancestry informative markers; SNP; Association studies

INTRODUCTION

Genetic association studies (GAS) are conducted to determine which genetic factors underlie susceptibility to complex diseases. Single nucleotide polymorphisms (SNPs) are the main genetic variations analyzed in GAS. In fact, SNPs have a great part in inter-individual differences (Sachidanandam et al., 2001). Moreover, SNP genotyping can be performed easily and inexpensively in a wide range of high-throughput technologies.

Generally, two groups of unrelated subjects are genotyped in GAS: one carrying a determined trait, such as a disease (cases), and another not carrying it (controls). Thus, if one of the alleles is more frequent in one group than the other, it is associated with the presence/absence of the trait, meaning that its presence may contribute to risk/protection in relation to the disease (Lewis, 2002).

A positive association may represent three situations: 1) the allele has indeed a causal role in the trait; 2) the allele is not causal, but is in linkage disequilibrium with the true causal polymorphism or 3) the observed association could be a spurious one (Cordell and Clayton, 2005). The last circumstance could be interpreted as a false-positive result.

A false-positive result, due to spurious association, may arise from undetected population substructure (stratification), For example, in an admixed population, individuals heterogeneous for their genetic backgrounds result from unequal genetic contribution from ancestral populations. Thus, during sampling, a subpopulation may be overrepresented in the cases and/or in the controls. This introduces bias on allele frequencies, which could result in a false positive during statistical analysis (Balding, 2006).

Population substructure thus is regarded as a confounding factor, and as a consequence, some methods were proposed to control for this issue. They all have in common the requirement for genotyping of unlinked polymorphisms not associated with the trait of interest in all sampled subjects (both cases and controls) and have been defined as “null SNPs” (Balding, 2006).

Some authors have proposed SNP panels for use in this context. Kosoy et al. (2009),

for instance, applied the "informativeness" algorithm developed by Rosenberg et al. (2003) to select 128 SNPs, designated ancestry informative markers (AIMs). They were used to infer European, African and Amerindian ancestry from ethnically diverse populations from United States (US) cities. It is important to clarify that we understand the term "ethnicity" as the social identity of a group or population with few or no relationship with its actual genetic background (Ali-Khan et al., 2011).

Brazilian populations were also founded by admixture from these three ancestral populations. This resulted in a peculiar population composition, where self-reported ancestry (based on skin color) does not correlate or reflect real ancestry, thus being a poor predictor of population substructure, when compared to more "homogeneous" populations in other countries (Pena et al., 2009).

Thus, our research group intended to explore these AIMs to estimate ancestry proportions in populations from the Recife metropolitan region, to avoid spurious results in GAS. Recife is the capital of Pernambuco State, which is located in Northeast Brazil, a region that received influence from three ancestral populations: native Amerindian, European (16th century settlers) and African (trafficked slaves), contributing to origin of the population of present times through admixture (Ribeiro, 1995; Parra et al., 2003).

However, genotyping of 128 AIMs would be costly and labor-intensive when analyzing a substantial number (usually more than 100 individuals) of cases and controls. Therefore, in trying to meet the necessity of ancestry identification encountered by our research group in the daily activity of simple association studies, we proposed a simple panel of AIMs, suitable for low-/medium-throughput laboratories with simple genotyping technologies and small financial resources.

Since Kosoy et al. (2009) claimed that the use of smaller (48 or 24 AIMs) subsets resulted in consistent estimates when compared to the larger datasets of the original 128 markers, we thus decided to examine the informativeness of these AIMs after further streamlining this panel, and using it to estimate the genetic ancestry and control for substructure in our admixed population during GAS.

MATERIAL AND METHODS

Ethics statement

The Research Ethics Committee of the Center of Health Sciences, Federal University of Pernambuco approved the study (protocol No. 257.941). Each subject gave written consent.

Ancestry informative marker subset selection, genotyping and sample data

Kosoy et al. (2009) selected 128 AIMs through the genome using allele frequency differences between populations (>45% difference between European-derived and African-derived populations and between European-derived and Amerindian-derived populations) and informativeness criteria (the ability to estimate the proportion of individual genetic ancestries) as elaborated by Rosenberg et al. (2003).

To streamline the number of AIMs, we selected the twelve most informative markers for substructure assignment of the three subpopulations (top 12) described by Kosoy et al. (2009). Briefly, the 12 AIMs were extracted from genotyping data of 643 subjects gathered by

the authors: 128 European Americans, 42 West African subjects, 105 Amerindians, 188 East Asian Americans, and 64 South Asian Americans. In addition, 60 subjects from European [Utah residents (CEPH) with Northern and Western European ancestry] and 56 from Yoruban (Ibadan, Nigeria) populations were included (The HapMap Consortium, 2003).

To test the feasibility of this streamlined AIM subset in admixed populations, we obtained the corresponding genotypes of both the 128 original and the 12 AIMs in an independent subset of unrelated individuals from populations with available data in 1000 Genomes database (Genomes Project Consortium et al., 2012).

This subset included 822 individuals from the following populations: Finland (N = 93), England and Scotland (N = 89), Spain (N = 14), Tuscany, Italy (N = 97), Yoruba from Ibadan, Nigeria (N = 85), Luhya from Webuye, Kenya (N = 96), African-American from Southwest US (N = 64), Colombia (N = 60), Mexico (N = 64), and Puerto Rico (N = 55). The 105 Amerindian subjects were from the original Kosoy et al. (2009) subset included in this procedure.

After *in silico* analysis using the independent subset described above, we genotyped the 12 AIMs in 192 samples of individuals from the metropolitan region of Recife. Each subject answered an epidemiological questionnaire, which required self-report in five ethnic categories as defined by the Instituto Brasileiro de Geografia e Estatística (IBGE - Brazilian Geography and Statistics Institute), the Brazilian census bureau, based on skin color. The five IBGE categories are: *branco* (white or Caucasian), *preto* (black), *pardo* (multiracial), *amarelo* ("yellow", Asian descent) and *indígena* (Amerindian) (IBGE, 2010).

Table 1 summarizes information about the top 12 AIMs (marker ID, chromosome location and allele frequencies, including the Recife frequencies).

Table 1. The twelve most informative markers for assignment in three subpopulations (K = 3), TaqMan probe assays, chromosome, reference allele frequencies as presented in Kosoy et al. (2009) paper and reference allele frequencies in Recife sample.

Marker	TaqMan probe assay	Location	EURA frequencies	AFR frequencies	AMI frequencies	Recife frequencies
rs4908343	C 2494120_10	1p36.11	0.82	0.04	0.95	0.59
rs7554936	C 26139689_10	1q21.3	0.34	0.99	0.12	0.44
rs6548616	C 29071253_10	3p12.3	0.25	0.96	0.05	0.43
rs7657799	C 29422763_10	4q24	0.05	0.86	0.01	0.21
rs10007810	C 1386349_10	4p13	0.25	0.96	0.05	0.41
rs6451722	C 2938090_10	5p12	0.24	0.90	0.01	0.36
rs1040045	C 8767011_10	6p25.1	0.73	0.10	0.98	0.67
rs10108270	C 30263561_10	8p23.2	0.35	0.97	0.03	0.39
rs772262	C 8340116_10	12q13.2	0.06	0.87	0.63	0.33
rs9530435	C 27192660_10	13q22.2	0.79	0.07	0.95	0.69
rs11652805	C 31084340_10	17q24.1	014	0.98	0.13	0.35
rs4891825	C 27956007_10	18q22.2	0.89	0.09	0.90	0.68

EURA = European American populations; AFR = African populations; AMI = Amerindian populations.

Since the self-reported ethnicity was recorded (Table 2), it was possible to observe how the estimated genetic ancestry would relate to the arbitrary skin-color categories, an approach followed by previous studies (Smith et al., 2004; Lins et al., 2010), although using a different set of markers. The Mann-Whitney test was used to make comparisons between the genetic ancestry estimates between self-reported ethnicities.

Table 2. Demographic characteristics of subjects from Recife, including self-reported race according to the classification of Instituto Brasileiro de Geografia e Estatística (IBGE - Brazilian Geography and Statistics Institute), the Brazilian census bureau.

Characteristic	N = 192	%
Gender		
Female	128	66.7
Male	64	33.3
Self-reported race		
Pardo (multiracial)	93	48.4
Branco (White or Caucasian)	81	42.2
Preto (Black)	18	9.4

All SNPs were genotyped using TaqMan SNP Genotyping Assays on ABI 7500 real-time PCR platform, following manufacturer instructions (Life Technologies, USA).

Population substructure assignment and comparison between AIM sets

All population substructure analyses were performed with the STRUCTURE software version 2.3.4 (Pritchard et al., 2000; Falush et al., 2003, 2007; Hubisz et al., 2009), as recommended by Kosoy et al. (2009). For runs with the original 128 AIMs, 10,000 burn-in cycles and 50,000 replicates in the admixture model were used. For models using the 12 AIMs, 20,000 burn-in cycles and 100,000 replicates also in the admixture model were used instead. Thus, different α (Dirichlet parameter, which in this case represents the degree of admixture) were calculated for each estimated population. Apart from this, all remaining options were set to their default, and each run had three iterations. All runs were performed with two to four ancestry clusters (K).

For ancestry proportion estimation in the Recife sample, the European American, African, and Amerindian top 12 genotypes (thus excluding the EAS and SAS populations) were inputted together with the genotypes from the new samples to serve as references. This was done to help determine which cluster generated by the STRUCTURE software corresponds to each ancestry, i.e., European, African and Amerindian (K = 3) ancestries, respectively.

Ancestry cluster plots were performed through the Distruct software version 1.1 (Rosenberg, 2004). The mean estimates of the 1000 Genomes populations resulting from the three iterations with both the original 128 and the top 12 AIMs were compared using Pearson correlation. Additionally, Bland-Altman plots were produced to help visualize the degree of concordance of the correlations, an approach similar to that used by Aldrich et al. (2008).

RESULTS

When analyzing the data of the 1000 Genomes populations, the ancestry estimates obtained by the two AIM sets, i.e., the original 128 AIMs genotyped by Kosoy et al. (2009) and the top 12, were similar for all three ancestries. The high correlation indices showed that the estimates had a high degree of concordance (Figure 1). Bland-Altman plots showed in more detail the degree of concordance between these estimates (Figure 2). Since these results showed that both estimates were remarkably similar in a population dataset regardless of the subset which was used for AIMs selection, the top 12 were used as an approximation to the 128 AIMs to genotype the sample from Recife.

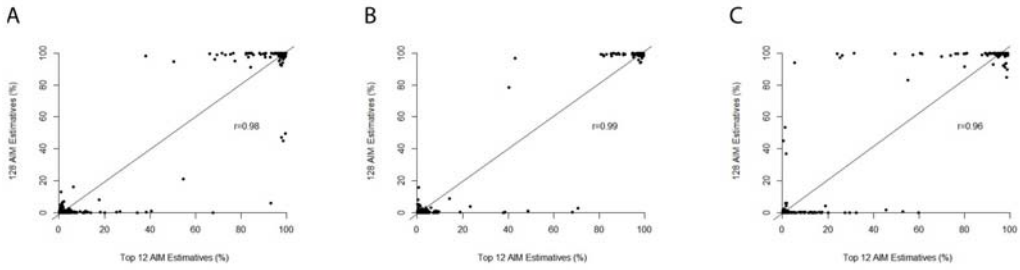


Figure 1. Correlation plots comparing estimates of 128 ancestry informative markers (AIMs) and a subset of its 12 most informative AIMs (top 12) in 822 ethnically diverse subjects derived from 1000 Genomes database. Comparison of **A.** European ancestry estimates, **B.** African ancestry estimates, and **C.** Amerindian ancestry estimates.

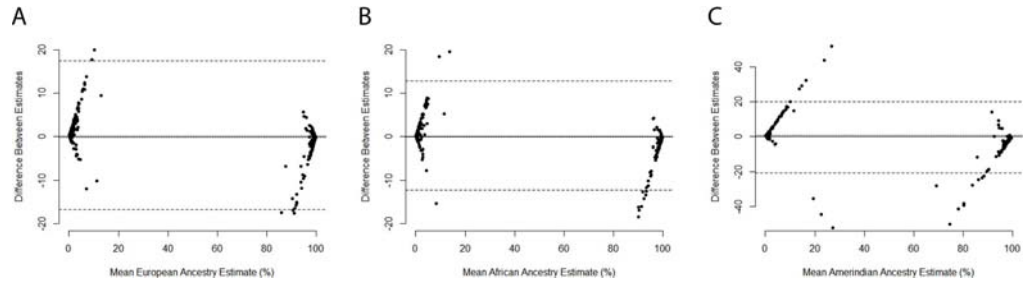


Figure 2. Bland-Altman plots showing the great concordance between estimates from the two ancestry informative marker subsets (128 and the 12 most informative markers) in 822 ethnically diverse subjects. **A.** European ancestry, **B.** African ancestry, and **C.** Amerindian ancestry comparisons. The solid line represents no difference between two estimates. The dashed lines represent 95% confidence intervals of the difference between estimates. Since the majority of data points are between these lines, it is demonstrated that the estimates did not differ significantly. The dotted line represents the mean estimates taking in consideration the two ancestry informative marker subsets (128 and the top 12).

All genotypes were in conformity with Hardy-Weinberg equilibrium. The output from the STRUCTURE software pointed out that K = 3 had the largest mean likelihood (L = -6095.2 against K = 2, L = -6311.0, and K = 4, L = -6110.6).

Considering individual *Q* (ancestry) estimates in the sample from Recife, Brazil, it was observed that all subjects had greater European influence when compared with African and Amerindian, without taking self-reported race into consideration (mean estimates: 59.7, 23.0 and 17.3%, respectively). However, there was considerable overlap of these estimates.

After stratifying according to self-reported ethnicity, subjects identified as black and *pardos* had marginally similar African proportions (27.9 *vs* 24.6%, Mann-Whitney U-test = 585, P = 0.04). Whites had more European ancestry proportion (61.9%) than did *pardos* (58.3%, U-test = 5237.5, P << 0.01) and blacks (56.6%, U-test = 1164, P << 0.01). Blacks and *pardos* had similar proportions of European ancestry (U-test = 998.5, P = 0.20).

Regarding Amerindian ancestry, *pardos* and whites had similar backgrounds (18.0 *vs* 17.1%, U-test = 4364, P = 0.07), with blacks having the lower proportion (15.5%) when compared with whites (U-test = 1086, P = 0.001) and *pardos* (U-test = 1107, P = 0.03). Table 3

summarizes these estimates and Figure 3 depicts the overall estimates of the reference populations (European American, African, and Amerindian) as well as Recife estimates.

Table 3. Comparisons between estimates of subjects from Recife (Northeast Brazil) according to self-reported races.

Self-reported race	European ancestry (%)		African ancestry (%)		Amerindian ancestry (%)	
	Mean ± SD	Min.-Max.	Mean ± SD	Min.-Max.	Mean ± SD	Min.-Max.
White (N = 81)	61.9 ± 5.5	41.9-68.4	20.1 ± 7.1	10.8-47.7	18.0 ± 3.0	10.4-24.6
*Pardo** (N = 93)	58.3 ± 5.5	42.2-68.5	24.6 ± 7.1	11.7-46.7	17.1 ± 3.0	10.6-25.8
Black (N = 18)	56.6 ± 5.6	49.0-65.8	27.9 ± 7.2	18.0-38.2	15.5 ± 3.0	11.2-19.3
All (N = 192)	59.7 ± 5.5	41.9-68.5	23.0 ± 7.1	10.8-47.7	17.3 ± 3.0	10.4-25.8

*Multiracial.

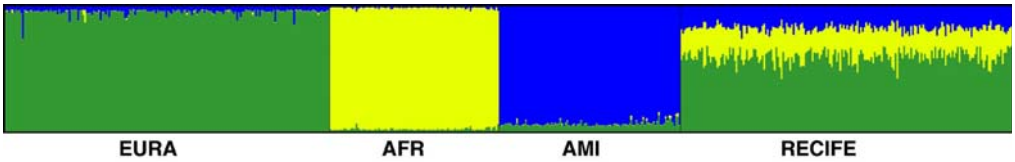


Figure 3. Plot of the overall top 12 estimates. The European American (EURA), African (AFR) and Amerindian (AMI) populations were included together with Recife samples to help ancestry assignment. The resulting Recife estimates demonstrate Brazilian populations admixture: a large contribution of European ancestry with less influence of African and Amerindian ancestries.

DISCUSSION

Population substructure and genetic ancestry is a fundamental issue to be considered when designing and developing GAS, because they could result in spurious association detection (Balding, 2006). The present study aimed to propose a streamlined panel of AIMs that could replace larger panels of markers, thus improving cost-benefit in the analysis of highly-admixed populations such as from Brazil, more specifically in Northeast Brazil, Recife city, Pernambuco State. Our results confirm the notion that the population from Recife had origin from three ethnically ancestral groups: Amerindians, Africans and Europeans (Alves-Silva et al., 2000).

Moreover, the analyses indicated that reduced numbers of markers were still capable of adequately estimating individual ancestry proportions. We compared the 12 most informative SNPs with the original 128, published elsewhere (Kosoy et al., 2009). There was remarkable concordance between the estimates, as revealed by Pearson correlation coefficients (Figure 1).

Currently, several genotyping methods are used to determine genetic ancestry. The most common are small insertion or deletion (indels) markers, which provide accurate information about ancestry since they are scattered throughout the genome (Mills et al., 2006). The standard methods of indel genotyping could be relatively costly and time-consuming, requiring expensive reagents for PCR, PCR product purification and special hardware, such as capillary gel electrophoresis instruments. Thus, some authors have reported ancestry estimation using SNP panels with real-time PCR platforms (Smith et al., 2004; Seldin and Price, 2008; Lins et al., 2010).

Estimates from Recife (Northeast Brazil) confirmed that Brazilian populations are admixed, since each individual had substantial genetic contribution from each subpopulation cluster. In summary, the European genomic contribution was most representative in our population, followed by the African and then the Amerindian contribution (average contributions of 59.7, 23.0 and 17.3% ancestry, respectively). This means that the Recife population has a low level of genetic substructure, since self-reported race, which is mainly based on skin-color, did not relate well to genetic proportions. Subjects identified as blacks, had even more European ancestry influence than African ancestry itself. Thus, skin-color does not correctly control for population substructure in admixed populations such as the Brazilian during GAS analysis.

These results confirm the observations of Pena et al. (2009), who also detected high European influence in the Brazilian genetic background. The authors also used samples from Pernambuco State. Our average European ancestry estimates in self-declared white subjects were somewhat lower than their own (61.9 *vs* 71.1%, respectively), and our African ancestry estimates were slightly higher (20.1 *vs* 14.2%, respectively).

Moreover, when comparing the estimates from white subjects from São Paulo State (Southeast Brazil) with white subjects from the Recife sample, it is evident how different regions of Brazil received different influences from the ancestral populations. Northeast Brazil received more influence from West African populations, whereas the Southeast received large numbers of European immigrants during the 19th century, a process that shaped these interregional differences in genetic background. The subjects from São Paulo had 77.9% European ancestry, 11.6% African and 10.5% Amerindian ancestry, whereas the subjects from Recife had 61.9, 20.1 and 18.0%, respectively (Pena et al., 2009). Similar results were obtained in another study by Lins et al. (2010), who used a set of 28 SNPs. In their sample from Northeast Brazil, the subjects had 77.4, 13.6 and 8.9% estimated ancestries.

These discrepancies may be related to differences in sample number and the markers employed to estimate the ancestries, since in one of them, a set of indel markers was used. Despite these differences, these studies and others performed in several Brazilian populations, e.g., Santos et al. (2010), show that, qualitatively, European ancestry makes up most of the genetic background of Brazilians, followed by minor contributions from African and Amerindian ancestral populations. Thus, we again highlight the importance of considering this admixture when performing GAS in Brazilian populations.

Keeping this in mind, we decided to further explore SNP usage for ancestry estimation in GAS, to obtain a streamlined panel and report a proof of concept that ancestry estimation could be simpler using a smaller number of markers than in previous studies.

We propose a rapid streamlined SNP panel based on 12 markers, spread through the genome, to be used for genetic ancestry estimation aimed at controlling the population substructure in GAS in Brazilian admixed populations. Our SNP panel, deriving from our daily experience in small-scale association studies with relatively small financial resources, has been specifically tailored for laboratories with low-/medium-throughput genotyping instrumentation and needs.

ACKNOWLEDGMENTS

Research supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES/PNPD program), Fundação de Amparo à Pesquisa do Estado de Pernambuco (APQ-1530-2.02/12), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and

IRCCS Burlo Garofolo (RC06/11). R.R. Moura is the recipient of a postgraduation fellowship from Fundação de Amparo à Pesquisa do Estado de Pernambuco (FACEPE, IBPG #0926-2.02/11).

REFERENCES

Aldrich MC, Selvin S, Hansen HM, Barcellos LF, et al. (2008). Comparison of statistical methods for estimating genetic admixture in a lung cancer study of African Americans and Latinos. *Am. J. Epidemiol.* 168: 1035-1046.

Ali-Khan S, Krakowski T, Tahir R and Daar A (2011). The use of race, ethnicity and ancestry in human genetic research. *HUGO J.* 5: 47-63.

Alves-Silva J, Santos MS, Guimaraes PE, Ferreira AC, et al. (2000). The ancestry of Brazilian mtDNA lineages. *Am. J. Hum. Genet.* 67: 444-461.

Balding DJ (2006). A tutorial on statistical methods for population association studies. *Nat. Rev. Genet.* 7: 781-791.

Cordell HJ and Clayton DG (2005). Genetic association studies. *Lancet* 366: 1121-1131.

Falush D, Stephens M and Pritchard JK (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567-1587.

Falush D, Stephens M and Pritchard JK (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol. Ecol. Notes* 7: 574-578.

Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, et al. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491: 56-65.

Hubisz MJ, Falush D, Stephens and Pritchard JK (2009). Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* 9: 1322-1332.

Instituto Brasileiro de Geografia e Estatística (IBGE) (2010). Atlas do Censo Demográfico 2010. Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro.

Kosoy R, Nassir R, Tian C, White PA, et al. (2009). Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Hum. Mutat.* 30: 69-78.

Lewis CM (2002). Genetic association studies: design, analysis and interpretation. *Brief. Bioinform.* 3: 146-153.

Lins TC, Vieira RG, Abreu BS, Grattapaglia D, et al. (2010). Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs. *Am. J. Hum. Biol.* 22: 187-192.

Mills RE, Luttig CT, Larkins CE, Beauchamp A, et al. (2006). An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res.* 16: 1182-1190.

Parra FC, Amado RC, Lambertucci JR, Rocha J, et al. (2003). Color and genomic ancestry in Brazilians. *Proc. Natl. Acad. Sci. U. S. A.* 100: 177-182.

Pena SD, Bastos-Rodrigues L, Pimenta JR and Bydlowski SP (2009). DNA tests probe the genomic ancestry of Brazilians. *Braz. J. Med. Biol. Res.* 42: 870-876.

Pritchard JK, Stephens M, Rosenberg NA and Donnelly P (2000). Association mapping in structured populations. *Am. J. Hum. Genet.* 67: 170-181.

Ribeiro D (1995). O Povo Brasileiro: A Formação e o Sentido do Brasil. Companhia das Letras, São Paulo.

Rosenberg NA (2004). Distruct: a program for the graphical display of population structure. *Mol. Ecol. Notes* 4: 137-138

Rosenberg NA, Li LM, Ward R and Pritchard JK (2003). Informativeness of genetic markers for inference of ancestry. *Am. J. Hum. Genet.* 73: 1402-1422.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, et al. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409: 928-933.

Santos NP, Ribeiro-Rodrigues EM, Ribeiro-Dos-Santos AK, Pereira R, et al. (2010). Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INSEL) ancestry-informative marker (AIM) panel. *Hum. Mutat.* 31: 184-190.

Seldin MF and Price AL (2008). Application of ancestry informative markers to association studies in European Americans. *PLoS Genet.* 4: e5.

Smith MW, Patterson N, Lautenberger JA, Truelove AL, et al. (2004). A high-density admixture map for disease gene discovery in African Americans. *Am. J. Hum. Genet.* 74: 1001-1013.

The HapMap Consortium (2003). The International HapMap Project. *Nature* 426: 789-796.

APÊNDICE C - ON THE USE OF CHINESE POPULATION AS A PROXY OF AMERINDIAN ANCESTORS IN GENETIC ADMIXTURE STUDIES WITH LATIN AMERICAN POPULATIONS

European Journal of Human Genetics (2015), 1–2

www.nature.com/ejhg

LETTER

On the use of Chinese population as a proxy of Amerindian ancestors in genetic admixture studies with Latin American populations

European Journal of Human Genetics advance online publication, 2 September 2015; doi:10.1038/ejhg.2015.184

Dear Editor,

We read the recent article of Magalhães da Silva et al[1] reporting the correlation between biogeographic ancestries, estimated using 30 ancestry informative markers (AIMs), and self-reported skin color in two different Brazilian Northeastern populations (Fortaleza and Salvador, capitals of the states of Ceará and Bahia, respectively). The authors observed that African ancestry is more correlated in the sample from Salvador than in the one from Fortaleza and that the use of different African populations as proxies of the Brazilian's African ancestors may influence the results.

One unusual point of this study was taking Han Chinese from Beijing (CHB) as pseudo-ancestors for Amerindians as there is no Native American population included in the HapMap Project (http://hapmap.ncbi.nlm.nih.gov/) and, in addition, because CHB population has been shown to have allele frequencies similar to those of Native Amerindians.[1]

It is well known that most Latin American countries are inhabited by tri-hybrid populations derived from African, Amerindian and European roots, in which their proportions show considerable variability.[2] As the vast majority of published studies aimed at inferring the admixture proportions of Latin American populations have not used CHB population as a proxy for Amerindian ancestors, we compared the admixture inference of some Latin American populations using Chinese or Native American ancestors as proxies.

In order to estimate admixture proportions, we used the following Latin American populations from 1000 Genomes Project Phase III[3]: Colombians from Medellin, Colombia (CLM); Mexican Ancestry from Los Angeles, USA (MXL); Peruvians from Lima, Peru (PEL); and Puerto Ricans from Puerto Rico (PUR). Besides the CHB population, we also used as proxies for African and European ancestors the Utah Residents with Northern and Western European ancestry (CEU) and Yoruba in Ibadan, Nigeria (YRI), respectively, also obtained from 1000 Genomes Project Phase III database. For Amerindian ancestors, a combination of Mayans, Quechuans and Nahua Natives (we call this group AMI) were used as pseudo-ancestors for Indigenous Americans as described by Kosoy et al.[4]

One hundred and twenty-seven SNPs were used as AIMs from a set of 128 SNPs validated by Kosoy et al[4] (the rs10954737 SNP is not present in the 1000 Genomes database). After merging 1000 Genomes with AMI data using PLINK version 1.90,[5] we employed STRUCTURE software v. 2.3.4 (Pritchard et al[6]) running into ParallelStructure R package.[7,8] The parameters applied were 10 independent runs with 100 000 burn-in steps and 100 000 Markov chain Monte Carlo replicates assuming three ancestral populations ($K = 3$) in admixture model, allele frequencies correlated and the parameter USEPOPINFO = 1. CLUMPAK Server allowed us to generate bar plots, referring to individual and population ancestry proportions.[9] We made two separated analyses: one using CEU, YRI and AMI as pseudoancestors and another using CEU, YRI and CHB.

The bar plot with the admixture estimate is shown in Figure 1 and the ancestry proportions for the CLM, MXL, PEL and PUR populations are displayed in Table 1. It is possible to observe a tendency towards increasing of Amerindian ancestry and a decreasing of European ancestry when the CHB population was used as proxy for Native American ancestors.

In general, the results obtained using AMI pseudo-ancestors are more similar to those found in the literature, also when using AMI pseudo-ancestors in studies concerning populations from the same city such as CLM, MXL, PEL and PUR (Table 1). The results published by Magalhães da Silva et al[1] predicted 54.7, 12.3 and 33% for EUR, AFR and CHB contributions, respectively, in a self-declared 'white' individuals from Fortaleza. Whereas in a study of Pena et al,[10] using 40 validated AIMs, the authors found 75.8, 13.3 and 10.9% in a group of 'white' individuals from the same city. The same tendency for increased AMR contribution, using CHB as pseudo-ancestors, occurred when analyzing 'brown' individuals. Curiously, this tendency was not confirmed when the results of the population from Salvador are compared with the findings from the study of Pena et al.[10]

We also evaluated whether the allele frequency of ancestry markers were different between AMI and CHB populations: in our set of 127 AIMs, 100 showed significant differences between allele frequencies (P-value from χ^2 test <0.05). Moreover, we observed a weak correlation between the allele frequencies of AMI and CHB populations ($r = 0.33$). Finally, average pairwise F_{ST} values were higher when comparing AMI with CEU or YRI (0.30 and 0.35, respectively) than when comparing CHB with CEU and YRI (0.16 and 0.25, in that order). As higher divergence corresponds to better Informativeness,[11] we could hypothesize that the use of AMI individuals as pseudo-ancestors of Native Americans should be preferred to that of CHB individuals.

In light of this, we believe that the results of Magalhães da Silva et al[1] may be biased owing to the use of CHB population instead AMI as reference samples of Amerindian ancestors. Some genotyping data from AMI populations such as those from Kosoy et al[4] are publically available as well as those from Human Genetic Diversity Project (http://www.hagsc.org/hgdp/files.html) and may serve as source of AIMs where a subset of the panel found by Galanter et al[12] can be downloaded.

In future studies dealing with Latin American tri-hybrid admixtures, our suggestion would be to consider genotyping data from AMI populations as first choice, eventually comparing the results with those obtained by using CHB as proxy as carried out by Magalhães da Silva et al.[1] Their comparison will allow a definitive choice based on the better representatives of Amerindian ancestry.

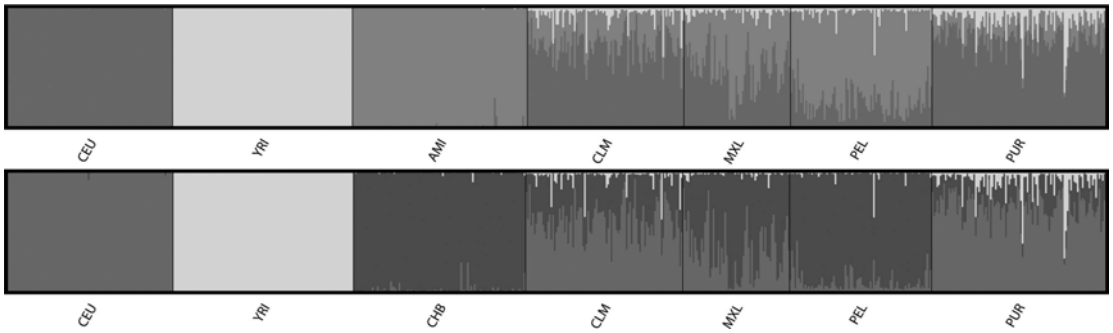


Figure 1 Bar plot with the admixture estimate of Latin American populations using AMI (top) and CHB (bottom) populations as proxies for the Amerindian ancestors. CEU = Utah Residents with Northern and Western European ancestry; YRI = Yoruba in Ibadan, Nigeria; AMI = combination of Mayans, Quechuans Amerindians and Nahua Native Americans; CHB = Han Chinese from Beijing; CLM = Colombians from Medellin, Colombia; MXL = Mexican Ancestry from Los Angeles, USA; PEL = Peruvians from Lima, Peru; PUR = Puerto Ricans from Puerto Rico.

Table 1 Proportions of European (EUR), African (AFR) and Native American (AMR) ancestors in Latin American populations

	Using AMI			Using CHB			Literature data			
Population	EUR	AFR	AMR	EUR	AFR	AMR	EUR	AFR	AMR	Authors
CLM	0.6770	0.0610	0.2620	0.5904	0.0390	0.3706	0.6000	0.1200	0.2800	Wang et al[13]
MXL	0.4769	0.0384	0.4848	0.3361	0.0150	0.6489	0.4580	0.1100	0.4290	Qu et al[14]
PEL	0.2020	0.0342	0.7638	0.0705	0.0138	0.9157	0.1400	0.0200	0.8400	Sandoval et al[15]
PUR	0.7422	0.1127	0.1450	0.7006	0.0890	0.2104	0.6400	0.2100	0.1500	Via et al[16]

Abbreviations: CLM = Colombians from Medellin, Colombia; MXL = Mexican Ancestry from Los Angeles, USA; PEL = Peruvians from Lima, Peru; PUR = Puerto Ricans from Puerto Rico.
The results using Amerindian (AMI) and Chinese individuals (CHB) are shown. Some findings retrieved from literature for similar populations are also displayed.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Ronald R de Moura*[,1,2], Valdir de Queiroz Balbino[1], Sergio Crovella[1,2] and Lucas AC Brandão[1,2,3]

[1]Department of Genetics, Federal University of Pernambuco (UFPE), Recife, Brazil;
[2]Laboratory of Immunopathology Keizo Asami (LIKA/UFPE), Recife, Brazil;
[3]Department of Pathology, Federal University of Pernambuco (UFPE), Recife, Brazil
E-mail: ronaldmoura1989@gmail.com

1 Magalhães da Silva T, Sandhya Rani M R, Nunes de Oliveira Costa G et al: The correlation between ancestry and color in two cities of Northeast Brazil with contrasting ethnic compositions. Eur J Hum Genet 2014; 23: 984–989.
2 Salzano FM, Sans M: Interethnic admixture and the evolution of Latin American populations. Genet Mol Biol 2014; 37: 151–170.
3 1000 Genomes Project Consortium: An integrated map of genetic variation from 1,092 human genomes. Nature 2012; 135: 0–9.
4 Kosoy R, Nassir R, Tian C et al: Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. Hum Mutat 2009; 30: 69–78.
5 Purcell S, Neale B, Todd-Brown K et al: PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559–575.
6 Pritchard JK, Stephens M, Donnelly P: Inference of population structure using multilocus genotype data. Genetics 2000; 155: 945–959.
7 Besnier F, Glover KA: ParallelStructure: a R package to distribute parallel runs of the population genetics program STRUCTURE on multi-core computers. PLoS One 2013; 8: 1–5.
8 R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing: Vienna, Austria, 2015. Available at ohttp://www.r-project.org/4.
9 Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I: Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour 2015; 15: 1179–1191.
10 Pena SDJ, Di Pietro G, Fuchshuber-Moraes M et al: The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. PLoS One 2011; 6: e17063.
11 Ding L, Wiener H, Abebe T et al: Comparison of measures of marker informativeness for ancestry and admixture mapping. BMC Genomics 2011; 12: 622–640.
12 Galanter JM, Fernandez-Lopez JC, Gignoux CR et al: Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. PLoS Genet 2012; 8: e1002554.
13 Wang S, Ray N, Rojas W et al: Geographic patterns of genome admixture in Latin American mestizos. PLoS Genet 2008; 4: 1–9.
14 Qu HQ, Li Q, Lu Y et al: Ancestral effect on HOMA-IR levels quantitated in an American population of Mexican origin. Diabetes Care 2012; 35: 2591–2593.
15 Sandoval JR, Salazar-Granara A, Acosta O et al: Tracing the genomic ancestry of Peruvians reveals a major legacy of pre-Columbian ancestors. J Hum Genet 2013; 58: 627–634.
16 Via M, Gignoux C, Roth LA et al: History shaped the geographic distribution of genomic admixture on the island of Puerto Rico. PLoS One 2011; 6: e16513.

APÊNDICE D - ANTIRETROVIRAL THERAPY IMMUNOLOGIC NON-RESPONSE IN A BRAZILIAN POPULATION: ASSOCIATION STUDY USING PHARMACO- AND IMMUNOGENETIC MARKERS

Antiretroviral Therapy Immunologic Non-response in a Northeast Brazil Population: Association Study using Pharmaco- and Immunogenetic Markers

Antonio Victor Campos Coelho[1*], Ronald Rodrigues de Moura[1], Rafael Lima Guimarães[1,2], Lucas André Cavalcanti Brandão[2,3], and Sergio Crovella[1,4]

[1]Department of Genetics, Federal University of Pernambuco. Avenida da Engenharia, Cidade Universitária, Recife, Pernambuco (Brazil). Postal code: 50740-600. Tel: +55 81 2126 8522.
[2]Laboratory of Immunopathology Keizo Asami (LIKA), Federal University of Pernambuco, Avenida da Engenharia, Cidade Universitária, Recife, Pernambuco (Brazil). Postal code: 50740-600.
[3]Department of Pathology, Federal University of Pernambuco, Avenida Prof. Moraes Rego, 1235, Cidade Universitária, Recife, Pernambuco (Brazil). Postal code: 50670-901.
[4]IRCCS Burlo Garofolo and University of Trieste, Via dell' Istria 65/1, Trieste (Italy). Postal code: 34137. Tel.: +39-040-378-5273.

*Corresponding author. Department of Genetics, Federal University of Pernambuco. Avenida da Engenharia, Cidade Universitária, Recife, PE (Brazil). Postal code: 50740-600. Tel: +55 81 2126 8522. Email: avccbio@gmail.com.

Running head: ART Non-response in a NE Brazil Population: Genetic Association Study

Abstract

Background: Antiretroviral therapy (ART) saved millions of lives from HIV-1 infection and AIDS, but some patients do not experience adequate CD4+ T cells gains even though viral suppression is achieved, phenomenon known as immunological non-response, which predisposes to non-AIDS systemic diseases. The genetic component of this condition is not yet completely elucidated. Thus, we aimed to discover predictive genetic markers of immune response delay through a case-control study. **Methods:** We recruited 135 patients from Recife, Northeast Brazil, among which 82 were non-responders, and genotyped a set of 94 selected single nucleotide polymorphisms (SNPs). Among those SNPs, 46 were located in genes involved in antiretroviral drugs pharmacodynamic pathways and immune system homeostasis, while the remaining 48 were ancestry informative markers (AIMs) for genetic ancestry estimation and controlling for eventual hidden population structure. **Results:** Male patients were overrepresented in non-responder group (p=0.01) and tended to have slower immune recovery than females, without reaching statistical significance (p=0.47). Non-responders also started with lower absolute CD4+ T cell counts (p<0.001). We found five SNPs significantly associated with the outcome, being three more frequent in non-responders than responders: rs3003596 (*NR1I3*) G allele (p=0.01); rs2243250 (*IL4*) A allele (p=0.02) and rs129081 (*ABCC1*) G allele (p=0.002), whereas the other two were less frequent in non-responders: rs2069762 (*IL2*) C allele (p=0.03) and rs10519613 (*IL15*) A allele (p=0.03). Patients had similar ancestry backgrounds. **Conclusions:** All significant associations were lost during multivariate survival analysis modeling. Therefore, more studies are needed to unravel the genetic basis of ART immunological non-response.

Keywords: HIV-1; pharmacodynamics; immunogenetics; genetic association study; survival analysis

1 Introduction

The introduction of antiretroviral therapy (ART) in the clinical practice saved millions of lives from acquired immunodeficiency syndrome (AIDS) related deaths, which is the result of chronic infection by the human immunodeficiency virus type 1 (HIV-1) (De Cock et al. 2011; World Health Organization 2015). Current ART regimens are combinations of three drugs. The first-line regimens usually include two nucleoside analog reverse transcriptase inhibitors (NRTIs) and a non-nucleoside reverse transcriptase inhibitor (NNRTI) or a protease inhibitor (PI), and their objective is to suppress viral replication (Vella et al. 2012).

With viral suppression, reestablishment of lost CD4+ T cells typically happens in a biphasic manner: a rapid proliferation during the first three to six months of ART, caused by memory T cells redistribution, followed by a slower proliferation phase conducted by naïve T cells production by the thymus (Li et al. 2011). However, some patients do not present optimal CD4+ T cells gains, even with persistent viral suppression. This phenomenon is named immunological failure or immunological non-response, and it is associated with higher risk of non-AIDS cardiovascular, liver, kidney disorders and early ageing (Teixeira et al. 2001; Tan et al. 2008; Pathai et al. 2014; Torres and Lewis 2014).

The immunological non-response has not yet been completely elucidated. Older age, male sex, advanced HIV-1 infection at treatment start and coinfections with other viruses are some known risk factors (Marziali et al. 2006; Li et al. 2011; Doitsh et al. 2014), but the influence of host genetic component is still debated. Promising genetic candidates include genes involved in antiretroviral drugs pharmacodynamic pathways (Telenti and Zanger 2008; Tozzi 2009) and genes involved in immune functions (inflammation, apoptosis) and homeostasis, such as the interleukins IL2, IL7 and IL15, which coordinate T cell proliferation (Levy 2006).

Our hypothesis was that genetic variation in these genes would affect the distribution of antiretroviral drugs, possibly favoring (undetected) residual virus replication, which in turn would drive immune activation and immune cell death, hampering CD4+ T cell recovery. Simultaneously, the SNPs in the immune system genes would favor this increased cell activation, increased apoptosis or decreased cell proliferation, leading to the suboptimal immunological response to ART. So, we aimed at identifying genetic variants associated with delay to proper CD4+ T cell counts recovery to quasi-normal or normal levels.

2 Material and methods

2.1 Subjects

We recruited 135 individuals (65 females, 48.2% and 70 males, 51.8%), coming from Recife and nearby cities, Pernambuco state, Northeast Brazil, living with HIV-1 attending healthcare at Instituto de Medicina Integral Professor Fernando Figueira (IMIP) for a case-control, observational study between 2011 and 2015.

Inclusion criteria were: age over than 18 years old, not reporting illicit drug use, and not being pregnant. Informed consent was obtained from all individual participants included in the study. Each patient then provided a peripheral blood sample after giving consent for participation in the research and authorization for review of their medical charts. IMIP Research Ethics Committee approved the study design (protocol number 3629-13).

Each patient answered a standard questionnaire about sex, age and age at ART start date, socioeconomic status, smoking and drinking habits. The data extracted from medical charts covered the whole period between the first and last plasma viral load (pVL) and CD4+ T cell absolute counts and percentage (relative to all white blood cells) measurements (retrospective follow-up).

Other extracted data included ART regimens received and their refill prescriptions and serological status for the following etiologic agents: hepatitis B and C viruses (HBV and HCV), cytomegalovirus (CMV, immunoglobulin G and M, IgG and IgM tests), human T-lymphotropic virus types 1 or 2 (HTLV-1/2), *Toxoplasma gondii* (toxoplasmosis agent, IgG and IgM tests) and *Treponema pallidum* (syphilis agent, VDRL test) and eventual AIDS-defining conditions following Center for Disease Control (CDC) 1993 Revised Classification System (CDC 1992). Table 1 displays patients' characteristics.

2.2 Patient classification

All patients achieved persistent viral suppression, which was defined as maintaining undetectable plasma viral load measurements (pVL<50 copies/mL) without viral load rebound, which was defined as two consecutive pVL>200 copies/mL at any time after initial suppression. In our setting, pVL measurements, CD4+ T cell counts and other laboratory tests are generally performed every three or four months, at the physician's discretion.

Immunologic response was defined as CD4+ T cell percentages achieving 30% or higher for two consecutive measurements during follow-up were classified as having immunological response, and immunological non-response if otherwise, following Brazilian Ministry of Health's guidelines (Brasil 2015). If a patient already had pre-ART CD4+ T cell percentages>30% (early start patients, for example), we instead considered immunological response as an absolute gain of 200 cells/μL, following a previous study (Li et al. 2011). We preferred CD4+ T cell percentages instead of absolute counts because percentages are less variable over time (Hulgan et al. 2005).

Thus, 82 patients were stratified into immunological non-response and 53 into immunological response groups and their follow-up duration was recorded for further survival analysis.

2.3 SNPs selection and genotyping

We selected genes and single nucleotide polymorphisms (SNPs) through literature search and functional criteria. We selected 46 SNPs distributed in 19 genes of antiretroviral pharmacodynamic pathways: *ABCB1* (rs1128503, rs2214102, rs2235048 and rs3842), *ABCC1* (rs129081, rs113264879, rs4148380, rs8056298, rs212091 and rs16967632), *ABCG2* (rs115770495, rs1448784 and rs2231142), *CYP1A2* (rs762551), *CYP2A6* (rs8192726), *CYP2B6* (rs8192709, rs28399499, rs34097093, rs28399502, rs707265 and rs1042389), *CYP3A4* (rs4646437), *SLC22A6* (rs11568629, rs11568628 and rs4149170) and *NR1I3* (rs3003596)(Fellay et al. 2002; Brumme et al. 2003; Cressey and Lallemant 2007; Gatanaga et al. 2007; Ingelman-Sundberg et al. 2007; Jung et al. 2008; Haas et al. 2009; Franke et al. 2010; Swart et al. 2012) and immunological activation and homeostasis: *CCL5* (rs2107538), *FAS* (rs2234767 and rs1800682), *IFNG* (rs2069709), *IL10* (rs2222202, rs1800871 and rs1800890), *IL10RA* (rs3135932 and rs9610), *IL15* (rs10519613 and rs10833), *IL1B* (rs16944), *IL2* (rs2069762), *IL4* (rs2243250), *IL7R* (rs1494555, rs11567762, rs6897932, rs3822731, rs987106 and rs3194051)(Nasi et al. 2005; Smith and Humphries 2009; Chew et al. 2011).

Additionally, 48 SNPs that served as ancestry informative markers (AIMs) were also genotyped (all variants are listed on Supplementary Table 1). Briefly, these SNPs help estimating ancestry proportions in admixed populations such as the one enrolled in our study, controlling for population structure and reducing bias during genetic association analysis (Coelho et al. 2015). Genomic DNA was extracted through Promega® Wizard Genomic DNA Purification Kit (Fitchburg, Wisconsin, USA), following manufacturer

instructions. Genotyping was performed through VeraCode® platform of GoldenGate® Illumina Inc (San Diego, California, USA) technology, following the manufacturer's instructions.

Raw genotyping data were extracted with Illumina® Genome Studio 2.0 software and imported in an Excel® worksheet. After processing the dataset, we imported it into PLINK software, version 1.90 (Chang et al. 2015) to perform quality control (QC) filtering. Samples and variants with less than 90% global call rates were removed from analysis. Variants with significant departure from Hardy-Weinberg equilibrium were also removed, using an exact test p-value<0.001 as threshold.

2.4 Ancestry proportion estimation

We used ADMIXTURE software (Alexander et al. 2009) to carry out a "supervised analysis" allowing estimation of ancestry proportions in our admixed samples, using 2000 bootstrap steps with the AIMs panel mentioned above. The calculations were made assuming three different ancestral populations (K=3; African, Amerindian and European).

2.5 Statistical analysis

Comparisons between immune non-response and immune response groups were performed through Fisher exact test or Chi-squared test for categorical variables (sex, ART regimens, genetic association tests) and Mann-Whitney test for numerical variables (age and pre-ART CD4+ T cell absolute counts). Age at ART start date was also treated as a categorical variable with four strata: *18 to 29 years*, *29 to 39 years*, *39 to 49 years* and *49 years or more*.

Additionally, univariate survival analyses were performed through Kaplan-Meier survival probability estimator. Estimators for each variable were then compared through Cox-Mantel log rank tests to assess if they exerted statistically significant influence on time to immune response.

Allele and genotypes counts and frequencies were obtained through direct counting. Compliance to Hardy-Weinberg equilibrium was also assessed through Chi-squared test. All tests were two-sided. Any variable with statistically significant association with immunologic outcome were included in a multivariate survival analysis Cox proportional hazards model along with individual African ancestry proportion (estimated by AIMs) to assess if any of these variables were associated with delayed time to immune response

(CD4+ T cell percentage reaching 30% or higher during treatment follow-up). All analyses were performed with R software, version 3.3.1 (R Core Team 2016).

Coinfections serological status, smoking and drinking habits (data not shown) and AIDS-defining conditions were not included in further analyses due to high prevalence of missing data (over 10% of data points) in the final dataset, in order to avoid the introduction of bias into the model. Table 1 also displays univariate comparison of patients' characteristics, their respective p-values, odds ratios (OR) and 95% confidence intervals (95% CI) whenever applicable.

3 Results

3.1 Univariate analysis: patients' data

We observed that males had more than two times increased risk of presenting immunological non-response (males were 61% of non-responders and 37.7% of responders; OR=2.56; 95% CI=1.20-5.60; p=0.01). Both groups had similar ages at treatment start (median 34.5 years for non-responders and 33 years for immunologic responders; p=0.11). As expected, non-responders started treatment with less absolute CD4+ T cell counts than responders (median 187.5 cells/μL versus 375.5 cells/μL, respectively, p<0.001), and the majority of non-responders started therapy with less than 200 cells/μL (48.8% versus 9.4%; p<0.001). Patients in both groups had similar ancestry backgrounds (mean African ancestry proportion 30.2% in non-responders versus 31.3% in responders, p=0.55).

Non-responders started treatment with more advanced disease than responders; 10.9% and 29.3% of non-responders presented B (symptomatic conditions) and C (AIDS-indicator conditions) CDC system stages, respectively, versus 1.9% and 13.2% of responders, although the difference did not reach statistical significance (p=0.11).

Patients started with similar first-line ART regimens (p=0.47), with the majority (68.9%) using zidovudine as the nucleoside analog reverse transcriptase inhibitor (NRTI) alongside lamivudine, and non-nucleoside reverse transcriptase inhibitors (NNRTIs), mostly efavirenz, instead of ritonavir-"boosted" protease inhibitors (PI/r) as the third option drug (58.5%). Table 1 also details each statistical comparison between the groups.

Most patients have no available serological tests results recorded in their medical charts, but it is reasonable to say that CMV and toxoplasmosis past/latent infections were somewhat prevalent (32.6% and 25.2%, respectively positive IgG/negative IgM tests). VDRL-positive tests were 14.2% of the total. Some patients were immune to HBV due to

vaccination (21.5%), a minority of them due to natural infection (6.7%) and others were susceptible (28.1%), but none presented chronic infection. We did not observe positive HCV and HTLV-1/2 serological tests within the available data. Table 2 details the numbers of reported serological tests results.

The median follow-up period extracted from medical charts was 33 months of treatment and laboratory tests history (interquartile range IQR=18-66,8; with a minimum of two months and maximum of 203 months).

The median time to achieve immunologic response was 91 months (95% CI=50-127). Men took longer to achieve immunologic response than women, needing a (median) time of 110 months (95% CI=42-127), while women needed 75 months (95% CI=42-upper bound not calculated), although it did not reach statistical significance (Cox-Mantel log rank X^2=0.50 on 1 degree of freedom; p=0.47). Age groups at ART start date also did not influence time until immunologic response (Cox-Mantel log rank X^2=1.6 on 3 degrees of freedom; p=0.67). Similarly, use of AZT instead of TDF (Cox-Mantel log rank X2=1.7 on 1 degree of freedom; p=0.19) or PIs instead of NNRTIs as third option (Cox-Mantel log rank X2=0.5 on 1 degree of freedom; p= 0.49) on ART regimens also did not influence time until response.

3.2 Genotyping quality control (QC), ancestry estimation and genetic association testing

Four candidate SNPs, rs16944 (*IL1B*), rs10833 (*IL15*) rs11568629 (*SLC22A6*) and rs16967632 (*ABCC1*) and six AIMs did not pass genotyping QC and were removed from analysis. Other variant, rs34097093 (*CYP2B6*), was also removed because all individuals in the sample had the same genotype (one of the alleles was fixed). The remaining allele and genotype frequencies were all in conformity to Hardy-Weinberg equilibrium according to PLINK software exact test. All frequencies and global call rates are displayed in Supplementary Table 2. Therefore, further analyses were performed using the remaining 41 candidate SNPs and 42 AIMs.

Five SNPs presented statistic association with immunological outcome. Three minor alleles were more frequent in non-responders than responders: rs3003596 (*NR1I3*) G allele (48.2% vs. 35.6%, p=0.01); rs2243250 (*IL4*) A allele (42.6% vs. 32.7%, p=0.02) and rs129081 (*ABCC1*) G allele (47.6% vs. 37.5%, p=0.02), whereas the other two gave inverse results: rs2069762 (*IL2*) C allele (20.1% vs. 30.2%; p=0.03) and rs10519613 (*IL15*) A allele (10.5% vs. 19.8%; p=0.03). Allele, genotype frequencies and statistical

analyses for these five SNPs are displayed in Table 3 and Supplementary Table 3 displays all genetic association results.

The individuals in our sample presented a major European ancestry contribution (mean proportion 55.1%±18.2), followed by the African (32.5%±16.1) and a minor Amerindian one (12.4%±9.8), as expected, due to our previous works with other samples coming from the same general population (Coelho et al. 2015). Since non-responders and responders had similar ancestry proportions (for example, mean African contribution 30.2%±15.9 vs. 31.3%±13.8, respectively; p=0.55), we believe that there is no hidden genetic structure biasing our genetic association analysis. As mentioned before, individual African genetic ancestry contributions were included in the multivariate Cox proportional hazards model for an additional "genomic control" (Balding 2006); however, all associations were lost (data not shown).

4 Discussion

ART, when followed correctly with good adherence, suppresses HIV-1 replication, decrease immune activation, favors immune recovery and protects against opportunist infections (Bartlett et al. 2001; Hutchinson 2001). However, some patients do not recover CD4+ T cell numbers to normal our quasi-normal levels (immunological failure or non-response), being at risk for non-AIDS diseases, such as cardiovascular, kidney and liver disorders (Li et al. 2011) and early ageing (Pathai et al. 2014; Torres and Lewis 2014). Thus, we performed a genetic association study through survival analysis to assess if polymorphisms in antiretroviral drugs pharmacodynamic pathways and immune system homeostasis were related with immunological failure in a sample from Recife, Northeast Brazil.

We found a high prevalence of immunological non-response (60.7%), which was higher than some estimates found in the literature, ranging between 10% and 40% (Peraire et al. 2014). However, it is important to notice that, since we focused our analysis exclusively on individuals with viral suppression to assure that viral replication is not the responsible for impaired CD4+ T cell numbers expansion, this estimate could be inflated simply due to sample stratification.

Our sample is generally comprised of people with lower socioeconomic status with less access to sexual education and healthcare. Thus, they usually receive late or very late HIV-1 infection diagnosis and, therefore, tend to start ART with low absolute CD4+ T cells counts, advanced disease and presenting opportunistic infections symptoms. A very

compromised immune system by the chronic HIV-1 infection is a predisposition factor to immunologic non-response (Gaardbo et al. 2012), and we indeed observed that non-responders begun treatment with lower absolute CD4+ T cell counts, with almost half of them starting treatment with less than 200 cells/μL.

Previous reviews also reported that male sex, older age (Peraire et al. 2014) and HCV coinfection are risk factors for immune non-response (Miller et al. 2005; Brites-Alves et al. 2015). We only found an association between male sex and immune failure during univariate analysis, but this association was lost during multivariate survival analysis. We also assessed HCV infection status alongside other agents (CMV, HBV, HTLV-1/2, toxoplasmosis and syphilis), but we could reach no conclusion due to the high prevalence of untested subjects for most of these infections, which hindered further analysis, since it could bias the results with sample size restriction. However, we believe that HCV and HTLV-1/2 had very low prevalence in our sample (no positive tests among those available), and therefore they are unlikely to be playing a role on immunological outcome of our sample, at least. We also did not observe any chronic HBV-infected subjects in our sample, but we detected some individuals with anti-HBV immunity due to past infections. We also found some individuals with latent CMV infection (positive IgG tests), a known causative agent of persistent immune activation (Sylwester et al. 2005), which can cause immune system exhaustion (Khaitan and Unutmaz 2011), but we cannot affirm if it was a factor favoring immune non-response in our sample, as it would be too speculative, since our serological data status is mostly lacking, as discussed above.

Sex and pre-ART CD4+ T cell counts were the only non-genetic differences between our study groups, since the individuals in our sample had similar ethnic backgrounds (as estimated by AIMs), and ART regimens types and distribution were alike between groups. Therefore, we expected to find genetic risk factors to immunologic non-response. We genotyped 46 candidate SNPs located in genes involved on antiretroviral drugs metabolism and transport and in genes involved on immune response. We found five SNPs associated - three with susceptibility to non-response while the remaining two with favorable response.

Two of the three SNPs associated with non-response are located in genes related to drug metabolism: *NR1I3* and *ABCC1*. The former is a nuclear receptor that senses foreign substances (such as antiretroviral drugs) and upregulates expression of other proteins that metabolize these substances aiming their excretion (Wada et al.) and the latter is a membrane active transport protein that ejects antiretroviral drugs from cells (Kis

et al. 2010). The third is located at *IL4* gene, which is involved in a polyfunctional immunoregulatory signal (Paul 2015). The two SNPs associated with favorable response are also located at genes related to immune system homeostasis, important for T and B cells proliferation, *IL2* (Liao et al. 2011), and T cell activation, *IL15* (Mishra et al. 2014)

We expected that SNPs in these genes would work in concert to affect gene function, altering distribution of antiretroviral drugs, and having deleterious consequences on immune function, leading to the suboptimal immunological response to ART. However, all genetic associations were lost after multivariate survival analysis modeling. As other authors did not find associations focusing on the same or similar genes (Fernandez et al. 2006) while others did (Haas et al. 2006) (for a review of previous genetic association studies, refer to (Peraire et al. 2014)), more studies are necessary to unravel the genetic component of ART immunological non-response.

5 Conclusion

We performed a genetic association study looking for genetic variants that would explain suboptimal gains CD4+ T cell counts in some individuals of a retrospective observational sample of individuals living with HIV-1 receiving ART from Northeast Brazil, but did not find any statistically significant association. Thus, more studies are necessary, so we could develop a personalized, predictive model of immunological non-response, consequently improving HIV-1 infection care, and perhaps preventing complications that came from immunological non-response, such as systemic diseases and/or early ageing.

Disclosure statement

The authors declare that they have no conflict of interest.

Acknowledgements

This work was supported by APQ-0568-2.02/10 and APQ-0599-2.02/14 grants from Fundação de Amparo à Ciência e Tecnologia de Pernambuco (FACEPE), 442225/2014-8 grant by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and BEX-7715/15-3; BEX-7711/15-8; BEX-7714/15-7 scholarships from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). The authors thank the IRCCS Burlo Garofolo (Trieste, Italy) for Illumina VeraCode® analysis financial support.

References

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655-1664

Balding DJ (2006) A tutorial on statistical methods for population association studies. Nat Rev Genet 7:781-791

Bartlett JA, DeMasi R, Quinn J, Moxham C, Rousseau F (2001) Overview of the effectiveness of triple combination therapy in antiretroviral-naive HIV-1 infected adults. AIDS 15:1369-1377

Brasil (2015) Protocolo clínico e diretrizes terapêuticas para manejo da infecção pelo HIV em adultos. Ministério da Saúde, Secretaria de Vigilância em Saúde, Programa Nacional de DST, Aids e Hepatites Virais, Brasília: Ministério da Saúde

Brites-Alves C, Netto EM, Brites C (2015) Coinfection by Hepatitis C Is Strongly Associated with Abnormal CD4/CD8 Ratio in HIV Patients under Stable ART in Salvador, Brazil. Journal of Immunology Research 2015:174215

Brumme ZL, Dong WW, Chan KJ, Hogg RS, Montaner JS, O'Shaughnessy MV, Harrigan PR (2003) Influence of polymorphisms within the CX3CR1 and MDR-1 genes on initial antiretroviral therapy response. AIDS 17:201-208

CDC (1992) 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. MMWR Recomm Rep 41:1-19

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4:7

Chew CS, Cherry CL, Kamarulzaman A, Yien TH, Aghafar Z, Price P (2011) A longitudinal study of the effects of ART on plasma chemokine levels in Malaysian HIV patients. Dis Markers 31:303-309

Coelho AV, De Moura RR, Da Silva RC, Kamada AJ, Guimaraes RL, Brandao LA, Coelho HF, Crovella S (2015) Meta-Analysis and Time Series Modeling Allow a Systematic Review of Primary HIV-1 Drug-Resistant Prevalence in Latin America and Caribbean. Curr HIV Res 13:125-142

Cressey TR, Lallemant M (2007) Pharmacogenetics of antiretroviral drugs for the treatment of HIV-infected patients: an update. Infect Genet Evol 7:333-342

De Cock KM, Jaffe HW, Curran JW (2011) Reflections on 30 years of AIDS. Emerg Infect Dis 17:1044-1048

Doitsh G, Galloway NL, Geng X, Yang Z, Monroe KM, Zepeda O, Hunt PW, Hatano H, Sowinski S, Munoz-Arias I, Greene WC (2014) Cell death by pyroptosis drives CD4 T-cell depletion in HIV-1 infection. Nature 505:509-514

Fellay J, Marzolini C, Meaden ER, Back DJ, Buclin T, Chave JP, Decosterd LA, Furrer H, Opravil M, Pantaleo G, Retelska D, Ruiz L, Schinkel AH, Vernazza P, Eap CB, Telenti A (2002) Response to antiretroviral treatment in HIV-1-infected individuals with allelic variants of the multidrug resistance transporter 1: a pharmacogenetics study. Lancet 359:30-36

Fernandez S, Rosenow AA, James IR, Roberts SG, Nolan RC, French MA, Price P (2006) Recovery of CD4+ T Cells in HIV Patients With a Stable Virologic Response to Antiretroviral Therapy Is Associated With Polymorphisms of Interleukin-6 and Central Major Histocompatibility Complex Genes. JAIDS Journal of Acquired Immune Deficiency Syndromes 41:1-5

Franke RM, Gardner ER, Sparreboom A (2010) Pharmacogenetics of drug transporters. Curr Pharm Des 16:220-230

Gaardbo JC, Hartling HJ, Gerstoft J, Nielsen SD (2012) Incomplete Immune Recovery in HIV Infection: Mechanisms, Relevance for Clinical Care, and Possible Solutions. Clinical and Developmental Immunology 2012:670957

Gatanaga H, Hayashida T, Tsuchiya K, Yoshino M, Kuwahara T, Tsukada H, Fujimoto K, Sato I, Ueda M, Horiba M, Hamaguchi M, Yamamoto M, Takata N, Kimura A, Koike T, Gejyo F, Matsushita S, Shirasaka T, Kimura S, Oka S (2007) Successful efavirenz dose reduction in HIV type 1-infected individuals with cytochrome P450 2B6 *6 and *26. Clin Infect Dis 45:1230-1237

Haas DW, Gebretsadik T, Mayo G, Menon UN, Acosta EP, Shintani A, Floyd M, Stein CM, Wilkinson GR (2009) Associations between CYP2B6 Polymorphisms and Pharmacokinetics after a Single Dose of Nevirapine or Efavirenz in African Americans. Journal of Infectious Diseases 199:872-880

Haas DW, Geraghty DE, Andersen J, Mar J, Motsinger AA, D'Aquila RT, Unutmaz D, Benson CA, Ritchie MD, Landay A, Group ACT (2006) Immunogenetics of CD4 lymphocyte count recovery during antiretroviral therapy: An AIDS Clinical Trials Group study. J Infect Dis 194:1098-1107

Hulgan T, Raffanti S, Kheshti A, Blackwell RB, Rebeiro PF, Barkanic G, Ritz B, Sterling TR (2005) CD4 Lymphocyte Percentage Predicts Disease Progression in HIV-Infected Patients Initiating Highly Active Antiretroviral Therapy with CD4 Lymphocyte Counts >350 Lymphocytes/mm3. Journal of Infectious Diseases 192:950-957

Hutchinson JF (2001) The Biology and Evolution of HIV. Annual Review of Anthropology 30:85-108

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

Jung N, Lehmann C, Rubbert A, Knispel M, Hartmann P, van Lunzen J, Stellbrink H-J, Faetkenheuer G, Taubert D (2008) Relevance of the Organic Cation Transporters 1 and 2 for Antiretroviral Drug Therapy in Human Immunodeficiency Virus Infection. Drug Metabolism and Disposition 36:1616-1623

Khaitan A, Unutmaz D (2011) Revisiting Immune Exhaustion During HIV Infection. Current HIV/AIDS reports 8:4-11

Kis O, Robillard K, Chan GNY, Bendayan R (2010) The complexities of antiretroviral drug-drug interactions: role of ABC and SLC transporters. Trends in Pharmacological Sciences 31:22-35

Levy Y (2006) Cytokine-based modulation of immune function in HIV infection. Current opinion in HIV and AIDS 1:69-73

Li T, Wu N, Dai Y, Qiu Z, Han Y, Xie J, Zhu T, Li Y (2011) Reduced thymic output is a major mechanism of immune reconstitution failure in HIV-infected patients after long-term antiretroviral therapy. Clinical Infectious Diseases 53:944-951

Liao W, Lin J-X, Leonard WJ (2011) IL-2 Family Cytokines: New Insights into the Complex Roles of IL-2 as a Broad Regulator of T helper Cell Differentiation. Current Opinion in Immunology 23:598-604

Marziali M, De Santis W, Carello R, Leti W, Esposito A, Isgro A, Fimiani C, Sirianni MC, Mezzaroma I, Aiuti F (2006) T-cell homeostasis alteration in HIV-1 infected subjects with low CD4 T-cell count despite undetectable virus load during HAART. AIDS 20:2033-2041

Miller MF, Haley C, Koziel MJ, Rowley CF (2005) Impact of Hepatitis C Virus on Immune Restoration in HIV-Infected Patients Who Start Highly Active Antiretroviral Therapy: A Meta-analysis. Clinical Infectious Diseases 41:713-720

Mishra A, Sullivan L, Caligiuri MA (2014) Molecular Pathways: Interleukin-15 Signaling in Health and in Cancer. Clinical cancer research : an official journal of the American Association for Cancer Research 20:2044-2050

Nasi M, Pinti M, Bugarini R, Troiano L, Lugli E, Bellodi C, Mussini C, Borghi V, Trenti T, Balli F, Esposito R, Cossarizza A (2005) Genetic polymorphisms of Fas (CD95) and Fas ligand (CD178) influence the rise in CD4+ T cell count after antiretroviral therapy in drug-naive HIV-positive patients. Immunogenetics 57:628-635

Pathai S, Bajillan H, Landay AL, High KP (2014) Is HIV a model of accelerated or accentuated aging? J Gerontol A Biol Sci Med Sci 69:833-842

Paul WE (2015) History of Interleukin-4. Cytokine 75:3-7

Peraire J, Viladés C, Pacheco YM, López-Dupla M, Domingo P, Gutiérrez M, Rosado I, Leal M, Richart C, Vidal F (2014) Evaluation of the pharmacogenetics of immune recovery in treated HIV-infected patients. Expert Opinion on Drug Metabolism and Toxicology 10:81-101

R Core Team (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/. In, Vienna

Smith AJ, Humphries SE (2009) Cytokine and cytokine receptor gene polymorphisms and their functionality. Cytokine Growth Factor Rev 20:43-59

Swart M, Whitehorn H, Ren Y, Smith P, Ramesar RS, Dandara C (2012) PXR and CAR single nucleotide polymorphisms influence plasma efavirenz levels in South African HIV/AIDS patients. BMC Med Genet 13:112

Sylwester AW, Mitchell BL, Edgar JB, Taormina C, Pelte C, Ruchti F, Sleath PR, Grabstein KH, Hosken NA, Kern F, Nelson JA, Picker LJ (2005) Broadly targeted human cytomegalovirus-specific CD4(+) and CD8(+) T cells dominate the memory compartments of exposed subjects. The Journal of Experimental Medicine 202:673-685

Tan R, Westfall AO, Willig JH, Mugavero MJ, Saag MS, Kaslow RA, Kempf MC (2008) Clinical Outcome of HIV-Infected Antiretroviral-Naive Patients With Discordant Immunologic and Virologic Responses to Highly Active Antiretroviral Therapy. JAIDS Journal of Acquired Immune Deficiency Syndromes 47:553-558 510.1097/QAI.1090b1013e31816856c31816855

Teixeira L, Valdez H, McCune JM, Koup RA, Badley AD, Hellerstein MK, Napolitano LA, Douek DC, Mbisa G, Deeks S, Harris JM, Barbour JD, Gross BH, Francis IR, Halvorsen R, Asaad R, Lederman MM (2001) Poor CD4 T cell restoration after suppression of HIV-1 replication may reflect lower thymic function. AIDS 15:1749-1756

Telenti A, Zanger UM (2008) Pharmacogenetics of anti-HIV drugs. Annu Rev Pharmacol Toxicol 48:227-256

Torres RA, Lewis W (2014) Aging and HIV/AIDS: pathogenetic role of therapeutic side effects. Lab Invest 94:120-128

Tozzi V (2009) Pharmacogenetics of antiretrovirals. Antiviral Res 85:190-200

Vella S, Schwartlander B, Sow SP, Eholie SP, Murphy RL (2012) The history of antiretroviral therapy and of its implementation in resource-limited areas of the world. AIDS 26:1231-1241

Wada T, Gao J, Xie W PXR and CAR in energy metabolism. Trends in Endocrinology & Metabolism 20:273-279

World Health Organization (2015) 2015 Progress Report on the Global Plan towards the elimination of new HIV infections among children and keeping their mothers alive.

Table 1. General characteristics of the recruited individuals living with HIV-1 and stratification according to antiretroviral therapy immunologic response.

Variables	Total (n=135)	Immunologic non-responders (n=82)	Immunologic responders (n=53)	Univariate analysis
Sex, n (%)				
Females	65 (48.2)	32 (39.0)	33 (62.3)	*reference*
Males	70 (51.8)	50 (61.0)	20 (37.7)	OR=2.56 (95% CI=1.20-5.60); **p=0.01**
Age at ART start date				
Median (IQR) [min.-max.]	33 (29-39) [21-54]	34.5 (29-40) [21-54]	33 (26-37) [21-53]	W=1813.5; p=0.11
Pre-ART CD4+ T cell counts, cells/µL of peripheral whole blood				
Median (IQR) [min.-max.]	33 (29-39) [21-54]	187.5 (86.8-281.5) [13-543]	375.5 (274.2-525.8) [48-1029]	W=669.5; **p<0.001**
By categories, n (%)				
Not available	39 (28.9)	8 (9.8)	7 (13.2)	*(not included in the comparison)*
Less than 200	65 (48.2)	40 (48.8)	5 (9.4)	
Between 200 and 350	25 (18.5)	23 (28.0)	15 (28.4)	X^2=34.9; df=3; **p<0.001**
Between 350 and 500	6 (4.4)	10 (12.2)	13 (24.5)	
Over 500	39 (28.9)	1 (1.2)	13 (24.5)	
Ancestry proportions estimated by AIMs, mean percentage (SD) [min.-max.]				
African	32.5 (16.1) [0.0-75.3]	30.2 (15.9) [0.0-70.2]	31.3 (13.8) [4.4-63.3]	
Amerindian	12.4 (9.8) [0.0-52.7]	13.2 (10.7) [0.0-52.7]	9.0 (9.2) [0.0-34.1]	W=1912; p=0.55
European	55.1 (18.2) [11.8-99.9]	56.2 (18.5) [11.8-99.9]	59.6 (15.7) [28.2-91.2]	
AIDS-defining conditions, CDC classification system				
Not available	93 (68.9)	49 (59.8)	44 (83.0)	*(not included in the comparison)*
A stage	1 (0.7)	0 (0.0)	1 (1.9)	
B stage	10 (7.4)	9 (10.9)	1 (1.9)	X^2=4.5; df=2; p=0.11
C stage	31 (23.0)	24 (29.3)	7 (13.2)	
First ART regimens				
Not available	9 (6.7)	5 (6.1)	4 (7.5)	*(not included in the comparison)*
Monotherapy	3 (2.2)	1 (1.2)	2 (3.8)	
Detailed regimens: [NRTI]+3TC [third option]				
ABC+3TC EFZ	1 (0.7)	1 (1.2)	0 (0.0)	
AZT+3TC ATV/r	4 (2.9)	3 (3.7)	1 (1.9)	
AZT+3TC EFZ	48 (35.7)	29 (35.4)	19 (35.8)	
AZT+3TC FPV/r	1 (0.7)	1 (1.2)	0 (0.0)	
AZT+3TC IDV/r	0 (0.0)	0 (0.0)	0 (0.0)	
AZT+3TC LPV/r	36 (26.7)	19 (23.2)	17 (32.1)	X^2=6.6; df=7; p=0.47
AZT+3TC NVP	4 (2.9)	2 (2.4)	2 (3.8)	
TDF+3TC ATV/r	3 (2.2)	1 (1.2)	2 (3.8)	
TDF+3TC EFZ	26 (19.3)	20 (24.4)	6 (11.3)	
TDF+3TC LPV/r	0 (0.0)	0 (0.0)	0 (0.0)	
ART regimens, stratified by [third option]				
NNTRI	79 (58.5)	52 (63.4)	27 (50.9)	*reference*
PI/r	44 (32.6)	24 (29.3)	20 (37.7)	OR=0.63 (95% CI=0.28-1.42); p=0.25
[NRTI] choice alongside 3TC				
ABC	1 (0.7)	1 (1.2)	0 (0.0)	*(not included in the comparison)*
AZT	93 (68.9)	54 (65.9)	39 (73.6)	*reference*
TDF	29 (21.5)	21 (25.6)	8 (15.1)	OR=1.89 (95% CI =0.71-5.46); p=0.19

3TC – lamivudine, 95% CI – 95% confidence interval, ABC – abacavir, AIDS – acquired immunodeficiency syndrome, ART – antiretroviral therapy, ATV/r – ritonavir-boosted atazanavir, CDC – Center for Disease Control (USA), df – degrees of freedom, EFZ – efavirenz, FPV/r – ritonavir-boosted fosamprenavir , IDV/r – ritonavir-boosted indinavir, IQR – interquartile range, LPV/r – ritonavir-boosted lopinavir, min.-max. – minimum and maximum values, NNRTI – non-nucleoside analog reverse transcriptase inhibitor, NRTI – nucleoside analog reverse transcriptase inhibitor, NVP – nevirapine, OR – odds ratio, p – p-value, PI – protease inhibitor, SD – standard deviation, TDF – tenofovir, W – Mann-Whitney test statistic, X2 – chi-squared test statistic.

Table 2. Coinfections serological status of the recruited patients living with HIV-1 stratification according to antiretroviral therapy immunologic response.

Etiologic agent	Total n=135 (%)	Immunologic non-responders n=82 (%)	Immunologic responders n=53 (%)
Cytomegalovirus (CMV)			
IgM test			
Untested	90 (66.7)	59 (72.0)	31 (58.5)
Negative	44 (32.6)	23 (28.0)	21 (39.6)
Positive	1 (0.7)	0 (0.0)	1 (1.9)
IgG test			
Untested	90 (66.7)	59 (72.0)	31 (58.5)
Negative	1 (0.7)	1 (1.2)	0 (0.0)
Positive	44 (32.6)	22 (26.8)	22 (41.5)
Hepatitis B virus (HBV)			
Untested	59 (43.7)	37 (45.1)	22 (41.5)
Susceptible	38 (28.1)	21 (25.6)	17 (32.1)
Chronic infection	0 (0.0)	0 (0.0)	0 (0.0)
Immune due to natural infection	9 (6.7)	6 (7.3)	3 (5.7)
Immune due to vaccination	29 (21.5)	18 (22.0)	11 (20.8)
Hepatitis C virus (HCV)			
Untested	68 (50.4)	41 (50.0)	27 (50.9)
Negative	67 (49.6)	41 (50.0)	26 (49.1)
Positive	0 (0.0)	0 (0.0)	0 (0.0)
Human T-lymphotropic virus type 1 or 2 (HTLV-1/2)			
Untested	114 (84.4)	76 (92.7)	38 (71.7)
Negative	21 (15.6)	6 (7.3)	15 (28.3)
Positive	0 (0.0)	0 (0.0)	0 (0.0)
Toxoplasma gondii			
IgM test			
Untested	90 (66.7)	59 (72.0)	31 (58.5)
Negative	45 (33.3)	23 (28.0)	22 (41.5)
Positive	0 (0.0)	0 (0.0)	0 (0.0)
IgG test			
Untested	90 (66.7)	59 (72.0)	31 (58.5)
Negative	11 (8.1)	7 (8.5)	4 (7.5)
Positive	34 (25.2)	16 (19.5)	18 (34.0)
***Treponema pallidum* (VDRL test)**			
Untested	47 (34.8)	30 (36.6)	17 (32.1)
Negative	68 (50.4)	38 (46.3)	30 (56.6)
Positive	20 (14.8)	14 (17.1)	6 (11.3)

Table 3. Allele and genotype frequencies of variants showing statistically significant genetic association with immunologic outcome.

#	SNP	Type	Alleles		Gene	Gene function/pathway	Alelle and genotype frequencies								X^2	DF	p
							Immunologic non-responders				Immunologic responders						
			A1	A2			A2 (%)	A1/A1 (%)	A1/A2 (%)	A2/A2 (%)	A2 (%)	A1/A1 (%)	A1/A2 (%)	A2/A2 (%)			
5	rs3003596	Downstream gene region	A	G	*NR1I3*	Regulation of drugs and endobiotic metabolism	79 (48.2)	21 (25.6)	43 (52.4)	18 (22.0)	37 (35.6)	25 (48.1)	17 (32.7)	10 (19.2)	10.10	2	***0.01***
26	rs2069762	Upstream gene region	A	C	*IL2*	Regulates T and B lymphocytes proliferation	33 (20.1)	49 (59.8)	33 (40.2)	0 (0.0)	32 (30.2)	25 (47.2)	24 (45.3)	4 (7.5)	73.13	2	***0.03***
27	rs10519613	3' untranslated region	C	A	*IL15*	T lymphocytes activation	17 (10.5)	67 (82.7)	11 (13.6)	3 (3.7)	21 (19.8)	34 (64.2)	17 (32.1)	2 (3.8)	67.10	2	***0.03***
38	rs2243250	Upstream gene region	G	A	*IL4*	Immunoregulation	69 (42.6)	20 (24.7)	53 (65.4)	8 (9.9)	32 (32.7)	23 (46.9)	20 (40.8)	6 (12.2)	8.02	2	***0.02***
74	rs129081	3' untranslated region	C	G	*ABCC1*	Membrane transport, antiretroviral drugs efflux	78 (47.6)	24 (29.3)	38 (46.3)	20 (24.4)	39 (37.5)	16 (30.8)	33 (63.5)	3 (5.8)	82.13	2	***0.02***

- order (by chromosome and genomic position) in which the candidate and ancestry informative markers are listed on the Supplementary Tables, DF – degrees of freedom, X^2 – value of chi-squared statistic from chi-square test of independence, p – p-value.

Supplementary Table 1. List of all analyzed single nucleotide polymorphisms (SNPs) in the recruited patients.

#	SNP	Chromosome	Position	Gene (if applicable)	Official gene symbol	Function or pathway	Variant type
1	rs4908343	1	27605187	(ancestry informative marker)			
2	rs1325502	1	41894599	(ancestry informative marker)			
3	rs3737576	1	101244007	(ancestry informative marker)			
4	rs7554936	1	151150013	(ancestry informative marker)			
5	rs3003596	1	161234427	Nuclear receptor subfamily 1 group I member 3	*NR1I3*	Regulation of drugs and endobiotic metabolism	Downstream gene region
6	rs2222202	1	206772036	Interleukin 10	*IL10*	Immunoregulation	Intronic
7	rs1800871	1	206773289	Interleukin 10	*IL10*	Immunoregulation	Upstream gene region
8	rs1800890	1	206776020	Interleukin 10	*IL10*	Immunoregulation	Upstream gene region
9	rs798443	2	7828144	(ancestry informative marker)			
10	rs4666200	2	29315545	(ancestry informative marker)			
11	rs4670767	2	37714253	(ancestry informative marker)			
12	rs13400937	2	79637797	(ancestry informative marker)			
13	rs260690	2	108963282	(ancestry informative marker)			
14	rs16944	2	112837290	Interleukin 1 beta	*IL1B*	Pro-inflammation cytokine	Upstream gene region
15	rs10496971	2	145012376	(ancestry informative marker)			
16	rs9809104	3	39104938	(ancestry informative marker)			
17	rs6548616	3	79350425	(ancestry informative marker)			
18	rs12629908	3	120803869	(ancestry informative marker)			
19	rs9845457	3	136195634	(ancestry informative marker)			
20	rs1513181	3	188857208	(ancestry informative marker)			
21	rs10007810	4	41552347	(ancestry informative marker)			
22	rs115770495	4	88090508	*ATP binding cassette subfamily G member 2*	*ABCG2*	Membrane transport, antiretroviral drugs efflux	3’ untranslated region
23	rs1448784	4	88091168	ATP binding cassette subfamily G member 2	*ABCG2*	Membrane transport, antiretroviral drugs efflux	3’ untranslated region
24	rs2231142	4	88131171	ATP binding cassette subfamily G member 2	*ABCG2*	Membrane transport, antiretroviral drugs efflux	Missense (Gln141Lys)
25	rs7657799	4	104454266	(ancestry informative marker)			
26	rs2069762	4	122456825	Interleukin 12	*IL2*	Regulates T and B lymphocytes proliferation	Upstream gene region
27	rs10519613	4	141732931	Interleukin 15	*IL15*	T lymphocytes activation	3’ untranslated region
28	rs10833	4	141733394	Interleukin 15	*IL15*	T lymphocytes activation	3’ untranslated region

#	SNP	Chromosome	Position	Gene (if applicable)	Official gene symbol	Function or pathway	Variant type
29	rs316598	5	2364512	(ancestry informative marker)			
30	rs870347	5	6844922	(ancestry informative marker)			
31	rs1494555	5	35871088	Interleukin 7 receptor	*IL7R*	Signal transducer, regulates T lymphocytes development	Missense (Val138Ile)
32	rs11567762	5	35873099	Interleukin 7 receptor	*IL7R*	Signal transducer, regulates T lymphocytes development	Intronic
33	rs6897932	5	35874473	Interleukin 7 receptor	*IL7R*	Signal transducer, regulates T lymphocytes development	Missense (Thr244Ile)
34	rs3822731	5	35875138	Interleukin 7 receptor	*IL7R*	Signal transducer, regulates T lymphocytes development	Intronic
35	rs987106	5	35875491	Interleukin 7 receptor	*IL7R*	Signal transducer, regulates T lymphocytes development	Intronic
36	rs3194051	5	35876172	Interleukin 7 receptor	*IL7R*	Signal transducer, regulates T lymphocytes development	Missense (Ile356Val)
37	rs6451722	5	43711276	(ancestry informative marker)			
38	rs2243250	5	132673462	Interleukin 4	*IL4*	Immunoregulation	Upstream gene region
39	rs6422347	5	178436082	(ancestry informative marker)			
40	rs1040045	6	4746925	(ancestry informative marker)			
41	rs2504853	6	12534879	(ancestry informative marker)			
42	rs4463276	6	144734195	(ancestry informative marker)			
43	rs731257	7	12629626	(ancestry informative marker)			
44	rs3842	7	87504050	ATP binding cassette subfamily B member 1	*ABCB1*	Membrane transport, antiretroviral drugs efflux	3’ untranslated region
45	rs2235048	7	87509195	ATP binding cassette subfamily B member 1	*ABCB1*	Membrane transport, antiretroviral drugs efflux	Intronic
46	rs1128503	7	87550285	ATP binding cassette subfamily B member 1	*ABCB1*	Membrane transport, antiretroviral drugs efflux	Synonymous (Gly412Gly)
47	rs2214102	7	87600185	ATP binding cassette subfamily B member 1	*ABCB1*	Membrane transport, antiretroviral drugs efflux	5’ untranslated region
48	rs4646437	7	99767460	Cytochrome P450 family 3 subfamily A member 4	*CYP3A4*	Antiretroviral drugs metabolism	Intronic
49	rs7803075	7	131057307	(ancestry informative marker)			
50	rs10236187	7	139747578	(ancestry informative marker)			
51	rs10108270	8	4333271	(ancestry informative marker)			
52	rs3943253	8	13501991	(ancestry informative marker)			
53	rs1471939	8	29083788	(ancestry informative marker)			
54	rs4746136	10	73541236	(ancestry informative marker)			
55	rs2234767	10	88989499	Fas cell surface death receptor	*FAS*	Apoptosis	Upstream gene region
56	rs1800682	10	88990206	Fas cell surface death receptor	*FAS*	Apoptosis	Upstream gene region
57	rs4918842	10	113557053	(ancestry informative marker)			
58	rs2946788	11	23988984	(ancestry informative marker)			
59	rs11568629	11	62984340	Solute carrier family 22 member 6	*SLC22A6*	Membrane transport, antiretroviral drugs influx/efflux	Synonymous (Pro117Pro)

#	SNP	Chromosome	Position	Gene (if applicable)	Official gene symbol	Function or pathway	Variant type
60	rs11568628	11	62984439	Solute carrier family 22 member 6	*SLC22A6*	Membrane transport, antiretroviral drugs influx/efflux	Synonymous (Pro84Pro)
61	rs4149170	11	62984817	Solute carrier family 22 member 6	*SLC22A6*	Membrane transport, antiretroviral drugs influx/efflux	5' untranslated region
62	rs3135932	11	117993348	Interleukin 10 receptor subunit alpha	*IL10RA*	Signal transducer, immunoregulation	Missense (Ser159Gly)
63	rs9610	11	118001371	Interleukin 10 receptor subunit alpha	*IL10RA*	Signal transducer, immunoregulation	3' untranslated region
64	rs2416791	12	11548554	(ancestry informative marker)			
65	rs772262	12	55769950	(ancestry informative marker)			
66	rs2069709	12	68159923	Interferon Gamma	*IFNG*	Immunoregulation	Upstream gene region
67	rs9319336	13	27050219	(ancestry informative marker)			
68	rs7997709	13	34273600	(ancestry informative marker)			
69	rs9530435	13	75419751	(ancestry informative marker)			
70	rs9522149	13	111174820	(ancestry informative marker)			
71	rs1760921	14	20349972	(ancestry informative marker)			
72	rs3784230	14	105212718	(ancestry informative marker)			
73	rs762551	15	74749576	Cytochrome P450 family 1 subfamily A member 2	*CYP1A2*	Antiretroviral drugs metabolism	Intronic
74	rs129081	16	16142082	ATP binding cassette subfamily C member 1	*ABCC1*	Membrane transport, antiretroviral drugs efflux	3' untranslated region
75	rs113264879	16	16142164	ATP binding cassette subfamily C member 1	*ABCC1*	Membrane transport, antiretroviral drugs efflux	3' untranslated region
76	rs4148380	16	16142574	ATP binding cassette subfamily C member 1	*ABCC1*	Membrane transport, antiretroviral drugs efflux	3' untranslated region
77	rs8056298	16	16142666	ATP binding cassette subfamily C member 1	*ABCC1*	Membrane transport, antiretroviral drugs efflux	3' untranslated region
78	rs212091	16	16142793	ATP binding cassette subfamily C member 1	*ABCC1*	Membrane transport, antiretroviral drugs efflux	3' untranslated region
79	rs16967632	16	16142926	ATP binding cassette subfamily C member 1	*ABCC1*	Membrane transport, antiretroviral drugs efflux	3' untranslated region
80	rs2107538	17	35880776	C-C motif chemokine ligand 5	*CCL5*	Immunoregulation	Upstream gene region
81	rs11652805	17	64991033	(ancestry informative marker)			
82	rs2125345	17	75786110	(ancestry informative marker)			
83	rs4891825	18	70200427	(ancestry informative marker)			
84	rs8192726	19	40848591	Cytochrome P450 family 2 subfamily A member 6	*CYP2A6*	Antiretroviral drugs metabolism	Intronic
85	rs8192709	19	40991369	Cytochrome P450 family 2 subfamily B member 6	*CYP2B6*	Antiretroviral drugs metabolism	Missense (Arg22Cys)
86	rs28399499	19	41012316	Cytochrome P450 family 2 subfamily B member 6	*CYP2B6*	Antiretroviral drugs metabolism	Missense (Ile328Thr)
87	rs34097093	19	41012465	Cytochrome P450 family 2 subfamily B member 6	*CYP2B6*	Antiretroviral drugs metabolism	Stop (Arg378*)
88	rs28399502	19	41016965	Cytochrome P450 family 2 subfamily B member 6	*CYP2B6*	Antiretroviral drugs metabolism	3' untranslated region
89	rs707265	19	41018182	Cytochrome P450 family 2 subfamily B member 6	*CYP2B6*	Antiretroviral drugs metabolism	3' untranslated region
90	rs1042389	19	41018248	Cytochrome P450 family 2 subfamily B member 6	*CYP2B6*	Antiretroviral drugs metabolism	3' untranslated region

#	SNP	Chromosome	Position	Gene (if applicable)	Official gene symbol	Function or pathway	Variant type
91	rs6104567	20	10214785	(ancestry informative marker)			
92	rs3907047	20	55384376	(ancestry informative marker)			
93	rs4821004	22	31970372	(ancestry informative marker)			
94	rs5768007	22	47812123	(ancestry informative marker)			

Supplementary Table 2. Allele and genotype frequencies of all variants analyzed.

#	SNP	Alleles		Gene (if applicable)	Immunologic non-responders					Immunologic responders					GCR	HWE p
		A1	A2		A1 (%)	A2 (%)	A1/A1 (%)	A1/A2 (%)	A2/A2 (%)	A1 (%)	A2 (%)	A1/A1 (%)	A1/A2 (%)	A2/A2 (%)		
1	rs4908343	A	G	(ancestry informative marker)	92 (56.1)	72 (43.9)	22 (26.8)	48 (58.5)	12 (14.6)	58 (55.8)	46 (44.2)	16 (30.8)	26 (50.0)	10 (19.2)	98.1	0.22
2	rs1325502	G	A	(ancestry informative marker)	104 (63.4)	60 (36.6)	33 (40.2)	38 (46.3)	11 (13.4)	69 (65.1)	37 (34.9)	23 (43.4)	23 (43.4)	7 (13.2)	100	0.85
3	rs3737576	A	G	(ancestry informative marker)	149 (90.9)	15 (9.1)	68 (82.9)	13 (15.9)	1 (1.2)	95 (89.6)	11 (10.4)	44 (83.0)	7 (13.2)	2 (3.8)	100	0.10
4	rs7554936	G	A	(ancestry informative marker)	94 (57.3)	70 (42.7)	31 (37.8)	32 (39.0)	19 (23.2)	56 (52.8)	50 (47.2)	13 (24.5)	30 (56.6)	10 (18.9)	100	0.49
5	rs3003596	A	G	*NR1I3*	85 (51.8)	79 (48.2)	21 (25.6)	43 (52.4)	18 (22.0)	67 (64.4)	37 (35.6)	25 (48.1)	17 (32.7)	10 (19.2)	98.1	0.3
6	rs2222202	G	A	*IL10*	107 (65.2)	57 (34.8)	39 (47.6)	29 (35.4)	14 (17.1)	75 (70.8)	31 (29.2)	28 (52.8)	19 (35.8)	6 (11.3)	100	0.03
7	rs1800871	G	A	*IL10*	103 (62.8)	61 (37.2)	33 (40.2)	37 (45.1)	12 (14.6)	64 (60.4)	42 (39.6)	20 (37.7)	24 (45.3)	9 (17.0)	100	0.71
8	rs1800890	T	A	*IL10*	120 (73.2)	44 (26.8)	44 (53.7)	32 (39.0)	6 (7.3)	73 (68.9)	33 (31.1)	27 (50.9)	19 (35.8)	7 (13.2)	100	0.4
9	rs798443	A	G	(ancestry informative marker)	84 (51.9)	78 (48.1)	19 (23.5)	46 (56.8)	16 (19.8)	62 (58.5)	44 (41.5)	17 (32.1)	28 (52.8)	8 (15.1)	98.8	0.23
10	rs4666200	A	G	(ancestry informative marker)	94 (57.3)	70 (42.7)	26 (31.7)	42 (51.2)	14 (17.1)	60 (56.6)	46 (43.4)	16 (30.2)	28 (52.8)	9 (17.0)	100	0.6
11	rs4670767	C	A	(ancestry informative marker)	136 (82.9)	28 (17.1)	56 (68.3)	24 (29.3)	2 (2.4)	96 (90.6)	10 (9.4)	43 (81.1)	10 (18.9)	0 (0.0)	100	1.00
12	rs13400937	C	A	(ancestry informative marker)	84 (52.5)	76 (47.5)	23 (28.8)	38 (47.5)	19 (23.8)	57 (53.8)	49 (46.2)	18 (34.0)	21 (39.6)	14 (26.4)	97.6	0.22
13	rs260690	A	C	(ancestry informative marker)	104 (63.4)	60 (36.6)	36 (43.9)	32 (39.0)	14 (17.1)	76 (71.7)	30 (28.3)	25 (47.2)	26 (49.1)	2 (3.8)	100	0.7
14	***rs16944***	***A***	***G***	***IL1B***	***79 (53.4)***	***69 (46.6)***	***22 (29.7)***	***35 (47.3)***	***17 (23.0)***	***48 (52.2)***	***44 (47.8)***	***14 (30.4)***	***20 (43.5)***	***12 (26.1)***	***77***	***0.37***
15	***rs10496971***	***A***	***C***	***(ancestry informative marker)***	***82 (89.1)***	***10 (10.9)***	***36 (78.3)***	***10 (21.7)***	***0 (0.0)***	***54 (90.0)***	***6 (10.0)***	***24 (80.0)***	***6 (20.0)***	***0 (0.0)***	***12.7***	***1.00***
16	rs9809104	A	G	(ancestry informative marker)	93 (58.1)	67 (41.9)	25 (31.3)	43 (53.8)	12 (15.0)	61 (57.5)	45 (42.5)	19 (35.8)	23 (43.4)	11 (20.8)	97.6	1.00
17	rs6548616	A	G	(ancestry informative marker)	85 (53.8)	73 (46.2)	23 (29.1)	39 (49.4)	17 (21.5)	60 (57.7)	44 (42.3)	13 (25.0)	34 (65.4)	5 (9.6)	94.4	0.21
18	rs12629908	G	A	(ancestry informative marker)	138 (84.1)	26 (15.9)	57 (69.5)	24 (29.3)	1 (1.2)	88 (86.3)	14 (13.7)	37 (72.5)	14 (27.5)	0 (0.0)	96.2	0.31
19	***rs9845457***	***NA***	***NA***	***(ancestry informative marker)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***0***	***NA***
20	rs1513181	G	A	(ancestry informative marker)	121 (73.8)	43 (26.2)	45 (54.9)	31 (37.8)	6 (7.3)	82 (77.4)	24 (22.6)	33 (62.3)	16 (30.2)	4 (7.5)	100	0.49
21	rs10007810	G	A	(ancestry informative marker)	92 (56.1)	72 (43.9)	26 (31.7)	40 (48.8)	16 (19.5)	56 (52.8)	50 (47.2)	13 (24.5)	30 (56.6)	10 (18.9)	100	0.73
22	rs115770495	G	A	*ABCG2*	161 (98.2)	3 (1.8)	79 (96.3)	3 (3.7)	0 (0.0)	101 (95.3)	5 (4.7)	49 (92.5)	3 (5.7)	1 (1.9)	100	0.10
23	rs1448784	A	G	*ABCG2*	162 (98.8)	2 (1.2)	80 (97.6)	2 (2.4)	0 (0.0)	103 (97.2)	3 (2.8)	50 (94.3)	3 (5.7)	0 (0.0)	100	1.00
24	rs2231142	C	A	*ABCG2*	150 (91.5)	14 (8.5)	68 (82.9)	14 (17.1)	0 (0.0)	93 (87.7)	13 (12.3)	40 (75.5)	13 (24.5)	0 (0.0)	100	0.36
25	rs7657799	A	C	(ancestry informative marker)	115 (71.9)	45 (28.1)	43 (53.8)	29 (36.3)	8 (10.0)	78 (73.6)	28 (26.4)	28 (52.8)	22 (41.5)	3 (5.7)	97.6	0.67
26	rs2069762	A	C	*IL2*	131 (79.9)	33 (20.1)	49 (59.8)	33 (40.2)	0 (0.0)	74 (69.8)	32 (30.2)	25 (47.2)	24 (45.3)	4 (7.5)	100	0.10
27	rs10519613	C	A	*IL15*	145 (89.5)	17 (10.5)	67 (82.7)	11 (13.6)	3 (3.7)	85 (80.2)	21 (19.8)	34 (64.2)	17 (32.1)	2 (3.8)	98.8	0.14
28	***rs10833***	***G***	***A***	***IL15***	***123 (79.9)***	***31 (20.1)***	***50 (64.9)***	***23 (29.9)***	***4 (5.2)***	***77 (77.0)***	***23 (23.0)***	***29 (58.0)***	***19 (38.0)***	***2 (4.0)***	***88.2***	***0.8***
29	rs316598	G	A	(ancestry informative marker)	82 (50.6)	80 (49.4)	17 (21.0)	48 (59.3)	16 (19.8)	56 (52.8)	50 (47.2)	15 (28.3)	26 (49.1)	12 (22.6)	98.8	0.3

#	SNP	Alleles		Gene (if applicable)	Immunologic non-responders					Immunologic responders					GCR	HWE p
		A1	A2		A1 (%)	A2 (%)	A1/A1 (%)	A1/A2 (%)	A2/A2 (%)	A1 (%)	A2 (%)	A1/A1 (%)	A1/A2 (%)	A2/A2 (%)		
30	***rs870347***	***NA***	***NA***	***(ancestry informative marker)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***0***	***NA***
31	rs1494555	A	G	*IL7R*	121 (73.8)	43 (26.2)	42 (51.2)	37 (45.1)	3 (3.7)	73 (68.9)	33 (31.1)	26 (49.1)	21 (39.6)	6 (11.3)	100	0.53
32	rs11567762	G	A	*IL7R*	142 (86.6)	22 (13.4)	60 (73.2)	22 (26.8)	0 (0.0)	86 (81.1)	20 (18.9)	34 (64.2)	18 (34.0)	1 (1.9)	100	0.2
33	rs6897932	G	A	*IL7R*	134 (82.7)	28 (17.3)	54 (66.7)	26 (32.1)	1 (1.2)	86 (82.7)	18 (17.3)	34 (65.4)	18 (34.6)	0 (0.0)	96.9	0.12
34	rs3822731	A	G	*IL7R*	139 (84.8)	25 (15.2)	58 (70.7)	23 (28.0)	1 (1.2)	87 (82.1)	19 (17.9)	35 (66.0)	17 (32.1)	1 (1.9)	100	0.53
35	rs987106	A	T	*IL7R*	82 (50.6)	80 (49.4)	20 (24.7)	42 (51.9)	19 (23.5)	55 (51.9)	51 (48.1)	16 (30.2)	23 (43.4)	14 (26.4)	98.8	0.73
36	rs3194051	A	G	*IL7R*	113 (69.8)	49 (30.2)	40 (49.4)	33 (40.7)	8 (9.9)	77 (74.0)	27 (26.0)	30 (57.7)	17 (32.7)	5 (9.6)	96.9	0.4
37	rs6451722	G	A	(ancestry informative marker)	87 (53.0)	77 (47.0)	24 (29.3)	39 (47.6)	19 (23.2)	61 (57.5)	45 (42.5)	17 (32.1)	27 (50.9)	9 (17.0)	100	0.86
38	rs2243250	G	A	*IL4*	93 (57.4)	69 (42.6)	20 (24.7)	53 (65.4)	8 (9.9)	66 (67.3)	32 (32.7)	23 (46.9)	20 (40.8)	6 (12.2)	91.3	0.04
39	rs6422347	A	G	(ancestry informative marker)	100 (61.0)	64 (39.0)	27 (32.9)	46 (56.1)	9 (11.0)	65 (61.3)	41 (38.7)	20 (37.7)	25 (47.2)	8 (15.1)	100	0.28
40	rs1040045	A	G	(ancestry informative marker)	87 (53.0)	77 (47.0)	25 (30.5)	37 (45.1)	20 (24.4)	64 (62.7)	38 (37.3)	18 (35.3)	28 (54.9)	5 (9.8)	96.2	1.00
41	rs2504853	G	A	(ancestry informative marker)	91 (55.5)	73 (44.5)	24 (29.3)	43 (52.4)	15 (18.3)	63 (59.4)	43 (40.6)	18 (34.0)	27 (50.9)	8 (15.1)	100	0.6
42	rs4463276	A	G	(ancestry informative marker)	85 (51.8)	79 (48.2)	24 (29.3)	37 (45.1)	21 (25.6)	60 (56.6)	46 (43.4)	18 (34.0)	24 (45.3)	11 (20.8)	100	0.3
43	rs731257	G	A	(ancestry informative marker)	140 (85.4)	24 (14.6)	61 (74.4)	18 (22.0)	3 (3.7)	90 (86.5)	14 (13.5)	38 (73.1)	14 (26.9)	0 (0.0)	98.1	0.73
44	rs3842	A	G	*ABCB1*	131 (79.9)	33 (20.1)	53 (64.6)	25 (30.5)	4 (4.9)	90 (84.9)	16 (15.1)	38 (71.7)	14 (26.4)	1 (1.9)	100	0.77
45	rs2235048	A	G	*ABCB1*	99 (60.4)	65 (39.6)	30 (36.6)	39 (47.6)	13 (15.9)	71 (68.3)	33 (31.7)	24 (46.2)	23 (44.2)	5 (9.6)	98.1	1.00
46	rs1128503	G	A	*ABCB1*	112 (68.3)	52 (31.7)	39 (47.6)	34 (41.5)	9 (11.0)	75 (70.8)	31 (29.2)	27 (50.9)	21 (39.6)	5 (9.4)	100	0.69
47	rs2214102	G	A	*ABCB1*	159 (97.0)	5 (3.0)	77 (93.9)	5 (6.1)	0 (0.0)	104 (98.1)	2 (1.9)	51 (96.2)	2 (3.8)	0 (0.0)	100	1.00
48	rs4646437	G	A	*CYP3A4*	112 (68.3)	52 (31.7)	40 (48.8)	32 (39.0)	10 (12.2)	68 (64.2)	38 (35.8)	25 (47.2)	18 (34.0)	10 (18.9)	100	0.05
49	***rs7803075***	***NA***	***NA***	***(ancestry informative marker)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***NA (NA)***	***0***	***NA***
50	rs10236187	A	C	(ancestry informative marker)	114 (70.4)	48 (29.6)	39 (48.1)	36 (44.4)	6 (7.4)	75 (73.5)	27 (26.5)	28 (54.9)	19 (37.3)	4 (7.8)	95	1.00
51	rs10108270	C	A	(ancestry informative marker)	90 (54.9)	74 (45.1)	29 (35.4)	32 (39.0)	21 (25.6)	54 (51.9)	50 (48.1)	14 (26.9)	26 (50.0)	12 (23.1)	98.1	0.16
52	rs3943253	A	G	(ancestry informative marker)	121 (73.8)	43 (26.2)	46 (56.1)	29 (35.4)	7 (8.5)	76 (74.5)	26 (25.5)	28 (54.9)	20 (39.2)	3 (5.9)	96.2	0.65
53	***rs1471939***	***A***	***G***	***(ancestry informative marker)***	***125 (76.2)***	***39 (23.8)***	***0 (0.0)***	***31 (88.6)***	***4 (11.4)***	***74 (71.2)***	***30 (28.8)***	***25 (48.1)***	***24 (46.2)***	***3 (5.8)***	***40.8***	***0.5***
54	rs4746136	G	A	(ancestry informative marker)	121 (74.7)	41 (25.3)	45 (55.6)	31 (38.3)	5 (6.2)	87 (82.1)	19 (17.9)	35 (66.0)	17 (32.1)	1 (1.9)	98.8	1.00
55	rs2234767	G	A	*FAS*	140 (85.4)	24 (14.6)	60 (73.2)	20 (24.4)	2 (2.4)	97 (91.5)	9 (8.5)	44 (83.0)	9 (17.0)	0 (0.0)	100	1.00
56	rs1800682	G	A	*FAS*	97 (59.1)	67 (40.9)	31 (37.8)	35 (42.7)	16 (19.5)	63 (59.4)	43 (40.6)	18 (34.0)	27 (50.9)	8 (15.1)	100	0.59
57	rs4918842	A	G	(ancestry informative marker)	134 (81.7)	30 (18.3)	53 (64.6)	28 (34.1)	1 (1.2)	87 (82.1)	19 (17.9)	36 (67.9)	15 (28.3)	2 (3.8)	100	0.57
58	rs2946788	A	C	(ancestry informative marker)	106 (64.6)	58 (35.4)	33 (40.2)	40 (48.8)	9 (11.0)	53 (51.0)	51 (49.0)	13 (25.0)	27 (51.9)	12 (23.1)	98.1	0.86
59	***rs11568629***	***A***	***G***	***SLC22A6***	***93 (98.9)***	***1 (1.1)***	***46 (97.9)***	***1 (2.1)***	***0 (0.0)***	***64 (97.0)***	***2 (3.0)***	***31 (93.9)***	***2 (6.1)***	***0 (0.0)***	***19.6***	***1.00***
60	rs11568628	C	A	*SLC22A6*	162 (98.8)	2 (1.2)	80 (97.6)	2 (2.4)	0 (0.0)	103 (97.2)	3 (2.8)	50 (94.3)	3 (5.7)	0 (0.0)	100	1.00

#	SNP	Alleles		Gene (if applicable)	Immunologic non-responders					Immunologic responders					GCR	HWE p
		A1	A2		A1 (%)	A2 (%)	A1/A1 (%)	A1/A2 (%)	A2/A2 (%)	A1 (%)	A2 (%)	A1/A1 (%)	A1/A2 (%)	A2/A2 (%)		
61	rs4149170	G	A	*SLC22A6*	128 (80.0)	32 (20.0)	52 (65.0)	24 (30.0)	4 (5.0)	82 (82.0)	18 (18.0)	35 (70.0)	12 (24.0)	3 (6.0)	91.9	0.25
62	rs3135932	A	G	*IL10RA*	151 (92.1)	13 (7.9)	69 (84.1)	13 (15.9)	0 (0.0)	100 (94.3)	6 (5.7)	47 (88.7)	6 (11.3)	0 (0.0)	100	1.00
63	rs9610	A	G	*IL10RA*	89 (54.9)	73 (45.1)	22 (27.2)	45 (55.6)	14 (17.3)	58 (54.7)	48 (45.3)	18 (34.0)	22 (41.5)	13 (24.5)	98.8	1.00
64	rs2416791	G	A	(ancestry informative marker)	97 (59.9)	65 (40.1)	28 (34.6)	41 (50.6)	12 (14.8)	63 (59.4)	43 (40.6)	15 (28.3)	33 (62.3)	5 (9.4)	98.8	0.11
65	rs772262	G	A	(ancestry informative marker)	92 (57.5)	68 (42.5)	26 (32.5)	40 (50.0)	14 (17.5)	63 (59.4)	43 (40.6)	17 (32.1)	29 (54.7)	7 (13.2)	97.6	0.48
66	rs2069709	C	A	*IFNG*	163 (99.4)	1 (0.6)	81 (98.8)	1 (1.2)	0 (0.0)	105 (99.1)	1 (0.9)	52 (98.1)	1 (1.9)	0 (0.0)	100	1.00
67	rs9319336	A	G	(ancestry informative marker)	133 (82.1)	29 (17.9)	54 (66.7)	25 (30.9)	2 (2.5)	90 (84.9)	16 (15.1)	40 (75.5)	10 (18.9)	3 (5.7)	98.8	0.53
68	rs7997709	A	G	(ancestry informative marker)	121 (73.8)	43 (26.2)	45 (54.9)	31 (37.8)	6 (7.3)	92 (86.8)	14 (13.2)	40 (75.5)	12 (22.6)	1 (1.9)	100	0.61
69	rs9530435	G	A	(ancestry informative marker)	105 (64.0)	59 (36.0)	36 (43.9)	33 (40.2)	13 (15.9)	64 (62.7)	38 (37.3)	17 (33.3)	30 (58.8)	4 (7.8)	96.2	0.85
70	rs9522149	A	G	(ancestry informative marker)	86 (54.4)	72 (45.6)	30 (38.0)	26 (32.9)	23 (29.1)	57 (55.9)	45 (44.1)	16 (31.4)	25 (49.0)	10 (19.6)	92.5	0.01
71	rs1760921	A	G	(ancestry informative marker)	110 (68.8)	50 (31.3)	36 (45.0)	38 (47.5)	6 (7.5)	79 (76.0)	25 (24.0)	29 (55.8)	21 (40.4)	2 (3.8)	95.7	0.39
72	rs3784230	G	A	(ancestry informative marker)	95 (58.6)	67 (41.4)	28 (34.6)	39 (48.1)	14 (17.3)	65 (61.3)	41 (38.7)	20 (37.7)	25 (47.2)	8 (15.1)	98.8	1.00
73	rs762551	A	C	*CYP1A2*	105 (66.5)	53 (33.5)	35 (44.3)	35 (44.3)	9 (11.4)	80 (75.5)	26 (24.5)	32 (60.4)	16 (30.2)	5 (9.4)	96.3	0.41
74	rs129081	C	G	*ABCC1*	86 (52.4)	78 (47.6)	24 (29.3)	38 (46.3)	20 (24.4)	65 (62.5)	39 (37.5)	16 (30.8)	33 (63.5)	3 (5.8)	98.1	0.48
75	rs113264879	G	A	*ABCC1*	162 (98.8)	2 (1.2)	80 (97.6)	2 (2.4)	0 (0.0)	103 (97.2)	3 (2.8)	50 (94.3)	3 (5.7)	0 (0.0)	100	1.00
76	rs4148380	G	A	*ABCC1*	158 (96.3)	6 (3.7)	76 (92.7)	6 (7.3)	0 (0.0)	98 (92.5)	8 (7.5)	45 (84.9)	8 (15.1)	0 (0.0)	100	1.00
77	rs8056298	C	A	*ABCC1*	155 (94.5)	9 (5.5)	73 (89.0)	9 (11.0)	0 (0.0)	99 (93.4)	7 (6.6)	46 (86.8)	7 (13.2)	0 (0.0)	100	1.00
78	rs212091	A	G	*ABCC1*	140 (86.4)	22 (13.6)	60 (74.1)	20 (24.7)	1 (1.2)	90 (86.5)	14 (13.5)	38 (73.1)	14 (26.9)	0 (0.0)	96.9	0.47
79	***rs16967632***	***G***	***A***	***ABCC1***	***107 (99.1)***	***1 (0.9)***	***53 (98.1)***	***1 (1.9)***	***0 (0.0)***	***59 (98.3)***	***1 (1.7)***	***29 (96.7)***	***1 (3.3)***	***0 (0.0)***	***22.5***	***1.00***
80	rs2107538	G	A	*CCL5*	124 (75.6)	40 (24.4)	47 (57.3)	30 (36.6)	5 (6.1)	81 (76.4)	25 (23.6)	32 (60.4)	17 (32.1)	4 (7.5)	100	0.64
81	***rs11652805***	***G***	***A***	***(ancestry informative marker)***	***44 (50.0)***	***44 (50.0)***	***11 (25.0)***	***22 (50.0)***	***11 (25.0)***	***35 (60.3)***	***23 (39.7)***	***9 (31.0)***	***17 (58.6)***	***3 (10.3)***	***8.4***	***0.64***
82	rs2125345	G	A	(ancestry informative marker)	87 (53.7)	75 (46.3)	23 (28.4)	41 (50.6)	17 (21.0)	67 (63.2)	39 (36.8)	21 (39.6)	25 (47.2)	7 (13.2)	98.8	0.86
83	rs4891825	A	G	(ancestry informative marker)	110 (67.1)	54 (32.9)	37 (45.1)	36 (43.9)	9 (11.0)	68 (64.2)	38 (35.8)	23 (43.4)	22 (41.5)	8 (15.1)	100	0.7
84	rs8192726	C	A	*CYP2A6*	157 (95.7)	7 (4.3)	75 (91.5)	7 (8.5)	0 (0.0)	99 (93.4)	7 (6.6)	46 (86.8)	7 (13.2)	0 (0.0)	100	1.00
85	rs8192709	G	A	*CYP2B6*	153 (96.8)	5 (3.2)	74 (93.7)	5 (6.3)	0 (0.0)	99 (95.2)	5 (4.8)	47 (90.4)	5 (9.6)	0 (0.0)	94.4	1.00
86	rs28399499	A	G	*CYP2B6*	155 (94.5)	9 (5.5)	73 (89.0)	9 (11.0)	0 (0.0)	104 (98.1)	2 (1.9)	51 (96.2)	2 (3.8)	0 (0.0)	100	1.00
87	rs34097093	G	A	*CYP2B6*	164 (100.0)	0 (0.0)	82 (100.0)	0 (0.0)	0 (0.0)	106 (100.0)	0 (0.0)	53 (100.0)	0 (0.0)	0 (0.0)	100	NA
88	rs28399502	C	A	*CYP2B6*	160 (97.6)	4 (2.4)	78 (95.1)	4 (4.9)	0 (0.0)	102 (100.0)	0 (0.0)	51 (100.0)	0 (0.0)	0 (0.0)	96.2	1.00
89	rs707265	G	A	*CYP2B6*	106 (65.4)	56 (34.6)	34 (42.0)	38 (46.9)	9 (11.1)	79 (74.5)	27 (25.5)	30 (56.6)	19 (35.8)	4 (7.5)	98.8	1.00
90	rs1042389	A	G	*CYP2B6*	129 (79.6)	33 (20.4)	53 (65.4)	23 (28.4)	5 (6.2)	86 (82.7)	18 (17.3)	35 (67.3)	16 (30.8)	1 (1.9)	96.9	0.57
91	rs6104567	A	C	(ancestry informative marker)	111 (68.5)	51 (31.5)	39 (48.1)	33 (40.7)	9 (11.1)	77 (72.6)	29 (27.4)	26 (49.1)	25 (47.2)	2 (3.8)	98.8	0.84

#	SNP	Alleles		Gene (if applicable)	Immunologic non-responders					Immunologic responders					GCR	HWE p
		A1	A2		A1 (%)	A2 (%)	A1/A1 (%)	A1/A2 (%)	A2/A2 (%)	A1 (%)	A2 (%)	A1/A1 (%)	A1/A2 (%)	A2/A2 (%)		
92	rs3907047	A	G	(ancestry informative marker)	143 (87.2)	21 (12.8)	62 (75.6)	19 (23.2)	1 (1.2)	96 (90.6)	10 (9.4)	44 (83.0)	8 (15.1)	1 (1.9)	100	0.68
93	rs4821004	A	G	(ancestry informative marker)	100 (61.0)	64 (39.0)	28 (34.1)	44 (53.7)	10 (12.2)	55 (53.9)	47 (46.1)	12 (23.5)	31 (60.8)	8 (15.7)	96.2	0.08
94	rs5768007	G	A	(ancestry informative marker)	132 (81.5)	30 (18.5)	53 (65.4)	26 (32.1)	2 (2.5)	94 (90.4)	10 (9.6)	43 (82.7)	8 (15.4)	1 (1.9)	96.9	1.00

GCR – global call rate; HWE p – p-value from the exact test to assess compliance to Hardy-Weinberg equilibrium, NA – not available.

Supplementary Table 3. Complete results of the genetic association tests with the immunologic outcome

#	SNP	Gene	X^2	Degrees of freedom	p-value
5	rs3003596	*NR1I3*	10.10	2	***0.01***
6	rs2222202	*IL10*	0.90	2	0.64
7	rs1800871	*IL10*	0.17	2	0.92
8	rs1800890	*IL10*	1.29	2	0.52
22	rs115770495	*ABCG2*	18.89	2	0.39
23	rs1448784	*ABCG2*	0.25	1	0.62
24	rs2231142	*ABCG2*	0.70	1	0.40
26	rs2069762	*IL2*	73.13	2	***0.03***
27	rs10519613	*IL15*	67.10	2	***0.03***
31	rs1494555	*IL7R*	30.92	2	0.21
32	rs11567762	*IL7R*	24.76	2	0.29
33	rs6897932	*IL7R*	0.71	2	0.70
34	rs3822731	*IL7R*	0.38	2	0.83
35	rs987106	*IL7R*	10.66	2	0.59
36	rs3194051	*IL7R*	0.96	2	0.62
38	rs2243250	*IL4*	8.02	2	***0.02***
44	rs3842	*ABCB1*	12.01	2	0.55
45	rs2235048	*ABCB1*	17.21	2	0.42
46	rs1128503	*ABCB1*	0.18	2	0.92
47	rs2214102	*ABCB1*	0.04	1	0.84
48	rs4646437	*CYP3A4*	12.08	2	0.55
55	rs2234767	*FAS*	25.21	2	0.28
56	rs1800682	*FAS*	0.96	2	0.62
60	rs11568628	*SLC22A6*	0.25	1	0.62
61	rs4149170	*SLC22A6*	0.57	2	0.75
62	rs3135932	*IL10RA*	0.24	1	0.63
63	rs9610	*IL10RA*	25.95	2	0.27
66	rs2069709	*IFNG*	6.22E-29	1	1.00
73	rs762551	*CYP1A2*	3.37	2	0.19
74	rs129081	*ABCC1*	82.13	2	***0.02***
75	rs113264879	*ABCC1*	0.25	1	0.62
76	rs4148380	*ABCC1*	13.42	1	0.25
77	rs8056298	*ABCC1*	0.01	1	0.91
78	rs212091	*ABCC1*	0.71	1	0.70
80	rs2107538	*CCL5*	0.34	2	0.84
84	rs8192726	*CYP2A6*	0.34	1	0.56
85	rs8192709	*CYP2B6*	0.13	1	0.72
86	rs28399499	*CYP2B6*	13.73	1	0.24
88	rs28399502	*CYP2B6*	11.65	1	0.28
89	rs707265	*CYP2B6*	27.77	2	0.25
90	rs1042389	*CYP2B6*	13.46	2	0.51

AENXO A: INSTRUÇÕES AOS AUTORES PARA PUBLICAÇÃO NA REVISTA PHARMACOGENETICS AND GENOMICS

Pharmacogenetics and Genomics

Online Submission and Review System

Author Resources

Patient Consent Form (PDF)
Reprint Ordering
Permissions Requests

Guidance for Authors on the Preparation and Submission of Manuscripts to Pharmacogenetics and Genomics

Note: These instructions comply with those formulated by the International Committee of Medical Journal Editors. For further details, authors should consult the "Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals" at www.icmje.org.

The Journal is a member of the Committee on Publication Ethics (COPE) which aims to define best practice in the ethics of scientific publishing. COPE has established a number of guidelines including a Code of Conduct, and created flow charts that help editor's process cases of suspected misconduct (www.publicationethics.org).

Appeals on editorial decisions should be sent to the Editor. Complaints related to how your paper was processed during peer-review and not resolved by the Editor, should be referred to the person named as publisher in "About the Journal" under "Journal Info" contacts (http://journals.lww.com/jpharmacogenetics/Pages/aboutthejournal.aspx)or if unsatisfied to COPE (www.publicationethics.org).

SCOPE

Pharmacogenetics and Genomics is devoted to the rapid publication of research papers, brief review articles and short communications on genetic determinants in response to drugs and other chemicals in humans and animals. The Journal brings together papers from the entire spectrum of biomedical research and science, including biochemistry, bioinformatics, clinical pharmacology, clinical pharmacy, epidemiology, genetics, genomics, molecular biology, pharmacology, pharmaceutical sciences, and toxicology. Under a single cover, the Journal provides a forum for all aspects of the genetics and genomics of host response to exogenous chemicals: from the gene to the clinic.

POINTS TO CONSIDER WHEN SUBMITTING YOUR PAPER TO ONE OF OUR JOURNALS

Please think carefully about the following points and make the appropriate declarations.

Redundant or duplicate publication

We ask you to confirm that your paper has not been published in its current form or a substantially similar form (in print or electronically, including on a web site or an abstract of your work), that it has not been accepted for publication elsewhere, and that it is not under consideration by another publication. The International Committee of Medical Journal Editors has provided details of what is and what is not duplicate or redundant publication (www.icmje.org). If you are in doubt (particularly in the case of material that you have posted on a web site), we ask you to proceed with your submission but to include a copy of the relevant previously published work or work under consideration by other journals. In your covering letter to the editors, draw attention to any published work that concerns the same patients or subjects as the present paper.

Authorship

First, we have (rarely) had problems when someone named as an author was not aware of the submission of a paper and, on occasion, did not support the findings published. All authors must acknowledge that they have read and approved the paper, they have met the criteria for authorship as established by the International Committee of Medical Journal Editors, that they believe that the paper represents honest work, and that they are able to verify the validity of the results reported. We also ask that one person is named and agreed upon as the corresponding author who will act as the sole contact between the editors and publishers. Changes to submitted papers will only be accepted through the corresponding author, unless he/she submits

in writing the authorization of an alternative contact. In addition to those from the ICJME the International Society for Medical Publication Professionals, ISMPP (www.ismpp.org) have produced some useful guidelines on authorship of studies sponsored by companies: Good Publication Practice (GPP3) (http://www.ismpp.org/gpp3).

Conflicts of interest

Authors must state all possible conflicts of interest in the manuscript, including financial, consultant, institutional and other relationships that might lead to bias or a conflict of interest. If there is no conflict of interest, this should also be explicitly stated as none declared. All sources of funding should be acknowledged in the manuscript. All relevant conflicts of interest and sources of funding should be included on the title page of the manuscript with the heading "Conflicts of Interest and Source of Funding:"

For example:
Conflicts of Interest and Source of Funding: A has received honoraria from Company Z. B is currently receiving a grant (#12345) from Organization Y, and is on the speaker's bureau for Organization X – the CME organizers for Company A. For the remaining authors none were declared.

Copyright

In addition, each author must complete and submit the journal's copyright transfer agreement, which includes a section on the disclosure of potential conflicts of interest based on the recommendations of the International Committee of Medical Journal Editors, (www.icmje.org/update.html).

A copy of the form is made available to the submitting author within the Editorial Manager submission process. Co-authors will automatically receive an Email with instructions on completing the form upon submission.

Open Access

Authors of accepted peer-reviewed articles have the choice to pay a fee to allow perpetual unrestricted online access to their published article to readers globally, immediately upon publication. Authors may take advantage of the open access option at the point of acceptance to ensure that this choice has no influence on the peer review and acceptance process. These articles are subject to the journal's standard peer-review process and will be accepted or rejected based on their own merit.

The article processing charge (APC) is charged on acceptance of the article and should be paid within 30 days by the author, funding agency or institution. Payment must be processed for the article to be published open access. For a list of journals and pricing please visit our Wolters Kluwer Open Health Journals page.

Authors retain copyright

Authors retain their copyright for all articles they opt to publish open access. Authors grant Wolters Kluwer an exclusive license to publish the article and the article is made available under the terms of a Creative Commons user license. Please visit our Open Access Publication Process page for more information.

Creative Commons license

Open access articles are freely available to read, download and share from the time of publication under the terms of the Creative Commons License Attribution-NonCommerical No Derivative (CC BY-NC-ND) license. This license does not permit reuse for any commercial purposes nor does it cover the reuse or modification of individual elements of the work (such as figures, tables, etc.) in the creation of derivative works without specific permission.

Compliance with funder mandated open access policies

An author whose work is funded by an organization that mandates the use of the Creative Commons Attribution (CC BY) license is able to meet that requirement through the available open access license for approved funders. Information about the approved funders can be found here: http://www.wkopenhealth.com/inst-fund.php

FAQ for open access
http://www.wkopenhealth.com/openaccessfaq.php

Permissions to reproduce previously published material
We ask you to send us copies of permission to reproduce material (such as illustrations) from the copyright holder. We cannot send your paper to press without these permissions.

Patient consent forms
The protection of a patient's right to privacy is essential. We ask you to send copies of patient consent forms on which patients or other subjects of your experiments clearly grant permission for the publication of photographs or other material that might identify them. If the consent form for your research did not specifically include this, please obtain it or remove the identifying material. A sample patient consent form is available from the Journal's website if required.

Ethics committee approval
You must state clearly in your submission in the Methods section that you conducted studies on human participants with the approval of an appropriate named ethics committee. Please also look at the latest version of the Declaration of Helsinki. Similarly, you must confirm that experiments involving animals adhered to ethical standards and must state the care of animal and licensing guidelines under which the study was performed.

Compliance with NIH and Other Research Funding Agency Accessibility Requirements
A number of research funding agencies now require or request authors to submit the post-print (the article after peer review and acceptance but not the final published article) to a repository that is accessible online by all without charge. As a service to our authors, LWW will identify to the National Library of Medicine (NLM) articles that require deposit and will transmit the post-print of an article based on research funded in whole or in part by the National Institutes of Health, Wellcome Trust, Howard Hughes Medical Institute, or other funding agencies to PubMed Central. The Copyright Transfer Agreement provides the mechanism.

SUBMISSIONS

All manuscripts must be submitted on-line through the web-based tracking system at http://pgen.editorialmanager.com. The site contains instructions and advice on how to use the system. Authors should NOT in addition then post a hard copy submission to the editorial office, unless have been instructed to do so by the editorial office.

Editorial Policy
The recipient editor will read each manuscript to determine whether it is appropriate for external peer review. Those deemed suitable will undergo external review; others will be rapidly returned to the corresponding author. Authors need not suggest referees, but the editors will respect the wishes of the corresponding author that the manuscript not be sent to a particular competing laboratory for review. Manuscripts will be judged solely on the basis of their content of original data and on the conclusions drawn from them. Manuscripts that report screening of mutant/variant allelic frequencies in an isolated population will no longer be considered for publication in Pharmacogenetics and Genomics, unless the mutant/variant is analysed with respect to an association with a particular disease or, in the special case, where comparisons of different populations add new insight into important clinical issues. In case-control studies, it is imperative that the allele(s) being analysed is (are) associated with a functional difference in expression of the gene product and that the gene product can be mechanistically linked to the disease being investigated. For example, is the disease mediated by chemicals or drugs that are metabolised by the gene product?

First-time users: Please click the Register button from the menu and enter the requested information. On successful registration, you will be sent an e-mail indicating your user name and password. Print a copy of this information for future reference. Note: If you have received an e-mail from us with an assigned user ID and password, or if you are a repeat user, do not register again. Just log in. Once you have an assigned ID and password, you do not have to re-register, even if your status changes (that is, author, reviewer, or editor).

Authors: Please click the log-in button from the menu at the top of the page and log in to the system as an Author. Submit your manuscript according to the author instructions. You will be able to track the progress of your manuscript through the system. If you experience any problems, please contact Christiana Adewale, Managing Editor, christiana.adewale@wolterskluwer.com. Requests for help and other questions will be addressed in the order received.

ARTICLE TYPES

Original Articles

Original articles should present new scientific research that represents a substantial advancement in the field. Original articles should contain a structured abstract with the following headings: Objectives, Methods, Results, and Conclusions. The main text of an original article should not exceed 3500 words. Original articles may contain up to 75 references and no more than a combined 12 figures and tables.

Mini-Reviews

Mini-Reviews should be timely reviews that provide a critical but balanced view of an important topic in the field. Mini-reviews are generally by invitation but proposals for mini-reviews may also be submitted to either Editor. An unstructured abstract of 250 words should be included with your mini-review. The main text of a mini-review should not exceed 2500 words, and it may contain up to 25 references and a combined 8 figures and tables. Personal communications and submitted manuscripts should not be cited.

Short Communications

Short communications can be used to submit preliminary results, concise negative studies, and/or succinct focused papers addressing methodological issues (e.g., genotyping assays). Short communications should contain an unstructured abstract of 250 words. The main text of a short communications should not exceed 1500 words, and it may contain up to 25 references and 3 figures and/or tables.

Opinion Pieces

Opinion pieces on any current topic will be considered. These may range from specific recommendations regarding individual drugs or genes, or general topics such as study design, implementation of pharmacogenomic testing, or public policy issues. Submitted manuscripts should have no more than 1200 words of text, 10 references, and 2 tables and/or figures. No abstract is required.

Letters to the Editor

Letters should be limited to 750 words or less, excluding references. One table or figure may be submitted with reader correspondence.

PRESENTATION OF PAPERS

Manuscripts that do not adhere to the following instructions will be returned to the corresponding author for technical revision before undergoing peer review.

Manuscript File

The Manuscript file must be submitted in a Word document. Double spacing should be used throughout the manuscript, which should include the following sections, each starting on a separate page: Title Page, abstract and keywords, text, acknowledgements, references, and figure legends (if applicable). Pages should be numbered consecutively, beginning with the Title Page, and the page number should be placed in the top right hand corner of each page. Abbreviations should be defined on their first appearance in the text; those not accepted by international bodies should be avoided.

Title Page

The Title Page should carry the full title of the paper and a short title, of no more than 45 characters and spaces, to be used as a 'running head' (and which should be so identified). The first name, middle initial and last name of each author should appear. If the work is to be

attributed to a department or institution, its full name should be included. Any disclaimers should appear on the Title Page, as should the name and address of the author responsible for correspondence concerning the manuscript and the name and address of the author to whom requests for reprints should be made. Finally, the Title Page should include a statement of conflicts of interest and source of funding, and when none state "none declared".

Abstracts

Abstracts are required for Original Articles, Mini-Reviews, and Short Communications. The abstract for Original Articles should state the Objective(s) of the study or investigation, basic Methods (selection of study subjects or laboratory animals; observational and analytical methods), main Results (giving specific data and their statistical significance, if possible), and the principal Conclusions. It should emphasize new and important aspects of the study or observations. Mini-Reviews and Short communications should include an unstructured summary of no more than 250 words. Opinion pieces do not have an abstract.

Key Words

The abstract should be followed by a list of 3 - 10 keywords or short phrases which will assist the cross-indexing of the article and which may be published. When possible, the terms used should be from the Medical Subject Headings list of the National Library of Medicine (http://www.nlm.nih.gov/mesh/meshhome.html).

Text

Full papers of an experimental or observational nature may be divided into sections headed Introduction, Methods (including ethical and statistical information), Results and Discussion (including a conclusion), although invited commentaries and reviews may require a different format. Please see the article type section above for more detailed requirements.

Acknowledgements

Acknowledgements should be made only to those who have made a substantial contribution to the study. Authors are responsible for obtaining written permission from people acknowledged by name in case readers infer their endorsement of data and conclusions.

References

References should be numbered consecutively in the order in which they first appear in the text. They should be assigned Arabic numerals, which should be given in brackets, e.g. [17]. References should include the names of all authors when six or fewer; when seven or more, list only the first six names and add et al. References should also include full title and source information. Journal names should be abbreviated as in MEDLINE (NLM Catalog, http://www.ncbi.nlm.nih.gov/nlmcatalog).

Articles in journals

Romerius P, Giwercman A, Moëll C, Relander T, Cavallin-Ståhl E, Wiebe T, et al. Estrogen receptor α single nucleotide polymorphism modifies the risk of azoospermia in childhood cancer survivors. Pharmacogenet Genomics. 2011;21:263-269.

Books

DeVita VT, Hellman S, Rosenberg SA (2005). Cancer, principles & practice of oncology. 7th ed. Philadelphia: Lippincott Williams & Wilkins.

Chapter in a book:

Restifo NP, Wunderlich JR. Cancer Immunology. In: DeVita VT, Hellman S, Rosenberg SA, eds. Cancer: principles & practice of oncology, 7th edn. Philadelphia: Lippincott Williams & Wilkins, 2005: 139-161

Personal communications and unpublished work should not feature in the reference list but should appear in parentheses in the text. Unpublished work accepted for publication but not yet released should be included in the reference list with the words 'in press' in parentheses beside the name of the journal concerned. References must be verified by the author(s) against the original documents.

Tables

Create tables using the table creating and editing feature of your word processing software (eg, Word, WordPerfect). Do not use Excel or comparable spreadsheet programs. Do not include any graphics or images in tables. Group all tables in a separate file. Cite tables consecutively in the text, and number them in that order. Each table should appear on a separate sheet and should include the table title, appropriate column heads, and explanatory legends (including definitions of any abbreviations used). Do not embed tables within the body of the manuscript.

Place explanatory matter in footnotes, not in the heading. Explain in footnotes all non-standard abbreviations that are used in each table. Identify statistical measures of variations, such as standard deviation and standard error of the mean. Be sure that each table is cited in the text. If you use a table or data from another published or unpublished source, obtain permission and acknowledge the source fully.

Figures/Illustrations

A) Creating Digital Artwork

1. Learn about the publication requirements for Digital Artwork: http://links.lww.com/ES/A42
2. Create, Scan and Save your artwork and compare your final figure to the Digital Artwork Guideline Checklist (below).
3. Upload each figure to Editorial Manager in conjunction with your manuscript text and tables.

B) Digital Artwork Guideline Checklist

Here are the basics to have in place before submitting your digital artwork:

- Artwork should be saved as JPEG, TIFF, EPS, or MS Office (DOC, PPT, XLS) files. High resolution PDF files are also acceptable.
- Crop out any white or black space surrounding the image.
- Please use either Arial or Helvetica font size 7 for any text or labels within illustrations.
- Diagrams, drawings, graphs, and other line art must be vector or saved at a resolution of at least 1200 dpi. If created in an MS Office program, send the native (DOC, PPT, XLS) file.
- Photographs, radiographs and other halftone images must be saved at a resolution of at least 300 dpi.
- Photographs and radiographs with text must be saved as postscript or at a resolution of at least 600 dpi.
- Each figure must be saved and submitted as a separate file. Figures should not be embedded in the manuscript text file.

Remember:

- References to figures and tables should be made in order of appearance in the text and should be in Arabic numerals in parentheses, e.g. (Fig. 2).
- Number figures in the figure legend in the order in which they are discussed.
- Upload figures consecutively to the Editorial Manager web site and enter figure numbers consecutively in the Description field when uploading the files.
- Illustrations should be presented to a width of 82 mm or, when the illustration demands it, to a width of 166 mm.

- Photomicrographs must have internal scale markers.
- If photographs of people are used, their identities must be obscured or the picture must be accompanied by written consent to use the photograph.
- If a figure has been published before, the original source must be acknowledged and written permission from the copyright holder for both print and electronic formats should be submitted with the material. Permission is required regardless of authorship or publisher, except for documents in the public domain.
- Figures may be reduced, cropped or deleted at the discretion of the editor.
- Colour illustrations are acceptable but authors will be expected to cover the extra reproduction costs (for current charges, contact the publisher).

Legends for illustrations

Captions should be typed in double spacing, beginning on a separate page of the manuscript file. Each figure should have an Arabic numeral corresponding to the illustration to which it refers. Internal scales should be explained and staining methods for photomicrographs should be identified.

Units of measurement

Measurements of length, height, weight, and volume should be reported in metric units (metre, kilogram, or litre) or their decimal multiples. Temperatures should be given in degrees Celsius. Blood pressures should be given in millimetres of mercury.
All haematologic and clinical chemistry measurements should be reported in the metric system in terms of the International System of Units (SI). Editors may request that alternative or non-SI units be added by the authors before publication.

Abbreviations and symbols

Use only standard abbreviations. Avoid abbreviations in the title and abstract. The full term for which an abbreviation stands should precede its first use in the text unless it is a standard unit of measurement.

Nomenclature

Authors are asked to confirm in their covering letter of submission, that their manuscript complies with the nomenclature guidelines developed by the HUGO nomenclature committee for human genes. The guidelines can be found at the following sites:

Human genes
Use genetic notation and symbols approved by the HUGO Nomenclature Committee. Before submission, approved gene symbols should be obtained from the HUGO Nomenclature Committee (http://www.genenames.org). Useful reference articles and forms: White et al. (1997), 'Guidelines for Human Gene Nomenclature', Genomics, 45, 468-471]; to submit new gene names, the Gene Name Proposal form may be completed on the nomenclature web page: (http://www.genenames.org/cgi-bin/request).

Human genetic variation
Designation of single nucleotide polymorphisms (SNPs), deletions, insertions and other gene mutations should follow the guidelines given in Hum Genet 2001; 109:121–124. The nomenclatures for allelic variations of human P450s should adhere to the recommendations given at http://www.imm.ki.se/CYPalleles/ those of N-acetyl transferases at http://www.louisville.edu/medschool/pharmacology/NAT.html and those of UDP glycosyltransferases in Pharmacogenetics 1997; 7:255–269.

Human cytogenics
Use ISCN nomenclature for cytogenetics notation [Mitelman, F. (ed.) ISCN 1995: An International System for Human Cytogenetic Nomenclature, S. Karger, Basel]. Human gene names and loci should be written in uppercase italics and Arabic numerals. Protein products are not italicised.
Mouse strain and genetic nomenclature:

International Committee on Standardised Genetic Nomenclature for Mice (http://www.informatics.jax.org/) new symbols and names for genes should be obtained before submission.

Supplemental Digital Content (SDC) Authors may submit SDC via Editorial Manager to LWW journals that enhance their article's text to be considered for online posting. SDC may include standard media such as text documents, graphs, audio, video, etc. On the Attach Files page of the submission process, please select Supplemental Audio, Video, or Data for your uploaded file as the Submission Item. If an article with SDC is accepted, our production staff will create a URL with the SDC file. The URL will be placed in the call-out within the article. SDC files are not copy-edited by LWW staff, they will be presented digitally as submitted. For a list of all available file types and detailed instructions, please visit http://links.lww.com/A142.

SDC Call-outs Supplemental Digital Content must be cited consecutively in the text of the submitted manuscript. Citations should include the type of material submitted (Audio, Figure, Table, etc.), be clearly labelled as "Supplemental Digital Content," include the sequential list number, and provide a description of the supplemental content. All descriptive text should be included in the call-out as it will not appear elsewhere in the article.
Example:
We performed many tests on the degrees of flexibility in the elbow (see Video, Supplemental Digital Content 1, which demonstrates elbow flexibility) and found our results inconclusive.

List of Supplemental Digital Content A listing of Supplemental Digital Content must be submitted at the end of the manuscript file. Include the SDC number and file type of the Supplemental Digital Content. This text will be removed by our production staff and not be published.
Example:
Supplemental Digital Content 1. wmv

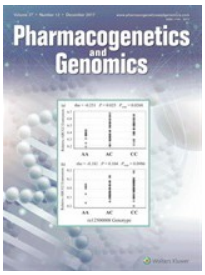
SDC File Requirements All acceptable file types are permissible up to 10 MBs. For audio or video files greater than 10 MBs, authors should first query the journal office for approval. For a list of all available file types and detailed instructions, please visit http://links.lww.com/A142.

AFTER ACCEPTANCE

Page proofs and corrections: Corresponding authors will receive electronic page proofs to check the copyedited and typeset article before publication. Portable document format (PDF) files of the typeset pages and support documents (e.g., reprint order form) will be sent to the corresponding author by e-mail. Complete instructions will be provided with the e-mail for downloading and printing the files and for returning the corrected page proofs to the publisher. It is the author's responsibility to ensure that there are no errors in the proofs. Changes that have been made to conform to journal style will stand if they do not alter the authors' meaning. Only the most critical changes to the accuracy of the content will be made. Changes that are stylistic or are a reworking of previously accepted material will be disallowed. The publisher reserves the right to deny any changes that do not affect the accuracy of the content. Authors may be charged for alterations to the proofs beyond those required to correct errors or to answer queries. Proofs must be checked carefully and corrections faxed within 24 to 48 hours of receipt, as requested in the cover letter accompanying the page proofs.

Reprints: Authors will receive a reprint order form and a price list with the page proofs. Reprint requests should be faxed to the publisher with the corrected proofs, if possible. Reprints are normally shipped 6 to 8 weeks after publication of the issue in which the item appears. Contact the Reprint Department, Lippincott Williams & Wilkins, Contact the Reprint Department, phone: 1.800.341.2258, email: reprintsgroup@lww.com, with any questions.

Permissions: For permission and/or rights to use content for which the copyright holder is LWW or the society, please go to the journal's website and after clicking on the relevant article, click on the "Request Permissions" link under the "Article Tools" box that appears on the right side of the page. Alternatively, send an e-mail to customercare@copyright.com.



ANEXO B: CURRÍCULO LATTES ATUALIZADO

Ronald Rodrigues de Moura

Curriculum Vitae

Novembro/2017

Ronald Rodrigues de Moura

Curriculum Vitae

Dados pessoais

Nome Ronald Rodrigues de Moura
Filiação Fernandes Ramos de Moura e Marinalva Rodrigues de Moura
Nascimento 26/02/1989 - Recife/PE - Brasil
Carteira de Identidade 7327287 SDS - PE - 19/07/2003
CPF 074.339.224-82

Formação acadêmica/titulação

2014 Doutorado em Genética.
Universidade Federal de Pernambuco, UFPE, Recife, Brasil
com <B>período sanduíche</B> em Universita degli Studi di Trieste (Orientador : Adamo Pio D'Adamo)
Título: Inferência da miscigenação genética em populações de Pernambuco: _Aplicação em estudos de associação
Orientador: Lucas André Cavalcanti Brandão
Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

2012 - 2014 Mestrado em Genética.
Universidade Federal de Pernambuco, UFPE, Recife, Brasil
Título: Avaliação de polimorfismos de genes da imunidade inata associados ao diabetes mellitus tipo 1, Ano de obtenção: 2014
Orientador: Lucas André Cavalcanti Brandão
Bolsista do(a): Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco

2008 - 2011 Graduação em Ciências Biológicas.
Universidade Federal Rural de Pernambuco, UFRPE, Recife, Brasil
Título: Estudo da base genética das cultivares de feijão preto (Phaseolus vulgaris L.) do Brasil
Orientador: Edson Ferreira da Silva

Ensino Profissional de nível técnico interrompido(a) .
Instituto Federal de Pernambuco, IFPE, Recife, Brasil
Ano de interrupção: 2009

2004 - 2006 Ensino Médio (2o grau) .
SESI - CAT DR. Diniz Passos, SESI, Brasil

Formação complementar

2014 - 2014 Curso de curta duração em Metacore e Integrity (Thomson Reuters). (Carga horária: 20h).
Universidade Federal de Pernambuco, UFPE, Recife, Brasil

Atuação profissional

1. Universidade Federal de Pernambuco - UFPE

Vínculo institucional

2012 - 2014 Vínculo: Mestrando , Enquadramento funcional: Mestrado, Regime: Dedicação exclusiva

Projetos

Projetos de pesquisaProjetos de pesquisa**2015 - Atual** Determinação do perfil genético, através de marcadores autossômicos e de linhagem, e desenvolvimento de estratégias metodológicas de inferências biogeográficas e fisionômicas a partir do DNA de populações do Nordeste para utilização em Genética Forense

Descrição: Para a determinação do perfil genético de amostras biológicas, utilizam-se principalmente marcadores em regiões especificas do DNA que são polimórficas na população, como as regiões de micro e minissatélites – short tandem repeats (STR). Nos casos em que ocorre mistura de material biológico, como nos crimes sexuais onde geralmente tem-se um homem como perpetrador e a mulher como vítima é possível recorrer à caracterização de polimorfismos específicos do cromossomo Y (Y-STR) para aumentar as chances de serem detectados, mesmo em pequenas quantidades de DNA, pois apenas o material genético de proveniência masculina é amplificado, sendo o perfil obtido podendo ainda ser comparado com diferentes suspeitos. Além disso, é necessário alterar o arcabouço estatístico necessário para comprovar ou refutar a presença do DNA do suspeito na amostra através da implementação de softwares que atuem nesse sentido. Para a identificação de uma amostra por meio do exame de DNA é necessário o confronto com uma amostra da vítima, do suspeito ou de um familiar (amostra referência). Na ausência da amostra referência, o material genético pode ainda ser identificado através da inferência de sua ancestralidade biogeográfica (Y-DNA, DNAs mitocondrial e autossômico) e de características externamente visíveis (CEV), como o aspecto facial, cor da pele, íris e cabelo. Visando aumentar o leque de possibilidades em que a análise de DNA pode ser utilizada no contexto forense, o presente projeto pretende: Criar um banco de dados genéticos a partir de marcadores autossômicos e de linhagens (mtDNA e marcadores no cromossomo Y) de indivíduos do nordeste; Identificar a origem biogeográfica através de marcadores autossômicos e de linhagem determinando sub-haplogrupos do mtDNA e cromossomo Y que permitem diferenciar as linhagens parentais presentes na população brasileira; Implementar sistemas de marcadores genéticos na identificação de características externamente visíveis, tais como modelagem facial, cor da íris, pele e cabelo; Desenvolver um software para realização de cálculos envolvidos em diversos contextos em genética forense baseado em uma plataforma web com a possibilidade de instalação em ambientes fechados; Aplicar os resultados obtidos com os casos reais recebidos nas instituições criminais afiliadas ao projeto. Nosso estudo será multicêntrico, incluindo populações do Ceará (CE), da Paraíba (PB) e Pernambuco (PE). Além da coleta de amostras, também será realizado o registro de características externamente visíveis. Serão realizadas reações de PCR e sequenciamento de nova geração para traçarmos o perfil genético através de marcadores autossômicos e de linhagem, bem como a inferência da ancestralidade biogeográfica dos indivíduos e predição de características externamente visíveis usando SNPs.
Situação: Em andamento Natureza: Projetos de pesquisa
Integrantes: Ronald Rodrigues de Moura; Sergio Crovella (Responsável); Lucas André Cavalcanti Brandao; Sérgio Paiva; Antônio Victor Campos Coelho; Kaynara Cecília Nery Rabêlo; Tatiana Costa de Oliveira; Manuela Barbosa Rodrigues de Souza; Klaudia Emanuela Ramos Tenorio; Simone Silva Santos Lopes; Silvana Magna Cavalcante do Monte; Sergio Marques de Lucena; Eliane dos Santos Pereira; Debora Menezes da Costa; Samara da Silva Cardoso Santiago
Financiador(es): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior-CAPES

2014 - Atual Rede InterSys: Biologia Sistêmica no Estudo de Função Gênica em Interações Bióticas

Descrição: O projeto envolve nove instituições e 17 subprojetos. Pretende estabelecer a rede INTERSYS, voltada para a formação de pessoal e geração de conhecimento científico de alto nível envolvendo interações bióticas a partir de abordagens multidisciplinares de biologia sistêmica (ômicas, biologia celular e bioinformática) através da integração de grupos nacionais e internacionais experientes..
Situação: Em andamento; Natureza: Pesquisa.
Situação: Em andamento Natureza: Projetos de pesquisa
Integrantes: Ronald Rodrigues de Moura; Ana M. Benko Iseppon (Responsável)

2012 - 2014 Avaliação de polimorfismos de genes da imunidade inata associados ao diabetes mellitus tipo 1

Descrição: De um modo geral, esse estudo visa investigar a influência genética de polimorfismos com o Diabetes Mellitus tipo 1 a fim de contribuir para o desenvolvimento de um perfil imunogenético de indivíduos que desenvolvem a doença, e especificamente: determinar a frequência de polimorfismos em genes IFH1 e PTPN2, previamente descritos na literatura como associados ao DM1, em uma população Brasileira; selecionar outros genes candidatos, previamente descritos na literatura como associados ao DM1; avaliar o grau de associação desses polimorfismos com o DM1 de um grupo de pacientes em relação a um grupo de indivíduos controle negativo e; ampliar o já formado banco de amostras de controles e pacientes com DM1. O grupo de estudo será formado por 200 pacientes com DM1 atendidos no Instituto de Medicina Integral Professor Fernando Figueira, Hospital das Clínicas e Hospital da Restauração. Após o consentimento livre e esclarecido pelo responsável do paciente, o mesmo será convidado a participar da pesquisa fornecendo 10 mL de sangue periférico total. O grupo controle será formado por adultos doadores saudáveis acima de 20 anos de idade, sem apresentar histórico familiar de doenças auto-imunes, incluindo o DM1, provenientes da Fundação HEMOPE. Será extraído o DNA genômico das células do sangue periférico anti-coagulado com EDTA de cada paciente e controle seguindo as instruções do fabricante do Kit comercial. O plasma será coletado a partir de sangue periférico total após a sua centrifugação. Inicialmente, as análises dos polimorfismos serão realizadas nos genes IFIH1 e PTPN2, porém ao longo do estudo, outros genes serão selecionados. A genotipagem será feita por meio de digestão do DNA genômico com enzimas de restrição seguida de uma PCR. Os fragmentos serão observados através de eletroforese em gel. O cálculo da frequência dos polimorfismos e do grau de associação como o DM1, serão feitos os testes chi quadrado e o teste exato de Fisher a partir dos resultados das genotipagens.
Situação: Concluído Natureza: Projetos de pesquisa
Integrantes: Ronald Rodrigues de Moura (Responsável); ;

2012 - 2015 Perfil genético de Marcadores de Ancestralidade (AIMs) e fatores da imunidade na população do estado de Pernambuco

Descrição: Um dos principais legados do término do Projeto Genoma Humano foi a descoberta de milhões de variações ao longo do DNA, das quais a maioria, acima de 1,5 milhões, se referia a pontuais modificações no genoma humano chamadas de SNPs (Polimorfismo de Base Única) ou inserções e deleções de nucleotídeos (INDELs). Essas variações gênicas representam a maior fonte de variações interindividuais genéticas e podem ser utilizadas como uma extraordinária ferramenta na análise de marcadores genéticos associados à susceptibilidade a uma gama de infecções e doenças genéticas, assim como de marcadores genéticos associados com a ancestralidade. Até poucos anos atrás, a cor da pele era o principal fator utilizado na classificação das diferentes raças em estudos de associação caso-controle. Embora a correlação entre cor da pele e outras características físicas estejam realmente relacionadas à ancestralidade, existe muita variação dentro de cada grupo e entre os diferentes grupos étnicos. Através da genotipagem de SNPs ou INDELs, pode-se chegar a um número representativo de marcadores autossômicos que sejam discriminantes de uma determinada população ou grupo étnico. Esses marcadores são chamados de "Ancestry Informative Markers" (AIMs). Atualmente, a estratificação populacional de acordo com a etnia é um dos primeiros passos para a realização de estudos de associação caso-controle. Esse procedimento minimiza a possibilidade de falsos positivos ou falsos negativos gerados pela distribuição desigual dos SNPs e INDELs entre grupos étnicos da amostra populacional.As variações SNPs e Indels também são usadas para o estudo de doenças multifatoriais, ou seja, as causadas por vários fatores, entre eles os genéticos e os ambientais. Várias doenças infecciosas e autoimunes já mostraram ter associação com variações genéticas em diversos genes da imunidade inata e adaptativa. Uma maneira de relacionar os SNPs que possam estar envolvidos no desenvolvimento das doenças autoimunes são os Estudos de Associação Genética (EAG), os quais tentam esclarecer a influência de fatores genéticos sobre doenças complexas. Desta forma, o presente projeto tem como objetivo avaliar a ancestralidade e o perfil imunogenético de uma amostra da população de doadores da Fundação HEMOPE do Recife do estado de Pernambuco através de marcadores informativos de ancestralidade (AIMs) e de SNPs/Indels de genes representativos da imunidade inata e adaptativa que possam estar envolvidos com a susceptibilidade a doenças infecciosas e/ou autoimunes. Com a caracterização da ancestralidade, os controles saudáveis poderão ser estratificados de acordo com a etnia, do ponto de vista genético, para que possam ser utilizados nos estudos de associação caso-controle das doenças infecciosas e autoimunes estudadas pelo grupo de pesquisa em Variabilidade Genética Humana da Universidade Federal de Pernambuco.
Situação: Concluído Natureza: Projetos de pesquisa
Integrantes: Ronald Rodrigues de Moura; Lucas André Cavalcanti Brandao (Responsável); Catarina Addobbati Jordão Cavalvanti

Financiador(es): Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco-FACEPE

Revisor de periódico

1. Genetica (Dordrecht. Online) -

Vínculo

2015 - 2016 Regime: Parcial

Áreas de atuação

1. Genética de Populações
2. Genética Humana e Médica
3. Biologia Evolutiva
4. Imunogenética

Idiomas

Inglês Compreende Bem , Fala Bem , Escreve Razoavelmente , Lê Bem

Prêmios e títulos

2013 Melhor apresentação oral da III Jornada de Pós-Graduação em Genética, UFPE

2012 Melhor apresentação oral da II Jornada de Pós-Graduação em Genética, UFPE

Producão

Produção bibliográfica
Artigos completos publicados em periódicos

1. CHAGAS, B. S.; GURGEL, A.; PAIVA, S.; LIMA, R.; CORDEIRO, M.; **MOURA, R. R.**; COELHO, A. V. C.; NASCIMENTO, K.; SILVA NETO, J.; CROVELLA, S.; FREITAS, A. C.
Synergic effect of oral contraceptives, GSTP1 polymorphisms, and high-risk HPV infection in development of cervical lesions. GENETICS AND MOLECULAR RESEARCH. , v.16, p.gmr160339742 - , 2017.

2. SILVA, J. A.; LIMA, S. C.; CAVALVANTI, C. A. J.; **MOURA, R. R.**; BRANDAO, L. A. C.; PANCOTTO, J. A. T.; DONADI, E. A.; CROVELLA, S.; SANDRIN-GARCIA, P.
Association of interferon-induced helicase C domain (IFIH1) gene polymorphisms with systemic lupus erythematosus and a relevant updated meta-analysis. Genetics and Molecular Research. , v.15, p.1 - 2, 2016.

3. COELHO, ANTONIO; **DE MOURA, RONALD**; KAMADA, ANSELMO; DA SILVA, RONALDO; GUIMARÃES, RAFAEL; BRANDÃO, LUCAS; DE ALENCAR, LUIZ; CROVELLA, SERGIO
Dendritic Cell-Based Immunotherapies to Fight HIV: How Far from a Success Story? A Systematic Review and Meta-Analysis. International Journal of Molecular Sciences (Online). , v.17, p.1985 - , 2016.

4. RABELO, K. C. N.; ALBUQUERQUE, C. M. R.; TAVARES, V. B.; SANTOS, S. M.; SOUZA, C. A.; OLIVEIRA, T. C.; **MOURA, R. R.**; BRANDAO, L. A. C.; CROVELLA, S.
Detecting multiple DNA human profile from a mosquito blood meal. GENETICS AND MOLECULAR RESEARCH. , v.15, p.1 - 8, 2016.

5. da SILVA, W. H. V. C.; **Moura, R.**; COELHO, A. V. C.; CROVELLA, S.; GUIMARAES, R. L.
Frequency of the CCR5-delta32 allele in Brazilian populations: A systematic literature review and meta-analysis. Infection, Genetics and Evolution (Print). , p.101 - 107, 2016.

6. **MOURA RODRIGUES, RONALD**; PLANA, MONSERRAT; GARCIA, FELIPE; ZUPIN, LUISA; KUHN, LOUISE; CROVELLA, SERGIO
Genome-wide scan in two groups of HIV-infected patients treated with dendritic cell-based immunotherapy. Immunologic Research. , v.64, p.1207 - 1215, 2016.

7. COELHO, A. V. C.; **MOURA, R. R.**; CROVELLA, S.; CELSI, F.
HLA-G genetic variants and hepatocellular carcinoma: a meta-analysis. GENETICS AND MOLECULAR RESEARCH. , v.15, p.1 - 8, 2016.

8. de LIMA JUNIOR, S.; de MACEDO, J.; TAVARES, M.; de OLIVEIRA, R.; HERACLIO, S.; MAIA, M. M.; **Moura, R.**; CROVELLA, S.; SOUZA, P. R. E.
Influence of IL-6, IL-8, and TGF-β1 gene polymorphisms on the risk of human papillomavirus-infection in women from Pernambuco, Brazil. Memórias do Instituto Oswaldo Cruz (Online). , v.111, p.663 - 669, 2016.

9. CROVELLA, S.; BIANCO, A. M.; VUCH, J.; ZUPIN, L.; **MOURA, R. R.**; TREVISAN, E.; SCHNEIDER, M.; BROLLO, A.; NICASTRO, E. M.; COSENZI, A.; ZABUCCHI, G.; BORELLI, V.
Iron signature in asbestos-induced malignant pleural mesothelioma: A population-based autopsy study. Journal of Toxicology and Environmental Health, Part A. , v.79, p.129 - 141, 2016.

10. COELHO, A.; **MOURA, R. R.**; CAVALVANTI, C. A. J.; GUIMARAES, R. L.; SANDRIN-GARCIA, P.; CROVELLA, S.; BRANDAO, L. A. C.
A rapid screening of ancestry for genetic association studies in an admixed population from Pernambuco, Brazil. Genetics and Molecular Research. , v.14, p.2876 - 2884, 2015.

11. TAVARES, N. A. C.; SANTOS, M. M. S.; **MOURA, R. R.**; ARAUJO, J.; GUIMARAES, R. L.; CROVELLA, S.; BRANDAO, L. A. C.
Association of TNF-α, CTLA4, and PTPN22 polymorphisms with type 1 diabetes and other autoimmune diseases in Brazil. Genetics and Molecular Research. , v.14, p.18936 - 18944, 2015.

12. CATAMO, E.; CAVALVANTI, C. A. J.; SEGAT, L.; FRAGOSO, T. S.; DANTAS, A. T.; MARIZ, H. A.; ROCHA JUNIOR, L. F.; DUARTE, A. L. B. P.; COELHO, A. V. C.; **MOURA, R. R.**; POLESELLO, V.; CROVELLA, S.; SANDRIN-GARCIA, P.
Comprehensive analysis of polymorphisms in the HLA-G 5¿ upstream regulatory and 3¿ untranslated regions in Brazilian patients with systemic lupus erythematosus. Tissue Antigens. , v.85, p.458 - 465, 2015.

13. SANTOS, S. M.; SOUZA, C. A.; RABELO, K. C. N.; SOUZA, P. R. E.; **MOURA, R. R.**; OLIVEIRA, T. C.; CROVELLA, S.
Distribution of forensic marker allelic frequencies in Pernambuco, Northestern Brazil. Genetics and Molecular Research. , v.14, p.4303 - 4310, 2015.

14. **MOURA, RONALD**; TRICARICO, PAOLA MAURA; CAMPOS COELHO, ANTONIO VICTOR; CROVELLA, SERGIO
GRID2 a novel gene possibly associated with mevalonate kinase deficiency. Rheumatology International (Berlin. Internet). , v.35, p.657 - 659, 2015.

15. COELHO, A. V. C.; **MOURA, R. R.**; Da SILVA, R. C.; KAMADA, A. J.; GUIMARAES, R. L.; BRANDAO, L. A. C.; COELHO, H. F. C.; CROVELLA, S.
Meta-analysis and time series modeling allow a systematic review of primary HIV-1 drug-resistant prevalence in Latin America and Caribbean. Current HIV Research (Print). , v.13, p.125 - 142, 2015.

16. **RODRIGUES DE MOURA, RONALD**; COELHO, ANTONIO VICTOR CAMPOS; DE QUEIROZ BALBINO, VALDIR; CROVELLA, SERGIO; BRANDÃO, LUCAS ANDRÉ CAVALCANTI
Meta-analysis of Brazilian genetic admixture and comparison with other Latin America countries. American Journal of Human Biology. , v.27, p.674 - 680, 2015.

17. SILVA, J. A.; TAVARES, N. A. C.; SANTOS, M. M. S.; **MOURA, R. R.**; GUIMARAES, R. L.; ARAUJO, J.; CROVELLA, S.; BRANDAO, L. A. C.
Meta-analysis of STAT4 and IFIH1 polymorphisms in type 1 diabetes mellitus patients with autoimmune polyglandular syndrome type III. Genetics and Molecular Research. , v.14, p.17730 - 17738, 2015.

18. BORELLI, VIOLETTA; **MOURA, RONAL R**; TREVISAN, ELISA; CROVELLA, SERGIO
NLRP1 and NLRP3 polymorphisms in mesothelioma patients and asbestos exposed individuals a population-based autopsy study from North East Italy. Infectious Agents and Cancer. , v.10, p.1 - 3, 2015.

19. **DE MOURA, RONALD R**; DE QUEIROZ BALBINO, VALDIR; CROVELLA, SERGIO; BRANDÃO, LUCAS A C
On the use of Chinese population as a proxy of Amerindian ancestors in genetic admixture studies with Latin American populations. European Journal of Human Genetics. , v.24, p.326 - 327, 2015.

20. SILVA, RONALDO CELERINO DA; TAVARES, NATHÁLIA DE ALENCAR CUNHA; **MOURA, RONALD**; COELHO, ANTÔNIO; GUIMARÃES, RAFAEL LIMA; ARAÚJO, JACQUELINE; CROVELLA, SERGIO; BRANDÃO, LUCAS ANDRÉ CAVALCANTI; SILVA, JAQUELINE DE AZEVÊDO
DC-SIGN polymorphisms are associated to Type 1 Diabetes Mellitus. Immunobiology (Jena. 1979). , v.219, p.859 - 865, 2014.

21. SEGAT, L.; ZUPIN, L.; **MOURA, RONALD**; COELHO, A. V. C.; CHAGAS, B. S.; FREITAS, A. C.; CROVELLA, S.
DEFB1 polymorphisms are involved in susceptibility to human papillomavirus infection in Brazilian gynaecological patients. Memórias do Instituto Oswaldo Cruz (Impresso). , v.109, p.918 - 922, 2014.

22. **Moura, R.**; PONTILLO, A.; DADAMO, P.; PIRASTU, N.; COELHO, A. V. C.; CROVELLA, SERGIO
Exome analysis of HIV patients submitted to dendritic cells therapeutic vaccine reveals an association of gene with response to the treatment. Journal of the International AIDS Society. , v.17, p.1 - 5, 2014.

23. PONTILLO, A.; Da SILVA, R. C.; **MOURA, R. R.**; CROVELLA, S.
Host genomic HIV restriction factors modulate the response to dendritic cell-based treatment against HIV-1. Human Vaccines & Immunotherapeutics. , v.10, p.26 - 27, 2014.

Artigos aceitos para publicação

1. DA SILVA, RONALDO; COELHO, ANTONIO; **MOURA, RONALD**; ARRAES, LUIZ; BRANDÃO, LUCAS; GUIMARÃES, RAFAEL; CROVELLA, SÉRGIO
CUL5 and APOBEC3G polymorphisms are partially implicated in HIV-1 infection and antiretroviral therapy in a Brazilian population. CURRENT HIV RESEARCH. , 2017.

2. CELERINO DA SILVA, RONALDO; **RODRIGUES DE MOURA, RONALD**; VICTOR CAMPOS COELHO, ANTONIO; CLÁUDIO ARRAES, LUIZ; ANDRÉ CAVALCANTI BRANDÃO, LUCAS; CROVELLA, SERGIO; LIMA GUIMARÃES, RAFAEL
HLA-C single nucleotide polymorphism associated with increased viral load level in HIV-1 infected individuals from Northeast Brazil. CURRENT HIV RESEARCH. , 2017.

Trabalhos publicados em anais de eventos (resumo)

1. ARAGAO, M. A. L.; **MOURA, RONALD**; BRANDAO, L. A. C.; CROVELLA, S.
Prevalence Of Sickle Cell Disease Allele In 1000 Genomes Populations And Its Relationship With Skin Color And Genetic Ancestry In: VI SINATER - International Symposium on Diagnosis and Therapy, 2015, Recife.
VI SINATER - International Symposium on Diagnosis and Therapy. , 2015.

2. **MOURA, R. R.**; PIMENTEL, L. F.; LIMA, S. C.; OLIVEIRA, J. R. M.; SANDRIN-GARCIA, P.; CROVELLA, S.; BRANDAO, L. A. C.
A meta-analyses of association studies involving SNPs from PSEN1 and PSEN2 genes and Alzheimer's Disease In: V Sinater - Simpósio Internacional de Diagnóstico e Terapêutica e VIII Jornada Científica do LIKA, 2014, Recife.
V Sinater - Simpósio Internacional de Diagnóstico e Terapêutica e VIII Jornada Científica do LIKA. , 2014.

3. da SILVA, W. H. V. C.; **MOURA, R. R.**; COELHO, A. V. C.; BRANDAO, L. A. C.; CROVELLA, S.; GUIMARAES, R. L.
Pharmacogenomic diversity in 1000 Genomes database of HAART-related genes In: V Sinater - Simpósio Internacional de Diagnóstico e Terapêutica e VIII Jornada Científica do LIKA, 2014, Recife.
V Sinater - Simpósio Internacional de Diagnóstico e Terapêutica e VIII Jornada Científica do LIKA. , 2014.

Produção técnica

Entrevistas, mesas redondas, programas e comentários na mídia

1. **Moura, R.**; COELHO, A. V. C.; CROVELLA, S.
Pesquisa da UFPE identifica genes que auxiliam ação de vacinas anti-HIV, 2014

2. **Moura, R.**; COELHO, A. V. C.; CROVELLA, S.
UFPE à frente em pesquisa de vacina contra Aids, 2014

Eventos

Eventos
Participação em eventos

1. Conferencista no(a) **I Simpósio sobre Local do Crime**, 2015. (Simpósio)
Perícia e a Universidade.

2. Conferencista no(a) **Quinta Ciência**, 2015. (Seminário)
Passado, Presente e Futuro de Vacinas Terapêuticas Usando Células Dendriticas.

3. Apresentação de Poster / Painel no(a) **VI SINATER - International Symposium on Diagnosis and Therapy**, 2015. (Simpósio)
Prevalence Of Sickle Cell Disease Allele In 1000 Genomes Populations And Its Relationship With Skin Color And Genetic Ancestry.

4. Conferencista no(a) **I Encontro Pernambucano de Genética Forense**, 2014. (Seminário)
Forensic DNA Phenotyping: determinando a aparência através do DNA.

5. Apresentação de Poster / Painel no(a) **V Sinater - Simpósio Internacional de Diagnóstico e Terapêutica e VIII Jornada Científica do LIKA**, 2014. (Simpósio)

A meta-analyses of association studies involving SNPs from PSEN1 and PSEN2 genes and Alzheimer's Disease.

Citações

Web of Science Total de citações : 27;Total de trabalhos : 25;Data : 08/11/2017; Fator H: 3;
Nome(s) do autor utilizado(s) na consulta para obter o total de citações:
Moura, Ronald R.

Google Acadêmico Total de citações : 54;Total de trabalhos : 19;Data : 21/06/2016
Nome(s) do autor utilizado(s) na consulta para obter o total de citações:
Ronald Moura

Totais de produção

Produção bibliográfica

Artigos completos publicados em periódico	27
Artigos aceitos para publicação	2
Trabalhos publicados em anais de eventos	16
Apresentações de trabalhos (Comunicação)	2
Apresentações de trabalhos (Conferência ou palestra)	2
Apresentações de trabalhos (Congresso)	6

Produção técnica

Curso de curta duração ministrado (extensão)	5
Programa de Rádio ou TV (entrevista)	2

Eventos

Participações em eventos (congresso)	4
Participações em eventos (seminário)	10
Participações em eventos (simpósio)	3
Participações em eventos (encontro)	1
Participações em eventos (outra)	1
Organização de evento (outro)	1

Outras informações relevantes

1 Monitoria em Genética Geral pela Universidade Federal Rural de Pernambuco, durante o período de Julho de 2010 a Dezembro de 2011.