



UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA

PEDRO RODRIGUES SOUSA DA CRUZ

AVALIAÇÃO DAS VARIAÇÕES NO NÚMERO DE CÓPIAS
(CNV_s) NA SUSCETIBILIDADE AO ACIDENTE VASCULAR
CEREBRAL EM PACIENTES COM ANEMIA FALCIFORME

COPY NUMBER VARIATIONS (CNV_s) IN THE ASSESSMENT
OF STROKE SUSCEPTIBILITY IN PATIENTS WITH SICKLE
CELL ANEMIA

CAMPINAS

2016

PEDRO RODRIGUES SOUSA DA CRUZ

**AVALIAÇÃO DAS VARIAÇÕES NO NÚMERO DE CÓPIAS
(CNVs) NA SUSCETIBILIDADE AO ACIDENTE VASCULAR
CEREBRAL EM PACIENTES COM ANEMIA FALCIFORME**

**COPY NUMBER VARIATIONS (CNVs) IN THE
ASSESSMENT OF STROKE SUSCEPTIBILITY IN PATIENTS
WITH SICKLE CELL ANEMIA**

*Tese apresentada ao Instituto de
Biologia da Universidade Estadual
de Campinas para obtenção de título
de Doutor em Genética e Biologia
Molecular, na área de Genética
Animal e Evolução.*

*Thesis presented to the Institute of
Biology of the University of
Campinas in partial fulfillment of the
requirements for the degree of
Doctor in Genetics and Molecular
Biology, in the area of Animal
Genetics and Evolution*

**ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL
DA TESE DEFENDIDA PELO ALUNO PEDRO
RODRIGUES SOUSA DA CRUZ, E ORIENTADA PELA
PROF. DRA. MÔNICA BARBOSA DE MELO**

Orientadora: PROFA. DRA. MÔNICA BARBOSA DE MELO

CAMPINAS

2016

Agência(s) de fomento e nº(s) de processo(s): FAPESP, 2012/06438-5; CNPq, 140967/2012-5

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

C889a Cruz, Pedro Rodrigues Sousa da, 1987-
Avaliação das variações no número de cópias (CNVs) na susceptibilidade ao acidente vascular cerebral em pacientes com anemia falciforme / Pedro Rodrigues Sousa da Cruz. – Campinas, SP : [s.n.], 2016.

Orientador: Mônica Barbosa de Melo.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Genética médica. 2. Anemia falciforme. 3. Acidente vascular cerebral. 4. Variações do número de cópias de DNA. 5. Genética de populações. I. Melo, Mônica Barbosa de, 1968-. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Informações para Biblioteca Digital

Título em outro idioma: Copy number variations (CNVs) in the assessment of stroke susceptibility in patients with sickle cell anemia

Palavras-chave em inglês:

Genetics, Medical
Sickle cell anemia
Stroke
DNA copy number variations
Populations genetics

Área de concentração: Genética Animal e Evolução

Titulação: Doutor em Genética e Biologia Molecular

Banca examinadora:

Mônica Barbosa de Melo [Orientador]
Mônica Vanucci Nunes Lipay
Daniela Sanchez Basseres
Nicola Conran Zorzetto
Felipe Bastos Rocha

Data de defesa: 31-08-2016

Programa de Pós-Graduação: Genética e Biologia Molecular

BANCA EXAMINADORA

Dra. Mônica Barbosa de Melo (Orientadora)

Dr. Felipe Bastos Rocha

Profa. Dra. Mônica Vanucci Nunes Lipay

Profa. Dra. Daniela Sanchez Basseres

Dra. Nicola Amanda Conran Zorzetto

Os membros da Comissão Examinadora acima assinaram a Ata de Defesa, que se encontra no processo de vida acadêmica do aluno.

“Ela vem no som da chuva
Dançando devagar seu ijexá
Senhora da Candelária, abá
Pra toda a sua nação iorubá”
Ponto de Nanã

Aos meus pais e avós Marlene e João.

AGRADECIMENTOS

Agradeço à minha orientadora Profa. Mônica pela minha formação científica, por sua paciência e dedicação. Sua contribuição foi inestimável ao longo de oito anos como mentora, professora e inspiração.

Agradeço à Dra. Galina Ananina por seu papel crucial na minha forma de enxergar e exercer a ciência, pelos ensinamentos, conselhos e paciência.

Agradeço à Daniela Stancato, pelas incontáveis ajudas, apoio, ensinamentos e compreensão. Sua presença no laboratório foi crucial durante todo o meu período aqui.

Agradeço às pessoas que consentiram em participar do nosso trabalho, pela convicção na produção de conhecimento e na pesquisa.

Agradeço aos colegas e amigos do laboratório de genética humana, pelo companheirismo e momentos de descontração. Obrigado Bruno, Ana Maria, Ana Luiza, Paulo, Gabriela e Mirta, pela companhia, amizade e paciência, principalmente.

Agradeço amigos do CBMEG, em especial à Rafa, que se tornou antes de tudo uma grande amiga e uma apoiadora imprescindível. Agradeço à Letícia e à Laura, pela capoeiragem e amizade pura. Agradeço ainda ao Passarinho por ser um modelo como profissional da educação e como transmissor de tradição popular (gratidão!). Tenho a agradecer ainda ao Bruno, ao Matheus, ao Surf, ao Chefe e à Sabrina, por muitos dos melhores momentos que vivi em Campinas. Agradeço aos amigos que aqui fiz, Mari, Cleo, Robson, Joice, Diego, Sara, Aline e Carol, por fazerem de Campinas uma segunda casa pra mim. Agradeço ainda ao Ricardo e Geovana pelo companheirismo que às vezes laço de sangue desconhece. Agradeço também à Angélica, o maior apoio da maior parte desta jornada.

Agradeço aos amigos de longa data, Fábio, Diogo, Sirota, Zalcman, Bitoca, Rachel, Juliana, Pamela, Marcelo, Hashi, Gustavo, Panda, Chaves e Eric, pelas pessoas incríveis que são.

Agradeço à minha mãe, que além de maior inspiração pela garra, vivacidade e afeto, me ensinou a apreciar o sublime. Ao meu pai, que me ensinou a olhar para as raízes e é um modelo intelectual. Agradeço a toda minha família, sem a qual nunca nem teria ousado sonhar.

Por fim, agradeço à FAPESP e ao CNPq pela bolsa e financiamento imprescindíveis para a realização deste projeto

RESUMO

Embora a anemia falciforme (AF) resulte da homozigose de uma única mutação, na posição 6 do *locus* da β -hemoglobina, fenotipicamente, é uma doença muito heterogênea, de modo que diferentes pacientes apresentam evoluções clínicas significativamente distintas. Praticamente todos os órgãos podem ser afetados pela oclusão vascular, merecendo destaque o Sistema Nervoso Central (SNC), onde são observados acidentes isquêmicos transitórios, infartos e hemorragia cerebral, os quais acometem aproximadamente 25% dos pacientes com AF. As complicações neurológicas são graves, podendo ser fatais em até 15% dos casos. A identificação precoce de pacientes com anemia falciforme suscetíveis ao Acidente Vascular Cerebral (AVC) poderia diminuir os riscos, possivelmente prevenir a recorrência de infartos e potencialmente reduzir sua incidência. Portanto, estudos que visem identificar novos grupos de risco para o desenvolvimento de AVC em pacientes com anemia falciforme seriam fundamentais para otimizar o controle clínico desta enfermidade, sendo uma dessas vias a abordagem molecular deste grupo de pacientes. Desta forma, propusemos a investigação da presença de variação no número de cópias alélicas ou “Copy Number Variation” (CNV), utilizando-se lâminas de alta densidade (*microarrays*), com o objetivo de identificar regiões do genoma, potencialmente envolvidas com o risco aumentado de AVC em pacientes com anemia falciforme. Este trabalho propôs ainda avaliar os aspectos populacionais relativos à população brasileira. É de grande interesse nos estudos de genética humana entender a manifestação de dado fenótipo em populações distintas, e se a associação de variantes em uma população pode ser extrapolada para outras. A anemia falciforme fornece um modelo nesse intento, passível de ser extrapolado para doenças complexas dado ser um grupo de origem bem conhecida, exemplo clássico de seleção balanceada em humanos, primeira doença a ter base molecular desvendada e ser de herança mendeliana. Observou-se marcante diferença na ancestralidade de indivíduos afetados por AF brasileiros em relação a norte-americanos, tanto no genoma geral, quanto no cromossomo 11. Chegou-se a uma listagem de regiões com variações no número de cópias potencialmente associadas ao AVC em ambas as populações.

ABSTRACT

Although sickle cell anemia results from homozygosity for a single mutation at position 6 of beta-hemoglobin locus, this disease presents high heterogeneity in phenotype, so that different patients may have significantly different clinical outcomes. Virtually all organs may be affected by vascular occlusion, with emphasis on the Central Nervous System (CNS), where are observed transient ischemic attacks, stroke and cerebral hemorrhage, which affect approximately 25% of patients with SCA. Neurological complications are serious and can be fatal in up to 15% of cases. Early identification of patients with sickle cell anemia, susceptible to stroke (CVA) could reduce the risk, possibly preventing the recurrence of heart attacks and potentially reduce their incidence. Therefore, studies aimed at identifying new risk groups for development of stroke in sickle cell patients would be essential to optimize the clinical management of this disease, one of these pathways the molecular approach of this group of patients. Thus, we propose to investigate the presence of copy number variation in allele or "Copy Number Variation" (CNV), using high density microarray in order to identify genomic regions potentially involved in the increased risk of stroke in sickle cell patients. The present work also proposes to assess the Brazilian population genetic aspects. It is highly desirable in human genetics to unveil the manifestation of a given phenotype in distinct populations, as well as whether an association identified on a specific population may hold to other populations. Sickle cell anemia provides a model for this purpose, and may be extrapolated to complex diseases given its well-known geographical region, as well as for being a classic case of balanced selection in humans, the first disease to have its molecular mechanism unraveled, and a Mendelian trait. Ancestral components markedly differ in Brazilian SCA patients compared to the United States subjects on both the genomic landscape and the local ancestry on chromosome 11. Here we present a list of CNV events potentially associated to stroke on both populations.

LISTA DE SÍMBOLOS E ABREVIATURAS

AF	anemia falciforme
AVC	acidente vascular cerebral
CDCV	doença comum/variante comum
CHCM	concentração de hemoglobina corpuscular média
CNS	central nervous system
CNV	variação de número de cópias
CVA	cerebrovascular disease
dl	decilitro
DL	desequilíbrio de ligação
DNA	ácido desoxirribonucleico
dNTP	didesoxinucleotídeos
DNA	ácido desoxirribonucléico
EDTA	ácido etilenodiaminotetracético
ET	endotelina
FAP	fator ativador de plaqueta
FT	fator tecidual
FvW	Fator de von Willebrand
g	grama
GM-CSF	fator estimulador de colônias de granulócitos-macrófago
GWAS	genomewide association study
HbA	hemoglobina A
HbA2	hemoglobina A2
HbAS	traço falciforme
HBB	gene da beta globina
HbC	hemoglobina C
HbF	hemoglobina F
HIV	virus da imunodeficiência humana
HbS	hemoglobina S
HU	hidroxiuréia
HWE	equilíbrio de Hardy-Weinberg
ICAM-1	molécula de adesão intercelular-1
indel	polimorfismo do tipo inserção/deleção
Kb	quilobase
kDa	quilodalton
LCR	Locus Control Region
mRNA	RNA mensageiro
mL	mililitro
mm	milímetro
mtDNA	DNA mitocondrial
NADPH	de dinucleótido de nicotinamida e adenina
M-CSF	fator estimulador de colônia de macrófagos
NO	óxido nítrico
PC	componente principal
PCA	análise de componente principal
PHHF	persistência hereditária de hemoglobina fetal
pb	pares de bases
PCR	reação em cadeia da polimerase

μl	microlitro
μm	micrômetro
r²	coeficiente de desequilíbrio de ligação
RBC	eritrócitos
RFLP	polimorfismo no comprimento do fragmento de restrição
RM	ressonância magnética
RNA	ácido ribonucleico
ROS	espécie reativa de oxigênio
SNP	polimorfismo de base única
TAT	trombina-antitrombina
TGF- β	fator de crescimento e transformação beta
TNF	fator de necrose tumoral
VCAM-1	molécula de adesão vascular-1
VCM	volume corpuscular médio
VEGF	fator de crescimento do endotélio vascular

SUMÁRIO

RESUMO.....	7
ABSTRACT	8
INTRODUÇÃO	12
Constituição das hemoglobinas saudáveis	13
Regiões codificadoras e <i>switchings</i> de hemoglobina	13
Variantes anômalas de hemoglobina.....	14
Hemoglobina S e polimerização.....	15
Perturbação da membrana celular	17
Hemólise e equilíbrio de óxido nítrico.....	18
Hemoglobina S e vaso-occlusão	19
Prevalência	22
Diagnóstico laboratorial	23
Diversidade fenotípica da anemia falciforme	24
Manifestações das doenças falciformes.....	25
Quadro clínico: complicações neurológicas.....	25
Ferramentas genéticas para a identificação de genes moduladores.....	30
Parâmetros populacionais	37
População brasileira	41
CAPÍTULO I	46
COPY NUMBER VARIATIONS AND THE RISK OF CEREBROVASCULAR DISEASE IN PATIENTS WITH SICKLE CELL ANEMIA	46
CAPÍTULO II.....	67
FINE SCALE GENETIC STRUCTURE OF THE POPULATIONS FROM THREE BRAZILIAN REGIONS	67
CAPÍTULO III	95
COMPARISON BETWEEN TWO DIFFERENT SICKLE CELL ANEMIA COHORTS: HIGH DIVERGENCE BETWEEN BRAZILIANS AND AMERICANS	95
DISCUSSÃO	118
CONCLUSÕES.....	123
REFERÊNCIAS BIBLIOGRÁFICAS	125
ANEXO I.....	149
ANEXO II.....	152
ANEXO III.....	153
ANEXO IV	155
ANEXO V.....	156

INTRODUÇÃO

Constituição das hemoglobinas saudáveis

Em 1926 por meio da técnica de ultra-centrifugação foi determinada a constituição da hemoglobina como um tetrâmero de peso molecular de 66,8 kDa. No entanto, a estrutura molecular da hemoglobina foi demonstrada apenas em 1970 (Perutz, 1970).

A hemoglobina é uma metaloproteína composta por um heterotetrâmero de duas subunidades de cadeias do tipo α -globina de 141 resíduos de aminoácidos cada e duas subunidades de cadeias do tipo não- α de 146 resíduos cada. Os monômeros encontram-se ligados a um grupamento prostético heme, responsável pela associação reversível ao oxigênio.

Os genes que codificam as diferentes cadeias globínicas são membros de uma mesma família ancestral. As cadeias do grupo α -globina são codificadas em um segmento de 35kb do cromossomo 16 pelos genes ξ , $\alpha 1$ e $\alpha 2$, enquanto os genes das cadeias do tipo β -globina encontram-se no cromossomo 11, ao longo de 60kb e são denominados ϵ , γ^G , γ^A , δ e β (MAVILIO Et al., 1983). Em ambos os *clusters* observa-se a disposição dos genes na ordem em que são expressos ao longo do desenvolvimento para a formação das hemoglobinas embrionárias ($\xi_2\epsilon_2$, $\xi_2\gamma_2$ e $\alpha_2\epsilon_2$), fetal ($\alpha_2\gamma_2$, HbF) e adultas ($\alpha_2\beta_2$, HbA e $\alpha_2\delta_2$, HbA₂).

Regiões codificadoras e *switchings* de hemoglobina

Em humanos, as hemoglobina embrionárias são produzidas majoritariamente em eritroblastos em desenvolvimento no saco vitelínico, até que o primeiro *switch* ocorre com a interrupção na produção, na 12^a semana de gestação, de cadeias ξ e ϵ e subsequente expressão de α e γ , concomitantemente à transferência da hematopoiese para o fígado do feto (WILBER; NIENHUIS; PERSONS, 2014).

A hemoglobina fetal (HbF) foi discriminada pela primeira vez da hemoglobina adulta (HbA) em 1866 por sua resistência diferencial à denaturação alcalina (KORBER, 1866). Aproximadamente 70 anos depois, Brinkan e colaboradores concluíram que a diferença na denaturação era devida à porção globínica da cadeia (BRINKMAN; WILDSCHUT; WITTERMANS, 1934). A hemoglobina fetal se diferencia de sua contraparte adulta por sua maior afinidade ao oxigênio, dada sua menor interação com o bisfosfoglicerato presente no interior do eritrócito, permitindo assim ao feto a absorção de oxigênio através do córion.

O segundo *switch*, presente apenas em seres humanos e primatas do velho mundo, envolve a redução na produção de globinas γ , conjuntamente com o gradativo aumento na síntese da cadeia β e, em menor grau, δ , do período perinatal até os primeiros dois anos de vida (WILBER; NIENHUIS; PERSONS, 2014), após o qual observa-se predominância das hemoglobinas adultas (HbA e HbA2) e entre 0.1 a 3% de HbF, processo representado esquematicamente na Figura 1.

Na década de 1960, foi descrita a permanência de níveis elevados de HbF em indivíduos adultos (WHEELER; KREVANS, 1961), fenômeno denominado persistência hereditária de hemoglobina fetal (PHHF).

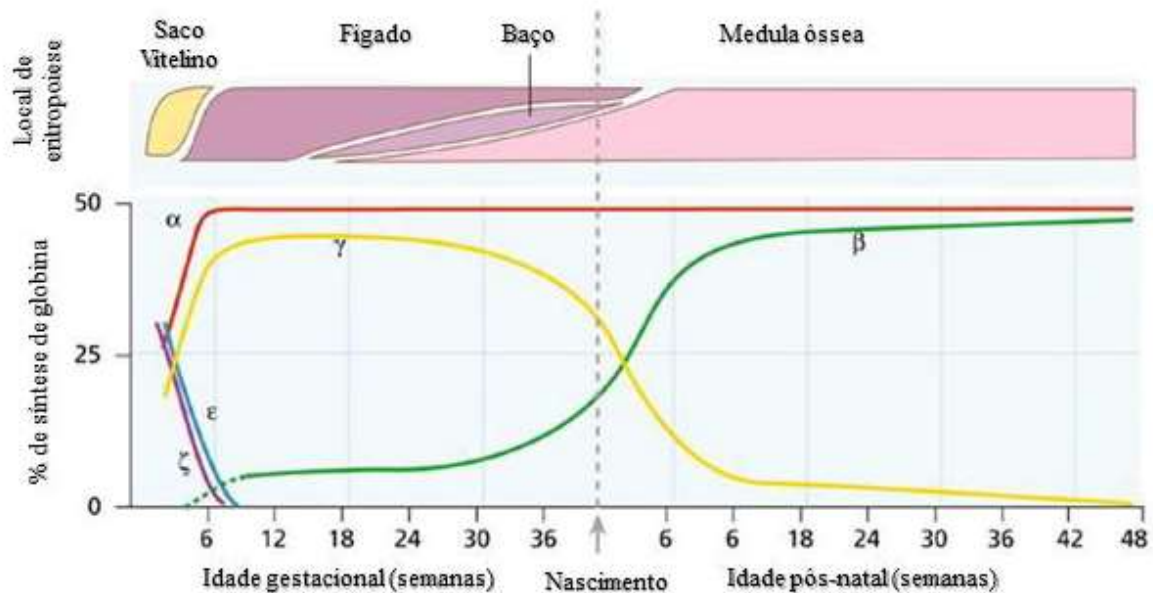


Figura 1. *Switching* de cadeias globínicas durante o desenvolvimento. (Adaptado de HOFFBRAND, CATOVSKY, TUDDENHAM, 2010).

Variantes anômalas de hemoglobina

Há mais de 700 variantes descritas da hemoglobina, em sua vasta maioria resultantes de mutações pontuais nas cadeias α , γ , β , ou δ , muitas das quais não acarretam alteração fenotípica (DRISS, 2008). As variantes advindas de mutação *missense* se manifestam em variação estrutural e são denominadas hemoglobinopatias, as doenças genéticas de maior distribuição global (entre as mais frequentes na maior parte das populações figuram a HbS, HbC, HbE, e HbD-Punjab). Mutações de ponto em regiões não codantes afetando a estrutura de mRNA e regiões regulatórias dos genes das cadeias de globina estão relacionadas às

diferentes talassemias, sendo outras alterações genéticas mais raras as responsáveis pelas hemoglobinas *Constant Spring* (substituição de códon de parada por sequência codificante) e *Lepore* (fusão dos genes δ e β).

Sendo uma variante de relativa alta frequência, a hemoglobina S (HbS, $\alpha_2\beta_2^S$) caracteriza o conjunto de hemoglobinopatias denominado Doenças Falciformes. A primeira descrição desta condição se deu em 1910 por Herrick ao observar eritrócitos de morfologia anômala em um estudante jamaicano portador de grave quadro anêmico (HERRICK, 1910), posteriormente Neel propôs que a manifestação da anemia falciforme é decorrente de um estado homozigoto, sendo a heterozigose responsável por falcização menos frequente e quadro assintomático (traço falciforme) (Neel, 1949). Devem-se a Pauling e cols. a distinção da HbS da forma adulta saudável (HbA) por seu ponto isoelétrico e migração eletroforética, permitindo identificar sua ocorrência em conjunto com a HbA de modo a discriminar traço e anemia falciforme (PAULING; ITANO, 1949).

Em 1956 revelou-se como fundamento genético da HbS a mutação pontual na sexta posição da cadeia da β -globina (GAG \rightarrow GTG), a qual resulta na substituição de um ácido glutâmico por uma valina ($\beta_6\text{Glu} \rightarrow \text{Val}$) (INGRAM, 1956).

Hemoglobina S e polimerização

A HbS apresenta cinco haplótipos atualmente reconhecidos como de origens étnica e geográfica distintas, CAR (República Centro Africana) ou Bantu no centro-sul e leste africano; o tipo Benin (BEN) originado no meio-oeste africano; o tipo Senegal (SEN) característico da África atlântica; o tipo Saudita ou Árabe-indiano presente na península arábica e na Índia; e o tipo Camarões encontrado nos limites geográficos deste país e numa pequena parte da costa oeste africana (ADEKILE, 1997; LAPOUMÉROULIE et al., 1992; NAGEL, 1994). Cada haplótipo corresponde a um evento independente de surgimento de HbS e a manutenção desta deveu-se à seleção balanceada, tendo em vista o caráter protetor do alelo para a infecção por *Plasmodium falciparum*, de alta incidência nas regiões citadas (do mesmo modo como ocorre para a HbC e a HbD) (CAVALLI-SFORZA, 1981).

A presença da valina na posição seis da globina β favorece a polimerização sob condições de baixo teor de oxigênio, redução de pH e/ou aumento de temperatura (Figura 2). Uma vez que as moléculas de hemoglobina encontram-se em concentração muito alta

intracelularmente (32-34 g/dl), tal compactação exige grande solubilidade (BUNN, 1997). Com a desoxigenação da hemoglobina, há a criação de um sítio hidrofóbico pela Val- β 6 exposta na superfície da molécula e os resíduos Phe-85 β e Leu-88 β das cadeias adjacentes cumprem o papel de regiões aceptoras (CHATTERJEE et al., 1982). A partir de um único ponto de nucleação, as moléculas de hemoglobina agregam em polímeros helicoidais de 14 unidades, os quais atuam como molde para a formação de longos feixes no interior do eritrócito (EATON; HOFRICHTER, 1995). Com a drástica redução da solubilidade ocasionada pela mutação estrutural e pelo mecanismo de polimerização em cadeia descrito, a concentração de hemoglobina corpuscular média (CHCM) adquire também importância para a falcização.

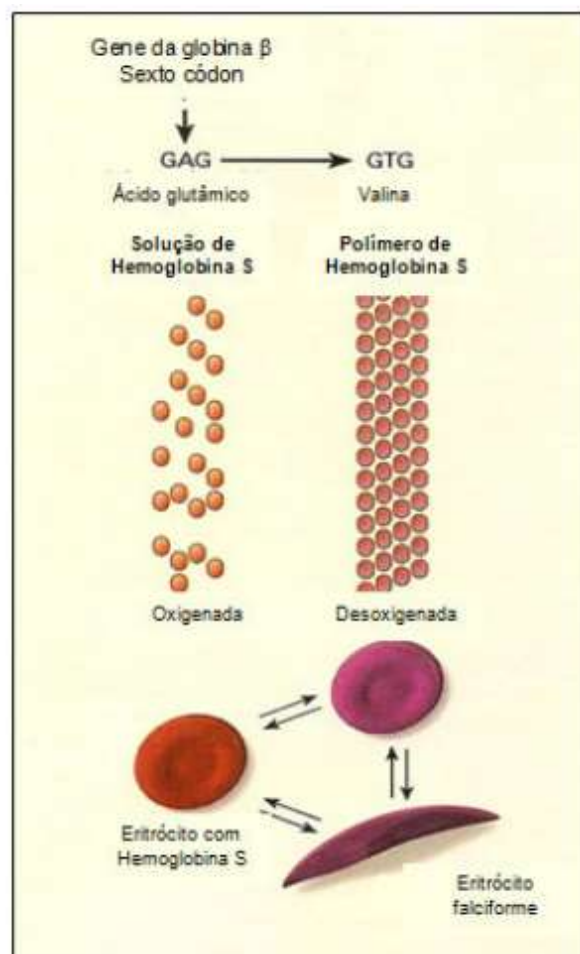


Figura 2. Processo de falcização do eritrócito por polimerização de HbS (adaptado de MAY, 2010).

A velocidade com que ocorre o início da polimerização depende da concentração de HbS elevada entre sua vigésima e quadragésima potência (EATON.; HOFRICHTER, 1987), com a redução da CHCM em 1g/dl ocasionando o atraso do início de formação de polímeros ao dobro do tempo (EMBURY, 1985).

A coocorrência com outros tipos de hemoglobina é ainda determinante para a tendência de polimerização da HbS (HEBBEL; OSAROGIAGBON; KAUL, 2004). A HbF ($\alpha_2\gamma_2$) e HbA₂ ($\alpha_2\delta_2$) limitam a polimerização em maior medida que HbC ($\alpha_2\beta_2^C$) e HbA ($\alpha_2\beta_2$), em virtude da ausência de cadeias β na composição das primeiras. Os haplótipos Senegal e Árabe-indiano estão associados com níveis de HbF elevados (superiores a 15%) em relação aos haplótipos Benin (de 5 a 15%) e CAR/Bantu (abaixo de 5%) (HORIUCHI et al., 1995; MUSUMECI et al., 1981), motivo pelo qual muitas vezes o haplótipo é aludido como fator modificador do potencial de polimerização.

Perturbação da membrana celular

A recorrência e a magnitude da polimerização perturbam a estabilidade da membrana plasmática e podem conduzir à deformação irreversível do eritrócito, o qual usualmente toma a forma de foice (JOINER, 1993).

Em um estudo clássico de 1952 foi demonstrado pela primeira vez que a falcização apresenta-se fortemente relacionada com o conteúdo iônico dos eritrócitos, nomeadamente depleção de K^+ e aumento de Na^+ (TOSTESON; SHEA; DARLING, 1952), com resultante perda de água e aumento na densidade dos eritrócitos, favorecendo o aumento de polímeros de HbS (BRUGNARA, 2003). Ao menos três vias de troca iônica encontram-se modificadas, a Na^+/K^+ -ATPase induzida por desoxigenação e promotora do influxo de Ca^{2+} , o qual afeta o canal de Gardos (LEW et al., 1985), além do cotransporte K^+-Cl^- , ativado em pH inferior a 7,4 (BRUGNARA, et al., 1989). Tais achados motivaram a interpretação de que a deformação do eritrócito por fibras de HbS seja a responsável pelo fluxo catiônico anormal, uma vez que a alteração de permeabilidade é tão maior quanto maior for a o grau de alteração morfológica do eritrócito (MOHANDAS; ROSSI; CLARK, 1986). Certas populações de eritrócitos falcizados exibem acentuada mudança de perfil catiônico e são os responsáveis primários nas alterações reológicas das doenças falciformes (BRUGNARA, 2003).

Este grupo celular irreversivelmente falcizado exhibe grande densidade celular e pode alcançar valores de CHCM próximos a 50 g/dl. Células densas são menos observadas quando há maior concentração de HbF nos eritrócitos ou quando a HbS coexiste com α -talassemia, porquanto reduzem a concentração de hemoglobina (FABRY et al., 1984). A proporção de células irreversivelmente falcizadas e a sobrevivência dos eritrócitos estão inversamente relacionados (SERJEANT, 1969).

Hemólise e equilíbrio de óxido nítrico

O óxido nítrico (NO) é produzido pelas células endoteliais a partir da arginina e tem a função de regular a vasodilatação por meio da ativação da guanilato ciclase solúvel, que produz 3',5'-monofosfato de guanosina cíclico na musculatura lisa vascular, responsável pela vasodilatação (STEVENS et al., 1981; GILL et al., 1995;), além de atuar como agente anti-inflamatório, anti-oxidante, modulador de moléculas de adesão, anticoagulante e inibidor da proliferação endotelial (KATO et al., 2006).

A meia-vida dos eritrócitos falcizados é, em média, sete dias, em contraste aos 60 dias de eritrócitos normais (HEBBEL, et al., 2009). A grande taxa de hemólise observada é em sua maior parte derivada da elevada eritrofagocitose por monócitos e neutrófilos nos tecidos extra-vasculares e é desencadeada principalmente pela ligação de anticorpos à membrana eritrocítica, mais sensível à lise mediada pelo sistema complemento (HEBBEL et al. 1984). No ambiente intravascular ocorre a ruptura mecânica das células falcizadas, e esta responde por um terço dos eventos de hemólise e como causa principal de liberação de hemoglobina no plasma (HEBBEL, et al., 2009). A hemoglobina plasmática consome NO, formando metahemoglobina e nitrato bioativo. Heme e íons de ferro liberados, bem como a hemoglobina, catalisam a produção de espécies reativas de oxigênio (ROS), já aumentada nas doenças falciformes por ação de xantina oxidase (HOUSTON et al., 1999), mieloperoxidase (BALDUS et al., 2002), NADPH oxidase (WOOD; HEBBEL; GRANGER, 2005) e leucócitos (HEBBEL et al., 2004), reduzindo a biodisponibilidade de NO (KATO; GLADWIN; STEINBERG, 2007).

Esta redução de NO aparece como um fator central na vasculopatia observada na anemia falciforme (Figura 3) (KATO et al., 2006). Os fenômenos que provocam esta redução são a reação química Fenton (devido a presença de hemoglobina e de ferro livres, levando à conversão de NO a nitrato e nitrito e transformação de hemoglobina a metahemoglobina (HOUSTON et al., 1999), a reação enzimática na qual NO é degradado a peróxido de nitrito pela ação das enzimas xantina oxidase e NADPH oxidase liberadas por células endoteliais e neutrófilos hipóxicos. Há ainda ação da arginase posta em circulação por eritrócitos lisados, que consome L-arginina (convertendo-a a ornitina), substrato da óxido nítrico sintetase e precursora de NO. Dado o papel deste na geração de guanosina monofosfato cíclico e relaxamento da musculatura lisa, sua depleção incorre em vasoconstrição e consequente

desequilíbrio da homeostase vascular, refletido em ativação endotelial e plaquetária (KATO; GLADWIN; STEINBERG, 2007).

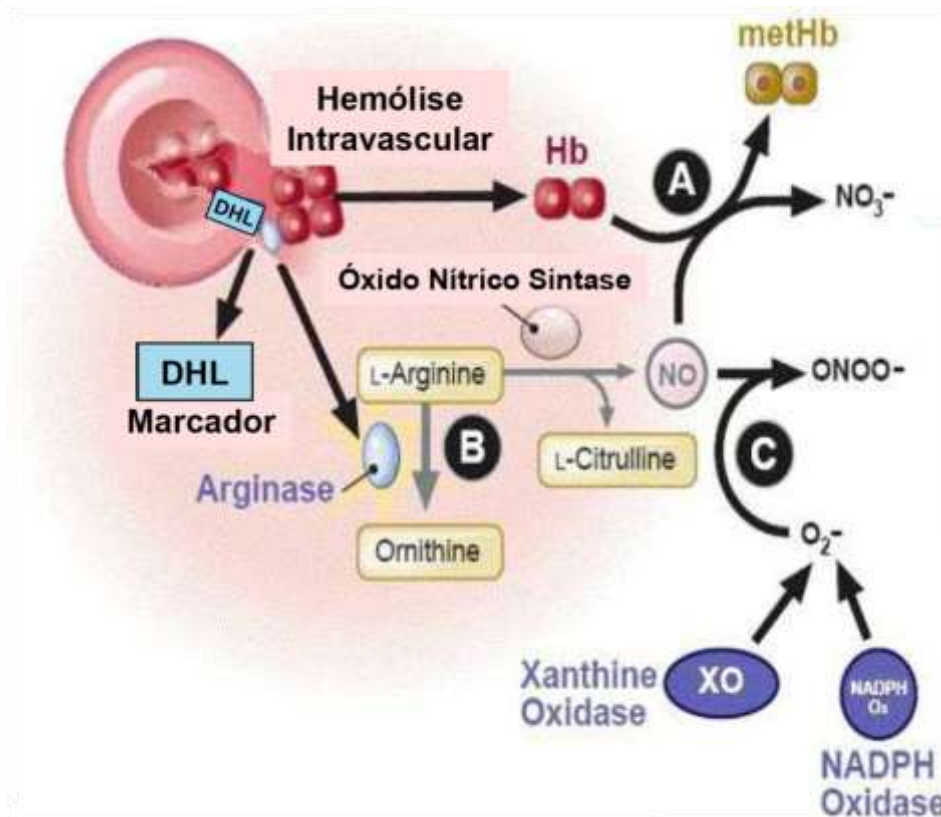


Figura 3. Hemólise intravascular reduz a atividade do ON adaptado de Kato, GJ (2007). Hemólise libera hemoglobina, arginase e LDH, limitando a biodisponibilidade do NO tanto diminuindo a sua síntese como a sua destruição. NO reage com hemoglobina formando metahemoglobina e nitrato (A). Arginina, o substrato para a síntese de NO pela sintetase do NO, é degradada a ornitina enzimaticamente pela arginase plasmática (B). LDH é um indicador do grau de hemólise: quantidade de hemoglobina e arginase liberadas. NO é inativado pelos radicais livres de oxigênio produzidos pela xantina oxidase, NADPH oxidase e sintetase do NO não acoplada a arginina produzindo peroxinitrito ONOO- (C) (Adaptado de KATO; GLADWIN; STEINBERG, 2007).

Hemoglobina S e vaso-oclusão

A retenção de eritrócitos densos em vênulas pós-capilares é o fenômeno classicamente atribuído à obstrução microvascular (KAUL et al., 1986) nas doenças falciformes. A extensão da adesão é inversa e intensamente relacionada ao diâmetro do vaso e se dá conforme a densidade do eritrócito (KAUL et al., 1989). Células mais densas são retidas ao acaso na rede pós-capilar, não obstante de maneira preferencial em vênulas com diâmetro entre 7,0 e 10,0 μ m, localidades em que já se encontram aderidas células menos densas em concentração de 1,0/100 μ m², próxima do limiar de bloqueio do vaso (KAUL et al., 1989).

Tanto a velocidade de fluxo quanto a tensão de cisalhamento (pressão tangencial do fluido na parede do vaso) têm sua taxa mais alta nas arteríolas, seguindo-se a abrupta queda de ambos conforme os eritrócitos alcançam os vasos imediatamente posteriores aos capilares (LIPOWSKY; KOVALCHECK; ZWEIFACH, 1978), ponto cuja aderência é mais acentuada. A retenção é reduzida de modo consistente com a retomada de velocidade e taxa de cisalhamento ao longo de vasos mais calibrosos e de ramificações mais esparsas (KAUL et al., 1989), dado que a topografia exerce influência. A combinação de obstrução pré-capilar e detenção de células densas induz hipóxia local e estimula a polimerização de HbS de tal modo que a contenção chegue potencialmente a abranger vasos contíguos em perímetro maior (KAUL et al., 1989).

A despeito de tradicionalmente atribuir-se protagonismo aos eritrócitos densos para a vaso-oclusão, observou-se que a proporção desses não prediz a gravidade das complicações decorrentes das doenças falciformes (BILLETT et al., 1986), exceção feita ao componente hemolítico (BUNN, 1997; KAUL, et al., 1996). A vaso-oclusão é fortemente relacionada à interação anormal entre os eritrócitos irreversivelmente falcizados e o endotélio vascular (BANDEIRA; SILVA, 2007), porém atualmente acredita-se que o fenômeno vaso-oclusivo compreende um processo que envolve, além de eritrócitos, leucócitos ativados, células endoteliais, plaquetas e proteínas do plasma (Figura 4) (ZAGO; FALCÃO; PASQUINI, 2013).

A liberação de hemoglobina pelos eritrócitos falcizados com resultante produção de espécies reativas de oxigênio (ROS) e redução de NO, aliada à isquemia e reperfusão, leva à ativação do endotélio vascular. Tal ativação se dá com a expressão de moléculas de adesão como VCAM-1 (molécula de adesão vascular -1), ICAM-1 (molécula de adesão intercelular-1), CD36 e E-selectina na superfície celular e a produção de citocinas e quimiocinas como Interleucina (IL)-8, IL-6 e GM-CSF (fator estimulador de colônias de granulócitos-macrófagos), além de fatores procoagulantes e vasoconstritores, como as Endotelinas 1 e 2 (ET-1, ET-2). Como consequência, há a captura de células brancas e vermelhas, ambas com propriedades adesivas também aumentadas (ZAGO; FALCÃO; PASQUINI, 2013). A E-selectina e a P-selectina modulam a rolagem na parede do vaso, enquanto VCAM-1 e ICAM-1 são responsáveis pela posterior aderência firme dos leucócitos (que por sua vez apresentam em sua superfície as integrinas Mac-1 e LFA-1, L-selectina e ligantes da E-selectina) ao endotélio (CAO et al., 1999). ICAM-1 liga-se ainda à integrina $\alpha 4\beta 1$ (VLA-4) de reticulócitos, os quais possuem também as moléculas de adesão CD36, ICAM-4 e BCAM/Lu atuando na aderência ao endotélio vascular (LUTTY; MCLEOD, 2008). Indivíduos com doença falciforme

apresentam taxas de neutrófilos, ICAM-1, VCAM-1 e P-selectina mais elevadas, o que corrobora com o modelo de que leucócitos e reticulócitos se aderem ao endotélio vascular, levando à obstrução vascular (CAO et al., 1999).

A ativação de leucócitos é favorecida por citocinas inflamatórias, observadas em maior número nestes pacientes, possivelmente em decorrência da lesão tecidual produzida por isquemia. Diversas moléculas inflamatórias se apresentam em níveis elevados na anemia falciforme, incluindo TNF- α (Fator de Necrose Tumoral- α), IL-1 β , proteína C reativa (potentes ativadores do endotélio), M-CSF (fator estimulador de colônia de macrófagos), IL-3, GM-CSF, IL-8 e IL-6 (ZAGO; FALCÃO; PASQUINI, 2013). Como resultado, populações de células brancas como de granulócitos e neutrófilos tornam-se mais aderentes ao endotélio, tendo sido demonstrado que o aumento no número de ambas as populações está associado a um quadro clínico mais grave (ANYAEGBU et al., 1998; ZAGO; FALCÃO; PASQUINI, 2013).

Atualmente considera-se que o processo vaso-oclusivo é desencadeado pela adesão de leucócitos (em especial dos neutrófilos, já que são células grandes, de 12 a 15 μm , relativamente rígidas e interagem com eritrócitos falcizados) (HOFSTRA et al., 1996), à parede da microvasculatura (ZAGO; FALCÃO; PASQUINI, 2013). Nesse cenário, a inflamação conduz o recrutamento de neutrófilos para as vênulas pós-capilares, os quais podem apresentar interações heterotípicas com hemácias falcizadas via Mac-1, além de ligar-se ao endotélio ativado, ocasionando estase na microcirculação, maior tempo de circulação (polimerização aumentada de hemoglobina S) e, por fim, obstrução vascular (ZHANG et al., 2016). Adicionalmente, leucócitos aderidos ao endotélio podem intermediar a adesão secundária de eritrócitos, com resultante obstrução física dos pequenos vasos da microcirculação (ZAGO; FALCÃO; PASQUINI, 2013).

Observa-se ainda na anemia falciforme um estado de hipercoagulabilidade, provavelmente pela ativação do endotélio e plaquetas, com níveis elevados de marcadores de ativação de trombina, plaquetas e células endoteliais, tais como Trombina-Antitrombina (TAT), dímero-D, ligante de CD40 solúvel (CD40L), Fator Tecidual (FT), Fator Ativador de Plaquetas (FAP) e Fator de von Willebrand (FvW) (ZAGO; FALCÃO; PASQUINI, 2013). Plaquetas de pacientes afetados por anemia falciforme encontram-se ativadas e podem ligar-se a eritrócitos, monócitos e neutrófilos, formando agregados (WUN et al., 1997; ZARBOCK; POLANOWSKA-GRABOWSKA; LEY, 2007).

O hematócrito também desempenha um importante papel fisiopatológico, devido a sua influência na viscosidade do sangue. Isso pode explicar em parte a discrepância entre a gravidade das doenças SS, SC e SThal. Estas últimas tendem a possuir maior hematócrito, contribuindo para uma viscosidade ainda mais elevada e, portanto, maior oclusão na microvasculatura durante o processo de falcização.

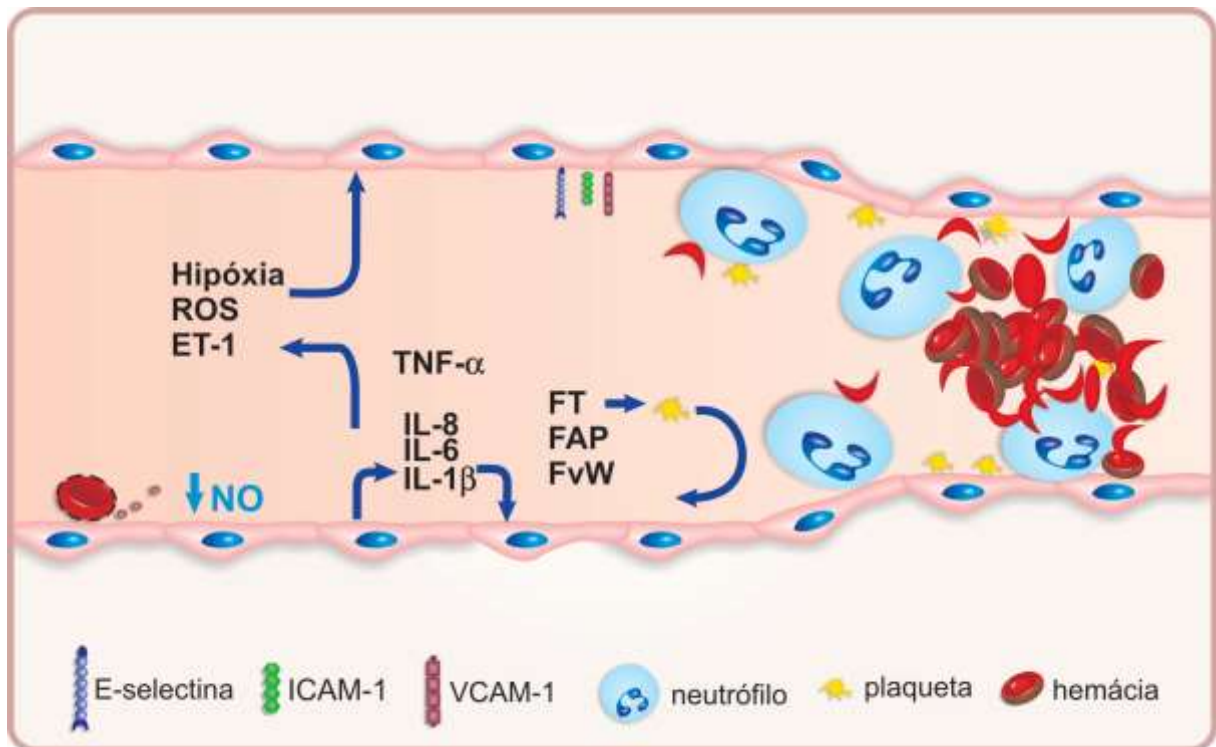


Figura 4. Representação esquemática do processo de vaso-oclusão. A adesão de eritrócitos e a hemólise intravascular levam à ativação das células endoteliais. Heme livre consome o óxido nítrico (NO) sintetizado pelas células endoteliais, esse mecanismo contribui para a ativação das células endoteliais, em conjunção com a presença de Espécies Reativas de Oxigênio (ROS) e vasoconstritores como Endotelina-1 (ET-1). O estado de hipercoagulabilidade leva a níveis aumentados de Fator Tecidual (FT), Fator Ativador de Plaquetas (FAP), Fator de von Willebrand (FvW) e ativação plaquetária. A ativação do endotélio também conta com a presença das plaquetas ativadas no vaso sanguíneo e a sua adesão à parede vascular. Após sua ativação, as células endoteliais expressam moléculas de adesão de (como ICAM-1, VCAM-1 e E-selectina) forma mais acentuada na superfície do vaso e também liberam mais citocinas e quimiocinas como o IL-8, IL-6 e IL-1 β , que com o TNF- α , contribuem para a inflamação vascular e a ativação das células sanguíneas. Os leucócitos e as hemácias aderem à parede vascular devido à sua ativação e expressão de moléculas de adesão e, também, podem haver interações heterotípicas entre leucócitos aderidos e hemácias. A vasoconstrição elevada e a obstrução física do vaso ocasionam uma redução no fluxo sanguíneo com conseqüente hipóxia e falcização das hemácias, dificultando a passagem do sangue, e finalmente resultando na vaso-oclusão (ZAGO; FALCÃO; PASQUINI, 2013).

Prevalência

A anemia falciforme é uma hemoglobinopatia que leva à formação de eritrócitos anômalos, sendo a doença monogênica mais comum no Brasil (MANFREDINI; CASTRO;

WAGNER, 2007). Atinge frequentemente, mas não exclusivamente, indivíduos de origem africana, chegando a acometer 0,1 a 0,3% da população negra, com tendência a atingir parcela cada vez mais significativa da população, devido ao alto grau de miscigenação do país (BANDEIRA et al., 1999).

A maior prevalência da anemia falciforme ocorre entre indivíduos negros por sua incidência elevada na África, embora seja também encontrada em países Mediterrâneos, principalmente Grécia e Itália, e ainda em Israel, Arábia Saudita e Índia. No Brasil, as frequências variam de 5 a 10% de heterozigotos HbAS em descendentes de africanos, sendo que no Estado de São Paulo, esta frequência é pouco menor que 2% (SALZANO, 1985; PAIXÃO ET AL., 2001).

A anemia falciforme é resultante da homozigose do gene β^S , correspondendo à forma mais grave das síndromes falciformes. Ocorre o predomínio da produção de HbS, ausência de HbA, acompanhadas de quantidade normais de HbA₂ e aumento moderado de HbF. Quando o gene β^S aparece em heterozigose (traço falciforme), usualmente não são observadas anormalidades hematológicas nos portadores. A morfologia eritrocitária, o nível de hemoglobina, assim como os números de leucócitos e plaquetas são normais. Estima-se que somente no Brasil haja aproximadamente 2 milhões de heterozigotos HbAS (STEINBERG, 1998).

Diagnóstico laboratorial

As técnicas mais utilizadas para o diagnóstico das síndromes falciformes incluem a eletroforese de hemoglobina em acetato de celulose com pH 8,4, em gel de ágar com pH 6,2 e teste de solubilidade da hemoglobina em tampão fosfato concentrado.

Os níveis de hemoglobina em pacientes na fase estável da anemia falciforme variam de 6 a 10g/dl, sendo a anemia normocrômica e normocítica, em geral, apesar dos níveis de reticulócitos elevados (entre 5 e 20%). Ocasionalmente podem ser observados eritroblastos circulantes (entre 5 e 20%). As hemácias em forma de foice são observadas nas doenças falciformes, embora outras formas alteradas também possam ser visualizadas.

Os clássicos dados laboratoriais de hemólise observados são: elevação de bilirrubina indireta, redução de haptoglobina, elevação de urobilinogênio urinário e hiperplasia eritróide na medula óssea. A leucocitose, alteração nem sempre relacionada com processo

infecioso, pode ser observada mesmo na fase estável. A contagem de plaquetas está em geral elevada, podendo atingir até 1.000.000 por μl . As provas de coagulação são em geral normais durante a fase estável, mas durante os episódios vaso-oclusivos alguns testes podem apresentar alterações características de hipercoagulabilidade (BALLAS, 2001; BUNN, 1997; STEINBERG, 1998).

Diversidade fenotípica da anemia falciforme

Embora a anemia falciforme resulte da homoziguidade de uma única mutação, na posição 6 do *locus* da β -hemoglobina, fenotipicamente, essa doença é muito heterogênea, de modo que os pacientes podem apresentar evoluções clínicas significativamente distintas. Dentre os moduladores conhecidos, o nível de HbF talvez seja o mais importante (Platt et al., 1994). Outros possíveis moduladores desta variabilidade fenotípica são os haplótipos do gene da β -globina, a região controladora da expressão do complexo do gene da β -globina (LCR – *Locus control region*), mutações que causam persistência hereditária de hemoglobina fetal e talassemia alfa (PLATT ET AL., 1994; OHENE-FREMPONG ET AL., 1998; SARNAIK & BALLAS, 2001; SEBASTIANI ET AL., 2005).

Com relação aos haplótipos do gene da β -globina, são definidos por polimorfismos do DNA em *cis* ligados ao complexo desse gene. A denominação dos haplótipos foi derivada da região geográfica onde cada haplótipo apresentou prevalência mais elevada, sendo eles: Benin, República Centro-Africana, Senegal, Camarões e Árabe-Indiano. Os haplótipos podem ser utilizados como marcadores de características genéticas em *cis*, herdadas juntamente com o gene β^S , podendo influenciar a expressão dos genes γ -globina, e consequentemente, os níveis de HbF. Os haplótipos Senegal e Árabe-Indiano estão associados a níveis elevados de HbF, enquanto que Benin e Banto a níveis menores de HbF. No Brasil, os haplótipos predominantes são Banto e Benin (BUNN, 1997; STEINBERG, 1998).

É frequente a associação entre anemia falciforme e talassemia-alfa, sendo que em negros, ela deriva quase exclusivamente da deleção de um dos genes da globina alfa. No Brasil, cerca de 20% dos negros são heterozigotos para talassemia-alfa. As consequências dessa associação são: redução do VCM, menor número de reticulócitos, menor grau de hemólise e maior concentração de hemoglobina, quando comparado com pacientes apresentando genótipo normal para os genes da globina alfa. Aparentemente, úlceras de perna, acidente vascular

cerebral e anormalidades da retina são menos frequentes em pacientes com anemia falciforme associada à talassemia alfa (STEINBERG, 1998).

Manifestações das doenças falciformes

As manifestações clínicas, tais como vaso-oclusão, necrose avascular da medula óssea e fibrose esplênica progressiva, se devem, principalmente, à hipóxia resultante da obstrução dos vasos sanguíneos por acúmulo de eritrócitos falcizados e da destruição precoce destes (BRUGNARA, 2003). Os órgãos que sofrem maiores riscos são aqueles com seio venoso, onde a circulação do sangue é lenta e a tensão de oxigênio e pH são baixos, ou aqueles com limitada suplementação de sangue arterial, como olhos e cabeça do fêmur (MANFREDINI et al., 2007).

Os sintomas da hipóxia podem ser agudos com crises dolorosas ou insidiosas. Esta última categoria envolve a necrose asséptica de quadris e retinopatia falciforme. Os efeitos dos danos teciduais agudos ou crônicos podem resultar na falência do órgão, principalmente em pacientes com idade avançada (MANFREDINI et al., 2007).

Quadro clínico: complicações neurológicas

O quadro clínico nas doenças falciformes é variável e derivado, primariamente, da oclusão vascular e, em menor grau, da anemia. Praticamente todos os órgãos podem ser afetados pela oclusão vascular (BALLAS, 1998).

Cerca de 25% dos pacientes com anemia falciforme são acometidos por algum tipo de comprometimento neurológico, sendo estes mais comuns em pacientes homozigotos HbSS, incluindo acidente vascular cerebral (AVC), hemorragia cerebral e ataques isquêmicos transitórios. Os infartos cerebrais são mais frequentes em crianças enquanto a hemorragia intracerebral é mais frequente em adultos. As complicações neurológicas são graves, podendo ser fatais em até 15% dos casos (PEGELow et al., 1997; WIERENGA et al., 2001).

Um estudo sobre a expectativa de vida de pacientes com anemia falciforme nos EUA na década de 80 demonstrou que a idade média de óbito entre estes pacientes foi de 42 anos para homens e 48 anos para mulheres, o que representa uma diminuição de 25 a 30 anos se comparado com a expectativa de vida da população negra americana em geral. Dezoito por

cento das mortes ocorreram em pacientes cronicamente doentes, com falência de múltiplos órgãos clinicamente óbvia (falência renal, cardíaca ou seqüela crônica e debilitante decorrente de AVC). Vinte e dois por cento dos pacientes estáveis com anemia falciforme faleceram devido a infarto agudo, sendo que a maioria ocorreu por hemorragia aguda (PLATT et al., 1994). O AVC em pacientes com anemia falciforme são potencialmente graves, podendo ser fatais ou afetar severamente as funções motoras e cognitivas nos afetados por anemia falciforme, grupo no qual que a probabilidade de AVC é aumentada em 300 vezes na infância em relação à população geral (STEVEN et al., 2016).

Aproximadamente 17% das crianças HbSS podem apresentar alterações na imagem por ressonância magnética (RM) do cérebro, sugestivas de infarto ou isquemia na ausência de história clínica de infarto. Esses pacientes são classificados como portadores de infarto silencioso e demonstram um desempenho significativamente menor quando submetidos a testes neuro-psicológicos se comparados a pacientes HbSS sem anormalidades na RM (KINNEY et al., 1999; MILLER et al., 2001; PEGELOW et al., 2001; WANG et al., 1998). Também se observou que crianças com infartos silenciosos apresentaram uma maior incidência de um infarto com sintomatologia clínica e de um infarto silencioso novo ou mais extenso em relação à incidência de infarto em crianças da população geral com anemia falciforme (PEGELOW et al., 2002).

AVCs são causados pela oclusão de grandes vasos, hipoperfusão da zona de transição entre a artéria cerebral média, a carótida distal interna e artérias cerebrais posteriores ou anteriores proximais e por diminuição do fluxo vascular em pequenos vasos (MONTALEMBERT et al., 1993). A vasculopatia resulta em acentuada estenose de grandes vasos, com hiperplasia de musculatura vascular lisa e formação de trombo no lúmen sendo os mecanismos fisiopatológicos mais marcantes, em detrimento de inflamação e aterosclerose (STEVEN et al., 2016). Tal processo se dá lentamente, conforme a formação de extensa vasculatura colateral evidencia, particularmente artérias lenticulostriatas ao longo dos gânglios basais (CRAFT et al., 1993). A aparência angiográfica de estenose com extensa rede de vasos colaterais foi denominada *moyamoya* por médicos japoneses, em referência ao aspecto difuso de fumaça (Figura 5). O suprimento sanguíneo é tênue, mesmo com a formação de colaterais, uma vez que essas apresentam alterações como paredes adelgadas e formação de microaneurismas, as predispondo a hemorragia e trombose (DIGGS; BROOKOFF, 1993).

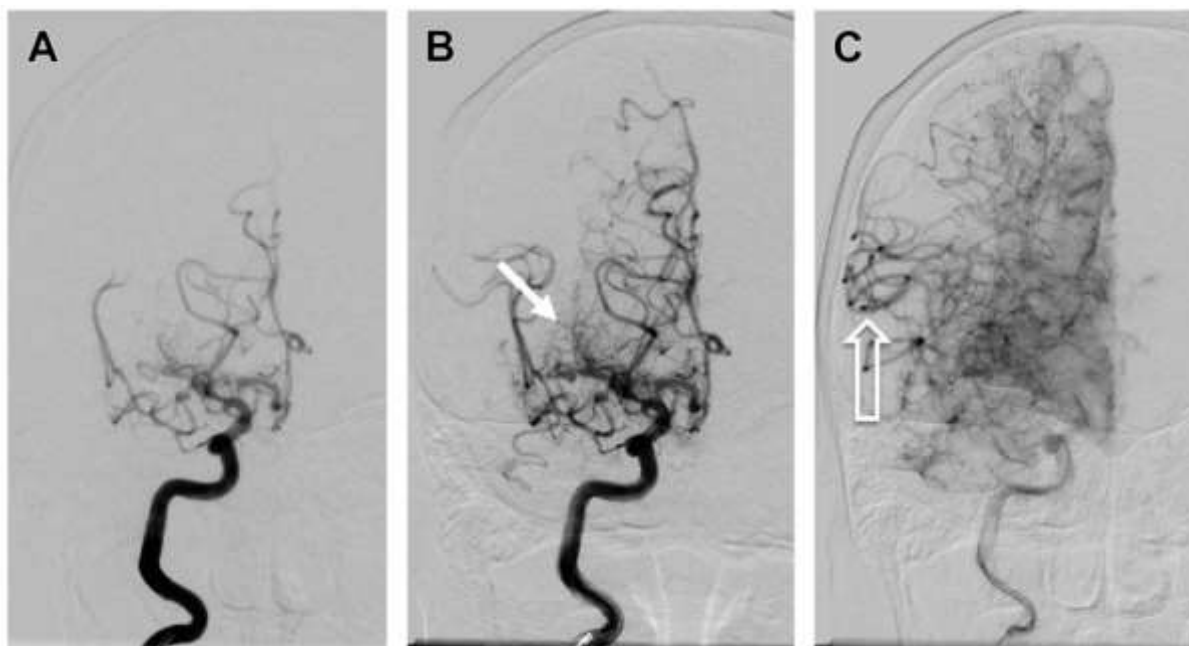


Figura 5. Imagens ântero-posteriores sequenciais de um angiograma de subtração digital após injeção na artéria carótida interna em paciente com estenose crônica de carótida interna. (A) Estreitamento grave da carótida supraclinóide com ausência do segmento M1 das artérias cerebrais médias. (B) Uma rede proeminente e definida de vasos colaterais na região da artéria cerebral média (seta) consistente com o fenômeno de *moyamoya*, juntamente com suprimento sanguíneo normal das artérias cerebrais anteriores. (C) Retardo de suprimento dos ramos distal artéria média (seta aberta) pela rede colateral e pial (STEVEN et al., 2016).

As taxas de aneurisma na anemia falciforme é aumentada (acima de 1% em crianças e de 10% em adultos), com alta prevalência de indivíduos com múltiplos episódios, provavelmente pela circulação aumentada (em resposta a anemia crônica e a hipoxemia) e dano crônico das camadas subjacentes da parede dos vasos. Ectasia arterial também já foi descrita, afetando comumente mas não exclusivamente a artéria basilar (DIGGS; BROOKOFF, 1993). Infartos isquêmicos se instalam pelos fenômenos oclusivos e fatores hemodinâmicos, incluindo redução da concentração arterial de oxigênio. Clinicamente, o AVC se manifesta como um déficit focal agudo com mais de 24 horas de duração (STEVEN et al., 2016). Os sintomas variam com o tamanho da lesão, mas podem incluir dificuldades motoras e sensoriais, perda de visão ou audição ou comprometimento da fala.

A conduta de prevenção é realizada pelo Doppler transcraniano, com uma velocidade média acima de 200 cm/s na carótida interna distal ou na artéria média proximal como sinal de alerta (ADAMS et al., 1992, 1990) e então instituição de terapia de transfusão crônica ou exsanguineotransfusão para evitar o aumento agudo de viscosidade, a contraindicação ou uso judicioso de corticosteroides e o controle da pressão arterial (STROUSE

et al., 2006). A conduta clínica no evento de um AVC isquêmico inclui o uso de transfusão e a indicação de hidroxiuréia (HU), geralmente em adultos, embora testes clínicos já tenham sido experimentados em pacientes pediátricos, com resultados positivos (WARE et al., 2004). Por fim, o transplante de medula óssea de irmão HLA-compatível com terapia mielosupressiva associada é um tratamento relativamente recente e eficaz, embora com riscos associados ainda elevados (WANG, 2007).

Os AVCs que se manifestam clinicamente afetam aproximadamente 6% a 17% dos pacientes com anemia falciforme (MONTALEMBERT et al., 1993; SEBASTIANI et al., 2005), sendo que a recorrência pode chegar a cerca de 80% nos 3 anos seguintes ao primeiro evento, se nenhuma terapia for instituída (TANG et al., 2001).

Vários fatores de risco para o AVC isquêmico em pacientes com anemia falciforme foram identificados. Dentre eles destacam-se a elevação da contagem de leucócitos e diminuição dos níveis de hemoglobina (BALKARAN et al., 1992; KINNEY et al., 1999), infecção pelo parvovírus humano B19 (WIERENGA et al., 2001), aumento da pressão sanguínea (PEGELow et al., 1997), tipo de haplótipo β^S e presença do gene alfa (OHENE-FREMPONG ET AL., 1998; SARNAIK & BALLAS, 2001) e aumento dos níveis de homocisteína (HOUSTON et al., 1997). O AVC hemorrágico, mais comum em pacientes adultos, está associado aos fatores de risco: ocorrência anterior de AVC isquêmico, aneurismas, baixos níveis de hemoglobina, alta contagem de células brancas, síndrome torácica aguda e hipertransfusão (OHENE-FREMPONG et al., 1998). De modo similar, infartos isquêmicos são mais comuns na infância, e os eventos da vida adulta geralmente representam a reincidência de episódios iniciados na infância; os escassos relatos de primeiro episódio de infarto isquêmico adulto apontam que esse pareça ter causas diversas daqueles ocorridos na vida infantil (CALVET et al., 2015; STROUSE et al., 2006).

Dados hematológicos em 17 de 310 crianças com anemia falciforme e que apresentaram infarto mostraram significativa elevação da contagem de leucócitos em relação aos indivíduos controle. O primeiro infarto coincidiu com uma queda aguda dos valores de hemoglobina em 5 pacientes (BALKARAN et al., 1992).

A infecção pelo parvovírus humano B19 causa a maioria das crises aplásticas clinicamente definidas na anemia falciforme. Com a terapia de transfusão, a evolução é geralmente benigna, contudo, as complicações cerebrovasculares em associação com crises

aplásticas induzidas por infecção por B19 foram descritas como 58% maiores do que o esperado no intervalo de 5 semanas após a infecção (WIERENGA et al., 2001).

A pressão sanguínea média para indivíduos com anemia falciforme é significativamente menor do que a encontrada na população geral. Valores que seriam considerados normais ou representariam uma discreta alteração em indivíduos normais devem ser vistos como perigosos na anemia falciforme, uma vez que existe uma associação entre o aumento da pressão sanguínea, infarto e mortalidade (PEGELow et al., 1997).

Os AVCs em crianças parecem ocorrer mais frequentemente no sexo feminino e em pacientes com um determinado haplótipo β^S . A deleção do gene alfa parece oferecer um efeito protetor contra esta complicação (OHENE-FREMPONG et al., 1998; SARNAIK & BALLAS, 2001). Recém-nascidos com 4 ou mais genes alfa e cujos halótipos β^S é Benin/CAR, atípico ou CAR/CAR parecem ter maior risco para AVC do que outros pacientes (SARNAIK & BALLAS, 2001). No estudo de KINNEY et al. (1999), observou-se que o haplótipo β^S Senegal, assim como níveis baixos de hemoglobina e aumento do número de leucócitos estavam associados à presença de infartos silenciosos.

A heterogeneidade fenotípica observada em pacientes afetados de anemia falciforme tem sido atribuída a genes modificadores, sendo que genes candidatos à suscetibilidade ao AVC incluem aqueles implicados em adesão de células endoteliais e inflamação (SWITZER et al., 2006; WANG et al., 2007). Uma das estratégias mais recentes para se compreender a interação entre genes, ambiente e fenótipo foi o uso de rede bayesiana, modelo que analisa a associação e interação simultâneas entre vários genes e sua relação com fatores clínicos e fisiológicos. Cento e oito “Single Nucleotide Polymorphisms” ou SNPs em 39 genes candidatos foram avaliados em 1398 pacientes. Os autores observaram que 31 SNPs em 12 genes estão associados aos níveis de hemoglobina fetal modulando o risco de infarto, dentre eles 3 genes na via do “Transforming Growth Factor-beta” (TGF-beta) e da “Selectin P” (SELP), os quais estão associados ao infarto na população geral. Os resultados foram validados em uma população diferente, com uma precisão de 98,2% (SEBASTIANI et al., 2005).

Em geral, as medidas terapêuticas aplicadas a casos de infarto na anemia falciforme não podem ser implementadas antes da ocorrência do primeiro infarto, sendo essa terapia direcionada à prevenção de infartos recorrentes por um programa de transfusão em longo prazo (BALKARAN et al., 1992). Alternativamente, o exame com Doppler ultra-sonográfico transcraniano, o qual seria capaz de detectar precocemente as lesões, identificando crianças

potencialmente suscetíveis ao AVC poderia ser utilizado, enfrentando, o alto custo de se disponibilizar o procedimento a todos os pacientes afetados (KINNEY et al., 1999; PEGELOW et al., 2001). No caso da presença de lesões, o programa de transfusão sanguínea por, no mínimo, cinco anos deve ser iniciado com o intuito de se manter o nível de HbS abaixo de 30% (MONTALEMBERT et al., 1993). A terapia de transfusão em longo prazo, contudo, está associada à toxicidade, esforço e custo substanciais, o que justifica sua utilização somente para pacientes considerados de alto risco (MILLER et al., 2001).

A identificação precoce de pacientes com anemia falciforme suscetíveis a AVCs poderia diminuir os riscos, possivelmente prevenir a recorrência de infartos e potencialmente reduzir sua incidência. Portanto, estudos que visem identificar novos grupos de risco para o desenvolvimento de AVC em pacientes com anemia falciforme seriam fundamentais para otimizar o controle clínico desta enfermidade, sendo uma dessas vias a abordagem molecular deste grupo de pacientes. Além disso, a abordagem genética pode conduzir ao esclarecimento das vias envolvidas no acometimento do SNC.

Ferramentas genéticas para a identificação de genes moduladores

A evolução dos marcadores moleculares, utilizados como ferramentas para a identificação de genes ocorre desde 1960 com a identificação de proteínas polimórficas, passando pela descoberta dos “Restriction Fragment Length Polymorphisms” (RFLPs), minissatélites, microssatélites e culminando com o Projeto HapMap (HAPMAP INTERNATIONAL CONSORTIUM, 2003), responsável pela genotipagem e mapeamento de cerca de um milhão e, posteriormente, de 4 milhões do total de 12 milhões de SNPs depositados em bancos de dados públicos. Hoje, os SNPs são a principal escolha para estudos de associação com o objetivo de se identificar genes de maior efeito ou modificadores em doenças complexas. Por meio destes marcadores, genes relacionados à degeneração macular relacionada à idade, diabetes, obesidade, doenças cardiovasculares e câncer têm sido identificados (BECKMANN et al., 2007).

Durante a última década vários grupos se dedicaram a confirmar o que já era postulado desde a década de 80: que o genoma humano também contém uma outra fonte abundante de polimorfismos, envolvendo deleções, inserções, duplicações e rearranjos complexos de regiões genômicas que variam de 1 kb a vários megabases de DNA em extensão, podendo compreender vários genes (IAFRATE et al., 2004; SEBAT et al., 2004; KOMURA et

al., 2006; REDON et al., 2006; EICHLER et al., 2007; WONG et al., 2007). Esse tipo de variação estrutural submicroscópica foi denominada “Copy Number Variation” ou CNV – variação no número de cópias. As estimativas mais conservadoras provenientes de estudos de centenas de indivíduos apontam que, no mínimo, 10% do genoma está sujeito à ocorrência de CNV (FEUK et al., 2006; REDON et al., 2006; WONG et al., 2007), as quais podem contribuir para a diversidade fenotípica em humanos assim como ajudar na elucidação dos mecanismos que originam várias doenças (WONG et al., 2007).

As variações no número de cópias foram documentadas, a princípio, em doenças herdadas, como a neuropatia do tipo 1 Charcot-Marie-Tooth (LUPSKI et al., 1991). Entretanto, hoje é sabido que as CNVs compreendem aproximadamente 12% do genoma humano, alterando potencialmente a dosagem gênica, desorganizando genes ou comprometendo a regulação de sua expressão, mesmo a longas distâncias. Logo, um número considerável de doenças com padrão mendeliano de herança pode ser afetado por CNVs, mas estas podem estar relacionadas também à etiologia de doenças complexas como resultado da variabilidade na dosagem gênica que pode afetar genes cuja função não possa ser restaurada (KNIGHT, 2005; BECKMANN et al., 2007; ESTIVILL & ARMENGOL, 2007; MCCARTHY et al., 2008). Além disso, deleções genômicas em indivíduos aparentemente saudáveis podem não causar diretamente uma doença monogênica, mas, na presença de fatores adicionais, tanto genéticos como ambientais, podem contribuir para o desenvolvimento de doenças poligênicas complexas de início tardio. As CNVs podem ter um papel na origem de doenças complexas pelo simples fato de que certas regiões que compreendem estas variações contêm vários genes (POLLEX & HEGELE, 2007).

Diversas características complexas têm sido associadas a CNVs, como por exemplo a suscetibilidade ao HIV-1, lúpus, doença de Chron, componentes de doenças cardiovasculares (BECKMANN et al., 2007; POLLEX & HEGELE, 2007). CNVs raras também parecem estar envolvidas na suscetibilidade a outras doenças comuns, como pancreatite crônica, autismo, doenças de Alzheimer e Parkinson, aumentando a chance de identificação das bases moleculares das formas hereditárias monogênicas destas doenças. A dispersão cromossômica das CNVs pode ter impacto nas variações individuais de resposta a drogas, assim como na suscetibilidade a infecções e câncer, diretamente ou modulando a penetrância ou variabilidade na expressão da característica em questão (BECKMANN et al., 2007).

Quanto à geração das CNVs, observam-se dois mecanismos prevalentes: recombinação homóloga (RH) e não-homóloga. Recombinação homóloga exige ampla identidade de sequência (até 300pb em humanos) (LISKAY; LETSOU; STACHELEK, 1987; REITER et al., 1998) e a enzima Rad51 em eucariotos para realizar a invasão da fita molde para o reparo (HASTINGS et al., 2009). Em contraste, a recombinação não-homóloga exhibe mecanismos independentes ou pouco dependentes de homologia. A RH é a base para diferentes mecanismos eficientes de reparo, mas pode ocasionar rearranjo cromossômico quando a sequência homóloga não está na mesma posição para as duas cromátides irmãs ou cromossomos homólogos, como é o caso de recombinação homóloga não-alélica (STANKIEWICZ; LUPSKI, 2002), que em geral ocorre pela presença frequente, ao longo do genoma, de sequências de repetição em baixo número de cópias (STANKIEWICZ; LUPSKI, 2006).

O mecanismo de RH mais estudado é a recombinação induzida por quebra de dupla fita que, por sua vez, compreende dois modelos: a junção dupla de Holliday (que pode levar à conversão gênica e ao *crossing over* dependendo da orientação em que as junções são resolvidas) e o anelamento dependente de síntese, que evita *crossing over* e perda de heteroziguidade, porém pode gerar rearranjos quando o DNA molde apresenta sequências repetidas (LIU et al., 2012); enquanto que no cenário da junção dupla de Holliday, variantes em número de cópias geralmente surgem em caso de *crossing over* desigual (HASTINGS et al., 2009). Este tipo de recombinação, dita homóloga não-alélica, é vista também quando, por meio de replicação induzida por quebra, região homóloga ectópica é utilizada como molde ou ainda, pela falta de anelamento à sequência homóloga em outra cromátide ou cromossomo, quando a religação das pontas seccionadas se dá a partir de regiões complementares de uma mesma cadeia, que se anelam com resultante perda do fragmento intermediário entre as duas sequências (LIU et al., 2012).

A recombinação não-homóloga (RNH), por seu turno, não utiliza homologia de sequência como meio de religar moléculas de DNA e se subdivide em replicativa e não-replicativa. A RNH não-replicativa é ainda estratificada em junção de extremidades (não-homóloga ou mediada por micro-homologia), levando a rearranjo quando ocorre quebra de fita dupla (causado por endonuclease, agentes exógenos ou ruptura da forquilha de replicação); e o modelo de quebra-fusão-conexão, que propõe um ciclo iniciado pela perda de telômero por cromátides irmãs, as quais se fundem, podendo sofrer ruptura na anáfase e gerar novamente extremidades com telômeros ausentes, repetindo o ciclo (MCCLINTOCK, 1951). O ciclo seria

responsável pela sequências duplicadas e invertidas observadas em células cancerosas, por exemplo (TANAKA et al., 2006).

A RNH replicativa, por fim, apresenta hoje três meios propostos de geração de alelos variantes em número de cópias. Há a derrapagem na replicação, na qual uma pequena extensão de sequência encontra identidade na parte do DNA em fita simples durante a replicação (levando à deleção ou duplicação da região entre as sequências homólogas) (ALBERTINI et al., 1982; FARABAUGH et al., 1978). De modo similar, fazendo também uso da forquilha de replicação, há o mecanismo denominado *FoSTeS* (interrupção da forquilha e troca de moldes) (LEE; CARVALHO; LUPSKI, 2007), em que a extremidade 3' de uma forquilha pode trocar de molde com um DNA simples fita de uma forquilha vizinha (SLACK et al., 2006). Alguns autores propuseram ainda que a replicação induzida por quebra possa ser mediado por micro-homologia (BAUTERS et al., 2008; PAYEN et al., 2008; VISSERS et al., 2007), enquadrando-a na categoria RNH replicativa e levando a criação de CNV de modo similar ao descrito acima para RH.

As CNVs raras e *de novo* têm papéis bem estabelecidos em doenças complexas (AUTON et al., 2015; STANKIEWICZ; LUPSKI, 2006, 2010), embora tipicamente com penetrância parcial e alta variabilidade em sua manifestação, como por exemplo as duplicações em 7q11.23 que conduzem ao autismo ou esquizofrenia (MALHOTRA et al., 2011; PINTO et al., 2010; SANDERS et al., 2011). A maior parte das CNVs em dado indivíduo, no entanto, é fruto de polimorfismo frequente, herdado e de surgimento remoto (MCCARROLL et al., 2008), em sua maioria bialélica e benigna, com algumas dessas variantes exercendo modesto efeito em fenótipos (CRADDOCK et al., 2010). Porém outra variedade de CNVs, menos explorada em estudos genéticos, são os multialélicos e complexos, provável resultado de múltiplos eventos de rearranjo durante a evolução humana. O projeto 1000 Genomes catalogou 1.356 desse tipo de CNV, dentre os 8.659 identificados em sua fase 1 (1000 GENOMES PROJECT CONSORTIUM et al., 2010). Dentre as 1.356, 121 CNVs mostram quatro ou mais alelos segregando, enquanto 45 tiveram cinco ou mais alelos identificados (CAMPBELL et al., 2011; CRADDOCK et al., 2010; HANDSAKER et al., 2015). Embora CNVs multialélicas sejam minoria dentre as variantes estruturais, elas são responsáveis por 88% da variação em dosagem gênica em humanos (HANDSAKER et al., 2015).

Um dos exemplos mais característicos de CNVs multialélicas compreende o gene *CCL3L1*, que codifica um ligante para o correceptor para o vírus HIV e encontra-se entre 0 e

14 cópias (GONZALEZ et al., 2005). Tal CNV e a seu efeito na susceptibilidade de infecção pelo vírus foram replicados em diferentes estudos, embora outros não tenham logrado replicação (AKLILLU et al., 2013; HUIK et al., 2010; LARSEN et al., 2012; LEE et al., 2010; RATHORE et al., 2009; SHOSTAKOVICH-KORETSKAYA et al., 2009).

Um estudo foi realizado com o objetivo de determinar a contribuição global das CNVs para os fenótipos complexos, por meio de análises de associação dos níveis de expressão de 14.925 transcritos com SNPs e CNVs em indivíduos que fazem parte do projeto HapMap. Como resultado, os SNPs e CNVs capturaram 83,6% e 17,7% da variação genética total detectada, respectivamente, com muito pouca sobreposição de sinais provenientes das duas variações (1,3%) (STRANGER et al., 2007). Esse estudo mais uma vez corrobora com a hipótese de que o estudo do genoma por meio de ambas as variações pode ser mais eficiente na elucidação de causas de fenótipos complexos e doenças em humanos. Os padrões de SNPs e CNVs, simultaneamente, aliados à dosagem das CNVs e fatores ambientais seriam combinados para produzir o fenótipo da doença ou outros caracteres fenotípicos (KOMURA et al., 2006; REDON et al., 2006; BECKMANN et al., 2007; ESTIVILL & ARMENGOL, 2007; HAMET & SEDA, 2007).

De fato, há diversas variações estruturais implicadas em doenças humanas (WEISCHENFELDT et al., 2013), como é o caso de doenças de herança mendelianas ou de alta penetrância geradas por polimorfismo de cópias recorrente, incluindo as síndromes de Smith–Magenis, Williams–Beuren e Potocki–Lupski (LUPSKI, 1998), resultantes da alteração de dosagem gênica, embora também tenha-se observado ainda efeito posicional nessas mesmas síndromes (MERLA et al., 2006; RICARD et al., 2010). Também derivadas de variantes recorrentes encontram-se as síndromes Digere e Velocardiofacial, envolvendo CNVs na região 22q11 (MAYNARD et al., 2002). Outros exemplos de doenças causadas por alteração na dosagem gênica são a leucodistrofia hipomielinizante ligada ao X (doença de Pelizaeus–Merzbacher), ligada a deleções, duplicações e triplicações no gene *PLP1* e Charcot–Marie–Tooth, em caso de duplicações em *PMP22*, ou neuropatia hereditária com suscetibilidade a paralisia por pressão, envolvendo deleção de *PMP22* (KELLER; CHANCE, 1999).

Além de alteração de dosagem, CNVs podem atuar revelando alelos recessivos ou alvos de *imprinting* (quando envolvem deleção), causando disfunção, como é o caso observado nas síndromes Angelman e Prader–Willi (WEISCHENFELDT et al., 2013) ou alterando a função de genes envolvidos em epistasia de modo a afetar elementos distantes e por fim resultar

em variação fenotípica, conforme observado na síndrome de Bardet–Biedl (BADANO et al., 2006) e proposto para doença de Crohn (ZUK et al., 2012). Variações cromossômicas estruturais podem ainda afetar elementos *cis*-regulatórios, os quais podem encontrar-se significativamente distantes (>100 kb) dos promotores que regulam e operam em determinado tecido e/ou estágio do desenvolvimento (WEISCHENFELDT et al., 2013). A β -talassemia causada por deleção de elemento regulatório *upstream* (LCR) (KIOUSSIS et al., 1983), holoprosencefalia gerada por translocações que separam o gene *sonic hedgehog* (*SHH*) de elementos regulatórios de sua expressão no prosencéfalo em desenvolvimento (BELLONI et al., 1996) são exemplos prototípicos, como de modo similar ocorre da síndrome de Pierre Robin, a qual decorre de microdeleções a 3 Mb de distância de *SOX9*, provavelmente afetando *enhancers*.

A Figura 6 ilustra os meios pelos quais as CNVs podem levar à alteração de função, pelo produto e/ou regulação de um ou mais genes.

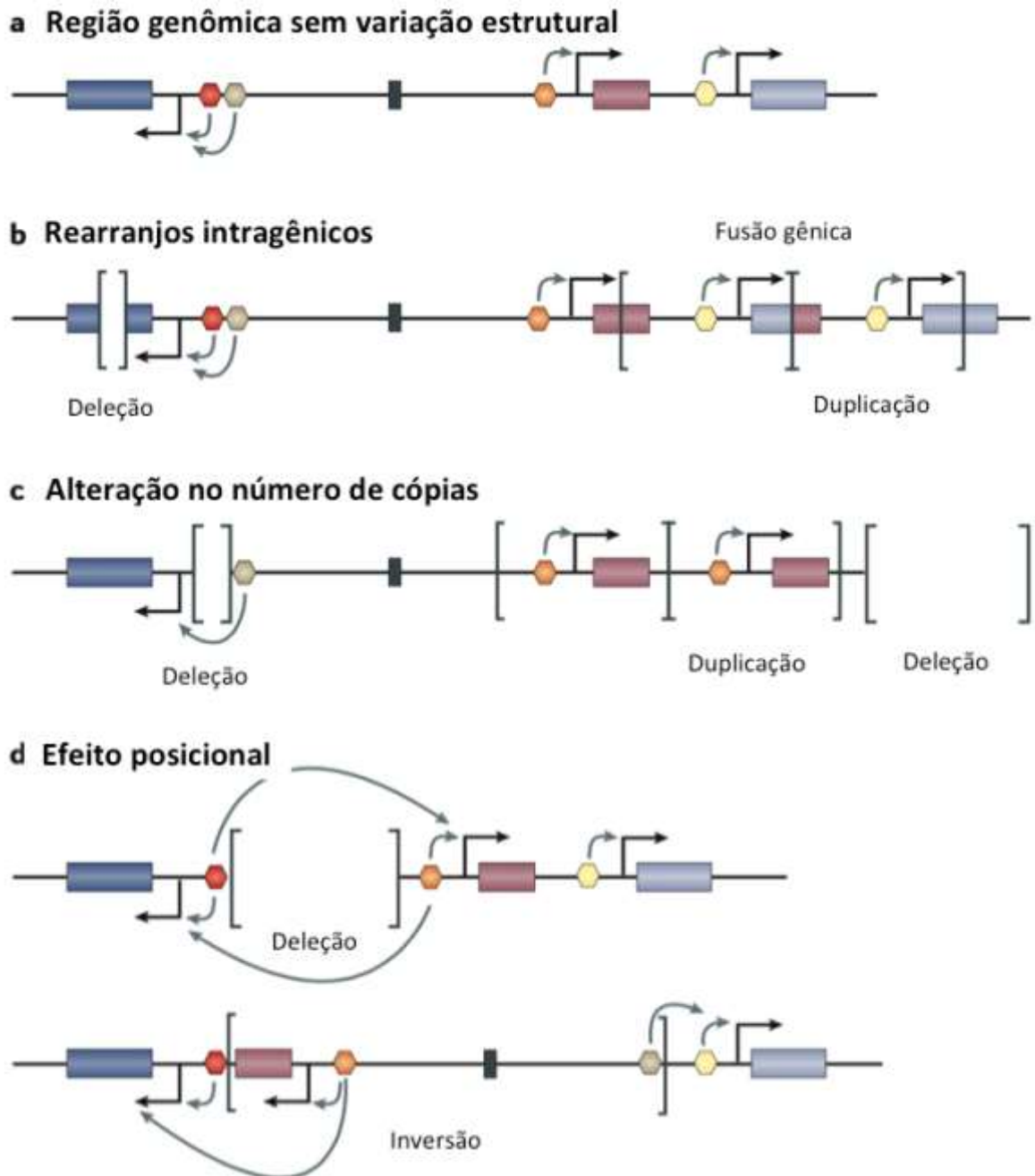


Figura 6. Representação de exemplos de impacto funcional gerado pela ocorrência de CNVs. Os genes (caixas) são regulados por *inputs* coletivos de elemento reguladores, incluindo *enhancers* específicos de tecidos (hexágonos, com diferentes cores indicando a especificidade do tecido e setas apontando para o gene alvo) e insuladores (retângulos pretos), que bloqueiam a atividade de elementos reguladores (A). (B-d) As variantes estruturais (mostradas entre colchetes) podem ter consequências fenotípicas, alterando regiões codificantes, como quando remove-se parte de uma região de codificação ou funde-se regiões codificadores após a duplicação, resultando em transcritos aberrantes (b). Alternativamente, deleções ou duplicações podem levar a alteração da dose de elemento regulatório (c), alterando a o sinal regulador (esquerda) ou número de cópias de determinado genes (direita). As variantes estruturais também podem afetar a expressão de genes em outros *loci*, isto é, causar efeito de posição (d), resultando assim em ganho ou perda de sinal regulador. Um evento sobreposto ao gene pode ocasionar a ruptura de sua porção codificante e/ou regulatória, causar uma fusão gênica com ou sem novo produto expresso, alterar uma

região regulatória em *cis*, ou levar à expressão do alelo recessivo ou sob *imprinting*. (Adaptado de WEISCHENFELDT et al., 2013).

Com a experiência proveniente do conhecimento das CNVs tem ficado claro que, apesar dos SNPs serem considerados os marcadores de maior poder no contexto dos estudos de associação, estas complementarão as análises de SNPs. Como os SNPs, as pequenas CNVs afetarão somente genes individuais e logo contribuirão, juntamente com os SNPs, para doenças monogênicas. Entretanto, ao contrário dos SNPs, as CNVs podem afetar dois ou mais genes e contribuir para doenças síndrômicas ou complexas causadas por defeitos em múltiplos genes. Finalmente, algumas CNVs envolvem regiões pobres em genes e podem ser tanto neutras funcionalmente como ter um impacto na suscetibilidade à doença por meio de seus efeitos em domínios não transcritos e que regulam a expressão gênica à distância (POLLEX & HEGELE, 2007). Apesar do projeto HapMap ter fornecido uma ferramenta de absoluta importância para os estudos de associação, está claro que o conjunto de marcadores analisados não cobre a variabilidade de todo o genoma. Apesar do grande número de SNPs, algumas regiões não são representadas, provavelmente devido à falta de informação sobre a sequência e, em grande parte, devido à presença de CNVs e duplicações de segmentos (ESTIVILL & ARMENGOL, 2007).

Desta forma, propusemos a avaliação de CNVs por meio de lâminas de alta densidade, provenientes da empresa Affymetrix (Affymetrix Genome-Wide Human SNP Array 6.0), com o objetivo de se identificar regiões do genoma, potencialmente envolvidas com o risco aumentado de AVC em pacientes com anemia falciforme. Estas lâminas contêm mais de 906.600 sondas para a detecção de SNPs e 946.000 de CNVs, separados por uma distância média de 700 bases ao longo do genoma, constituindo uma ferramenta robusta para o presente estudo.

Parâmetros populacionais

A identificação de ancestralidade fornece informações de grande valia sobre o entendimento da história genética dos indivíduos, sendo de ampla importância para mapear eventos de miscigenação e introgressão, e identificar alvos recentes de seleção (GOMPERT et al., 2012; JIN et al., 2012; TANG et al., 2007). Há crescente entendimento a respeito da diversidade em componentes ancestrais em populações como os afro-americanos e latinos, com

proporções de ancestralidade variáveis ao longo do genoma (BRYC et al., 2010a; TANG et al., 2007).

Compreender a ancestralidade em nível cromossômico em indivíduos miscigenados é importante para explorar os eventos de miscigenação, incluindo o tempo do evento, em gerações (POOL; NIELSEN, 2009; PUGACH et al., 2011; STAM, 1980), o crescimento populacional a partir de dado evento de migração (GOMPERT et al. 2012; JIN et al. 2011; TANG et al. 2007), melhorar a imputação de genótipos por meio de painéis de referência baseados em ancestralidade local, além de permitir o mapeamento de miscigenação em escala fina, conjuntamente com o mapeamento de desequilíbrio de ligação (DL) (PASANIUC et al., 2011). Tais aprimoramentos permitem maior precisão na identificação de variantes genéticas associadas a caracteres fenotípicos.

Até recentemente, vigorou quase incontestemente a hipótese de que doenças comuns são causadas por variantes comuns (*common disease/common variant*, CDCV) (REICH; LANDER, 2001). O modelo inicialmente propõe que diferentes combinações de alelos comuns determinam o risco para uma condição clínica não-rara. A hipótese CDCV foi de grande influência para a criação do banco de dados HapMap (International Haplotype Mapping, INTERNATIONAL HAPMAP CONSORTIUM, 2003) e a proliferação de estudos de associação de amplitude genômica (genome-wide association studies, GWAS) (LUPSKI et al., 2011), iniciativas dedicadas ao estudo do espectro de variação genética considerada comum (convencionalmente, de frequência alélica igual ou superior a 5%).

Manolio et al. (2009) defendem que explicar a variação genética em si traz conhecimento a respeito de potencial efeito funcional. Avaliar a variabilidade, no entanto, não se traduz de forma direta em entendimento de impacto (LUPSKI et al., 2011). Genes identificados por meio de seu efeito aditivo trouxeram esclarecimentos a respeito de vias importantes nos caracteres considerados, mas apesar do volumoso conteúdo de dados que os GWAS geraram, relativamente pouca proporção da variabilidade fenotípica foi satisfatoriamente explicada e, de modo geral, os dados não são consistentes, ao menos, com uma versão meramente aditiva da hipótese CDCV (LUPSKI et al., 2011). Alelos que levam ao risco a doenças estão sobre-representados na faixa de baixas frequências (MARTH et al., 2011), na qual observam-se alelos de surgimento mais recente. Tal fenômeno se dá pois a seleção natural purificadora reduz a frequência de variantes altamente deletérias, de modo a mantê-las a baixas frequências alélicas.

Os 3 Gb do genoma humano haplóide torna a tarefa de sequenciamento completo custosa e de difícil abordagem estatística. Além disso, os testes de associação contemporâneos são derivados dos testes empregados previamente à era do sequenciamento paralelo em massa e baseados na seleção de marcadores que acessem dada porção da variabilidade entre indivíduos. Portanto, é importante o trabalho de seleção de marcadores (dado a dificuldade de sequenciar e avaliar genomas completos). O uso de marcadores é subjacente à ideia de que estes estão ligados ao *locus* mecanicamente implicado no fenótipo (BALDING, 2006). Em um estudo de associação típico a variante causal provavelmente não será genotipada, possivelmente porque não é uma variante de mesma classe que os marcadores, como por exemplo no caso do ensaio utilizado genotipar SNPs e a variante causal ser uma inserção, deleção, inversão ou variante no número de cópias. Além disso, como discutido acima, uma variante rara pode não estar contida em uma técnica de genotipagem baseada em hibridização, dado que os marcadores são previamente desenhados e contidos em sondas, projetadas na maioria das vezes a partir da anotação do projeto HapMap e selecionados com frequência acima de 5%, para permitir tratabilidade estatística. Alelos não descritos, geralmente raros e de maior efeito, tendem a ser negligenciados por essas abordagens.

De maneira geral, estudos de associação são adequados apenas para as variantes comuns. Argumenta-se, a favor da hipótese CDCV, que os tamanhos efetivos nas populações humanas são baixos, que a pressão seletiva sobre condições clínicas de início tardio é reduzida e, ainda, que variantes ligadas a predisposição a doenças podiam ter papel adaptativo no passado (como é o caso das vias do metabolismo de lipídios, positivamente selecionadas para favorecer estoque e hoje ligadas à obesidade) (BALDING, 2006). Embora importantes variantes tenham sido associadas a doenças complexas, há ainda pouca clareza sobre em qual proporção as variantes comuns contribuem para a herdabilidade das doenças humanas não-raras (MANOLIO et al., 2009).

A seleção de marcadores visando o estudo genético leva em conta o potencial informativo deste, como o haplótipo em que está inserido, que por sua vez, está intimamente ligado à estrutura populacional. Estudos de associação do tipo caso-controle comparam indivíduos ditos não-aparentados, quando a rigor o parentesco está presente, mas é presumido distante. Indivíduos com uma mesma variante causal partilham um ancestral comum relativamente próximo, e o número de gerações, isso é, de recombinações, determina a proporção de marcadores adjacentes à variante que será comum a esses indivíduos. Espera-se que portadores da variante causal estejam mais representados pelo grupo caso de um estudo de

associação, e que os marcadores fortemente ligados a essa variante encontrem-se também sobre-representados em relação ao grupo controle (NUSSBAUM; MCINNES; WILLARD, 2008).

É importante destacar que há uma fonte sistemática de variação alélica entre casos e controles e de base populacional, denominada estratificação ou estrutura genética. A divergência nas frequências alélicas de um dado marcador ou conjunto de marcadores em casos e controles pode ser atribuída a diferenciação populacional entre os grupos e não necessariamente ao fenótipo que conduziu o agrupamento. A estratificação é um importante fator para se levar em conta em um estudo de associação.

Dois métodos de correção de estratificação prevaleceram por muito tempo: o controle genômico e a associação estrutural (DEVLIN; BACANU; ROEDER, 2004; DEVLIN; ROEDER, 1999; PRITCHARD et al., 2000; REICH; GOLDSTEIN, 2001; SATTEN; FLANDERS; YANG, 2001; SETAKIS; STIRNADEL; BALDING, 2006). O controle genômico corrige a estratificação ajustando a estatística de associação para cada marcador por um mesmo fator de inflação geral, portanto não abordando de modo específico regiões com diferenciação mais marcada entre populações, enquanto a associação estrutural (como a utilizada pelo programa STRUCTURE; PRITCHARD; STEPHENS; DONNELLY, 2000) atribui amostras a agrupamentos subpopulacionais e então investiga sinais de associação para cada grupo. Se, de modo mais realista, for permitido o fracionamento de indivíduo a mais de uma subpopulação, então este recurso torna-se computacionalmente proibitivo para estudos com alta densidade genômica de marcadores (LANDER; SCHORK, 1994; SETAKIS; STIRNADEL; BALDING, 2006).

Criada pelo proeminente matemático Karl Pearson (1901), a análise de componentes principais (PCA) tornou-se prevalente como técnica padrão de avaliação da distância genética entre indivíduos. Tal técnica aplica transformação linear ortogonal da matriz de correlação ou de covariância de uma série de dados normalizados (genótipos, por exemplo), de modo a conservar o máximo de variância em componentes principais (PCs) independentes (ortogonais) entre si. Tais componentes são ordenados decrescentemente por sua pontuação de fatores (*scores*), resultantes do produto entre cada variável pontual e seu carregamento (*loading*), um vetor unitário representativo da máxima variância daquela variável (JACKSON, 1991).

Price et al. (2006) propuseram uma metodologia, implementada no *software* EIGENSTRAT, que consiste na geração de eixos de variância genética (componentes principais), sobre os quais se decompõem, por meio de regressão logística, fenótipos e genótipos de modo a atribuir a proporção correspondente a ancestralidade nessas duas categorias. Por fim, a estatística de associação é computada usando os valores de genótipos e fenótipos ajustados para ancestralidade, equivalendo a um teste de casos e controles pareados quanto a seus componentes ancestrais.

A plotagem de indivíduos no plano cartesiano, com os PCs como eixos, provê uma maneira intuitiva de interpretar a estrutura populacional. Agrupamentos de indivíduos podem ser compreendidos como populações discretas, enquanto indivíduos miscigenados se estendem ao longo das populações ancestrais (REICH et al., 2009). É corrente a aplicação de ambos os métodos, PCA (como EIGENSTRAT e PCAdmix; BRISBIN et al., 2012) e baseados em modelos (STRUCTURE e congêneres) em um mesmo conjunto de dados para interpretar parâmetros populacionais.

A estrutura populacional tem duplo impacto em estudos genéticos: no desenho de modo a abranger marcadores informativos e na possível correção de sinais de significância induzidos pela estratificação amostral e não pelo fenótipo de interesse.

População brasileira

A população brasileira é a quinta maior população do globo, com mais de 200 milhões de habitantes (IBGE 2015). Tal população destaca-se ainda pelo alto grau de miscigenação, dada a mistura das populações nativas, e das ondas de imigração europeia e africana. Estima-se que 2,5 milhões de autóctones de diversas etnias viviam no que é hoje o território brasileiro, e entre 1500 e 1808 aproximadamente 500 mil portugueses, na maioria homens, chegaram ao país (FREIRE-MAIA; SALZANO, 1967). Entre 1550 e 1870, o Brasil ainda recebeu, estima-se, quatro milhões de africanos de diferentes partes do continente, sendo o maior destino do tráfico humano no período da escravidão (KEHDY et al., 2015). Tal histórico demográfico torna a população brasileira de especial interesse no entendimento do papel da migração e miscigenação para a constituição do genoma humano moderno e suas aplicações historiográficas, populacionais e clínicas.

Os primeiros estudos a abordar diretamente a genética populacional brasileira datam da década de 1990 e dedicaram-se a avaliar os cromossomos de herança uniparental (BATISTA DOS SANTOS et al., 1999; BORTOLINI et al., 1997; SANTOS et al., 1996; WARD et al., 1996) em populações indígenas e amazônicas. Os haplótipos de DNA mitocondrial (mtDNA) revelaram alta contribuição relativa de haplótipos ameríndios e africanos na região norte do país.

Em 2000 o primeiro estudo avaliando a população de modo mais amplo foi publicado, no qual se avaliou 247 indivíduos das regiões Norte, Nordeste, Sudeste e Sul quanto aos haplótipos de mtDNA por meio da técnica de restrição enzimática (RFLP). Demonstrou-se que indivíduos autodeclarados brancos apresentam 28% de haplogrupos africanos e 33% foram identificados como portadores de haplogrupos ameríndios (ALVES-SILVA et al., 2000). Uma subamostra dos indivíduos desse estudo anterior, portanto também autodeclarados brancos, foi avaliada quanto à patrilinearidade (CARVALHO-SILVA et al., 2001). Em contraste com o estudo anteriormente publicado, há baixa contribuição de cromossomos Y de origem subsaariana (haplogrupo E3a*, frequência: 2%) ou ameríndia (haplogrupo Q3*, não encontrado), enquanto entre africanos e indígenas as frequências de E3a* e Q3* são da ordem de 53% e 78%, respectivamente (BORTOLINI et al. 2002), com grande representatividade de haplogrupos no cromossomo Y contendo variantes tipicamente europeias.

Os dados moleculares demonstram um padrão de reprodução assimétrico, com os brasileiros brancos exibindo patrinhagens europeias, enquanto a maioria de suas matrinhagens é ameríndia ou africana. Tais achados estão em consonância com o histórico de povoamento do Brasil. Na época da colônia, a coroa portuguesa tolerava relacionamentos entre portugueses e mulheres indígenas, passando a estimular ativamente casamentos desse tipo por meio de um Alvará de Lei promulgado em 4 de abril de 1755 por Marquês do Pombal (PENA; BORTOLINI, 2004), com vistas a garantir a ocupação territorial. Há ainda consistente eixo de variação geográfica, com frequências ao longo das regiões significativamente variáveis: a maioria das linhagens mitocondriais no Norte é de origem ameríndia (54%), enquanto a ancestralidade africana é predominante no Nordeste (44%) e a europeia no Sul (66%), em relação a marcadores no cromossomo Y.

Utilizando um painel de dez polimorfismos autossômicos de ampla diferença em frequências alélicas (> 45%) entre africanos e europeus (PARRA et al, 1998, PENA et al., 2004) verificou-se a ancestralidade nuclear em 173 indivíduos no Vale do Jequitinhonha, região

nordeste de Minas Gerais (30 indivíduos classificados como pretos, 29 brancos e 114 pardos), além de duzentos indivíduos autotclassificados como brancos supracitados (CARVALHO-SILVA et al., 2001). Averiguou-se que a ancestralidade africana foi mais prevalente nas regiões nordeste e sudeste, de modo condizente com os dados de mtDNA. As regiões sul e norte, por sua vez, mostram valores menores de ancestralidade africana e diferiram significativamente entre si nas proporções de origem ameríndia.

Os autores concluíram, baseados nos marcadores autossômicos, que na amostra brasileira analisada, o alto índice de miscigenação faz com que características físicas, em especial pigmentação, sejam marcadores pouco confiáveis na determinação da origem geográfica dos ancestrais de um indivíduo.

Em 2007 foi realizado o estudo de 120 indivíduos negros da cidade de São Paulo quanto aos haplogrupos no mtDNA e cromossomo Y (GONÇALVES et al., 2007), concluindo que 48% dos haplogrupos do cromossomos Y e 85% do mtDNA eram característicos da África Sub-Saariana, reforçando o padrão de reprodução direcional.

Em 2011, Pena e cols. empregaram 40 variantes *indel* autossômicas como marcadores informativos de ancestralidade em 934 indivíduos das regiões Norte, Nordeste, Sudeste e Sul (PENA et al., 2011). Tal investigação encontrou predominância de ancestralidade europeia em todas as regiões amostradas (69.7%, 60.6%, 73.7% e 77.7%, respectivamente), apesar do censo demográfico sugerir maior variabilidade interregional, amparado na auto-declaração de cor de pele. Tal constatação de que a auto-declaração no país não se correlaciona com a ancestralidade genética dos indivíduos foi ainda explorada e reforçada pelo estudo de um outro grupo, abrangendo 5.851 indivíduos do Sul, Sudeste e Nordeste do país (LIMA-COSTA et al., 2015). Em publicação do ano de 2012, 138 indivíduos amostrados na cidade de São Paulo foram genotipados por técnica de microarranjos de alta densidade de marcadores (GIOLO et al., 2012), observando-se grande miscigenação e, interessantemente, que painéis que retêm a variabilidade encontrada nos cinco primeiros PCs da análise de componentes principais (de um total de 365.116 marcadores) variam em tamanho entre 250 e 750 marcadores. Tal achado é consistente com recente publicação do consórcio EpiGen, na qual 6.487 indivíduos (Rio Grande do Sul: 3.736, Minas Gerais: 1.442 e Bahia: 1.309) genotipados em 370.539 *loci* para a construção de um painel mínimo. Foram selecionados 192 marcadores com alta correlação com o total quanto à informação de ancestralidade (91%, 92%, e 74% respectivamente, para as ancestralidades africana, europeia e ameríndia) (SANTOS et al., 2015).

Em uma outra publicação do grupo, 2,2 milhões de SNPs foram genotipados em 6.487 indivíduos, dentre os quais, ainda, há 30 amostras submetidas a sequenciamento de genoma completo (KEHDY et al., 2015). Foi constatado que indivíduos de Salvador/BA apresentam a maior ancestralidade africana dentre as regiões amostradas (50,8%; sendo 42,9% europeia e 6,4% ameríndia); de Bambuí/MG apresenta 78,5% de ancestralidade europeia, 14,7% africana e 6,7% ameríndia. Por fim, na amostra de Pelotas/RS, há 76,1% de componente europeu, 15,9% africano e 8% indígena. Ainda nesse estudo, reafirma-se a reprodução direcional e que a contribuição genética nativa foi reduzida pouco após a chegada de europeus. Indivíduos de diferentes regiões parecem, ainda, diferir na origem europeia, com o Sul e o Sudeste do país tendo origens europeias mais diversas, enquanto no Nordeste a origem europeia é mais restrita à Península Ibérica. De modo análogo, o componente de ancestralidade africana mais presente no Sul e Sudeste está relacionado com a etnia banto e a África oriental, enquanto no nordeste a ancestralidade africana é mais frequentemente associada com a porção ocidental do continente, não-banto.

Em decorrência do acima exposto, o presente trabalho dedica-se a explorar os aspectos genéticos das doenças falciformes e a estratificação genética em população geral e sua relação entre pacientes brasileiros e norte-americanos, tendo por objetivo final de entender o impacto das CNVs na suscetibilidade ao AVC. Para tanto, sugere-se explorar a população brasileira a fundo, como ponto de partida para a comparação tanto entre a população geral e os afetados de AF, como entre indivíduos afetados por AF de diferentes nacionalidades, a norte-americana – melhor estudada – e a brasileira, altamente miscigenada e descrita em menor nível de detalhamento. Como objetivo secundário, portanto, está o detalhamento da estrutura populacional da população brasileira em relação a outras populações. Tais objetivos refletem a estrutura da presente tese, organizada em cinco partes. Esta primeira consiste de uma introdução geral em que se expõe as bases moleculares da anemia falciforme, as vias de vaso-oclusão, hemólise e adesão celular e então as manifestações clínicas, com ênfase no acidente vascular cerebral e considerações a respeito dos marcadores genéticos utilizados em estudos de associação e a importância da estrutura genética em estudos genéticos e, em particular, na anemia falciforme. No Capítulo I, seção que se segue à presente, aborda-se a associação entre AVC e variação no número de cópias em pacientes afetados por AF. Para tanto foram utilizadas duas coortes, uma brasileira e uma norte-americana, avaliadas por microarranjos de genotipagem e foram observadas regiões em associação significativa ou sugestivas em ambas. Na terceira parte, amostragens urbanas de indivíduos supostamente saudáveis de diferentes

regiões do Brasil (Recife/PE, Joinville/SC e Campinas/SP) são descritos em sua estrutura genética fina e comparados, observando-se sub-estrutura não descrita anteriormente na literatura da população urbana brasileira. Tal trabalho foi desenvolvido pelo cuidado necessário, no Capítulo I, em avaliar indivíduos de regiões geográficas distintas do Brasil. A quarta parte, como as duas anteriores, é um manuscrito destinado a publicação e deve-se ao interesse em entender diferenças populacionais entre coortes de anemia falciforme nacionalmente distintas, preocupação decorrida igualmente do capítulo primeiro. Foi observado que as populações afetadas por anemia falciforme brasileira e norte-americana diferem em sua ancestralidade proporcional tanto em nível autossômico geral quanto localmente ao cromossomo 11. Por fim, a última parte destina-se à discussão e conclusões gerais, que abarcam os principais achados do período em que a tese foi desenvolvida, conectando as partes anteriores naquilo que lhes é correlato.

CAPÍTULO I**COPY NUMBER VARIATIONS AND THE RISK OF CEREBROVASCULAR
DISEASE IN PATIENTS WITH SICKLE CELL ANEMIA**

A seção que se segue constitui o manuscrito que será submetido como artigo original ao periódico “European Journal of Human Genetics”.

Resumo

Embora a anemia falciforme (AF) resulte da homozigose de uma única mutação, na posição 6 do *locus* da β -hemoglobina, fenotipicamente, essa doença é muito heterogênea, de modo que diferentes pacientes podem apresentar evoluções clínicas significativamente distintas. Praticamente todos os órgãos podem ser afetados pela oclusão vascular, merecendo destaque o Sistema Nervoso Central (SNC), onde são observados acidentes isquêmicos transitórios, infartos e hemorragia cerebral, os quais acometem aproximadamente 25% dos pacientes com AF. As complicações neurológicas são graves, podendo ser fatais em até 15% dos casos. A identificação precoce de pacientes com anemia falciforme, suscetíveis à Acidente Vascular Cerebral (AVC) poderia diminuir os riscos, possivelmente prevenir a recorrência de infartos e potencialmente reduzir sua incidência. Dessa forma, foi realizada a análise de CNVs por meio de microarranjos de DNA em duas coortes de anemia falciforme independentes, uma norte-americana, do Hospital Infantil da Filadélfia (CHOP) e uma brasileira, derivada do Centro de Hematologia e Hemoterapia (HEMOCENTRO) e da Fundação de Hematologia e Hemoterapia de Pernambuco (HEMOPE). Após a avaliação criteriosa de histórico médico e controle de qualidade de dados, 99 pacientes pediátricos afetados por doenças cerebrovasculares foram comparados com 301 controles, no grupo norte-americano. Em relação aos pacientes brasileiros, após a filtragem, 44 casos de AVC isquêmico foram submetidos à análise juntamente a 46 indivíduos sem qualquer registro de doença cerebrovascular. Sinais de associação foram detectados em ambas populações.

**COPY NUMBER VARIATIONS AND THE RISK OF CEREBROVASCULAR
DISEASE IN PATIENTS WITH SICKLE CELL ANEMIA**

*Cruz PRS¹, Ananina G¹, Fertrin KY², Bezerra MAC³, Araújo AS⁴, Cendes F⁵, Mena F¹,
Costa FF², Hakonarson H⁶, Melo MB¹*

1. Laboratory of Human Genetics, Center for Molecular Biology and Genetic Engineering (CBMEG), University of Campinas – UNICAMP, Campinas, São Paulo, Brazil
2. Hematology and Hemotherapy Center, University of Campinas – UNICAMP, Campinas, São Paulo, Brazil
3. Department of Genetics, Federal University of Pernambuco, Recife, PE, Brazil.
4. Hematology and Haemotherapy Foundation of Pernambuco – HEMOPE, Recife, PE, Brazil.
5. Neuroimaging Laboratory, Department of Neurology, University of Campinas - UNICAMP, Campinas, SP, Brazil
6. Center for Applied Genomics, Abramson Research Center, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.

Correspondence to: Mônica Barbosa de Melo, Laboratory of Human Genetics, Center of Molecular Biology and Genetic Engineering (CBMEG), University of Campinas - UNICAMP, P.O.Box: 6010, Campinas, SP, Brazil, 13083-875.

ABSTRACT

Although sickle cell anemia (SCA) results from the homozygosity for a single mutation in amino acid 6 of the β -globin locus, this disease presents high phenotypic heterogeneity, so that different patients may have significantly different clinical outcomes. Vascular occlusion occurs systemically, and transient ischemic attacks, stroke and cerebral hemorrhage affect approximately 25% of patients with SCA. Early identification of patients with sickle cell anemia susceptible to cerebrovascular accident (CVA) could reduce the risk, possibly preventing the recurrence of infarcts and potentially reducing their incidence. Thus, we conducted copy number variation (CNV) analysis using genotyping microarrays in two independent sickle cell anemia cohorts, an American, from Children's Hospital of Philadelphia (CHOP) and a Brazilian, derived from Hematology and Hemotherapy Center (HEMOCENTRO) and from Pernambuco Hematology and Hemotherapy Foundation (HEMOPE). After medical records evaluation and data quality control, 99 pediatric subjects affected by cerebrovascular disease (CVD; transient ischemic attacks, arterial occlusion and stroke) were compared to 301 controls for the American cohort. Regarding the Brazilian patients, 44 ischemic stroke cases underwent analysis along with 46 individuals without any record of CVD, after filtering. We found association signals on both populations.

INTRODUCTION

Hemoglobinopathies are among the most common inherited diseases, globally. Sickle cell anemia (SCA) is a prevalent condition, and lacks a safe drug therapy. More than 500,000 newborns worldwide are estimated to be affected per year. Although SCA results from homozygosis of a single mutation at position 6 of beta-hemoglobin locus, this disease presents high phenotypic heterogeneity, so that different patients may have significantly different clinical outcomes. Virtually all organs may be affected by vascular occlusion. The most life-threatening vaso-occlusive episodes affect the central nervous system, where transient ischemic attacks, stroke and cerebral hemorrhage are observed. Cerebrovascular disease (CVD) affects approximately 25% of patients with SCA (PEGELOW et al., 1997; WIERENGA; SERJEANT; SERJEANT, 2001).

Copy number variations (CNVs) are defined as duplication, deletions, inversions, or complex rearrangements in genomic segments ranging from 100 bp to few Mb (ALMAL; PADH, 2012; REDON et al., 2006). CNVs account for a significant proportion of human genetic variation and have been expected to complement SNPs in the contribution to genetic susceptibility loci for common diseases. Candidate gene and GWAS studies have unveiled an abundant number of SNPs associated with cerebrovascular disease in SCA (BELISÁRIO et al., 2015; FLANAGAN et al., 2011, 2013; HOPPE et al., 2007; SEBASTIANI et al., 2005a, 2010). This is, to date, the first report on copy number variants on a sickle cell anemia related CVD.

We propose to investigate CNV occurrence by genotyping two cohorts to access CNVs potentially associated to cerebrovascular disease on SCA patients. Pediatric SCA patients were recruited at the Children's Hospital of Philadelphia (CHOP), comprising 123 subjects with cerebrovascular disease (transient ischemic attacks, stroke, arterial occlusion and/or cerebral hemorrhage) and 605 subjects without brain occlusion episodes. We also compared this group to a Brazilian SCA cohort, composed by 47 ischemic stroke affected subjects and 51 SCA controls.

RESULTS

Individuals were stratified according to cerebrovascular disease (cohorts' demographics is depicted in Table 1). Brazilian affected group includes young adults, so older

controls were intentionally selected to diminish the chances of adult stroke occurrence in this particular group.

We next assessed the overall population structure by ADMIXTURE using approximately 44,000 independent markers, derived exclusively from Affymetrix 6.0 genotyping data to control for cross-platform bias (CHOP subjects selected for PCA had both Illumina and Affymetrix 6.0 genotypes available). Sickle cell anemia patients from the US and from Brazil differ, as expected, as depicted in Figure 1.

We also performed PCA for each cohort separately (Figures 3 and 4), observing that case and control groups fairly overlap in both situations. American sickle cell patients cluster into two groups in agreement to the specific Illumina chip used. We, thus, underscore the importance of applying logistic regression with the three principal components as the suited method for correcting systematic sources of variance, as the PCA plot indicates to be the case between different array types.

After quality control (QC) of genotyping data, 400 samples were retained in the analysis for the American SCA cohort (99 cases and 301 controls). With respect to Brazilians, after quality control screening this group consisted of 90 subjects (44 cases and 46 controls).

The descriptive characteristics of CNVs called in cases and controls after QC are shown in Table 2. An average number of 29.8 CNVs including 14.9 deletions and 14.36 duplications were called for each American case, whereas Brazilian cases have on average 67.7 events, divided in 36.9 deletions and 30.9 duplications. For the American controls, we detected 15.1 deletions and 15.0 duplications per individual, which results in a total of 30.1 CNVs per individual. Brazilian controls shown 81.4 CNVs, which are composed by 40.9 deletions and 40.6 duplications. African Americans are still under evaluation for total and average per sample events.

We tested CNVs for association using an unbiased segment-based scoring approach to identify CNV regions (CNVRs) associated with the CVD status (GLESSNER; LI; HAKONARSON, 2013). For each probe, CNV frequency between cases and controls were compared by Fisher exact test. Then neighboring probes were collapsed into CNVRs, which is constituted by a genomic span of consecutive probes and have comparable significance when comparing case to control status. The local lowest probe p-value in Fisher exact test was used to represent the association of CNVR with disease status.

ParseCNV detected a total of 120 CNVRs (26 deletion and 94 duplication) with probes enriched in American SCA cases. Therefore, the resulting experiment-wide significance threshold is $p = 4.12 \times 10^{-04}$ with multiple testing correction.

There were no CNVR surpassing the significance level after multiple testing correction, among the nominally significant loci, there were 6 regions that comprises genes (Table 3; chr11:51221009-51293595 is 6.6 kb away from TRNA_Lys). Note in Table 3 that a column for the CNVR association confidence is present. ParseCNV uses number of probes, segmental duplications, number DGV entries, proximity to centromere/telomere, population frequency and average GC as criteria for evaluating the reliability of each association result.

Further, we inspected CNVRs covering gene regions in respect to the function of these genes, and also the region susceptibility and frequency of CNV through the number of DGV entries and proximity to segmental duplications, given that rare events are more likely to be causal, additionally to the potential of detrimental impact (evaluating DECIPHER entries, conserved elements and overlapping to exons and/or introns). This approach highlights the *HMOX1* hemizygous deletions as candidates to further investigation, as they abolish at least 3 exons of the gene and are not yet described. The validation for the *HMOX1* variant yield negative results, and the deletions seem to be an artifact.

ParseCNV association result for Brazilian patients is summarized in Table 4. Differently from the American cohort, no CNVR stands to the program's quality check. We then adopted an alternative approach by filtering variants that were consensus for two independent algorithms (BRLMM and the hidden Markov model implemented in PennCNV); following that, we further filtered CNVs exclusive of either cases or controls (regarding the PennCNV calling genotypes) and intersecting genes. Finally, permutation was performed in plink as an association test. The permutation results were evaluated in UCSC genome browser for the significant CNVRs regarding DGV entries, genes functions, conservation, clinical impact (by DECIPHER) and regional segmental duplications or another repetitive elements that might indicate recurrent events.

By following this approach, we were able to detect eight genome-wide significant CNVs in Brazilians (chr9:113,081,566-113,398,562, chr12:49,011,335-49,146,574, chr13:40,778,613-41,056,050, chr16:1,947,414-2,021,056, chr16:62596315-62701713, chr5:97,973,326-98,026,873, chr6:14,681,298-14,956,983 and chr6:37,063,020-37,110,473), with corrected p-values varying from 0.0018 to 0.0406. Table 5 summarizes these findings

along with the genes affected, when there is any (four CN variants are intergenic). Genes affected encompasses SVEP1, CCNT1, KANSL2, LINC00598, MSRB1, RPL3L and HS3ST6, mostly committed to cell cycle and maintaining (CCNT1 encodes for a cyclin C protein; KANSL2 is related to chromatin organization; MSRB1 is related to translation termination; and RPL3L encodes a protein that shares sequence similarity with ribosomal protein L3, whereas LINC00598 is a non-coding RNA class), except for SVEP1, related to cell attachment process. Knockdown of SVEP1 via small interfering RNA reduces E-selectin expression on the surface of human umbilical vein endothelial cells (HUVECs), and SVEP1 knockdown increased soluble ICAM1 and E-selectin concentrations in supernatant of stimulated HUVECs (SCHWANZER-PFEIFFER et al., 2010).

Although not genome-wide significant (nominal $p = 0.0173$), hemizygous deletions on *FLT3* gene stands as good candidates to validation (Figure 6), since the gene encodes a tyrosine-protein kinase that acts on the regulation of differentiation, proliferation and survival of hematopoietic progenitor cells and of dendritic cells, besides being a paralog of *FLT1*, implicated in the production of HbF (SOLOVIEFF et al., 2010a). This variant was present in three cases in the PennCNV calling algorithm results, and six cases in the BRLMM typing (Figure 6), despite failing to reach genome-wide significance, the inspection revealed that this variant is not yet described and, moreover the genotyping algorithms agreed in two individuals.

DISCUSSION

In the present work, we investigated the potential role of copy number variants in cerebrovascular disease related to sickle cell anemia by genotyping two sickle cell anemia populations. Since sickle-cell heterozygotes (HbAHbS) has higher fitness in malaria-endemic regions, due to a lower susceptibility to host *Plasmodium malariae* relative to the HbA homozygote, HbS mutation was preserved after five different events that took place around 5 kya. Present haplotypes are named after their putative geographical origin (Benin, CAR or Bantu, Cameroon, Senegal and Arab) and contribute to the overwhelming phenotypic heterogeneity observed in SCA. American is the best-studied population in respect to SCA, whereas Brazilian population is poorly described and differs from the Americans regarding its origin.

We find that Brazilian and US sickle cell anemia patients differ in association signals, consistently with historical records, although we cannot rule out the experimental bias, given that the arrays were handled by different researchers. Africans transported to the United States were originated mainly from Central West African ports, where the Benin haplotype prevails, while the 2.4 million African Negroes transported to Brazil, from 1701 to 1843, was predominantly from Angola, Congo, and Mozambique, areas where the CAR haplotype is more frequent. This dissimilarity has clinical importance, mainly because fetal hemoglobin levels greatly varies among haplotypes. Moreover, recently was found that European ancestry has a effect on variants affecting fetal hemoglobin levels in Brazilian patients namely *BCL11A* and *HMP-2*, and patients bearing these polymorphisms showed a decreased odd of having stroke (LEONARDO et al., 2016).

HMOX1 posed as an interesting candidate for validation, the deletions potentially affects Heme oxygenase gene (*HMOX1*), an inducible enzyme expressed in the endothelium that catabolizes hemoglobin and heme, its inhibition increases vascular stasis (BELCHER et al., 2006). Moreover, polymorphisms on *HMOX1* gene has been shown to modulate HbF levels in sickle cell anemia patients (GIL et al., 2013). The deletions did not stand validation by qPCR, though (Supplementary figure 1). *FLT3* deletions, found in Brazilians, was elected the most promising variant of this group, because it was called by two different algorithms, it seemed rare and abolishing the last exon of the *FLT3* gene. This gene codes the Fms-like tyrosine kinase 3 (FLT-3), receptor-type tyrosine-protein kinase FLT3, or fetal liver kinase-2 (Flk2), a protein that acts as a cytokine receptor and is involved in the survival, apoptosis and differentiation of myeloid lineages (KARSUNKY et al., 2003). This deletion may, thus, alter the cell development of cell lines as granulocytes or macrophages, affecting inflammation pathways and contributing to induce stroke. This variant was also not validated by qPCR (Supplementary figure 2).

The perspectives of the current work involve to further explore the list of variants generated in each cohort. Interestingly, in both cases genes' functions involve transcription factor, cell cycle and signaling. This enrichment could indicate a complex effect of CNVs on cell signaling network, and gene ontology analysis may elucidate phenotypic impact. Combining CNVs and SNPs in the same analysis may also lead to a better understanding of stroke genetic basis in sickle cell anemia patients, as conditions like the thrombocytopenia-absent radius (TAR) syndrome involve both structural variant and SNP in the *RBM8A* gene (WEISCHENFELDT et al., 2013).

METHODS

Subjects

Primary data were retrieved from and verified in Epic Care Clinical System (Epic, Verona, WI) for CHOP individuals. Subjects with ambiguity regarding sickle cell status were removed from further analysis, as well as vascular diseases, diabetes and anemia.

Brazilian SCA patients with and without stroke were recruited at two Brazilian hematology centers, HEMOPE (Recife, PE) and HEMOCENTRO (Campinas, SP). Patients' recruitment and blood collection were approved by the Ethics Committee of FCM-UNICAMP (Campinas-SP, Brazil). This study was conducted according to the tenets of the Declaration of Helsinki. Demographic data on these patients are depicted on Table 1.

Genotyping

Genotyping of both African American and American SCA was performed at the Center for Applied Genomics, Children's Hospital of Philadelphia. Genotyping was carried out on the Illumina HumanHap550 BeadChip or Illumina Human610-Quad Bead-Chip (Illumina Inc., CA, USA).

The genotyping for Brazilian samples was performed by Affymetrix Genome-Wide Human SNP 6.0 Array (Affymetrix Inc., CA, USA) according to manufacturer's recommendations and executed at the Center for Molecular Biology and Genetic Engineering (CBMEG, SP, Brazil). Schematic analysis pipeline is represented in Figure 2.

CNV calling

For the American cohorts, 535,752 SNPs common to the Illumina Human610-Quad BeadChip or the Illumina HumanHap550 BeadChip were used for CNV detection. The PennCNV(WANG et al., 2007) software was used to call CNVs based on the intensity data, log R ratio (LRR) and the B allele frequency (BAF) from all arrays in this study. For Affymetrix data, LRR and BAF values were extracted by Affy2sv package (HERNANDEZ-FERRER et al., 2015) prior to CNV calling. The Brazilian cohort were also genotyped by the BRLMM

algorithm (AFFYMETRIX, 2006), the overlapping calls between algorithms were analyzed to filter candidates of association.

Population analysis

We filtered probes and samples using PLINK (PURCELL et al., 2007). Population structure and admixture were analyzed applying ADMIXTURE (ALEXANDER; NOVEMBRE; LANGE, 2009) software, including reference populations from the HapMap consortium and a sample of allegedly healthy Brazilians (HbAA).

CNV quality control

CNV metrics were reviewed and samples were removed if presenting outlier values in any of the following categories: call rate, intensity noise measured by the standard deviation of Log R Ratio (LRR SD), cryptic relatedness between samples, intensity waviness measured by GC base pair wave factor (GCWF), and high number of CNVs. The distribution of the CNV metrics was plotted, the distribution plots typically display a linear phase and an exponential phase, and samples with CNV metrics falling into the exponential phase were removed.

For Illumina data, data quality metric thresholds we have set were: call rate > 98%, SD LRR < 0.3, |GCWF| < 0.05, and count CNV < 100. For Affymetrix 6.0 data, these measures were: call rate > 96%, SD LRR < 0.35, |GCWF| < 0.02, and count CNV < 80. Furthermore, we used PLINK for genome-wide identity-by-descent analysis and duplicated or related samples which identity-by-descent score (PI_HAT) > 0.125 were detected. One sample from each pair of such cryptic related or duplicated samples was removed.

CNV association statistical analysis

We adopted the CNV association software ParseCNV (GLESSNER; LI; HAKONARSON, 2013) to detect the association between CNV region (CNVR) and cerebrovascular disease. ParseCNV uses a genome-wide segment-based scoring approach.

For logistic regression analysis, we first generated ped files for CNV status. In the deletion ped files, CNV genotype status was defined in the following way: 1 1 for CN = 0, 1 2 for CN = 1, and 2 2 for others. In the duplication ped file, CNV genotype status was defined as 1 1 for CN = 4, 1 2 for CN = 3, and 2 2 for others. This strategy allows for logistic regression analysis including the principal components to adjust for population stratification.

Permutation

Using PLINK, we performed 10,000 permutations using CNV calls present in either cases or controls only and overlapping genes, then the top hits evaluated. This analysis was conducted on Brazilians only, because ParseCNV yielded low confidence score results. Besides probe p-value on permutation, we also evaluated low DGV entries, absence of segmental duplications nearby and gene function as criteria for selecting candidates for validation.

CNV validation

Universal Probe Library (Roche, Indianapolis, IN, USA) assay was carried out for candidate validation on the CHOP sample.

FUNDING

This work was supported by Brazilian Synchrotron Light Laboratory; Children's Hospital of Philadelphia; São Paulo State Research Foundation [2008/57441-0, 2012/06438-5, 2015/13152-9];

REFERENCES

1. Pegelow, C. H. *et al.* Natural history of blood pressure in sickle cell disease: risks for stroke and death associated with relative hypertension in sickle cell anemia. *Am. J. Med.* **102**, 171–177 (1997).
2. Wierenga, K. J., Serjeant, B. E. & Serjeant, G. R. *Cerebrovascular complications and parvovirus infection in homozygous sickle cell disease.* *J. Pediatr.* **139**, 438–442 (2001).

3. Redon, R. *et al.* Global variation in copy number in the human genome. *Nature* **444**, 444–454 (2006).
4. Almal, S. H. & Padh, H. Implications of gene copy-number variation in health and diseases. *J. Hum. Genet.* **57**, 6–13 (2012).
5. Hoppe, C. *et al.* Confirmation of an association between the TNF(-308) promoter polymorphism and stroke risk in children with sickle cell anemia. *Stroke.* **38**, 2241–6 (2007).
6. Sebastiani, P. *et al.* Genetic modifiers of the severity of sickle cell anemia identified through a genome-wide association study. *Am. J. Hematol.* **85**, 29–35 (2010).
7. Sebastiani, P., Ramoni, M. F., Nolan, V., Baldwin, C. T. & Steinberg, M. H. Genetic dissection and prognostic modeling of overt stroke in sickle cell anemia. **37**, 435–440 (2005).
8. Flanagan, J. M. *et al.* Genetic mapping and exome sequencing identify 2 mutations associated with stroke protection in pediatric patients with sickle cell anemia. *Blood* **121**, 3237–45 (2013).
9. Belisário, A. R. *et al.* Association of alpha-thalassemia, TNF-alpha (-308G>A) and VCAM-1 (c.1238G>C) gene polymorphisms with cerebrovascular disease in a newborn cohort of 411 children with sickle cell anemia. *Blood Cells, Mol. Dis.* **54**, 44–50 (2015).
10. Flanagan, J. M. *et al.* Genetic predictors for stroke in children with sickle cell anemia. *Blood* **117**, 6681–4 (2011).
11. Wang, K. *et al.* PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17**, 1665–1674 (2007).
12. Hernandez-Ferrer, C. *et al.* affy2sv: an R package to pre-process Affymetrix CytoScan HD and 750K arrays for SNP, CNV, inversion and mosaicism calling. *BMC Bioinformatics* **16**, 167 (2015).
13. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–75 (2007).
14. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS*

Genet. **2**, e190 (2006).

15. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
16. Glessner, J. T., Li, J. & Hakonarson, H. ParseCNV integrative copy number variation association software with quality tracking. *Nucleic Acids Res.* **41**, e64–e64 (2013).
17. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
18. Belcher, J. D. *et al.* Heme oxygenase-1 is a modulator of inflammation and vaso-occlusion in transgenic sickle mice. *J. Clin. Invest.* **116**, 808–16 (2006).
19. Gil, G. P. *et al.* Polymorphism in the HMOX1 gene is associated with high levels of fetal hemoglobin in Brazilian patients with sickle cell anemia. *Hemoglobin* **37**, 315–24 (2013).
20. Solovieff, N. *et al.* Fetal hemoglobin in sickle cell anemia: genome-wide association studies suggest a regulatory region in the 5' olfactory receptor gene cluster. *Blood* **115**, 1815–22 (2010).

TABLES AND FIGURES

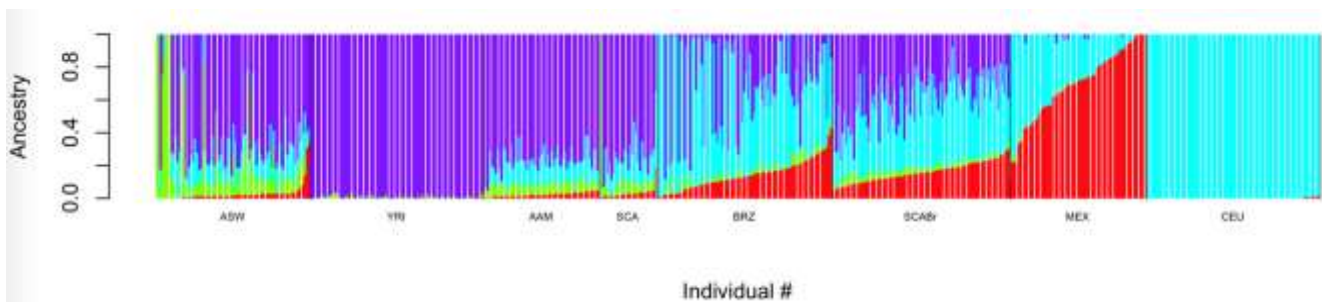


Figure 1. population structure assessed by ADMIXTURE ($k=4$). Labels: AAM: Afro Americans, BRZ: Brazilian healthy individuals, CEU: European ancestry, MEX: Mexican ancestry, SCA: sickle cell anemia patients in the US, YRI: Yoruba in Nigeria, LWK: : Luhya in Webuye, Kenya, SCABr: Brazilian sickle cell anemia.

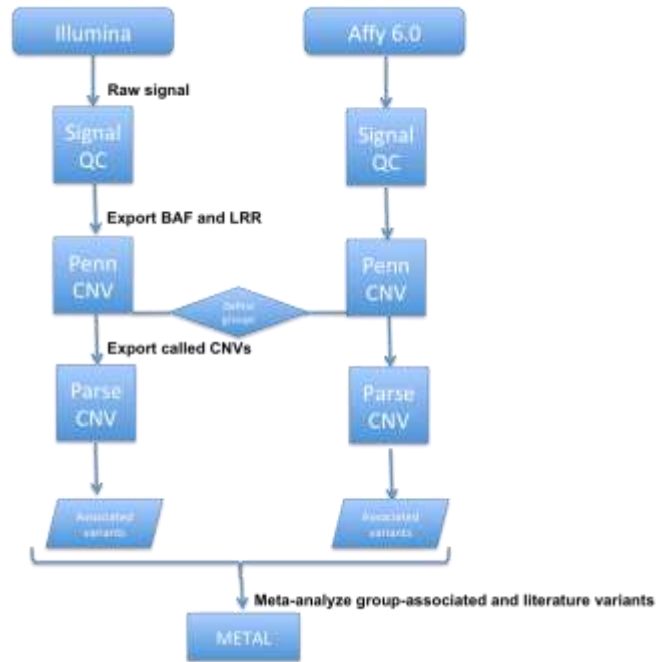


Figure 2. Schematic overview of the analysis.

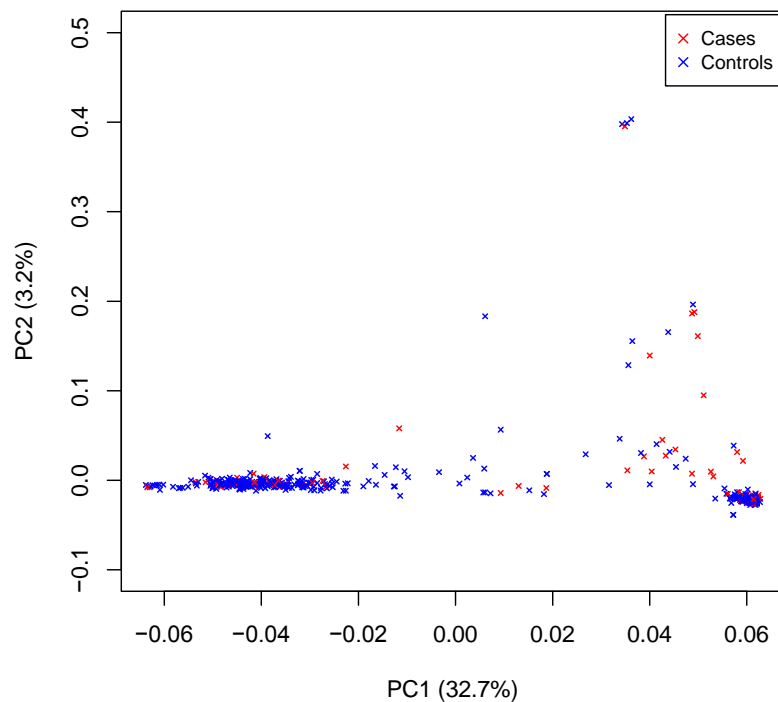


Figure 3. PCA on American sickle cell anemia patients, stratified in case and control groups. Variance capture by principal components is depicted within parenthesis. PC: principal component.

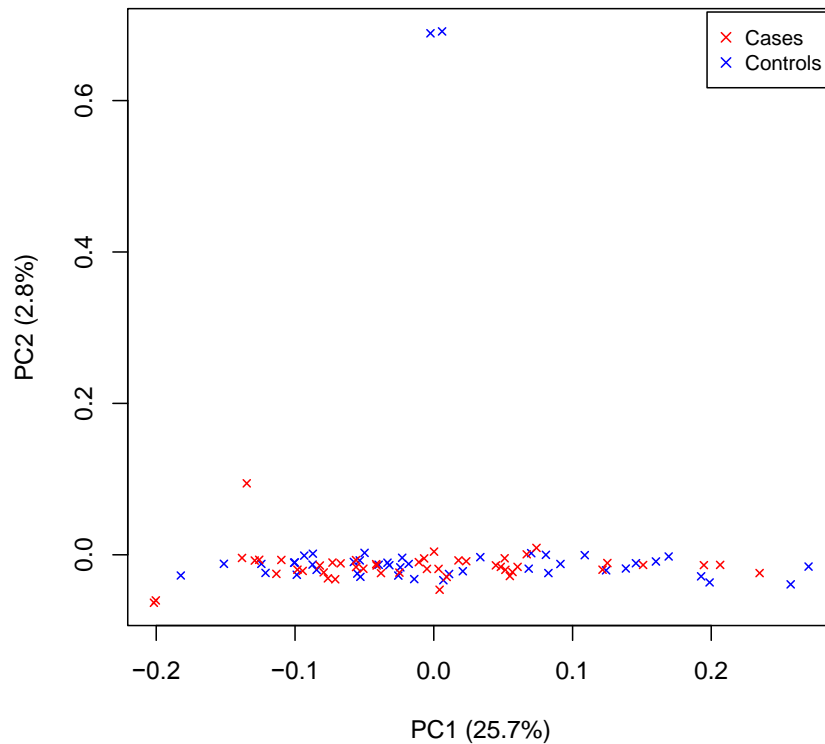


Figure 4. PCA on Brazilian sickle cell anemia patients, stratified in case and control groups. Variance capture by principal components is depicted within parenthesis. PC: principal component.

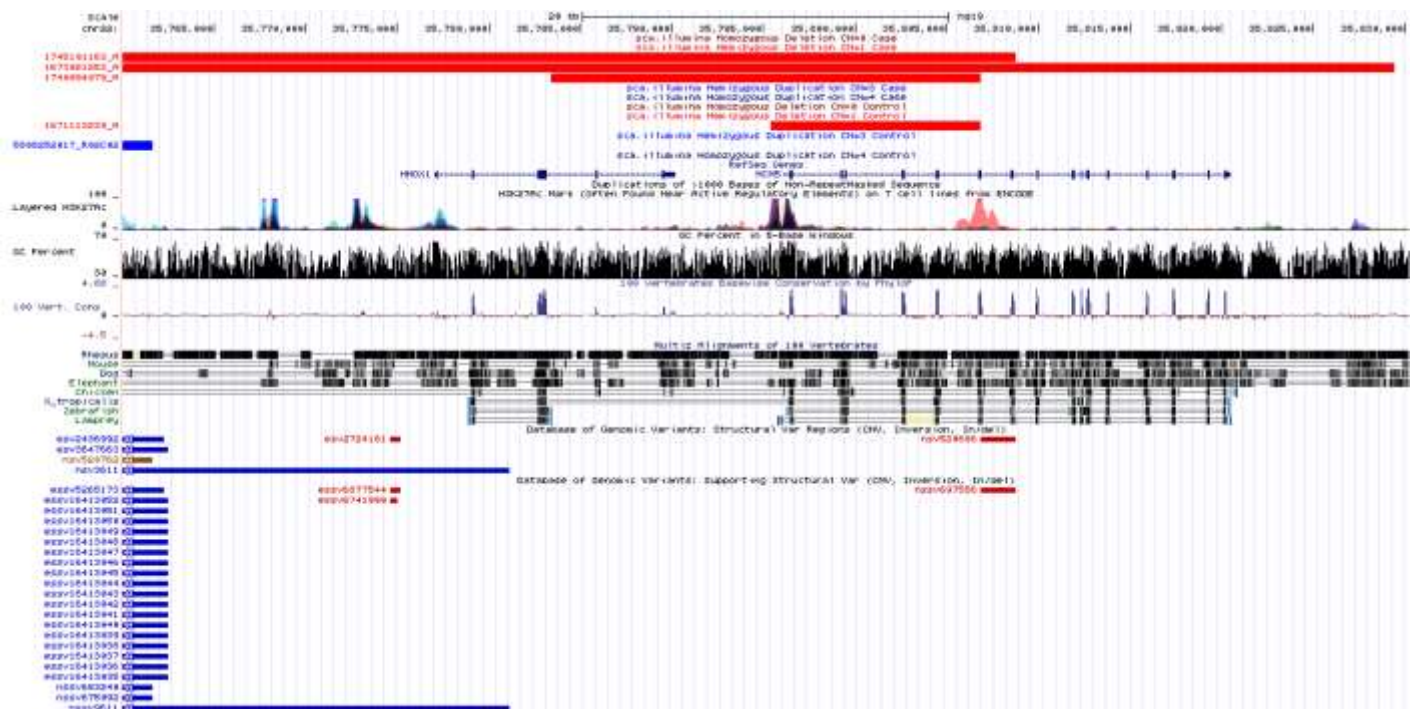


Figure 5. Deletions in gene HMOX1. Red rectangles indicate the individual deletions observed among American cerebrovascular disease cases.

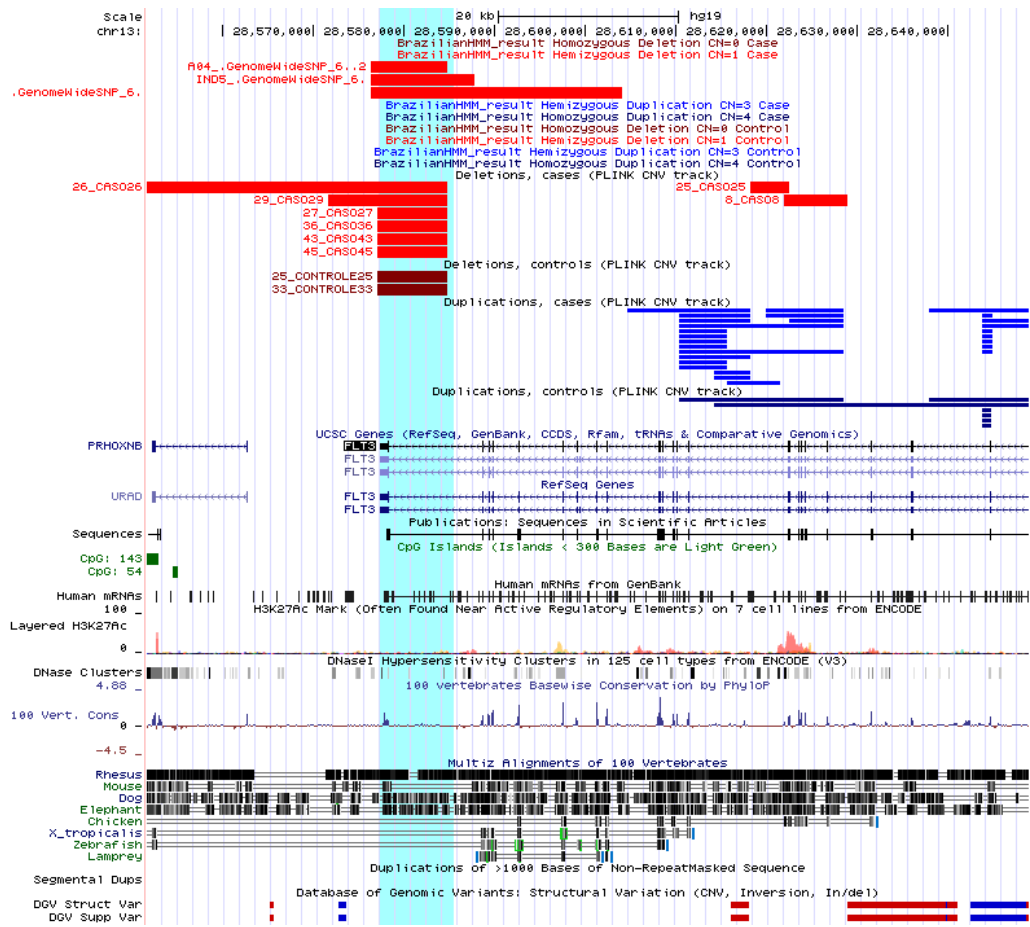


Figure 6. Deletions in the *FLT3* gene. Red rectangles indicate the individual deletions observed among Brazilian stroke cases.

Table 1. Sample demographics.

	US Cohort		Brazilian Cohort		African Americans (n = 4,328)
	Affected (n = 123)	Unaffected (n = 605)	Affected (n = 47)	Unaffected (n = 51)	
Average age (\pm SD)	9.3 (\pm 5.9)	10.2 (\pm 6.0)	21.3 (\pm 12.4)	33.4 (\pm 8.8)	8.4 (\pm 5.9)
Male (%)	60.5	48.0	37.3	42.6	57.0
Ethnicity					
European ancestry	65	124	10	12	0
African ancestry	58	481	37	39	4,328

Table 2. CNV profiles of American and Brazilian cases and controls.

	American cohort		Brazilian cohort	
	Cases (n = 99)	Controls (n = 301)	Cases (n = 44)	Controls (n = 46)
Total no. CNVs				
Deletions	1477	4538	1659	1881
Duplications	432	1518	1389	1870
Average no.CNV/subject				
Deletions	14.92	15.08	36.87	40.89
Duplications	14.36	15.04	30.87	40.65
Average size of CNVs (kb)				
Deletions	62.54	65.54	19.67	25.71
Duplications	380.68	188.13	53.21	47.05

Table 3. CNVRs nominally significant enriched in SCA American cases. Only nominally significant results passing confidence test on ParseCNV is shown here.

CNVR(hg19)	Gene	P-value	Type	Cases	Controls	Pass/Fail
chr20:6217226-9-62198348	C20orf195,PRIC285,S RMS	9.48×10^{-04}	Del	8	3	PASS
chr1:6318069-6379607	ACOT7,GPR153	0.0117	Del	5	2	PASS
chr22:35783413-35806815	HMOX1,MCM5	0.0148	Del	3	0	PASS
chr9:3901415-3916263	GLIS3	0.0148	Del	3	0	PASS
chr19:42001210-42035716	LOC100505495	0.0351	Del	4	2	PASS
chr11:51221009-51293595	TRNA_Lys	0.0451	Del	5	4	PASS
chr21:20074007-20078424	BC028044,BC051441	0.0211	Dup	7	6	PASS

Table 4 . CNVRs nominally significant enriched in SCA Brazilian cases by PennCNV.

CNVR(hg19)	Gene	P-value	Type	Cases	Controls	Pass/Fail
chr12:33299791-33302201	SYT10	0.0262	Del	5	0	FAIL
	AJ004954,LO C727924,OR4 M2,OR4N3P, OR4N4,abPart					
chr15:22351239-22483127	s	0.0141	Del	1	9	FAIL

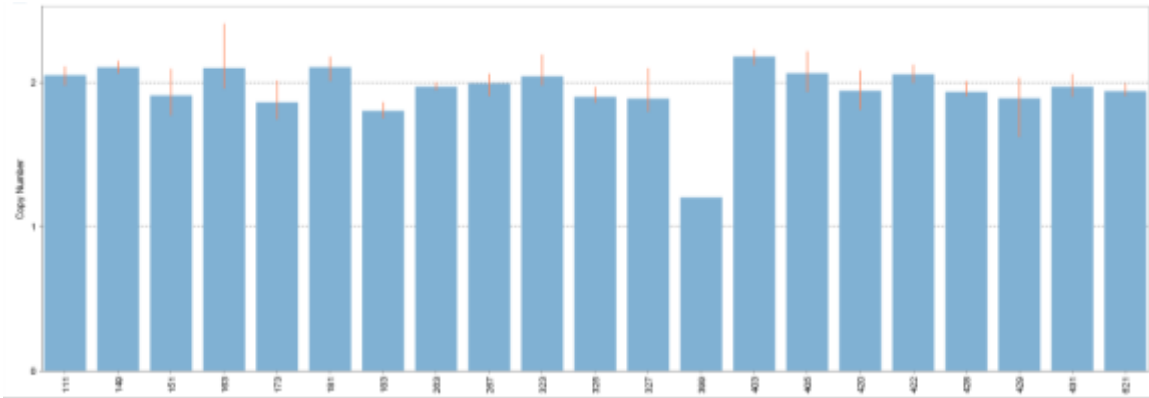
Table 5 . CNVRs genome-wide significant in SCA Brazilian by overlapping PennCNV and GTC calls.

CNVR(hg19)	Gene	Corrected p
chr9:113,081,566-113,398,562	SVEP1	0.0018
chr12:49,011,335-49,146,574	CCNT1, KANSL2	0.0406
chr13:40,778,613-41,056,050;	LINC00598	0.0055
chr16:1,947,414-2,021,056	MSRB1, RPL3L, HS3ST6	0.0406
chr16:1,947,414-2,021,056	intergenic	0.0406
chr5:97,973,326-98,026,873	intergenic	0.0406
chr6:14,681,298-14,956,983	intergenic	0.0406
chr6:37,063,020-37,110,473	intergenic	0.0406

SUPPLEMENTARY MATERIAL

Copy Number Call	chr22:35789054-35789113	chr22:35796205-35796265	chr22:35804539-35804600	Notes	Subject Status
	SCA_CNV_1	SCA_CNV_4	SCA_CNV_6		
1064899236	2	2	2	SCA_CNV_4 calculated at 1.48	Normal
1068867748	2	2	2		Potential loss of heterozygosity
1248076699	2	2	2		Called CNV
1450009662	2	2	2		Normal
183305413	2	2	2		Normal
2933042257	2	2	2	SCA_CNV_4 calculated at 2.51	Potential loss of heterozygosity
3990840712	2	2	2		Potential loss of heterozygosity
4896287245	2	2	2		Called CNV
4920388137	2	2	2	SCA_CNV_1 calculated at 2.52; SCA_CNV_4 calculated at 2.85	Normal
6714539219	2	2	2		Potential loss of heterozygosity
683351829	2	2	2		Potential loss of heterozygosity
7218320492	2	2	2		Normal
7894667033	2	2	2		Potential loss of heterozygosity
7922536202	2	2	2		Normal
9009669987	2	2	2		Normal
9720669845	2	2	2		Potential loss of heterozygosity
1188719965	2	2	2		Normal
1445831880	2	2	2		Normal
1831833456	2	2	2		Potential loss of heterozygosity
245270400	2	2	2		Normal
3253633085	2	2	2		Normal
3529855245	2	2	2		Potential loss of heterozygosity
7038009109	2	2	2		Called CNV
7043068801	2	2	2		Potential loss of heterozygosity
779737473	2	2	2		Potential loss of heterozygosity
807771420	2	2	2		Normal
8203432254	2	2	2		Normal
8376273423	2	2	2		Potential loss of heterozygosity
8697469412	2	2	2		Potential loss of heterozygosity

Supplementary figure 1. Validation results for *HMOX1* putative copy number variant. Three independent qPCR probes were used, gray lines show the individuals that had deletions called (copy number for all three was normal).



Supplementary figure 2. qPCR showing copy number for both cases (four first columns) and most controls (fifth to 21st column) for *FLT3* last exon. There is only one control seeming to have a deletion.

CAPÍTULO II**FINE SCALE GENETIC STRUCTURE OF THE POPULATIONS FROM THREE
BRAZILIAN REGIONS**

A seção que se segue constitui o manuscrito que será submetido como artigo original ao periódico “Plos Genetics”.

RESUMO

Populações miscigenadas se colocam como um desafio a ser tratado em estudos de genética médica e carecem de exploração aprofundada. Com mais de 200 milhões de habitantes, o Brasil figura como a quinta maior população do mundo e é altamente miscigenada. Existem poucos estudos explorando aspectos da genética de populações do país. Apresenta-se aqui a estrutura genética em alta resolução de brasileiros amostrados em três regiões do país, nas quais reside cerca de 84,5% do total da população. Foram analisadas 185 amostras individuais genotipadas pelas plataformas de microarranjo GenomeWide SNP 5.0 e 6.0, juntamente com os dados dos projetos HapMap e 1000 Genomes, publicamente disponíveis. Observou-se que brasileiros diferem de outras populações latino-americanas nos padrões de desequilíbrio de ligação, demonstrando valores reduzidos de r^2 e maior decaimento de DL, desviando-se do padrão observado para europeus e outras populações latino-americanas, situando-se como intermediário entre esses e o padrão africano. Consistentemente com estudos anteriores, nossa abordagem de análise de componentes principais em noventa e um brasileiros, em conjunto com populações de referência mostrou distribuição da nossa população em uma faixa de variabilidade que se estende entre os continentes africano e europeu. O método ADMIXTURE mostrou que brasileiros portam maior variabilidade de componentes europeus e africanos, porém não de componentes ameríndios, quando comparados com populações hispânicas. Também foi observado que as três regiões geográficas analisadas são discrimináveis pelo método bayesiano empregado, contrastando com estudos anteriores sugerindo uniformidade intrapopulacional.

**FINE SCALE GENETIC STRUCTURE OF THE POPULATIONS FROM THREE
BRAZILIAN REGIONS**

Ananina G¹, Cruz PRS¹, Lopes-Cendes IT², Gil-da-Silva-Lopes VL², Passos Lima C³, França PHC⁴, Cabral NL⁴, Donatti A², Secolin R², Lourenço GJ³, Bezerra MAC⁵, Araújo AS⁶, Simioni M², Costa FF⁷, Melo MB¹

1. Laboratory of Human Genetics, Center for Molecular Biology and Genetic Engineering (CBMEG), University of Campinas – UNICAMP, Campinas, SP, Brazil.
2. Department of Medical Genetics, School of Medical Science, University of Campinas - UNICAMP. The Brazilian Institute of Neuroscience and Neurotechnology (BRAINN), Campinas, SP, BRAZIL.
3. Department of Clinical Medicine, University of Campinas – UNICAMP, Campinas, SP,, Brazil.
4. Department of Medicine, Regional University of Joinville - UNIVILLE, Joinville Stroke Registry, Joinville, SC, Brazil.
5. Department of Genetics, Federal University of Pernambuco, Recife, PE, Brazil.
6. Hematology and Haemotherapy Foundation of Pernambuco – HEMOPE, Recife, PE, Brazil.
7. Hematology and Hemotherapy Center, University of Campinas - UNICAMP, Campinas, SP, Brazil.

Correspondence to: Galina Ananina, Laboratory of Human Genetics, Center of Molecular Biology and Genetic Engineering (CBMEG), University of Campinas - UNICAMP, P.O.Box: 6010, Campinas, SP, Brazil, 13083-875. Telephone number +55 019 3521-1135, email address: g.ananina@gmail.com.

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

ABSTRACT

Admixed populations stand as challenge on assessment by medical genetics studies and are still not fully explored. With more than 200 million inhabitants, Brazil figures as the fifth largest population in the world. There are few studies exploring the country's population genetics aspects. Here we present high-resolution genetic structure of Brazilians sampled on three regions of the country that account for approximately 84.5% of the overall population. We analyzed 185 individual Brazilian samples genotyped with the *Genome Wide SNP Array 5.0* and the *Genome Wide SNP Array 6.0* platforms along with data from HapMap International and 1000 Genomes Projects, publicly available. We found that Brazilians differ from other Latin American populations in linkage disequilibrium patterns, showing the lowest r^2 values and higher LD decay, deviating from the pattern observed for European and Latin American and posing as intermediate between these and African outline. Consistent with previous reports on general ancestral composition, our model-based approach to principal component analysis on ninety-one Brazilians along with reference populations shows the distribution of our population on an African-European axis of variability. ADMIXTURE method shows that Brazilians bear greater variability in European and African components, but not Amerindian compared to Hispanic populations. We also observed that the three geographical regions analyzed are discernable by the Bayesian method we applied, contrasting with previous reports suggesting uniformity within this population.

INTRODUCTION

Modern day medicine is more and more based on genetic and genomic data in such a way that incomplete knowledge of population genetic makeup may affect quality of medical research. Highly heterogeneous recently admixed populations represent a special challenge itself.

The Brazilian population is the fifth largest in the world and has 204,450,649 residents with reference date on 1th July 2015 (1). However, its genetic profile is still poorly understood, especially at the whole genomic scale.

Demographic history notes. The Brazilian population is highly admixed and ethnically heterogeneous; it resulted from a number of migration events that was accompanied by really intensive admixture. This population was formed by mixing three main ancestral roots: Amerindians, Europeans and Africans. Initially, all the area of the country was inhabited by indigenous Amerindians. There is little information about the size of the indigenous population at the time the Portuguese arrived in 1500, although frequently mentioned number is approximately 2.5 million (2). It is estimated that between 1500 and 1808, around 500,000 Portuguese settlers, mostly men, arrived in Brazil. Of course, it gave impetus to unions with Amerindian women.

In the mid-16th century the slave trade began, extending to 1850 and taking around 4 million Africans to Brazil (2), the highest number throughout American countries. During these three centuries, the demography promoted introgression of African genetic material to the “Brazilian genome”. Genetic studies showed the occurrence of preferential mating in relation to skin pigmentation, and directional mating, given that European men joined with African women (3).

About 1.7 million slaves (approximately 43% of the total quantity) arrived in Brazil in the early 19th century. During this period, the Amerindian population has suffered a drastic reduction in size due to conflicts with settlers and the spread of European diseases, to which Amerindians were not resistant. Probably the biggest contribution of Amerindians in the formation of the Brazilian population occurred in the first two or three centuries of colonization and had no accented importance from the 19th century, when the displacement of people to urban areas started. With the maintenance of the slave trade (until 1850) and European

immigration (20th century) the Amerindian genetic component was eroding from the nuclear genome of Brazilians.

In the year 1888, slavery was banned, and the government adopted a policy to encourage immigration from Europe to Brazil. During the period of approximately 50 years, 1872-1925, Brazil received about 5.5 million immigrants from Europe and the Middle East (2), being approximately 34% Italian, 29% Portuguese, 14% Spanish, 5% Japanese, 4% Germans, 2% Lebanese and Syrians, and 12% of other nationalities. This dramatic demographic event was probably responsible for concealing the regional differences in ancestry, because the European ancestry component became uniformly dominant, constituting in similar proportions to the North (69.7%), Northeast (60.6%), Southeast (73.7) and South (77.7%) of Brazil (4).

Genetic profile. Naturally, mitochondrial DNA haplogroups distribution in Brazil reveals great heterogeneity, showing coexistence of maternal lineages from three phylogeographic regions: Amerindian, African and European. A study of 247 Brazilians from different urban regions of Brazil reported that around a third (33%) of the mitochondrial lineages of individuals who identified themselves as "white" is of Amerindian origin. Other haplotypes were of African (28%) and European (39%) origin (5). A higher proportion of Amerindian matrilineage strains are normally observed in Northern Brazil when compared to Southern Brazil; the Amerindian component being at the level of mtDNA higher than for nuclear markers (5, 6). Here we present high resolution nuclear genetic structure of Brazilians sampled on the three most populated regions of the country (Southeast: 43%, Northeast: 27% and South: 14,5% of the total inhabitants) (1).

Aims. To evaluate the genetic diversity in the samples from the Brazilian States of São Paulo, Pernambuco and Santa Catarina in comparison with other worldwide populations.

RESULTS

Subjects and genotyping. After quality control filtering, 91 Brazilian samples genotyped with Genome Wide SNP Arrays 6.0 and 94 samples genotyped with Genome Wide SNP Arrays 5.0 were analyzed. For the data obtained from HapMap and 1000 Genomes projects, a few individuals showed IBD higher than 0.1875, and one of each pair was removed from further analysis.

Overall, the number of SNPs in the samples successfully genotyped with 6.0 array was 861,474. Of these, only 840,856 SNPs were present in the 1000genomes sample; 112 SNPs were removed due to the strands mismatch (strands flip) resulting in 840,744 markers left. After application of quality filters (Mendelian errors, low frequency and poorly genotyped SNPs) we kept 286,825 SNPs. Genotyping with *Genome Wide SNP Arrays 5.0* platform produced 427,171 SNPs' genotypes, of which 422,875 SNPs are common to the *Genome Wide SNP Array 6.0* platform. Taking into account the 1000genomes data, there were 415,248 common SNPs among all samples, resulting in 367,972 SNPs kept after Mendelian errors, low frequency and poorly genotyped SNPs removal (Figure S1).

LD decay. To estimate LD, we analyzed 185 individual Brazilian samples genotyped with *Genome Wide SNP Arrays 5.0* and *6.0* platforms. We extracted common markers among these platforms and markers present in the data obtained from HapMap and 1000 Genomes Projects. At this stage, we had 158,609 common SNPs among all genotyping platforms remaining after QC pruning. Before LD analysis, we also removed HWE-inconsistent SNPs. LD (r^2), estimated by PLINK, had highest values in the European populations and the lowest in the Africans. Brazilian population showed intermediate LD values between these two extremes. Comparing to other Latin American populations, Brazilian population showed lower values of r^2 (Figure 1).

Admixture analysis. Admixture and all further analysis were based only on the Brazilian samples genotyped with *Genome Wide SNP Arrays 6.0* platform and data obtained from HapMap and 1000 Genomes projects. After QC filtering and platforms merging (286,825 SNPs retained), we removed markers localized in the regions involved in long range LD (24) (280,820 SNPs retained) and pruned out SNPs in LD (131,365 remaining SNPs). To perform the last step, we calculated LD between each pair of SNPs in the window of 50 consecutive SNPs and removed one of a pair if r^2 was greater than 0.5; than the window was shifted 5 SNPs and the procedure repeated (14). We ran ADMIXTURE 10 times with different seed values for the number of clusters (K) from 1 to 12 taking cross validation error into account. The minimum cross validation error (0.6178) was achieved at K=7; K6=0.6183, K8=0.6179 (Figure S2).

Further, the runs for the selected K value were evaluated with CLUMPP software. All replicate runs of ADMIXTURE produced very stable solution; the maximum pairwise similarity coefficient was 0.999. Final bar plot and plots with summarizing statistics are shown on Figure 2.

FineStructure analysis. In our analysis we applied the linked model approach. The PCA plots of the resulting coancestry matrix are shown on Figure 3. Eigenvalues sharply drop after the fourth value, therein the first four values account for 90.6% of the variation. (Figure S3). Selected eigenvectors plots are depicted on Figures 3 and 4.

FineStructure software also implements a model-based, namely Bayesian, approach to infer population structure. Population membership assignment is performed using Markov chain Monte Carlo (MCMC) algorithm. Thus, we assessed the convergence of the runs comparing the graphics of pairwise coincidence matrices produced by two runs initialized with different random seeds. We systematically increased the number of iterations until the solution converges and the differences between two runs were too small to allow discrimination. Finally, the software builds a tree plot. Our tree showing the similarity among the groups and relationship of the Brazilian samples with other populations is represented on Figure 6.

DISCUSSION

This is, to date, the only report on the fine scale resolution of Brazilian genetic structure. We took advantage of cohorts from different geographical regions, which span considerable population of the country. Our data is consistent with other publications on Brazilian population regarding ancestry composition, with this population being highly admixed (7) and derived mostly from European and African populations (8), as show on Figure 2 and Figure 3. As it has been already shown, LD in African populations is markedly less than in Europeans (9), whereas LD in Latino-American populations was only slightly different from that in non-African samples (10, 11).

However, we found that LD patterns for Brazilians were different from other Latin American populations and more similar to the pattern observed in the populations with African ancestry. We observed the pattern being more diverse and harboring smaller haplotypes, since it has lower linkage disequilibrium and greater LD decay. This may be explained by the relatively rapid expansion of the Brazilian population by means of multiple migrations and demographic growth. We do not disregard the major contribution of African genomes to the Brazilian genetic heritage as it was shown by the ADMIXTURE analysis (Figure 2), allowing this population to preserve greater diversity.

Moreover, ADMIXTURE analysis provides information on the variability of each ancestral component, since we were able to compare medians and quantiles among populations. We found that Brazilians bear greater heterogeneity on supposedly African and European axis of ancestry regarding Latin American populations presented here. This does not hold for the Amerindian component, for which Brazilians have the least contribution and variation. This is supported by the history of the county colonization, with the coast being prioritized for settlement alongside to the contentious relation with natives. The fact that the first Portuguese settlers were men ought to have influenced the interbreeding between Europeans and natives or Africans in the initial centuries of colonization. Gonçalves and cols. (12) found that Brazilian Y-chromosome harbors footprints mostly of European origin and only 2.5% of lineages were from sub-Saharan Africa, and none were Amerindian, whereas mtDNA data in white Brazilians showed 60% of Amerindian or African matrilineages.

Our PCA data shows that Brazilians, as expected, form a continuum between European and African individuals, where we show these proxy ancestral individuals on a better structure resolution than any other previously published work with Brazilian population. Interestingly, the plot of PC2 vs. PC3 highlights relationship between Brazilian population and the populations of African origin. One can observe that the genetic composition of this sample of Brazilian population differs from the ASW population (African Ancestry in SW USA). ASW population demonstrates higher similarity to YRI (Yorubian) population, wherein Brazilians are slightly distributed in the direction of MKK (Maasai in Kinyawa, Kenya) group (Figure 4). Quite likely, it does indicate diverse sources of slave traffic to Brazil in comparison to the USA.

Model-based Bayesian clustering approach subdivides analyzed dataset in 110 fine scale groups (Figure 6). The two major clusters separate groups with pronounced African ancestry from all others. However, Brazilians can be found in both of this groups being much more variable than other Latino-American populations. Some fine-scale groups are formed by clustering Brazilian individuals with CLM or PUR indicating the higher number of shared haplotypes in these groups.

We were able, additionally, to discriminate the three different analyzed geographic regions, diversely from previously stated (8). Brazilian populations seem to have heterogeneity among different regions to some extent (Figure 5). Northeastern population has a greater distribution on the corrected PCA plot, whereas Southern population distributes more uniformly and Southeastern population behaves as an intermediate between these two. The greater

diversity towards septentrional parts of the country seems to reflect an augmented African descent. Southeast Brazil, specially, has observed additional waves of immigration of Europeans on the early 20th century, complying with the govern initiatives. Although high admixture rates were observed, there are still differences among regions, meaning that genetic flux is not as broad as suggested before (8).

In conclusion, we have shown fine scale population structure along the three most populous regions of Brazil and demonstrated they are not as uniform as initially reported. Nevertheless, Brazilians altogether are highly heterogeneous even when compared with other admixed populations. The variability seems to be owed to European and, in a greater extent, to African genetic contribution. This is reflected in smaller haplotypes sizes and relaxed general linkage disequilibrium, which, in turn, has impact in association studies in Brazilians as in any other highly admixed population. Association and linkage studies rely on the linkage between causal variants and their markers, meaning that haplotypes variability and size have direct impact on how the tests are designed and conducted. Moreover, genetic association is expected to display low correlation between populations too discrepant in genetic structure.

Since it is still not fully ascertained how variants linked to complex traits will manifest across different genetic backgrounds, the present work provides valuable information through the description of a highly admixed population in a detailed fashion, as well as by demonstrating the diversity within this population. Further studies are encouraged to comprehend how association, linkage and meta-analysis studies are better carried out for assigning impact on traits for different human populations.

MATERIALS AND METHODS

Subjects. The present research followed the principles of the Declaration of Helsinki, international and Brazilian ethical guidelines for biomedical research involving human subjects. The research protocol was approved by the Ethics Committee of the State University of Campinas Medical School (Campinas, São Paulo State, Brazil). All patients received the materials about the study and signed an informed consent.

DNA samples of four Brazilian groups of individuals were analyzed in the present study: (i) control cohort of the project “Copy number variations (CNVs) in the assessment of stroke susceptibility in patients with sickle cell anemia”, N=45, non-carriers of the sickle cell mutation; (ii)

control cohort of the project “High-density micro-array technique in the assessment of Copy Number Variation in Congenital Defects of Complex Inheritance: Oral Clefts as a Model”, N=20; (iii) control cohort of the project “Joinville Stroke Registry, Joinville, SC, Brazil”, N=26; and (iv) the participants of the project “Identification of susceptibility genes for squamous cell carcinoma of the tongue by large scale genotyping”, N=98. Geographically sampling process occurred in three Brazilian cities: Recife (1,617,183 inhabitants; Pernambuco State), Campinas (1,164,098 inhabitants; São Paulo State), and Joinville (562,151 inhabitants; Santa Catarina State) (1).

Besides the Brazilian samples, we also used publicly available data from the HapMap International Project and 1000 Genomes Project (13, 14). HapMap raw data of the samples genotyped with *Genome Wide SNP Arrays 6.0* from Affymetrix (15)(6) was downloaded in the Affymetrix *.cel format (14, 16). We obtained 11 HapMap populations: ASW, CEU, CHB, CHD, GIH, JPT, LWK, MEX, MKK, TSI, YRI; detailed description of these populations can be found on the HapMap Homepage (14, 17). Four populations, CLM, FIN, GBR, PUR, were acquired from the database of 1000 Genomes Project, where detailed description of the populations and samples can also be found (13, 18) in the form of individual’s genotype data. All populations can be conditionally divided into several regional groups according to their ancestry: (EUR) European – CEU, FIN, GBR, TSI; (AFR) African – ASW, LWK, MKK, YRI; (EAS) East Asian – CHB, CHD, JPT; (SAS) South Asian – GIH; and (ALA) Admixed Latin American – CLM, MEX, PUR, and Brazilian population (*hereinafter referred to as “BR”*), as shown on Table 1.

Samples genotyped with *Genome Wide SNP Arrays 5.0* were used only once in our analyses, to estimate linkage disequilibrium (LD) decay. They were pooled together with other Brazilian samples and the common SNPs were extracted. All other analyses involved only the samples genotyped with *Genome Wide SNP Arrays 6.0*.

Genotyping. We collected 5 ml of peripheral blood of each participant and extracted the whole genomic DNA with QIAamp DNA Blood Midi Kit - Qiagen, Hilden, Germany.

We used two genotyping platforms: *Genome Wide SNP Arrays 6.0* and *5.0* from Affymetrix (Affymetrix Inc, Santa Clara, CA), interrogating 906,600 and 500,568 polymorphic markers, respectively. Assay protocols are essentially the same for the two array platforms and can be consulted on the manufacturer homepage (19–22). Preparing our DNA samples, we rigorously followed the manufacturer instructions. Finally, the samples were hybridized onto

the chips, stained and scanned. Genotypes were generated using Affymetrix Genotyping Console software (GTC) version 4.1.3 with default settings. We applied the implemented algorithms: BRLMM that uses perfect-match probes (BRLMM-P) algorithm for the Affymetrix Genome-Wide SNP Array 5.0 and Birdseed v2 algorithm for the SNP Array 6.0. More detailed information on these algorithms can be found elsewhere. The assembly version of human genome used for genotyping was GRCh37/hg19.

Quality Control (QC) of genotypes. Basic manipulation with row genotypes data and the basic quality control procedure was performed with the aid of the PLINK software (versions 1.07 and 1.9) (23–26). Each population was examined at two levels: per-marker and per-individual (27).

Sample quality control. Each individual sample was checked for the discordance in relation to the sex register, outlying missing genotype call rate; presence of duplicated or genetically related samples. Relatedness was estimated with the use of PLINK software. The software permits calculate genome-wide identity-by-descent (IBD) based on identity-by-state IBS information observed at a given locus (two different alleles, one or two common alleles). For this analysis, we first extracted a subset of SNPs in approximate linkage equilibrium ($r^2 < 0.5$), applying the implementation on PLINK, namely linkage disequilibrium (LD) pruning procedure. The extracted set of markers was used in IBD estimation. We removed one individual of each pair with $IBD > 0.1875$ (18); samples with $IBD > 0.98$ were considered as duplicated.

Autosomal SNPs. Only autosomal SNPs were included in our analysis.

Marker quality control. For each population, we removed markers with too high missing genotypes (> 0.05); those that demonstrated deviation from Hardy-Weinberg equilibrium (HWE), $p\text{-value} < 1e\text{-}6$; and SNPs with low minor allele frequency (MAF) < 0.05 . Also, we composed a list of SNPs for which at least one Mendelian inconsistency was observed in the populations that had trios' information; afterwards, this SNPs were filtered out from the analysis.

Merging datasets. All datasets in use were based on the same GRCh37/hg19 assembly version of human genome. Merging datasets we always extracted only common SNPs.

LD decay. We compared LD in the Brazilian populations to the populations of European, African, and Latino American ancestry. LD-decay was measured as a function of physical distance and was estimated using resources of PLINK software. We calculated squared

correlation coefficient (r^2) for each pair of SNPs within 100 kb window. The results were plotted with our custom R script (28). All SNP pairs were divided into bins of 1 kb size, and for each bin the mean r^2 was plotted.

Population structure. To explore population stratification, we applied several analytical tools: FineStructure version 2 (29), and ADMIXTURE version 1.3 (30) programs. Detailed description of the underlying methodology can be found in the mentioned above references. In brief, FineStructure captures information provided by patterns of haplotype similarity and summarizes it as a co-ancestry matrix. We used the linked model which takes advantage of the information enclosed in the linkage disequilibrium data and, accordingly to the authors, outperforms the unlinked analysis (29). ADMIXTURE software clusters individuals on the bases of their allelic profile using differences in allelic frequencies; so that SNPs in linkage disequilibrium can have a confounding effect. Therefore, marker data set used with the ADMIXTURE program was previously thinned to remove SNPs in LD. We also removed known regions involved in long range LD in human populations (31). ADMIXTURE employs a model-based manner, namely the likelihood model, to estimate ancestry. First, we identified the optimal value of K and then we performed ten independent runs with different seeds. Afterward, ADMIXTURE output files were analyzed with CLUMPP software version 1.1.2 (32) to identify common modes among replicates. CLUMPP (Cluster Matching and Permutation Program) permutes the clusters output of independent runs and outputs the best permuted matrix. For $K=7$, we used *Greedy* algorithm with 100 random input orders to be tested and G' pairwise matrix similarity statistics. The output clustering matrix was plotted applying in-house-made R script.

As already mentioned, FineStructure software explores the haplotype structure and takes as input only haplotype data. Accordingly, all our samples were simultaneously phased with the aid of SHAPEIT program version 2 (33, 34). We used only marker data set that passed the quality control. Phasing was performed with the reference panel: “1,000 Genomes haplotypes – Phase I integrated variant set release (SHAPEIT2) in NCBI build 37 (hg19) coordinates”, the version with haplotypes with singleton sites removed (33, 35). Besides, to work in the framework of “linked model”, FineStructure requires a recombination map; we used `genetic_map_HapMapII_GRCh37.tar.gz` (36).

FUNDING

The work was supported by Brazilian Synchrotron Light Laboratory; São Paulo State Research Foundation [2008/57441-0, 2012/06438-5]; and the Counsel of Technological and Scientific Development [8367/2011-1, 150398/2013-1].

ACKNOWLEDGEMENTS

We would like to thank all the volunteers who participated in this study. We acknowledge François Artiguenave, Cristiane S. Rocha and Murilo G. Borges for the helpful comments, Maria E. R. Camargo and Daniela Stancato for their assistance in the experiment preparation.

Conflict of Interest Statement. *None declared.*

REFERENCES

1. IBGE - Instituto Brasileiro de Geografia e Estatística Estimativas de população para 1º de julho de 2015. [Internet]. [cited 2016 Jan 26]. Available from: http://ibge.gov.br/home/estatistica/populacao/estimativa2015/estimativa_tcu.shtm
2. IBGE (2007) Brasil - 500 anos de povoamento Revised edition. IBGE - Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro. [cited 2016 Jan 26]. 237 p. Available from: <http://biblioteca.ibge.gov.br/visualizacao/livros/liv6687.pdf>
3. Carvalho-Silva,D.R., Santos,F.R., Rocha,J. and Pena,S.D. (2001) The phylogeography of Brazilian Y-chromosome lineages. *Am. J. Hum. Genet.*, **68**, 281–286.
4. Pena,S.D.J., Di Pietro,G., Fuchshuber-Moraes,M., Genro,J.P., Hutz,M.H., Kehdy,F. de S.G., Kohlrausch,F., Magno,L.A.V., Montenegro,R.C., Moraes,M.O., et al. (2011) The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PloS One*, **6**, e17063.
5. Alves-Silva,J., da Silva Santos,M., Guimarães,P.E., Ferreira,A.C., Bandelt,H.J., Pena,S.D. and Prado,V.F. (2000) The ancestry of Brazilian mtDNA lineages. *Am. J. Hum. Genet.*, **67**, 444–461.
6. Pena,S.D.J., Bastos-Rodrigues,L., Pimenta,J.R. and Bydlowski,S.P. (2009) DNA tests probe the genomic ancestry of Brazilians. *Braz. J. Med. Biol. Res. Rev. Bras. Pesqui. Médicas E Biológicas Soc. Bras. Biofísica Al*, **42**, 870–876.
7. Giolo,S.R., Soler,J.M.P., Greenway,S.C., Almeida,M.A.A., de Andrade,M., Seidman,J.G., Seidman,C.E., Krieger,J.E. and Pereira,A.C. (2012) Brazilian urban population genetic structure reveals a high degree of admixture. *Eur. J. Hum. Genet. EJHG*, **20**, 111–116.
8. Pena,S.D.J., Di Pietro,G., Fuchshuber-Moraes,M., Genro,J.P., Hutz,M.H., Kehdy,F. de S.G., Kohlrausch,F., Magno,L.A.V., Montenegro,R.C., Moraes,M.O., et al. (2011) The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PloS One*, **6**, e17063.
9. Reich,D.E., Cargill,M., Bolk,S., Ireland,J., Sabeti,P.C., Richter,D.J., Lavery,T., Kouyoumjian,R., Farhadian,S.F., Ward,R., et al. (2001) Linkage disequilibrium in the human genome. *Nature*, **411**, 199–204.

10. Silva-Zolezzi,I., Hidalgo-Miranda,A., Estrada-Gil,J., Fernandez-Lopez,J.C., Uribe-Figueroa,L., Contreras,A., Balam-Ortiz,E., del Bosque-Plata,L., Velazquez-Fernandez,D., Lara,C., et al. (2009) Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 8611–8616.
11. 1000 Genomes Project Consortium, Auton,A., Brooks,L.D., Durbin,R.M., Garrison,E.P., Kang,H.M., Korbelt,J.O., Marchini,J.L., McCarthy,S., McVean,G.A., et al. (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.
12. Gonçalves,V.F., Carvalho,C.M.B., Bortolini,M.C., Bydlowski,S.P. and Pena,S.D.J. (2008) The phylogeography of African Brazilians. *Hum. Hered.*, **65**, 23–32.
13. 1000 Genomes Project Consortium 1000 Genomes Homepage. [Internet]. [cited 2016 Jan 26]. Available from: <http://www.1000genomes.org/>
14. International HapMap Project Consortium HapMap Homepage. [Internet]. [cited 2016 Jan 26]. Available from: <http://hapmap.ncbi.nlm.nih.gov/>
15. Affymetrix Inc, Santa Clara, CA. Affymetrix: Genetic Analysis: SNP Analysis. [Internet]. [cited 2016 Feb 5]. Available from: http://www.affymetrix.com/estore/catalog/131533/AFFY/Genome-Wide+Human+SNP+Array+6.0#1_1
16. International HapMap Project Consortium Index of /downloads/raw_data. [Internet]. [cited 2016 Feb 5]. Available from: http://hapmap.ncbi.nlm.nih.gov/downloads/raw_data/?N=D
17. International HapMap Project Consortium HapMap: Documentation: Samples/Individuals: Phase 3 Summary Table. Summary Table [Internet]. [cited 2016 Feb 5]. Available from:<http://hapmap.ncbi.nlm.nih.gov/downloads/>
18. 1000 Genomes Project Consortium 1000 Genomes: Data: Sample lists. [Internet]. [cited 2016 Feb 5]. Available from: http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20130606_sample_info/20130606_sample_info.xlsx
19. Affymetrix Inc, Santa Clara, CA. Affymetrix - Help - FAQ - Genome-Wide Human SNP

- Array 5.0 Index. [Internet]. [cited 2016 Jan 26]. Available from: http://www.affymetrix.com/support/help/faqs/gw_human_snp5/index.jsp
20. Affymetrix Inc, Santa Clara, CA. Affymetrix - Help - FAQ - Genome-Wide Human SNP Array 6.0 Index. [Internet]. [cited 2016 Jan 26]. Available from: http://www.affymetrix.com/support/help/faqs/gw_human_snp6/index.jsp
 21. Affymetrix Inc, Santa Clara, CA. Affymetrix - Help - FAQ - How many SNPs and Copy Number Probes are on the SNP 6.0 array? [Internet]. [cited 2016 Jan 26]. Available from: http://www.affymetrix.com/support/help/faqs/gw_human_snp6/faq_20.jsp
 22. Affymetrix Inc, Santa Clara, CA. Affymetrix - Help - FAQ - What is the content of the SNP5.0 array? [Internet]. [cited 2016 Jan 26]. Available from: http://www.affymetrix.com/support/help/faqs/gw_human_snp5/faq_2.jsp
 23. Purcell,S., Neale,B., Todd-Brown,K., Thomas,L., Ferreira,M.A.R., Bender,D., Maller,J., Sklar,P., de Bakker,P.I.W., Daly,M.J., et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
 24. Purcell,S. PLINK: Whole genome data analysis toolset. Packages: PLINK 1.07 and PLINK 1.9.
 25. Purcell,S. and Chang,C. (2015) PLINK. [Internet]. [cited 2016 Jan 26]. Available from: <http://pngu.mgh.harvard.edu/~purcell/plink/>
 26. Chang,C.C., Chow,C.C., Tellier,L.C., Vattikuti,S., Purcell,S.M. and Lee,J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, **4**, 7.
 27. Anderson,C.A., Pettersson,F.H., Clarke,G.M., Cardon,L.R., Morris,A.P. and Zondervan,K.T. (2010) Data quality control in genetic case-control association studies. *Nat. Protoc.*, **5**, 1564–1573.
 28. R Core Team (2014) R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria. [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2014. Available from: <http://www.R-project.org/>
 29. Lawson,D.J., Hellenthal,G., Myers,S. and Falush,D. (2012) Inference of population structure using dense haplotype data. *PLoS Genet.*, **8**, e1002453.

30. Alexander,D.H., Novembre,J. and Lange,K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.*, **19**, 1655–1664.
31. Price,A.L., Weale,M.E., Patterson,N., Myers,S.R., Need,A.C., Shianna,K.V., Ge,D., Rotter,J.I., Torres,E., Taylor,K.D., et al. (2008) Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.*, **83**, 132-135-139.
32. Jakobsson,M. and Rosenberg,N.A. (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinforma. Oxf. Engl.*, **23**, 1801–1806.
33. Delaneau,O., Marchini,J. and The 1000 Genomes Project Consortium (2014) Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat. Commun.*, **5**, 3934.
34. O’Connell,J., Gurdasani,D., Delaneau,O., Pirastu,N., Ulivi,S., Cocca,M., Traglia,M., Huang,J., Huffman,J.E., Rudan,I., et al. (2014) A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet.*, **10**, e1004234.
35. 1000 Genomes Project Consortium 1000GP Phase 1 haplotypes. Release: 9 Dec 2013. Update: 16 Jun 2014.
36. International HapMap Project Consortium Index of /downloads/recombination/2011-01_phaseII_B37.

Table 1. Populations' description.

Populations' labels	Description
CEU	Utah residents with Northern and Western European ancestry from the CEPH collection
FIN	Finnish in Finland
GBR	British in England and Scotland
TSI	Toscani in Italia
ASW	Americans of African Ancestry in SW USA
LWK	Luhya in Webuye, Kenya
MKK	Maasai in Kinyawa, Kenya
YRI	Yoruba in Ibadan, Nigeria
CHB	Han Chinese in Beijing, China
CHD	Chinese in Metropolitan Denver, Colorado
JPT	Japanese in Tokyo, Japan
GIH	Gujarati Indians in Houston, Texas
CLM	Colombians from Medellin, Colombia
MEX	Mexican Ancestry from Los Angeles USA
PUR	Puerto Ricans from Puerto Rico
BR	Brazilians from Pernambuco, São Paulo and Santa Catarina; Brazil

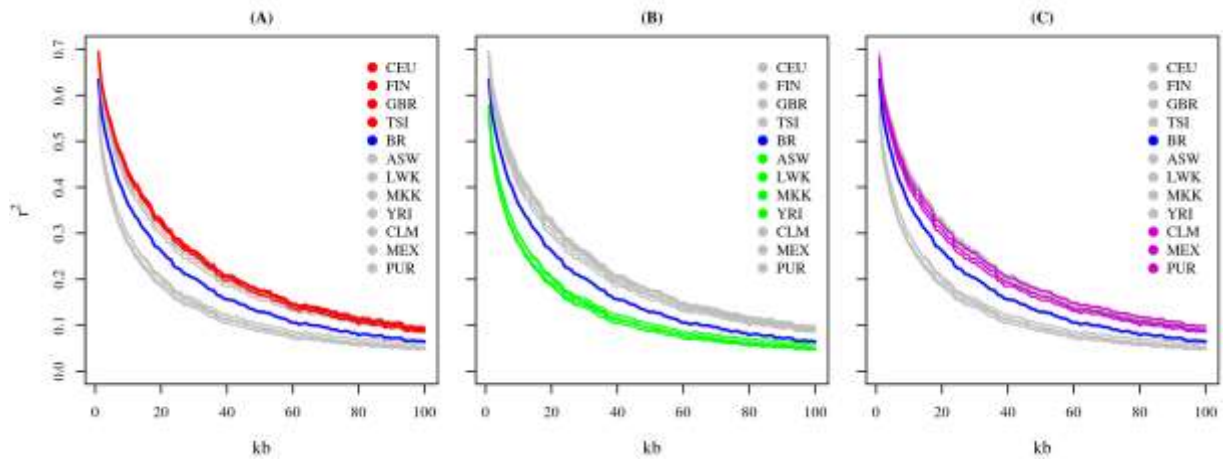


Figure 1. Linkage disequilibrium (r^2) estimated with PLINK software, v. 1.9, and plotted by population as a function of physical distance (kb). (A) LD plot of the Brazilian population (blue line) versus populations of European (red lines) ancestry; (B) Brazilian population (blue line) versus populations of African (green lines) ancestry; (C) Brazilian population (blue line) versus Latin American populations (pink lines).

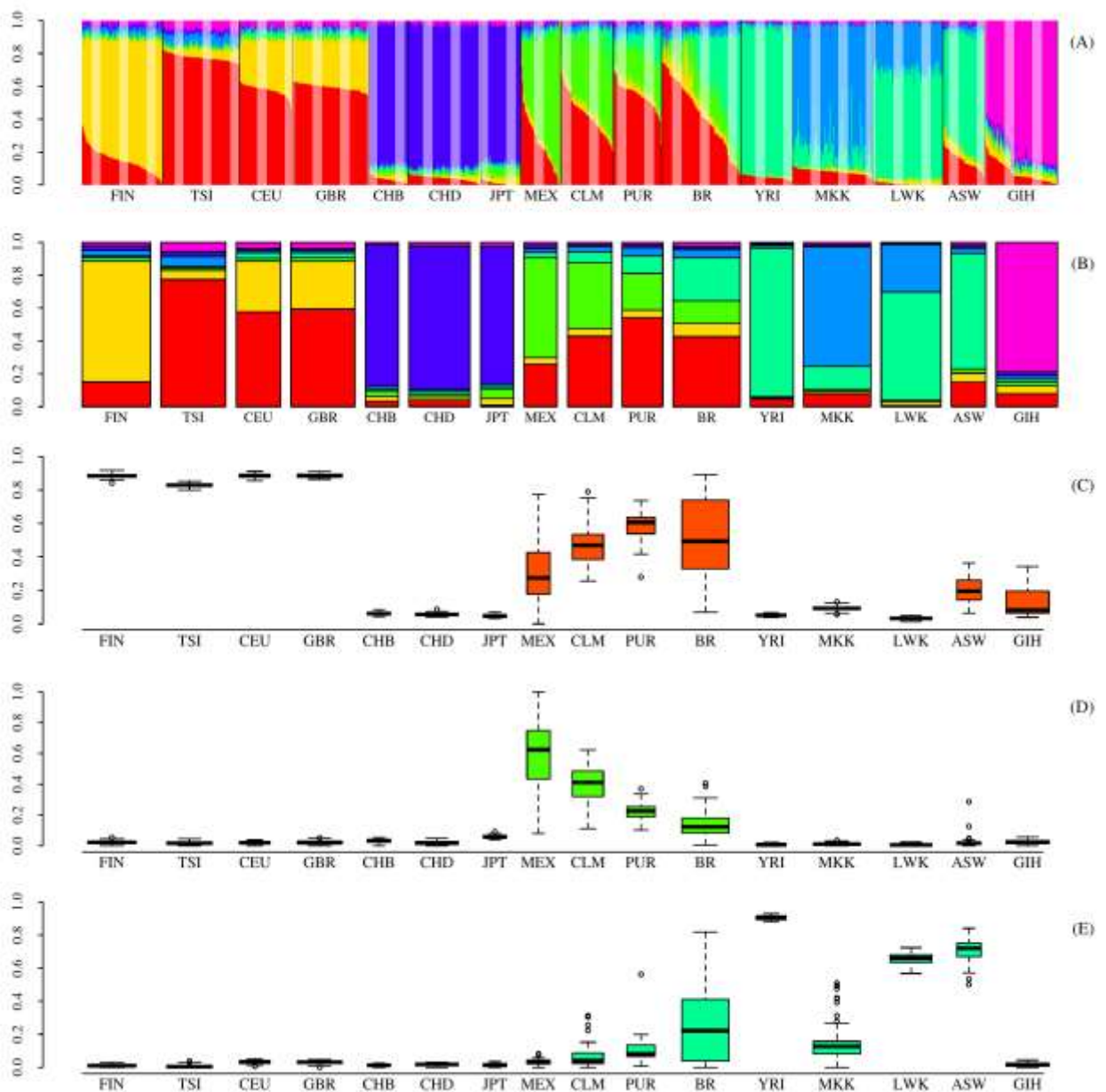


Figure 2. ADMIXTURE plot, $K=7$. (A) Individuals are represented by vertical bars partitioned into segments; the length of the segments is proportional to the estimated membership in the clusters. (B) Bar plot of the estimated membership in the clusters averaged over each population. Box plots of estimated components of membership in different clusters: (C) European membership component; (D) Yorubian component; (E) Amerindian component. BR stands for the Brazilian population. Other populations maintain the same initials as in HapMap and 1000 genomes projects.

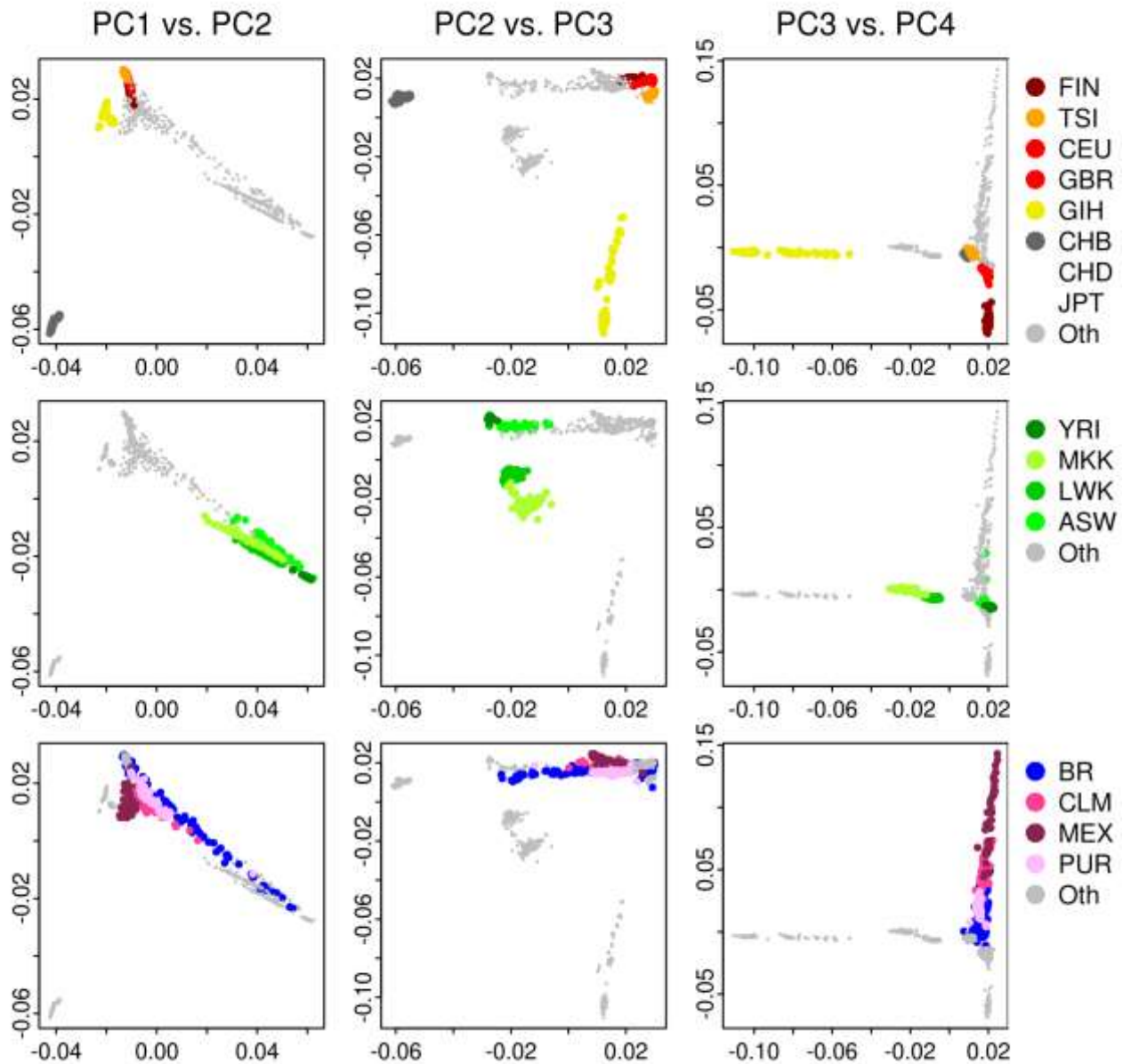


Figure 3. Plot of the selected Principal Components. Each row of the figure highlights the particular group of populations. The first row shows the extended group consisting of the Euroasian populations. The second row highlights populations of African ancestry, while the third row emphasizes Latino-American populations.

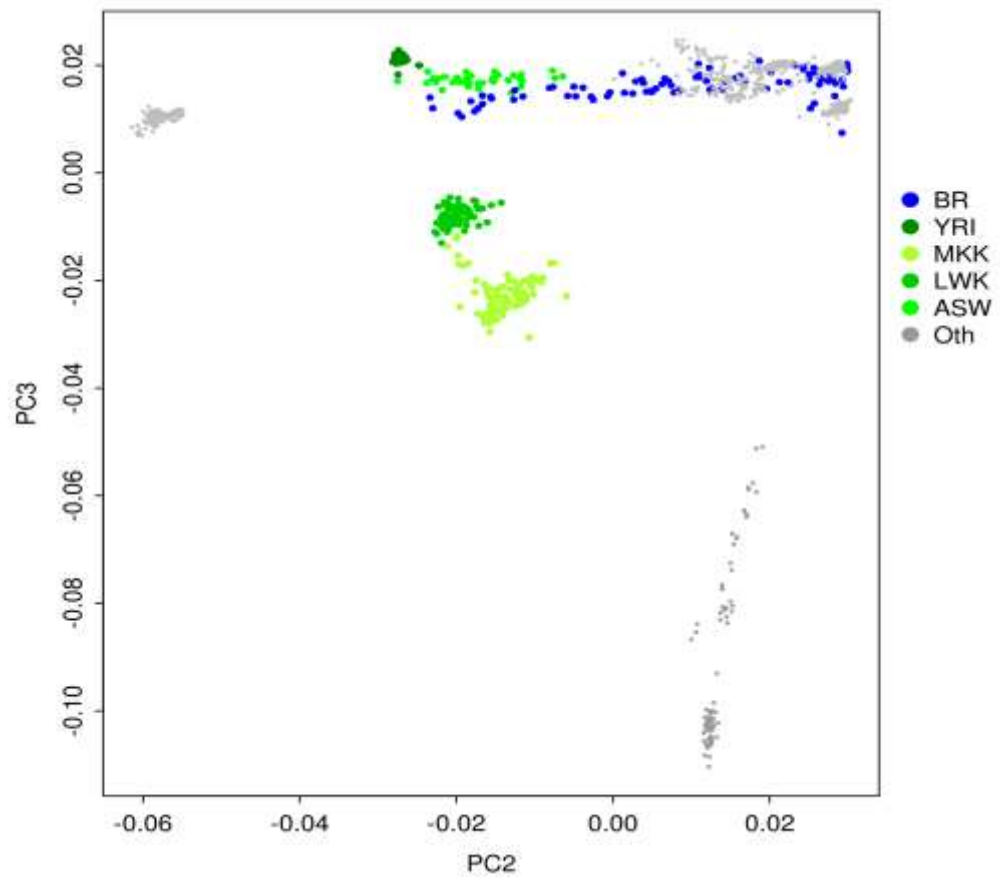


Figure 4. PC2 vs.PC3 plot. The plot highlights relationship between the Brazilian population and the populations of African origin. The plot demonstrates divergence between ASW and Brazilian populations.

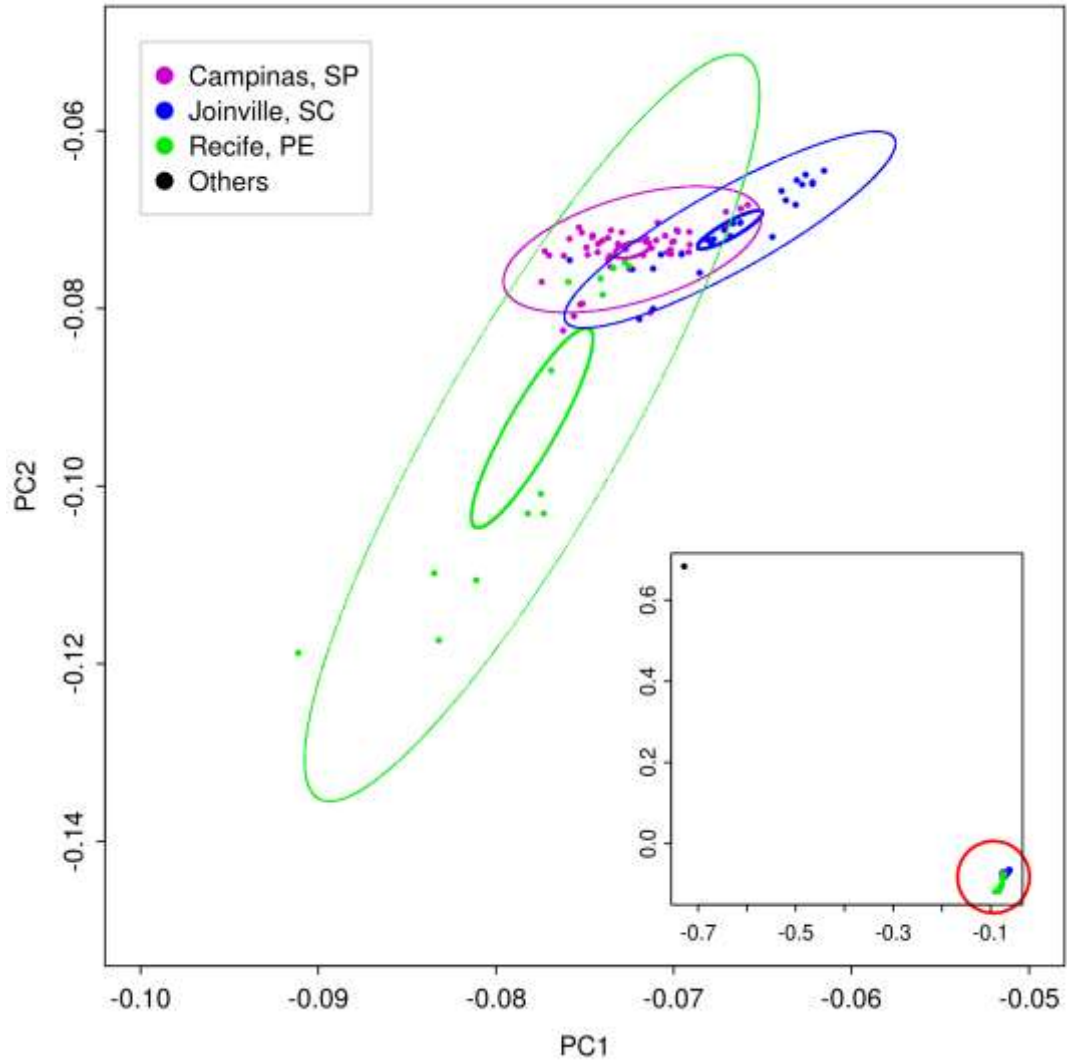


Figure 5. PC1 vs. PC2 plot stratifying Brazilians by site of collection: Campinas, Joinville and Recife. The plot shows differential distribution among regions.

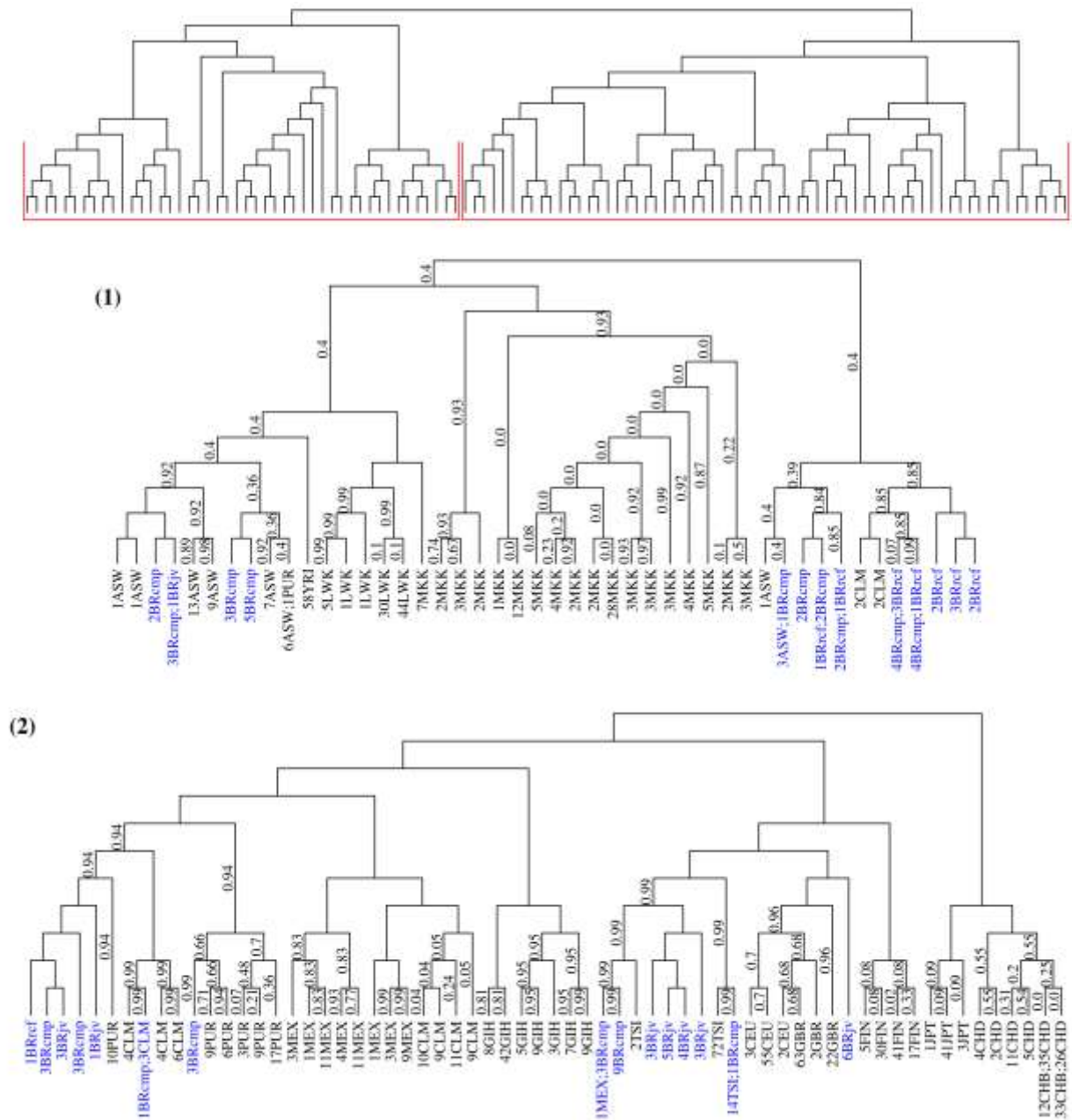
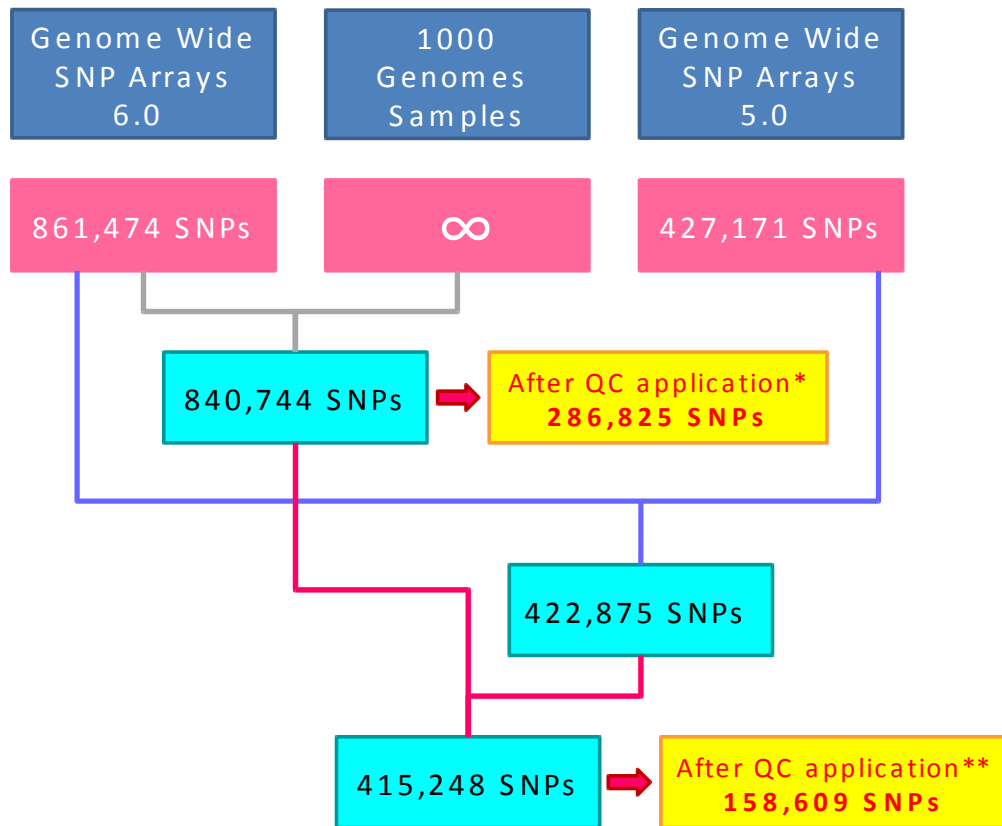


Figure 6. Dendrogram of relationship between populations. Top figure represents overall structure of the dendrogram. (1) zoom in the leftmost cluster of the dendrogram; (2) Zoom in of the rightmost cluster. Posterior probability values below 1 are shown as edge labels. Leaf labels shows population membership and number of individuals in the particular group. Blue color highlights leaves with Brazilian individuals participation.

Supplemental figures and legends



*Mendelian errors, low frequency , poorly genotyped SNPs, and HWE inconsistent SNPs

**Mendelian errors, low frequency and poorly genotyped SNPs

Figure S1. The scheme depicts the numbers of SNPs genotyped with Genome Wide SNP Arrays 6.0, Genome Wide SNP Arrays 5.0, and obtained from 1000 Genomes Project samples. It also presents the numbers of SNPs after merging and QC pruning.

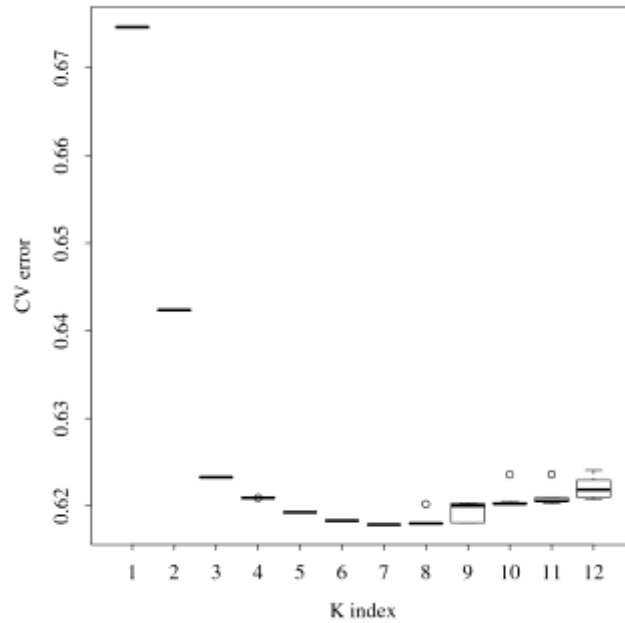


Figure S2. Box plot of cross validation error estimated over K from 1 to 12 in 10 ADMIXTURE runs with different random seeds. CV error reaches the minimum (0.6178) at K=7.

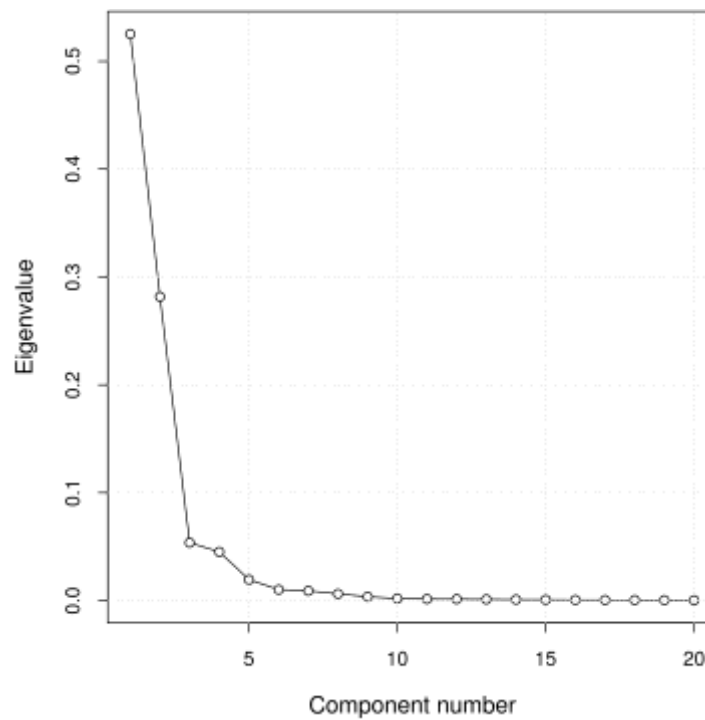


Figure S3. Screen plot of the scaled eigenvalues; the first 20 values are shown. One can observe that the eigenvalues are quickly drop down after the fourth value, therein the first four values explain 90.6% of the variation.

ABBREVIATIONS

LD: linkage disequilibrium

QC: quality control

R^2 : LD coefficient

HWE: Hardy-Weinberg Equilibrium

mtDNA: mitochondrial DNA

CAPÍTULO III**COMPARISON BETWEEN TWO DIFFERENT SICKLE CELL ANEMIA
COHORTS: HIGH DIVERGENCE BETWEEN BRAZILIANS AND AMERICANS**

A seção que se segue constitui o manuscrito que será submetido como artigo original ao periódico “Human Molecular Genetics”.

Resumo

A anemia falciforme (AF) é uma condição hematológica causada pela homoziguidade de uma mutação de ponto no gene da beta globina, resultando em uma forma anômala de hemoglobina (hemoglobina S). AF é caracterizada principalmente por anemia crônica e crises vaso oclusivas dolorosas, contudo, a gravidade é consideravelmente variável entre indivíduos. Diversos moduladores genéticos foram associados à heterogeneidade clínica da doença. Expressão de globina gama é considerada um alvo terapêutico promissor, dado que a hemoglobina fetal (HbF) geralmente apresenta impacto benéfico no fenótipo. A produção de HbF está ligada aos haplótipos do *cluster* da globina beta (HBB), que por sua vez têm origem em diferentes populações. Propusemos o estudo de duas coortes derivadas do Brasil (Pernambuco e São Paulo) e Estados Unidos da América (Pensilvânia) quanto à ancestralidade genômica e local, referente ao cromossomo 11. O *software* ADMIXTURE foi aplicado utilizando 178,159 marcadores e mostrou que as proporções de ancestralidade variam marcadamente em indivíduos brasileiros, mais heterogêneos, em relação aos indivíduos norte-americanos. Quanto à ancestralidade no cromossomo 11, norte-americanos afetados por AF apresentaram médias de 76%, 18.2% e 5.8% das ancestralidades africana, europeia e ameríndia, respectivamente. As médias para os pacientes de AF brasileiros foram 43.9%, 39.4% e 16.7%, respectivamente. No presente trabalho foi demonstrado que indivíduos falciformes das duas nacionalidades amostradas divergem consideravelmente no plano genômico e no cromossomo 11 em relação às proporções de componentes ancestrais.

**COMPARISON BETWEEN TWO DIFFERENT SICKLE CELL ANEMIA
COHORTS: HIGH DIVERGENCE BETWEEN BRAZILIANS AND AMERICANS**

Cruz PRS¹, Ananina G¹, Lopes-Cendes IT², Simioni M², Gil-da-Silva-Lopes VL², França PHC³, Cabral NL³, Donatti A³, Secolin R², Bezerra MAC⁴, Araújo AS⁵, Hakonarson H⁶, Costa FF⁷, Melo MB¹

1. Laboratory of Human Genetics, Center for Molecular Biology and Genetic Engineering (CBMEG), University of Campinas – UNICAMP, Campinas, SP, Brazil.
2. Department of Medical Genetics, School of Medical Science, University of Campinas - UNICAMP. The Brazilian Institute of Neuroscience and Neurotechnology (BRAINN), Campinas, SP, BRAZIL.
3. Department of Medicine, Regional University of Joinville - UNIVILLE, Joinville Stroke Registry, Joinville, SC, Brazil.3. Department of Clinical Medicine, University of Campinas – UNICAMP, Campinas, SP,, Brazil.
4. Department of Genetics, Federal University of Pernambuco, Recife, PE, Brazil.
5. Hematology and Haemotherapy Foundation of Pernambuco – HEMOPE, Recife, PE, Brazil.
6. Center for Applied Genomics, Abramson Research Center, Children’s Hospital of Philadelphia, Philadelphia, PA 19104, USA.
7. Hematology and Hemotherapy Center, University of Campinas – UNICAMP, Campinas, São Paulo, Brazil

Correspondence to: Mônica Barbosa de Melo, Laboratory of Human Molecular Genetics, Center of Molecular Biology and Genetic Engineering (CBMEG), University of Campinas - UNICAMP, P.O.Box: 6010, Campinas, SP, Brazil, 13083-875.

ABSTRACT

Sickle cell anemia (SCA) is a blood disorder caused by the homozygous state of a single point mutation in the beta globin gene resulting in an abnormal form of hemoglobin (HbS). SCA is characterized primarily by chronic anemia and painful vaso-occlusive episodes. However, severity varies considerably among individuals. Several genetic modulators have been proposed to affect the clinical outcome. Expression of gamma-globin, mapped to chromosome 11, is regarded as a promising therapeutic target, since fetal hemoglobin (HbF) generally have beneficial impact on the symptoms. Production of HbF is tightly connected to the haplotypes of the beta-globin gene locus (HBB) that, in turn, are well known to have arisen in different human populations. Here we propose to compare SCA patients of urban populations from Brazil (Pernambuco, São Paulo states) and the United States (Pennsylvania) regarding general and local ancestry on the chromosome 11, where the HBB cluster is located. ADMIXTURE method on 178,159 markers presented different proportions of European and African ancestries for these two cohorts, and Brazilians displayed much more variability. On the level of chromosome 11, American SCA were calculated to have average 76% of African ancestry, while we observed averages of 18.2% and 5.8% for European and Native American components, respectively. The average proportions for Brazilians were 43.9% (African), 39.4% (European) and 16.7% (Native American) on the same chromosome. In the present work we found that, beyond the African component, SCA patients from Brazil are different in proportions of general continental ancestries compared to American SCA patients both genome-wise and, especially, locally on the chromosome 11.

INTRODUCTION

Sickle cell anemia (SCA) is the first monogenic disease ever described, being responsible for the coining of the Pauling's concept of "molecular disease". Although resulting from a point mutation located in the chromosome 11, in the beta-globin gene cluster (HBB), at the phenotypic level, sickle cell anemia is not a monogenic disease; it is multigenic one. This mutation causes red blood cells (RBCs) to acquire a sickle shape under conditions of hypoxia. The large range of phenotypes encompasses anemia, cell adhesion, vaso-occlusion, stroke, severe pain and organ failure. Besides the beta-globin locus, there is a number of pleiotropic genes (involved in pathophysiologic events secondary to the primary disease, or complications) and epistatic genes that altogether account to the course of the disease (DRISS, A., ASARE, K.O. J.M. HIBBERT¹, B.E. GEE², T.V. ADAMKIEWICZ³, 2008; NAGEL, 2001). To date, over 100 SNPs that might be associated with specific subphenotypes are mapped to genes that act as inflammatory mediators (BELCHER et al., 2006; BOWERS et al., 2013; HEBBEL; OSAROGIAGBON; KAUL, 2004), modulators of oxidant injury, Nitric Oxide (NO) biology (DE CATERINA et al., 1995; ELAGOUZ et al., 2010; KATO et al., 2006; KATO; GLADWIN; STEINBERG, 2007; PAWLOSKI; HESS; STAMLER, 2005), vaso-regulatory molecules (ADEWOYE et al., 2006; BALDWIN et al., 2005; FLANAGAN et al., 2011; NOLAN et al., 2005) and cell adhesion factors (ASAHARA et al., 1998; BLANN et al., 2007; CARLSON et al., 2004; GAO et al., 2001; ZHANG et al., 2006).

Special interest have been driven to fetal hemoglobin (HbF) levels, since it has in most cases an ameliorative effect on SCA patients, as the gamma-chains of HbF do not polymerize with the sickle hemoglobin (HbS). Among the several genetic determinants shown to contribute to the baseline HbF levels, the HBB haplotypes are the most well established (NAGEL; LABIE, 1989). High HbF levels are associated with Senegal and Asian-Indian haplotypes, which generate a milder hematological and clinical phenotype compared with the Benin, Bantu and Cameroon haplotypes. Individuals with Bantu haplotypes presents the most severe phenotype, with the lowest HbF level (STEINBERG; SEBASTIANI, 2012), while individuals with the Benin haplotype usually have intermediate features (STEINBERG; SEBASTIANI, 2012). SCA patients from the Southwestern Province of Saudi Arabia have HbF levels twice as high as African Americans despite having similar HBB haplotypes (ALSULTAN et al., 2011), while patients from the Eastern Province have even higher levels of HbF (AKINSHEYE et al., 2011).

Creary et al (CREARY et al., 2009) have reported an association between European ancestry and the proportion of erythrocytes containing HbF, although Solovieff et al(SOLOVIEFF et al., 2010b) did not find an association between fetal hemoglobin concentration and European ancestry.

Despite the importance of ancestry and its clear effect on the course of sickle cell disease, there are still few studies on the ancestry composition of SCA cohorts outside the United States. In the last few years we have observed an increasing awareness of recent admixing process as a tool for studying variation in phenotypes among human populations (BARAN et al., 2012; BRYC et al., 2010b; KEHDY et al., 2015; MONTINARO et al., 2015; REICH et al., 2005; TANG et al., 2007; WINKLER; NELSON; SMITH, 2010; ZHU; TANG; RISCH, 2008). Still, to date, only one population genetics report is dedicated to the Brazilian SCA cohort. Silva et al. (DA SILVA et al., 2011) found this population to be highly heterogeneous and depicts a surprising 13.8% proportion of patients with predominant European ancestry. Here we compare the SCA patients from the Childrens' Hospital of Philadelphia (USA), Hematology-Hemotherapy Center in Campinas (Brazil) and Hematology and Hemotherapy Foundation of Pernambuco (Brazil) on genomic level, as well as regarding local ancestry on the chromosome 11, aiming to describe ancestral composition differences that might be explored regarding clinical variance.

RESULTS

We evaluated the ancestry by performing genomic and local ancestry on SCA patients. Complementary data of general population for Brazilians (collected across the three major geographical Brazilian regions) and African descendants from Philadelphia, collected at the same center of the SCA patients. Their genomic data is summarized in the 178,159 SNPs-based PCA, depicted on Figure 1. As expected, Brazilians behave widely heterogeneously, and are distributed between European and African clusters, a pattern demonstrated before (GIOLO et al., 2012; SANTOS et al., 2015). Sickle cell anemia patients from Brazil, on the other hand, do not vary at the same extent as general population, although distributes roughly the same. African-Americans, also regarded as an admixed population, are far more uniformly scattered, and clusters together with SCA.

We further looked to the general ancestral composition by simulating different components (K) on ADMIXTURE software (Figure 2). SCA subjects from US and from Brazil have mainly European and African ancestry, though varying in proportions, especially for the South-European component. Brazilian cosmopolitan reference shows striking heterogeneity, being very close to its SCA counterpart, except that the latter seems to be scarcely composed of individuals more of European than African ancestry while African Americans and SCA are not visually distinguishable by this analysis, and both have little European contribution into their genomes compared to Brazilians. One would expect that African Americans behave closely to SCA in the US, especially for being collected in the same locality, but Brazilian SCA, also considered blacks/mulattos, diverge a larger degree from the latter in the ancestral composition than one would suppose them to do. That is a clear evidence of the higher degree of admixture in Brazilians in the past few generations. We provide a Hudson's F_{st} as a measure of genetic distance between populations (Table 2), showing greater proximity between Brazilian SCA and European populations than the American SCA and Europeans. Nonetheless, sickle cell anemia patients are closer to each other ($F_{st} = .018$) than to Europeans (.033, .028 and .031 for Brazilians; .085, .082 and .083 for Americans), yet maybe not as evidently as expected.

Moreover, we investigated local ancestry on the chromosome 11, where the HBB cluster is located, responsible for both defective beta-globin and beneficial gamma-globin genes. Here we show ten randomly selected patients from each population (Figure 3), and although both are highly variable in composition, Americans typically show more than 50% of African haplotypes, differently from Brazilians. For the sake of veracity, a three-way admixing scenario was considered for both populations, and we computed the average contribution of the three inferred ancestral populations to the SCA patients in Brazil and in the US. While American 11 chromosomes are mainly constituted of African haplotypes (76%), Brazilians had a more balanced constitution with 43.9% of African and 39.3% of European haplotype inferred from phased data. Not only SCA patients from US and Brazil varied in European haplotypic blocks (18.2% vs. 39.3%, respectively), as they also diverged in supposedly Native American proportions (5.7% vs 16.7%, respectively), although it is important to note that when using East Asian reference data as proxy of ancestral Native American we end up inflating the estimative of haplotypes contribution of that particular population (BARAN et al., 2012).

DISCUSSION

Although it is well recognized that 85%–95% of human genetic variation is attributable to variation within a population, while 5%–15% is due to variation among populations (BARBUJANI et al., 1997; ROSENBERG et al., 2002) it is still uncertain whether similar levels of within-versus among- population components of variation will extend to phenotypes.

We found that Brazilian and US sickle cell anemia patients markedly differ in ancestry proportions. The African origin on these two populations are distinct; Africans forced to migrate into the United States were mainly from Central West Africa, where the Benin haplotype predominates, while Africans transported to Brazil, were mainly from Angola, Congo, and Mozambique, areas where the CAR haplotype is more frequent, as shown by HBB haplotypes (BEZERRA et al., 2008; CABRAL et al., 2011; FIGUEIREDO et al., 1996). This dissimilarity has been attributed to clinical features, since clinical evolution is correlated to fetal hemoglobin levels. HbS haplotypes and variations in cis-acting elements associated with different haplotypes are partially responsible for the variation seen in HbF levels among sickle cell disease patients, along with eQTL and modulatory elements acting in trans (CHANG et al., 1995; DOVER et al., 1992; THEIN et al., 1994). Senegal and Asian-Indian haplotypes, more prone to overproduction of HbF, share the –158(C>T) polymorphism in the promoter of the G-gamma gene, detectable with the XmnI restriction enzyme (LABIE et al., 1985; NAGEL et al., 1985; STEINBERG et al., 1995). The beta-globin gene cluster haplotype, independently of the HbF levels, may also be correlated with survival of SCA patients (BAKANAY et al., 2005).

In the present work we found that, beyond the African component, SCA patients from Brazil are different in proportions of general continental ancestries compared to American SCA patients both genome-wise and, specially, locally on the chromosome 11. This has important implications, since this is the chromosome where the gamma-globin locates, together with beta-globin, locus control region and fetal hemoglobin hereditary persistence variants. It has been advised that admixture mapping may only be applicable when ancestral populations differ in the phenotype of interest (DA SILVA et al., 2011), and that seems to be the case for SCA patients regarding HbF production. Admixed mapping, nonetheless, has been applied when the ancestral populations are European and African (ADLER et al., 2010; CREARY et al., 2009; REICH et al., 2005; WINKLER; NELSON; SMITH, 2010; ZHU; TANG; RISCH, 2008). It is still matter of debate whether HbF levels are influenced by European ancestry (DA

SILVA et al., 2011), whereas different ancestries inside the African continent have already been proven to alter gamma-globin expression. Virtually, if we were able to discriminate haplotypes of different African origins, it would be possible to perform admixture mapping for loci involved in gamma-globin expression. Moreover, it is still unclear how different levels of admixture will translate to HbF production and other phenotypic traits. The present work underlines the importance of understanding the ancestries interplay, since SCA varies greatly on local level, especially for Brazilians, which shows that European blocks highly contribute in the overall haplotypic structure.

Considered mostly an Africa-derived condition, sickle cell anemia still has population aspects to be ascertained, as an example, on the MTHFR SNPs and its implication in the avascular necrosis. This gene has been clearly found to be associated with vascular necrosis in the Cooperative Study of Sickle Cell Disease (CSSCD) population, and still could not be confirmed in the Kuwaiti population (ADEKILE et al., 2001). This association, however, has been replicated for Brazilians (MOREIRA NETO et al., 2006), as BCL11A and HBS1L-MYB loci relation to HbF production (BEZERRA et al., 2008). Other associations failed to be replicated in Brazilian cohorts, like the TNF-alpha (-308)G/A and pediatric stroke (BELISÁRIO et al., 2015). It seems to have some level of overlap in causative genes and networks across populations, in conjunction with population-specific markers, as the MPO on infection susceptibility in Brazilian SCA patients (COSTA et al., 2005). Leonardo et cols. evaluated five SNPs in the *HMIP-2* and one in *BCL11A* genes in 244 Brazilian patients and found rs9399137, a relatively common European polymorphism, variant significantly associated with HbF levels, and, similarly, rs11886868 in *BCL11A* was also associated with HbF (LEONARDO et al., 2016).

These variances explain the extreme interindividual differences in the severity of the disease and the role that population genetics seem to play in sickle cell anemia complications. In the last years a number of epistatic genes and pleiotropic genes have been found, and many others are candidates (NAGEL, 2001). The new genomic era upholds the idea of personalized medicine, as high-throughput sequencing promises to accelerate our multigenic understanding of diseases. Acknowledging ancestry in admixed populations is an important task for the efforts towards individualized clinical practice. Genetic studies require careful consideration about ancestry in the study design, not only to avoid population stratification bias, but also to maximize the power of finding novel variants.

CONCLUSIONS

Admixture mapping has been effectively used to discover novel variants for phenotypes that are more common on certain ethnic groups and are related to admixture (FREEDMAN et al., 2006). By discriminating population-specific variants and haplotypes, is possible to compare populations and/or undentify alleles from diferent origins in a same population that explain phenotypic variability. Sickle cell disease ancestry studies could lead to novel loci associated with phenotypic variability. Here we demonstrated that SCA subjects from different localities vary on the genomic and local ancestry on chromosome 11, thus, SCA seems a promising target for admixture mapping.

MATERIALS AND METHODS

Subjects

The present research followed the principles of the Declaration of Helsinki; all patients were presented to the aims and details of the study and signed an informed consent.

The American SCA cohort data (30 individuals) was kindly provided by Dr. Hakon Hakonarson, and filtered out from the Epic Care Clinical System (Epic, Verona, WI), along with auto-declared African American (AAM) not affected by SCA. Ninety-one Brazilian SCA patients were recruited at two Brazilian hematology centers, HEMOPE (Recife, PE) and HEMOCENTRO (Campinas, SP). Patients' recruitment and blood collection were approved by the Ethics Committee of FCM-UNICAMP (Campinas-SP, Brazil), while Brazilian control groups from different studies were used as broad Brazilian reference.

Genotyping

Genotyping of both American and Brazilian samples was carried out on Affymetrix Genome-Wide Human SNP 6.0 array platform (Affymetrix Inc., CA, USA), accordingly with to manufacturer's protocol.

These data was analyzed along with reference populations from the HapMap International Project and 1000 Genomes Project (DURBIN et al., 2010; INTERNATIONAL; CONSORTIUM, 2003). We selected six HapMap populations: ASW, CEU, LWK, MXL, TSI and YRI, while other two populations, FIN and CLM, were acquired from the database of 1000 Genomes Project, as shown on Table 1.

Quality Control (QC)

Manipulation of row genotypes data and the basic quality control procedure was performed with the aid of the PLINK software (versions 1.07) (PURCELL et al., 2007). Each individual sample was checked for the discordance in relation to the sex register, outlying missing genotype call rate; presence of duplicated or genetically related samples ($IBD > 0.98$ and 0.1875 , respectively). Each population was evaluated for genotyping quality and markers consistency throughout the sample, removing markers with high missing genotypes (> 0.05); or demonstrating deviation from Hardy-Weinberg equilibrium (HWE), p -value $< 1e-6$; and SNPs with low minor allele frequency (MAF) < 0.05 . Also, we composed a list of SNPs for which at least one Mendelian inconsistency was observed in the populations that had trios' information; afterwards, this SNPs were drop out from the analysis.

LD

We have controlled the data for regions of high LD through the extraction of local and long range markers in LD ($r^2 < 0.5$) at first place. We also excluded known regions of extensive LD (PRICE et al., 2008) across the genome.

Population structure

Population structure and admixture were analyzed applying EIGENSOFT (PATTERSON; PRICE; REICH, 2006) and ADMIXTURE (ALEXANDER; NOVEMBRE; LANGE, 2009) software, including reference populations from the HapMap consortium and a sample of allegedly healthy Brazilians (HbAA). To explore population genome-wide stratification, we applied EIGENSOFT *version 2* (PATTERSON; PRICE; REICH, 2006), and ADMIXTURE *version 1.3* (ALEXANDER; NOVEMBRE; LANGE, 2009) programs. EIGENSOFT applies principal component analysis (PCA) to the genotypic data, a

non-parametric approach of reducing the multidimensionality to orthogonal eigenvectors that enclose the maximum variance. We also calculated the F-statistic among populations, by the Hudson Fst method, implemented in the EIGENSOFT package.

ADMIXTURE software clusters individuals on the bases of their allelic profile using differences in allelic frequencies by a likelihood model, to estimate ancestry. We identified the optimal value of K (5) by the least error.

Local ancestry inference

We phased genotypes of the chromosome 11 on both Brazilian and American SCA patients using the SHAPEIT software(DELANEAU; ZAGURY; MARCHINI, 2012). We then analyzed local ancestry in the chromosome 11 using the software SABER(TANG et al., 2006), which implements a Markov-Hidden Markov Model for inferring locus-specific ancestry in admixed individuals (TANG et al., 2006). We modeled the SCA as an admixture of chromosomes from three ancestral populations with various global proportions of European, Native American and West African ancestries. To accomplish that we consider admixed haplotypes as mosaics of segments taken from three of the HapMap phase 3 haplotype panels (INTERNATIONAL HAPMAP CONSORTIUM, 2003): phased haplotypes from the CEU (117 haplotypes), CHB+JPT (169) and YRI (115) panels, as proxy haplotype data from Europeans, Native American and African ancestors.

Conflict of Interest Statement. *None declared.*

REFERENCES

1. Driss, A., Asare, K.O. J.M. Hibbert¹, B.E. Gee², T.V. Adamkiewicz³, and J. K. S. Sickle Cell Disease in the Post Genomic Era: A Monogenic Disease with a Polygenic Phenotype. *Brain. Behav. Immun.* **22**, 629 (2008).
2. Nagel, R. L. Pleiotropic and epistatic effects in sickle cell anemia. *Curr. Opin. Hematol.* **8**, 105–10 (2001).
3. Hebbel, R. P., Osarogiagbon, R. & Kaul, D. The endothelial biology of sickle cell disease: inflammation and a chronic vasculopathy. *Microcirc. New York NY 1994* **11**, 129–151 (2004).
4. Bowers, A. S., Reid, H. L., Greenidge, A., Landis, C. & Reid, M. Blood viscosity and the expression of inflammatory and adhesion markers in homozygous sickle cell disease subjects with chronic leg ulcers. *PLoS One* **8**, e68929 (2013).
5. Belcher, J. D. *et al.* Heme oxygenase-1 is a modulator of inflammation and vaso-occlusion in transgenic sickle mice. *J. Clin. Invest.* **116**, 808–16 (2006).
6. Kato, G. J. *et al.* Lactate dehydrogenase as a biomarker of hemolysis-associated nitric oxide resistance, priapism, leg ulceration, pulmonary hypertension, and death in patients with sickle cell disease. *Blood* **107**, 2279–85 (2006).
7. De Caterina, R. *et al.* Nitric oxide decreases cytokine-induced endothelial activation. Nitric oxide selectively reduces endothelial expression of adhesion molecules and proinflammatory cytokines. *J. Clin. Invest.* **96**, 60–8 (1995).
8. Kato, G. J., Gladwin, M. T. & Steinberg, M. H. Deconstructing sickle cell disease: reappraisal of the role of hemolysis in the development of clinical subphenotypes. *Blood Rev.* **21**, 37–47 (2007).
9. Pawloski, J. R., Hess, D. T. & Stamler, J. S. Impaired vasodilation by red blood cells in sickle cell disease. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 2531–6 (2005).
10. Elagouz, M., Jyothi, S., Gupta, B. & Sivaprasad, S. Sickle cell disease and the eye: old

- and new concepts. *Surv. Ophthalmol.* **55**, 359–77 (2010).
11. Baldwin, C. *et al.* Brief report Association of klotho , bone morphogenic protein 6 , and annexin A2 polymorphisms with sickle cell osteonecrosis. **106**, 372–375 (2005).
 12. Nolan, V. G. *et al.* Association of single nucleotide polymorphisms in klotho with priapism in sickle cell anaemia. *Br. J. Haematol.* **128**, 266–72 (2005).
 13. Flanagan, J. M. *et al.* Genetic predictors for stroke in children with sickle cell anemia. *Blood* **117**, 6681–4 (2011).
 14. Adewoye, A. H. *et al.* Association of polymorphisms of IGF1R and genes in the transforming growth factor- beta /bone morphogenetic protein pathway with bacteremia in sickle cell anemia. *Clin. Infect. Dis.* **43**, 593–8 (2006).
 15. Zhang, S. X. *et al.* Pigment epithelium-derived factor (PEDF) is an endogenous antiinflammatory factor. *FASEB J.* **20**, 323–5 (2006).
 16. Blann, a. D., Mohan, J. S., Bareford, D. & Lip, G. Y. H. Soluble P-selectin and vascular endothelial growth factor in steady state sickle cell disease: relationship to genotype. *J. Thromb. Thrombolysis* **25**, 185–189 (2007).
 17. Carlson, C. S. *et al.* Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am. J. Hum. Genet.* **74**, 106–120 (2004).
 18. Gao, G. *et al.* Unbalanced expression of VEGF and PEDF in ischemia-induced retinal neovascularization. *FEBS Lett.* **489**, 270–276 (2001).
 19. Asahara, T. *et al.* Tie2 Receptor Ligands, Angiopoietin-1 and Angiopoietin-2, Modulate VEGF-Induced Postnatal Neovascularization. *Circ. Res.* **83**, 233–240 (1998).
 20. Nagel, R. L. & Labie, D. DNA haplotypes and the beta s globin gene. *Prog. Clin. Biol. Res.* **316B**, 371–93 (1989).
 21. Steinberg, M. H. & Sebastiani, P. Genetic modifiers of sickle cell disease. *Am. J. Hematol.* **87**, 795–803 (2012).
 22. Alsultan, A. *et al.* Fetal hemoglobin in sickle cell anemia: Saudi patients from the Southwestern province have similar HBB haplotypes but higher HbF levels than African

- Americans. *Am. J. Hematol.* **86**, 612–4 (2011).
23. Akinsheye, I. *et al.* Fetal hemoglobin in sickle cell anemia. *Blood* **118**, 19–27 (2011).
 24. Creary, L. E. *et al.* Ethnic differences in F cell levels in Jamaica: a potential tool for identifying new genetic loci controlling fetal haemoglobin. *Br. J. Haematol.* **144**, 954–60 (2009).
 25. Solovieff, N. *et al.* Fetal hemoglobin in sickle cell anemia: genome-wide association studies suggest a regulatory region in the 5' olfactory receptor gene cluster. *Blood* **115**, 1815–22 (2010).
 26. Reich, D. *et al.* A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. *Nat. Genet.* **37**, 1113–8 (2005).
 27. Baran, Y. *et al.* Fast and accurate inference of local ancestry in Latino populations. *Bioinformatics* **28**, 1359–1367 (2012).
 28. Winkler, C. a, Nelson, G. W. & Smith, M. W. Admixture mapping comes of age. *Annu. Rev. Genomics Hum. Genet.* **11**, 65–89 (2010).
 29. Tang, H. *et al.* Recent genetic selection in the ancestral admixture of Puerto Ricans. *Am. J. Hum. Genet.* **81**, 626–33 (2007).
 30. Bryc, K. *et al.* Genome-wide patterns of population structure and admixture in West Africans and African Americans. **107**, (2010).
 31. Zhu, X., Tang, H. & Risch, N. Admixture Mapping and the Role of Population Structure for Localizing Disease Genes. *Adv. Genet.* **60**, 547–569 (2008).
 32. Kehdy, F. S. G. *et al.* Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proc. Natl. Acad. Sci.* **112**, 8696–8701 (2015).
 33. Montinaro, F. *et al.* Unravelling the hidden ancestry of American admixed populations. *Nat. Commun.* **6**, 6596 (2015).
 34. da Silva, M. C. F. *et al.* Extensive admixture in Brazilian sickle cell patients: implications for the mapping of genetic modifiers. *Blood* **118**, 4493–5; author reply 4495 (2011).
 35. Giolo, S. R. *et al.* Brazilian urban population genetic structure reveals a high degree of

- admixture. *Eur. J. Hum. Genet.* **20**, 111–6 (2012).
36. Santos, H. C. *et al.* A minimum set of ancestry informative markers for determining admixture proportions in a mixed American population: the Brazilian set. *Eur. J. Hum. Genet.* 1–7 (2015). doi:10.1038/ejhg.2015.187
 37. Barbujani, G., Magagni, A., Minch, E. & Cavalli-Sforza, L. L. An apportionment of human DNA diversity. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 4516–9 (1997).
 38. Rosenberg, N. A. *et al.* Genetic structure of human populations. *Science* **298**, 2381–5 (2002).
 39. Bezerra, C. *et al.* Globin Loci Associate With Fetal Hemoglobin Levels and Pain Crises in Sickle Cell Disease. *Pain* **105**, 1–6 (2008).
 40. Cabral, C. *et al.* Determination of b haplotypes in patients with sickle-cell anemia in the state of Rio Grande do Norte , Brazil. **424**, 421–424 (2011).
 41. Figueiredo, M. S. *et al.* Effect of alpha-thalassemia and beta-globin gene cluster haplotypes on the hematological and clinical features of sickle-cell anemia in Brazil. *Am. J. Hematol.* **53**, 72–6 (1996).
 42. Dover, G. J. *et al.* Fetal hemoglobin levels in sickle cell disease and normal individuals are partially controlled by an X-linked gene located at Xp22.2. *Blood* **80**, 816–24 (1992).
 43. Thein, S. L. *et al.* Detection of a major gene for heterocellular hereditary persistence of fetal hemoglobin after accounting for genetic modifiers. *Am. J. Hum. Genet.* **54**, 214–28 (1994).
 44. Chang, Y. C., Smith, K. D., Moore, R. D., Serjeant, G. R. & Dover, G. J. An analysis of fetal hemoglobin variation in sickle cell disease: the relative contributions of the X-linked factor, beta-globin haplotypes, alpha-globin gene number, gender, and age. *Blood* **85**, 1111–7 (1995).
 45. Labie, D. *et al.* Common haplotype dependency of high G gamma-globin gene expression and high Hb F levels in beta-thalassemia and sickle cell anemia patients. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 2111–4 (1985).
 46. Nagel, R. L. *et al.* Hematologically and Genetically Distinct Forms of Sickle Cell

- Anemia in Africa. *N. Engl. J. Med.* **312**, 880–884 (1985).
47. Steinberg, M. H. *et al.* Gender and haplotype effects upon hematological manifestations of adult sickle cell anemia. *Am. J. Hematol.* **48**, 175–81 (1995).
 48. Bakanay, S. M. *et al.* Mortality in sickle cell patients on hydroxyurea therapy. *Blood* **105**, 545–7 (2005).
 49. Adler, S. *et al.* Mexican-American admixture mapping analyses for diabetic nephropathy in type 2 diabetes mellitus. *Semin. Nephrol.* **30**, 141–149 (2010).
 50. Adekile, A. D., Kutlar, F., Haider, M. Z. & Kutlar, A. Frequency of the 677 C→T mutation of the methylenetetrahydrofolate reductase gene among Kuwaiti sickle cell disease patients. *Am. J. Hematol.* **66**, 263–6 (2001).
 51. Moreira Neto, F. *et al.* The clinical impact of MTHFR polymorphism on the vascular complications of sickle cell disease. *Braz. J. Med. Biol. Res.* **39**, 1291–5 (2006).
 52. Belisário, A. R. *et al.* Association of alpha-thalassemia, TNF-alpha (-308G→A) and VCAM-1 (c.1238G→C) gene polymorphisms with cerebrovascular disease in a newborn cohort of 411 children with sickle cell anemia. *Blood Cells, Mol. Dis.* **54**, 44–50 (2015).
 53. Costa, R. N. P. *et al.* Association of the G-463A myeloperoxidase polymorphism with infection in sickle cell anemia. *Haematologica* **90**, 977–9 (2005).
 54. Freedman, M. L. *et al.* Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 14068–73 (2006).
 55. Admixture mapping of white cell count: genetic locus responsible for lower white blood cell count in the Health ABC and Jackson Heart studies. - PubMed - NCBI. at <<http://www.ncbi.nlm.nih.gov/pubmed/18179887>>
 56. Durbin, R. M. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
 57. International, T. & Consortium, H. The International HapMap Project. *Nature* **426**, 789–796 (2003).
 58. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based

- linkage analyses. *Am. J. Hum. Genet.* **81**, 559–75 (2007).
59. Price, A. L. *et al.* Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.* **83**, 132–5; author reply 135–9 (2008).
 60. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
 61. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
 62. Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10**, 5–6 (2012).
 63. Tang, H., Coram, M., Wang, P., Zhu, X. & Risch, N. Reconstructing genetic ancestry blocks in admixed individuals. *Am. J. Hum. Genet.* **79**, 1–12 (2006).

Table 1. Populations' tags.

Populations' labels	Description
CEU	Utah residents with Northern and Western European ancestry from the CEPH collection
FIN	Finnish in Finland
TSI	Toscani in Italia
ASW	Americans of African Ancestry in SW USA
LWK	Luhya in Webuye, Kenya
YRI	Yoruba in Ibadan, Nigeria
CLM	Colombians from Medellin, Colombia
MXL	Mexican Ancestry from Los Angeles USA
AAM	African American from Philadelphia; USA
SCA	Sickle cell anemia patients from Philadelphia USA
BRZ	Brazilians from Pernambuco, São Paulo and Santa Catarina; Brazil
SCB	Sickle cell anemia patients from Campinas and Recife; Brazil

Table 2. Fst statistics calculated between each pair of populations using all 178,159 common SNPs

	FIN	TSI	CEU	MXL	CLM	BRZ	SCB	AAM	SCA	ASW	LWK
TSI	0.012										
CEU	0.008	0.006									
MXL	0.031	0.032	0.033								
CLM	0.017	0.014	0.016	0.009							
BRZ	0.02	0.015	0.018	0.023	0.01						
SCB	0.033	0.028	0.031	0.031	0.019	0.003					
AAM	0.09	0.084	0.088	0.084	0.07	0.034	0.02				
SCA	0.085	0.079	0.083	0.08	0.066	0.031	0.018	0.001			
ASW	0.082	0.077	0.081	0.077	0.063	0.03	0.018	0.002	0.003		
LWK	0.116	0.11	0.115	0.108	0.094	0.054	0.037	0.007	0.009	0.011	
YRI	0.127	0.121	0.126	0.118	0.104	0.062	0.043	0.006	0.009	0.011	0.009

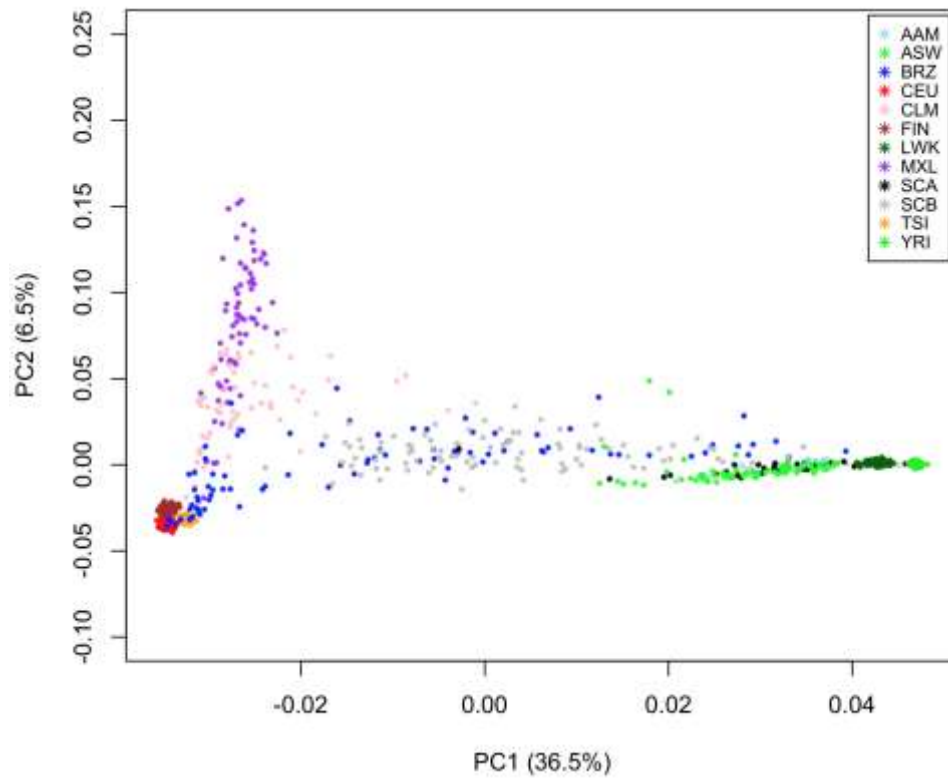


Figure 1. Plot of the selected Principal Components Analysis performed on 178,159 markers. Variability accounted by each PC depicted is in parenthesis.

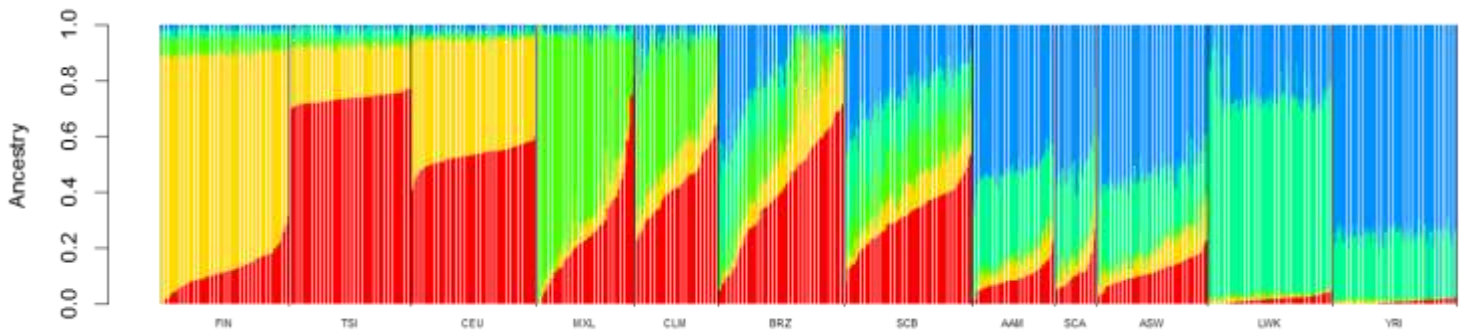
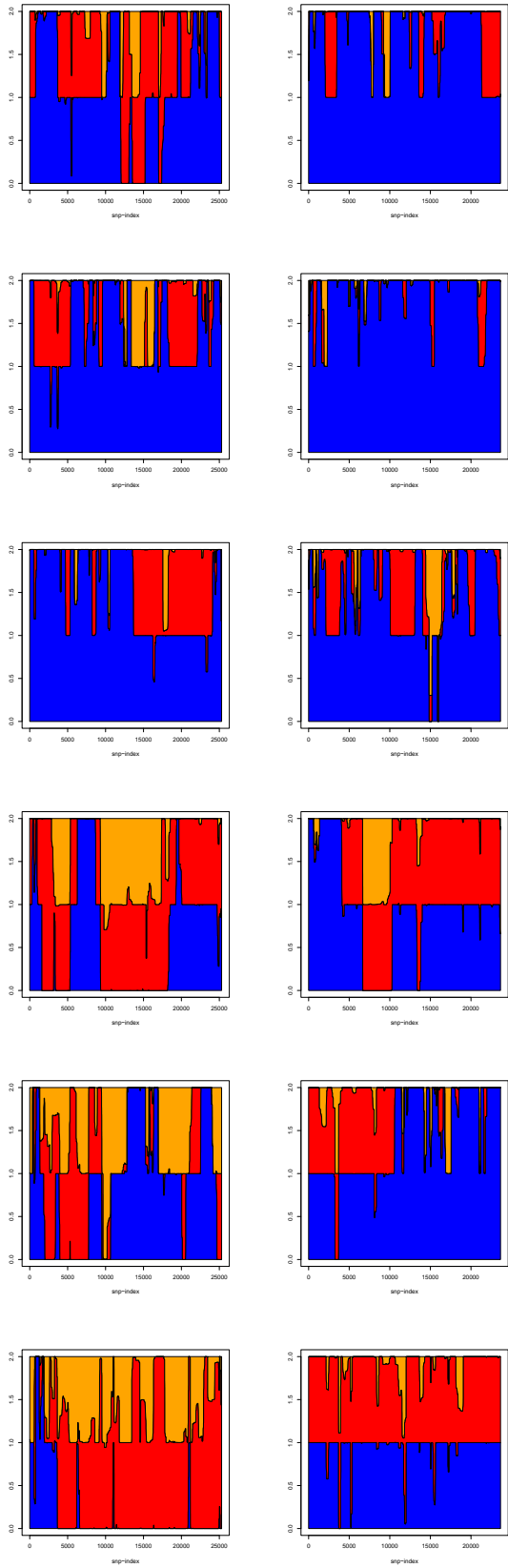


Figure 2. Barplot ADMIXTURE analyses (K=5) performed in the genome. Each row is an individual, populations tags are shown on the X-axis, while Y axis is the proportional ancestry.



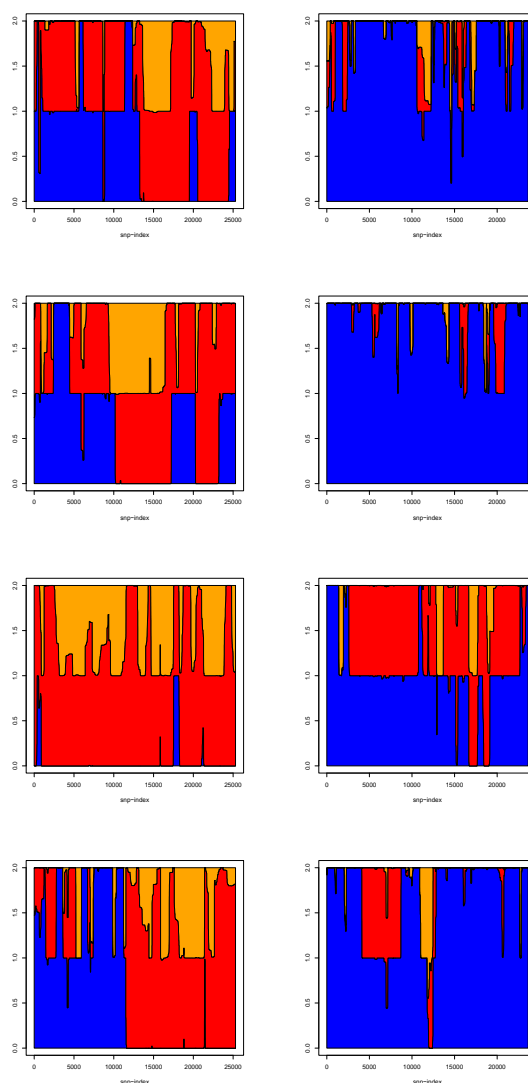


Figure 3. Inferred Ancestry for admixed individuals on chromosome 11. Left column is from Brazilian SCA, while the right column is from US SCA patients. The Y-axis represents the posterior probability that one allele is derived from a specific ancestry; the X-axis indicates the order of the markers. Blue is African ancestry, red is European and yellow Native American.

FUNDING

This work was supported by Brazilian Synchrotron Light Laboratory; Childrens' Hospital of Philadelphia; São Paulo State Research Foundation [2008/57441-0, 2012/06438-5, 2015/13152-9]; and the Council of Technological and Scientific Development [8367/2011-1, 150398/2013-1].

DISCUSSÃO

Estudos de associação baseiam-se na premissa de que uma variante causal será predominante em um grupo caso (para determinado traço fenotípico), e depletada no grupo controle. Apesar de sustentar um postulado simples, a forma como se configura a transmissão genética e a dinâmica populacional impõe aos testes de associação um dado grau de incerteza. A idade da mutação funcional (e, portanto, a pressão seletiva), o tamanho do bloco em desequilíbrio de ligação (logo, a taxa de recombinação na região, possíveis gargalos populacionais, tamanho efetivo da população) são importantes variáveis determinantes na frequência alélica e na seleção de marcadores (ou confiabilidade de um marcador já encontrado em associação). Esses, por sua vez, são aspectos indispensáveis ao bom desenho de um teste de associação, da mesma forma como a avaliação cuidadosa acerca da estrutura populacional, para evitar falsos sinais de associação.

Este trabalho propõe avaliar os aspectos populacionais relativos aos brasileiros que, embora já tenham sido descritos, ainda não foram satisfatoriamente investigados e encontram-se na literatura em nível menor de detalhamento que outras populações, como a britânica (HELLENTHAL et al., 1885) e os norte-americanos de origem africana (BRYC et al., 2010a). Observou-se que os indivíduos de regiões distintas do Brasil apresentam diferenciação pela PCA corrigida pelo software fineStructure (LAWSON et al., 2012), reforçou-se a variabilidade da população brasileira, frente inclusive a outras populações latinas. A variabilidade não foi alta, no entanto, quando consideramos o componente ameríndio nos brasileiros amostrados.

É possível que a variabilidade nos componentes europeu e africano se devam as diferentes origens locais nessas duas localidades. O Brasil observou grandes ondas de migração europeia nos séculos XIX e XX, juntamente com, em menor grau, japonesa e sírio-libanesa. Quanto à origem africana, estudo recente mostrou que existe a contribuição genética tanto da África ocidental, quanto oriental, onde predomina a etnia banto (KEHDY et al., 2015). Os brasileiros nativos, embora de diferentes etnias, pertenciam a grupamentos populacionais relativamente pequenos, havendo menor variabilidade genética. Populações como a mexicana, em contraste, têm origem nativo-americana de povos pré-colombianos de grande distribuição na América Central. A miscigenação no Brasil tem marcante perfil direcional, com cromossomos Y (CARVALHO-SILVA et al., 2001) de origem predominantemente europeia, enquanto haplogrupos de mtDNA são enriquecidos em haplogrupos ameríndios e africanos (ALVES-SILVA et al., 2000).

A presença de haplogrupos de mtDNA ameríndios contrasta com a representatividade no DNA genômico de indivíduos brasileiros. Uma hipótese para tal fenômeno é que o relativo baixo número efetivo das populações ameríndias na introgressão de genomas africanos e europeus resultou em rápida erosão uma vez em recombinação com esses genomas mais diversos. É possível que regiões de baixa recombinação no genoma brasileiro ainda estejam permeados por haplótipos ameríndios, mas faltam estudos que testem esta hipótese.

O decaimento de DL observado em brasileiros, mais acentuado que outras populações latinas, revela relativa pequena extensão dos haplótipos. Isso tem implicações importantes, em conjunto com a intensa miscigenação, na montagem de painéis e na escolha de marcadores para estudos de associação, ou replicação de achados derivados de estudos em outras populações. Marcadores encontrados em uma população de haplótipos mais extensos pode não servir para marcar uma variante causal na nossa população, dado que a recombinação pode ter atuado na quebra de DL entre o marcador e a variante marcada. Tal achado ressalta ainda a importância de utilizar marcadores próprios na população brasileira e painéis que reflitam a estrutura haplotípica característica da última.

Indivíduos afetados por anemia falciforme são candidatos desejáveis para se avaliar a diferenciação de grupos de afetados em grupos populacionais distintos, dado ser um grupo de origem bem conhecida, classicamente utilizado como exemplo de seleção balanceada em humanos, e de herança mendeliana. A população afetada por anemia falciforme a que mais trabalhos foram dedicados, a norte-americana, conta ainda com rica descrição populacional, inclusive com literatura abordando o grupo populacional em que a frequência do alelo HbS é mais expressiva, os afro-americanos. É de grande interesse demonstrar como diferentes condições se manifestam em populações distintas, principalmente ao que tange grupamentos humanos altamente miscigenados, tal qual o brasileiro; e a anemia falciforme se mostra um potencial modelo nesse intento, passível de ser extrapolado para doenças complexas. Há ainda o acréscimo em conhecimento a respeito da aplicabilidade e/ou sobreposição dos diversos achados encontrados em norte-americanos para a população brasileira afetada pela AF.

O mapeamento por miscigenação (*admixture mapping*) tem se mostrado uma ferramenta promissora na descoberta de marcadores associados a doenças humanas (WINKLER; NELSON; SMITH, 2010; ZHU; TANG; RISCH, 2008). A técnica baseia-se no mapeamento genômico entre grupos caso e controle, da variação em um determinado eixo de

ancestralidade, o qual acredita-se que seja mais propenso a desenvolver a condição. Regiões distintas entre grupos indicaria que um mais *loci* estão relacionados ao fenótipo observado naquela região.

Aventou-se o uso dessa ferramenta na produção de HbF em populações falciformes (DA SILVA et al., 2011), embora não seja clara a variabilidade continental (europeus em relação a africanos) para esse fenótipo. Há uma premissa de que as populações ancestrais difiram marcadamente no fenótipo de interesse, juntamente com a condição de que a miscigenação seja recente. Observou-se que falciformes brasileiros diferem marcadamente de falciformes norte-americanos, na composição relativa ancestral em nível genômico e, ainda, em nível local, para o cromossomo 11, em que está localizado o *cluster* da beta-globina, HBB, em seu braço curto.

É sabido que 85% a 95% da variação genética em humanos está contida intrapopulacionalmente, enquanto 5% a 15% se deve as diferenças entre as populações (BARBUJANI et al., 1997; ROSENBERG et al., 2002). Ainda assim, não se sabe como interações entre essas duas classes de variação são extrapoladas para o fenótipo. O Capítulo III exhibe importante descrição da população brasileira afetada por AF, em comparação com uma população também afetada, porém de país de origem distinto.

É importante notar que os haplótipos de HBB exercem grande influência na produção de HbF (ADEKILE, 1997; ALSULTAN et al., 2011), e portanto espera-se que elementos atuem em *cis* na expressão da beta globina. O haplótipo predominante em pacientes de AF brasileiros é o CAR (FIGUEIREDO et al., 1996), considerado o de menor produção de HbF. Não se sabe, no entanto, como esse haplótipo porta-se em conjunção com regiões flangeadoras de diferentes origens ancestrais, além do fato de que, como vimos no Capítulo II, os haplótipos gerais em população brasileira são relativamente pequenos, sendo possível que os haplótipos de hemoglobina no Brasil não marque as variantes funcionais do mesmo modo que em outras populações, mais homogêneas.

Tal demonstração da variabilidade brasileira em indivíduos falciformes pode servir de incentivo para estudos de mapeamento por miscigenação em fenótipos diferentes de HbF, em que se tenha mais clara a distinção entre populações. Outro desdobramento do Capítulo III é que diversos achados em população norte-americanas podem não ser aplicáveis aos brasileiros, dada a marcante diferença em ancestralidade.

Encontrou-se sugestiva diferenciação entre as populações distintas em relação aos achados de CNVs (Capítulo I), embora não possamos descartar o papel das plataformas distintas de genotipagem, bem como as diferenças nos atributos demográficos e clínicos entre as duas cortes, uma vez que a idade difere, bem como os fenótipos. A população norte-americana afetada consiste de indivíduos com qualquer descrição de episódio oclusivo na circulação cerebral, enquanto a população brasileira afetada foi definida apenas por AVC isquêmico. Além disso, a população norte-americana é inteiramente pediátrica, enquanto brasileiros são não apenas pediátricos, mas também contêm alguns jovens adultos.

Ainda assim, é interessante notar que as CNVs potencialmente envolvidas em AVC em indivíduos falciformes encontram-se em vias diferentes daquelas observadas em SNPs associados, indicando o caráter complementar destes dois marcadores.

CONCLUSÕES

Os dados derivados da presente tese demonstram que existe diferenciação genética importante entre indivíduos brasileiros de regiões distintas do Brasil. A população brasileira se mostrou mais heterogênea que as populações referência empregadas nas comparações populacionais, inclusive outras populações latinas, também miscigenadas. A contribuição de genomas ameríndios, no entanto, é maior nessas populações que na população brasileira, embora não possamos desconsiderar que a recombinação atue erodindo os haplótipos ameríndios. Os haplótipos dos brasileiros utilizados são menores em relação a todas as populações referência, excetuando-se as populações africanas. O menor desequilíbrio de ligação pode ser explicado pela ampla contribuição africana à formação populacional do Brasil, o rápido crescimento demográfico e a repetidos eventos de migração.

A comparação entre falciformes de diferentes nacionalidades sugere que esses difiram importantemente em ancestralidade, conforme o histórico de origem demográfica na África aponta, e, ainda, que os componentes europeus e ameríndios sejam também distintos entre norte-americanos e brasileiros. Na ancestralidade local no cromossomo 11, indivíduos brasileiros se mostraram mais heterogêneos que os norte-americanos, com mais rupturas haplotípicas e maior presença europeia e ameríndia.

Por fim, as estratégias seguidas para associação entre CNVs e o fenótipo de AVC são sugestivas e os *loci* podem se tornar alvo de mais estudos para o entendimento das vias de AVC nestes pacientes, que demonstraram genes em associação sugestiva distintos, possivelmente em virtude da diferenciação populacional, explorada no estudo comparativo entre as populações (Capítulo III).

REFERÊNCIAS BIBLIOGRÁFICAS

1000 GENOMES PROJECT CONSORTIUM et al. A map of human genome variation from population-scale sequencing. **Nature**, v. 467, n. 7319, p. 1061–73, 28 out. 2010.

ADAMS, R. et al. The Use of Transcranial Ultrasonography to Predict Stroke in Sickle Cell Disease. **New England Journal of Medicine**, v. 326, n. 9, p. 605–610, 27 fev. 1992.

ADAMS, R. J. et al. Cerebral vessel stenosis in sickle cell disease: criteria for detection by transcranial Doppler. **The American journal of pediatric hematology/oncology**, v. 12, n. 3, p. 277–82, 1990.

ADEKILE, A. D. Historical and anthropological correlates of beta S haplotypes and alpha- and beta-thalassemia alleles in the Arabian Peninsula. **Hemoglobin**, v. 21, n. 3, p. 281–296, 1997.

ADEKILE, A. D. et al. Frequency of the 677 C->T mutation of the methylenetetrahydrofolate reductase gene among Kuwaiti sickle cell disease patients. **American journal of hematology**, v. 66, n. 4, p. 263–6, abr. 2001.

ADEWOYE, A. H. et al. Association of polymorphisms of IGF1R and genes in the transforming growth factor- beta /bone morphogenetic protein pathway with bacteremia in sickle cell anemia. **Clinical infectious diseases : an official publication of the Infectious Diseases Society of America**, v. 43, n. 5, p. 593–8, 1 set. 2006.

ADLER, S. et al. Mexican-American admixture mapping analyses for diabetic nephropathy in type 2 diabetes mellitus. **Seminars in Nephrology**, v. 30, n. 2, p. 141–149, 2010.

Admixture mapping of white cell count: genetic locus responsible for lower white blood cell count in the Health ABC and Jackson Heart studies. - PubMed - NCBI. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/18179887>>.

AFFYMETRIX. BRLMM: an Improved Genotype Calling Method for the GeneChip® Human Mapping 500K Array Set. **Affymetrix**, p. 1–18, 2006.

AKINSHEYE, I. et al. Fetal hemoglobin in sickle cell anemia. **Blood**, v. 118, n. 1, p. 19–27, 7 jul. 2011.

AKLILLU, E. et al. CCL3L1 copy number, HIV load, and immune reconstitution in sub-Saharan Africans. **BMC Infectious Diseases**, v. 13, n. 1, p. 536, 12 dez. 2013.

ALBERTINI, A. M. et al. On the formation of spontaneous deletions: The importance of short sequence homologies in the generation of large deletions. **Cell**, v. 29, n. 2, p. 319–328, 1982.

ALEXANDER, D. H.; NOVEMBRE, J.; LANGE, K. Fast model-based estimation of ancestry in unrelated individuals. **Genome Research**, v. 19, n. 9, p. 1655–1664, 2009.

ALMAL, S. H.; PADH, H. Implications of gene copy-number variation in health and diseases. **Journal of human genetics**, v. 57, n. 1, p. 6–13, jan. 2012.

ALSULTAN, A. et al. Fetal hemoglobin in sickle cell anemia: Saudi patients from the Southwestern province have similar HBB haplotypes but higher HbF levels than African Americans. **American journal of hematology**, v. 86, n. 7, p. 612–4, jul. 2011.

ALVES-SILVA, J. et al. The ancestry of Brazilian mtDNA lineages. **American journal of human genetics**, v. 67, n. 2, p. 444–61, 2000.

ANYAEGBU, C. C. et al. Peripheral blood neutrophil count and candidacidal activity correlate with the clinical severity of sickle cell anaemia (SCA). **European journal of haematology**, v. 60, n. 4, p. 267–8, abr. 1998.

ASAHARA, T. et al. Tie2 Receptor Ligands, Angiopoietin-1 and Angiopoietin-2, Modulate VEGF-Induced Postnatal Neovascularization. **Circulation Research**, v. 83, n. 3, p. 233–240, 10 ago. 1998.

AUTON, A. et al. A global reference for human genetic variation. **Nature**, v. 526, n. 7571, p. 68–74, 2015.

BADANO, J. L. et al. Dissection of epistasis in oligogenic Bardet–Biedl syndrome. **Nature**, v. 439, n. 7074, p. 326–330, 19 jan. 2006.

BAKANAY, S. M. et al. Mortality in sickle cell patients on hydroxyurea therapy. **Blood**, v. 105, n. 2, p. 545–7, 15 jan. 2005.

BALDING, D. J. A tutorial on statistical methods for population association studies. **Nature**

reviews. **Genetics**, v. 7, n. 10, p. 781–91, out. 2006.

BALDUS, S. et al. Spatial mapping of pulmonary and vascular nitrotyrosine reveals the pivotal role of myeloperoxidase as a catalyst for tyrosine nitration in inflammatory diseases. **Free Radical Biology & Medicine**, v. 33, n. 7, p. 1010, 2002.

BALDWIN, C. et al. Brief report Association of klotho , bone morphogenic protein 6 , and annexin A2 polymorphisms with sickle cell osteonecrosis. v. 106, n. 1, p. 372–375, 2005.

BALKARAN, B. et al. Stroke in a cohort of patients with homozygous sickle cell disease. **The Journal of pediatrics**, v. 120, n. 3, p. 360–366, 1992.

BALLAS, S. K. Sickle cell disease: clinical management. **Baillière's clinical haematology**, v. 11, n. 1, p. 185–214, mar. 1998.

BANDEIRA, D. M.; SILVA, M. A. E. Alterações oculares nas doenças falciformes. **Management**, v. 29, n. 3, p. 285–287, 2007.

BANDEIRA, F. M. G. C. et al. Características de recém-nascidos portadores de hemoglobina " S " detectados através de triagem em sangue de cordão umbilical Hemoglobin " S " positive newborn detected by cord blood and its characteristics. **Jornal de Pediatria**, v. 75, p. 167–171, 1999.

BARAN, Y. et al. Fast and accurate inference of local ancestry in Latino populations. **Bioinformatics**, v. 28, n. 10, p. 1359–1367, 2012.

BARBUJANI, G. et al. An apportionment of human DNA diversity. **Proceedings of the National Academy of Sciences of the United States of America**, v. 94, n. 9, p. 4516–9, 29 abr. 1997.

BATISTA DOS SANTOS, S. E. et al. Differential contribution of indigenous men and women to the formation of an urban population in the Amazon region as revealed by mtDNA and Y-DNA. **American journal of physical anthropology**, v. 109, n. 2, p. 175–80, jun. 1999.

BAUTERS, M. et al. Nonrecurrent MECP2 duplications mediated by genomic architecture-driven DNA breaks and break-induced replication repair. **Genome research**, v. 18, n. 6, p. 847–

58, jun. 2008.

BECKMANN, J. S.; ESTIVILL, X.; ANTONARAKIS, S. E. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. **Nature reviews. Genetics**, v. 8, n. 8, p. 639–46, ago. 2007.

BELCHER, J. D. et al. Heme oxygenase-1 is a modulator of inflammation and vaso-occlusion in transgenic sickle mice. **The Journal of clinical investigation**, v. 116, n. 3, p. 808–16, mar. 2006.

BELISÁRIO, A. R. et al. Association of alpha-thalassemia, TNF-alpha (-308G>A) and VCAM-1 (c.1238G>C) gene polymorphisms with cerebrovascular disease in a newborn cohort of 411 children with sickle cell anemia. **Blood Cells, Molecules, and Diseases**, v. 54, n. 1, p. 44–50, 2015.

BELLONI, E. et al. Identification of Sonic hedgehog as a candidate gene responsible for holoprosencephaly. **Nature genetics**, v. 14, n. 3, p. 353–6, nov. 1996.

BEZERRA, C. et al. Globin Loci Associate With Fetal Hemoglobin Levels and Pain Crises in Sickle Cell Disease. **Pain**, v. 105, n. 33, p. 1–6, 2008.

BILLETT, H. H. et al. The percentage of dense red cells does not predict incidence of sickle cell painful crisis. **Blood**, v. 68, n. 1, p. 301–303, 1986.

BLANN, A. D. et al. Soluble P-selectin and vascular endothelial growth factor in steady state sickle cell disease: relationship to genotype. **Journal of Thrombosis and Thrombolysis**, v. 25, n. 2, p. 185–189, 14 dez. 2007.

BORTOLINI, M. C. et al. Genetic variability in two Brazilian ethnic groups: a comparison of mitochondrial and protein data. **American journal of physical anthropology**, v. 103, n. 2, p. 147–56, jun. 1997.

BOWERS, A. S. et al. Blood viscosity and the expression of inflammatory and adhesion markers in homozygous sickle cell disease subjects with chronic leg ulcers. **PloS one**, v. 8, n. 7, p. e68929, jan. 2013.

BRINKMAN, R.; WILDSCHUT, A.; WITTERMANS, A. On the occurrence of two kinds of

haemoglobin in normal human blood. **The Journal of Physiology**, v. 80, n. 4, p. 377–387, 28 fev. 1934.

BRISBIN, A. et al. PCAdmix: Principal Components-Based Assignment of Ancestry along Each Chromosome in Individuals with Admixed Ancestry from Two or More Populations. **Human biology**, v. 84, n. 4, p. 343–364, 2012.

BRUGNARA, C. Sickle cell disease: from membrane pathophysiology to novel therapies for prevention of erythrocyte dehydration. **Journal of pediatric hematologyoncology official journal of the American Society of Pediatric HematologyOncology**, v. 25, n. 12, p. 927–933, 2003.

BRUGNARA, C.; VAN HA, T.; TOSTESON, D. C. Acid pH induces formation of dense cells in sickle erythrocytes. **Blood**, v. 74, n. 1, p. 487–495, 1989.

BRYC, K. et al. Genome-wide patterns of population structure and admixture in West Africans and African Americans. **Proceedings of the National Academy of Sciences**, v. 107, n. 2, p. 786–791, 2010a.

BRYC, K. et al. Genome-wide patterns of population structure and admixture in West Africans and African Americans. v. 107, n. 2, 2010b.

BUNN, H. F. Pathogenesis and treatment of sickle cell disease. **The New England journal of medicine**, v. 337, n. 11, p. 762–9, 11 set. 1997.

CABRAL, C. et al. Determination of b haplotypes in patients with sickle-cell anemia in the state of Rio Grande do Norte, Brazil. v. 424, p. 421–424, 2011.

CALVET, D. et al. First Ischemic Stroke in Sickle-Cell Disease. **Stroke**, v. 46, n. 8, 2015.

CAMPBELL, C. D. et al. Population-genetic properties of differentiated human copy-number polymorphisms. **American journal of human genetics**, v. 88, n. 3, p. 317–32, 11 mar. 2011.

CAO, J. et al. Angiogenic factors in human proliferative sickle cell retinopathy. **The British journal of ophthalmology**, v. 83, n. 7, p. 838–46, jul. 1999.

CARLSON, C. S. et al. Selecting a maximally informative set of single-nucleotide

polymorphisms for association analyses using linkage disequilibrium. **American journal of human genetics**, v. 74, n. 1, p. 106–120, 2004.

CARVALHO-SILVA, D. R. et al. The phylogeography of Brazilian Y-chromosome lineages. **American journal of human genetics**, v. 68, n. 1, p. 281–6, 2001.

CAVALLISFORZA, L. L. Citation Classic - the Genetics of Human-Populations. **Current ContentsLife Sciences**, n. 18, p. 17, 1981.

CHANG, Y. C. et al. An analysis of fetal hemoglobin variation in sickle cell disease: the relative contributions of the X-linked factor, beta-globin haplotypes, alpha-globin gene number, gender, and age. **Blood**, v. 85, n. 4, p. 1111–7, 15 fev. 1995.

CHATTERJEE, R. et al. Mechanism for the increase in solubility of deoxyhemoglobin S due to cross-linking the beta chains between lysine-82 beta 1 and lysine-82 beta 2. **Biochemistry**, v. 21, n. 23, p. 5901–9, nov. 1982.

COSTA, R. N. P. et al. Association of the G-463A myeloperoxidase polymorphism with infection in sickle cell anemia. **Haematologica**, v. 90, n. 7, p. 977–9, jul. 2005.

CRADDOCK, N. et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. **Nature**, v. 464, n. 7289, p. 713–20, 1 abr. 2010.

CRAFT, S. et al. Neuropsychologic effects of stroke in children with sickle cell anemia. **The Journal of pediatrics**, v. 123, n. 5, p. 712–7, nov. 1993.

CREARY, L. E. et al. Ethnic differences in F cell levels in Jamaica: a potential tool for identifying new genetic loci controlling fetal haemoglobin. **British journal of haematology**, v. 144, n. 6, p. 954–60, mar. 2009.

DA SILVA, M. C. F. et al. Extensive admixture in Brazilian sickle cell patients: implications for the mapping of genetic modifiers. **Blood**, v. 118, n. 16, p. 4493–5; author reply 4495, 20 out. 2011.

DE CATERINA, R. et al. Nitric oxide decreases cytokine-induced endothelial activation. Nitric oxide selectively reduces endothelial expression of adhesion molecules and proinflammatory cytokines. **The Journal of clinical investigation**, v. 96, n. 1, p. 60–8, jul. 1995.

DELANEAU, O.; ZAGURY, J.-F.; MARCHINI, J. Improved whole-chromosome phasing for disease and population genetic studies. **Nature Methods**, v. 10, n. 1, p. 5–6, 27 dez. 2012.

DEVLIN, B.; BACANU, S.-A.; ROEDER, K. Genomic Control to the extreme. **Nature Genetics**, v. 36, n. 11, p. 1129–1130, nov. 2004.

DEVLIN, B.; ROEDER, K. Genomic control for association studies. **Biometrics**, v. 55, n. 4, p. 997–1004, dez. 1999.

DIGGS, L. W.; BROOKOFF, D. Multiple cerebral aneurysms in patients with sickle cell disease. **Southern medical journal**, v. 86, n. 4, p. 377–9, abr. 1993.

DOVER, G. J. et al. Fetal hemoglobin levels in sickle cell disease and normal individuals are partially controlled by an X-linked gene located at Xp22.2. **Blood**, v. 80, n. 3, p. 816–24, 1 ago. 1992.

DRISS, A., ASARE, K.O. J.M. HIBBERT¹, B.E. GEE², T.V. ADAMKIEWICZ³, AND J. K. S. Sickle Cell Disease in the Post Genomic Era: A Monogenic Disease with a Polygenic Phenotype. **Brain, Behavior, and Immunity**, v. 22, n. 5, p. 629, jul. 2008.

DURBIN, R. M. et al. A map of human genome variation from population-scale sequencing. **Nature**, v. 467, n. 7319, p. 1061–1073, 28 out. 2010.

EATON, W. A.; HOFRICHTER, J. Hemoglobin S gelation and sickle cell disease. **Blood**, v. 70, n. 5, p. 1245–1266, 1987.

EATON, W. A.; HOFRICHTER, J. The biophysics of sickle cell hydroxyurea therapy. **Science (New York, N.Y.)**, v. 268, n. 5214, p. 1142–3, maio 1995.

EICHLER, E. E. et al. Completing the map of human genetic variation. **Nature**, v. 447, n. 7141, p. 161–165, 2007.

ELAGOZ, M. et al. Sickle cell disease and the eye: old and new concepts. **Survey of ophthalmology**, v. 55, n. 4, p. 359–77, 2010.

EMBURY, S. H. The Interaction of Coexistent α -Thalassemia and Sickle Cell Anemia: A Model for the Clinical and Cellular Results of Diminished Polymerization? **Annals of the New**

York Academy of Sciences, v. 445, n. 1 Fifth Cooley', p. 37–44, jun. 1985.

ESTIVILL, X.; ARMENGOL, L. Copy Number Variants and Common Disorders: Filling the Gaps and Exploring Complexity in Genome-Wide Association Studies. **PLoS Genetics**, v. 3, n. 10, p. 13, 2007.

FABRY, M. E. et al. Dense cells in sickle cell anemia: the effects of gene interaction. **Blood**, v. 64, n. 5, p. 1042–1046, 1984.

FARABAUGH, P. J. et al. Genetic studies of the lac repressor: VII. On the molecular nature of spontaneous hotspots in the lacI gene of Escherichia coli. **Journal of Molecular Biology**, v. 126, n. 4, p. 847–863, 1978.

FEUK, L.; CARSON, A. R.; SCHERER, S. W. Structural variation in the human genome. **Nature Reviews Genetics**, v. 7, n. 2, p. 85–97, 2006.

FIGUEIREDO, M. S. et al. Effect of alpha-thalassemia and beta-globin gene cluster haplotypes on the hematological and clinical features of sickle-cell anemia in Brazil. **American journal of hematology**, v. 53, n. 2, p. 72–6, out. 1996.

FLANAGAN, J. M. et al. Genetic predictors for stroke in children with sickle cell anemia. **Blood**, v. 117, n. 24, p. 6681–4, 16 jun. 2011.

FLANAGAN, J. M. et al. Genetic mapping and exome sequencing identify 2 mutations associated with stroke protection in pediatric patients with sickle cell anemia. **Blood**, v. 121, n. 16, p. 3237–45, 18 abr. 2013.

FREEDMAN, M. L. et al. Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men. **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 38, p. 14068–73, 19 set. 2006.

FREIRE-MAIA, N.; SALZANO, F. M. **Populações Brasileiras; Aspectos Demográficos, Genéticos E Antropológicos ... - Newton Freire-Maia - Google Books**. 1. ed. São Paulo: [s.n.].

GAO, G. et al. Unbalanced expression of VEGF and PEDF in ischemia-induced retinal neovascularization. **FEBS Letters**, v. 489, n. 2–3, p. 270–276, 2001.

- G R SERJEANT, B E SERJEANT, P. F. M. The irreversibly sickled cell; a determinant of haemolysis in sickle cell anaemia. **British Journal of Haematology**, v. 17, n. 6, p. 527–533, 1969.
- GIL, G. P. et al. Polymorphism in the HMOX1 gene is associated with high levels of fetal hemoglobin in Brazilian patients with sickle cell anemia. **Hemoglobin**, v. 37, n. 4, p. 315–24, jan. 2013.
- GILL, F. M. et al. Clinical events in the first decade in a cohort of infants with sickle cell disease. Cooperative Study of Sickle Cell Disease. **Blood**, v. 86, n. 2, p. 776–83, 15 jul. 1995.
- GIOLO, S. R. et al. Brazilian urban population genetic structure reveals a high degree of admixture. **European journal of human genetics : EJHG**, v. 20, n. 1, p. 111–6, jan. 2012.
- GLESSNER, J. T.; LI, J.; HAKONARSON, H. ParseCNV integrative copy number variation association software with quality tracking. **Nucleic Acids Research**, v. 41, n. 5, p. e64–e64, 2013.
- GOMPERT, Z. et al. Genomics of isolation in hybrids. **Philosophical transactions of the Royal Society of London. Series B, Biological sciences**, v. 367, n. 1587, p. 439–50, 5 fev. 2012.
- GON??ALVES, V. F. et al. The phylogeography of African Brazilians. **Human Heredity**, v. 65, n. 1, p. 23–32, 2007.
- GONZALEZ, E. et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. **Science (New York, N.Y.)**, v. 307, n. 5714, p. 1434–40, 4 mar. 2005.
- HAMET, P.; SEDA, O. Current status of genome-wide scanning for hypertension. **Current Opinion in Cardiology**, v. 22, n. 4, p. 292–297, 2007.
- HANDESAKER, R. E. et al. Large multiallelic copy number variations in humans. **Nature Genetics**, v. 47, n. 3, p. 296–303, 26 jan. 2015.
- HASTINGS, P. et al. Mechanisms of change in gene copy number. **Nature reviews. Genetics**, v. 10, n. 8, p. 551–564, 2009.

HEBBEL, R. P. et al. A systems biology consideration of the vasculopathy of sickle cell anemia: the need for multi-modality chemo-prophylaxis. **Cardiovascular hematological disorders drug targets**, v. 9, n. 4, p. 271–292, 2009.

HEBBEL, R. P.; MILLER, W. J. Phagocytosis of sickle erythrocytes: immunologic and oxidative determinants of hemolytic anemia. **Blood**, v. 64, n. 3, p. 733–741, 1984.

HEBBEL, R. P.; OSAROGIAGBON, R.; KAUL, D. The endothelial biology of sickle cell disease: inflammation and a chronic vasculopathy. **Microcirculation New York NY 1994**, v. 11, n. 2, p. 129–151, 2004.

HELLENTHAL, G. et al. Fine scale genetic structure of the British population. **forthcoming in Nature**, 1885.

HERNANDEZ-FERRER, C. et al. affy2sv: an R package to pre-process Affymetrix CytoScan HD and 750K arrays for SNP, CNV, inversion and mosaicism calling. **BMC bioinformatics**, v. 16, n. 1, p. 167, jan. 2015.

HERRICK, J. Peculiar elongated and sickle-shaped red blood corpuscles in a case of severe anemia. **Archives of Internal Medicine**, v. 15, p. 490–493, 1910.

HOFFBRAND AV, CATOVSKY D, TUDDENHAM EG, G. A. Postgraduate Haematology. In: **Postgraduate Haematology**. Oxford, UK: Wiley-Blackwell, 2010. p. 989–1061.

HOFSTRA, T. C. et al. Sickle erythrocytes adhere to polymorphonuclear neutrophils and activate the neutrophil respiratory burst. **Blood**, v. 87, n. 10, p. 4440–7, 15 maio 1996.

HOPPE, C. et al. Confirmation of an association between the TNF(-308) promoter polymorphism and stroke risk in children with sickle cell anemia. **Stroke; a journal of cerebral circulation**, v. 38, n. 8, p. 2241–6, ago. 2007.

HORIUCHI, K. et al. Estimation of fetal hemoglobin levels in individual red cells via fluorescence image cytometry. **Cytometry**, v. 20, n. 3, p. 261–267, 1995.

HOUSTON, M. et al. Binding of xanthine oxidase to vascular endothelium. Kinetic characterization and oxidative impairment of nitric oxide-dependent signaling. **The Journal of Biological Chemistry**, v. 274, n. 8, p. 4985–4994, 1999.

HOUSTON, P. E. et al. Homocysteine in sickle cell disease: relationship to stroke. **The American Journal of Medicine**, v. 103, n. 3, p. 192–196, 1997.

HUIK, K. et al. CCL3L1 copy number is a strong genetic determinant of HIV seropositivity in Caucasian intravenous drug users. **The Journal of infectious diseases**, v. 201, n. 5, p. 730–9, mar. 2010.

IAFRATE, A. J. et al. Detection of large-scale variation in the human genome. **Nature Genetics**, v. 36, n. 9, p. 949–951, 2004.

INGRAM, V. M. A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. **Nature**, v. 178, n. 4537, p. 792–794, 1956.

INTERNATIONAL, T.; CONSORTIUM, H. The International HapMap Project. **Nature**, v. 426, n. 6968, p. 789–796, 2003.

JACKSON, J. E. Front Matter -- A User's Guide to Principal Components. **A User's Guide to Principal Components**, p. i--3, 1991.

JIN, W. et al. Genome-wide detection of natural selection in African Americans pre- and post-admixture. **Genome Research**, v. 22, n. 3, p. 519–527, 1 mar. 2012.

JOINER, C. H. Cation transport and volume regulation in sickle red blood cells. **American Journal of Physiology**, v. 264, n. 2 Pt 1, p. C251–C270, 1993.

KARSUNKY, H. et al. Flt3 Ligand Regulates Dendritic Cell Development from Flt3⁺ Lymphoid and Myeloid-committed Progenitors to Flt3⁺ Dendritic Cells In Vivo. **The Journal of Experimental Medicine**, v. 198, n. 2, p. 305–313, 21 jul. 2003.

KATO, G. J. et al. Lactate dehydrogenase as a biomarker of hemolysis-associated nitric oxide resistance, priapism, leg ulceration, pulmonary hypertension, and death in patients with sickle cell disease. **Blood**, v. 107, n. 6, p. 2279–85, 15 mar. 2006.

KATO, G. J.; GLADWIN, M. T.; STEINBERG, M. H. Deconstructing sickle cell disease: reappraisal of the role of hemolysis in the development of clinical subphenotypes. **Blood reviews**, v. 21, n. 1, p. 37–47, jan. 2007.

KAUL, D. K.; FABRY, M. E.; NAGEL, R. L. Vaso-occlusion by sickle cells: evidence for selective trapping of dense red cells. **Blood**, v. 68, n. 5, p. 1162–1166, 1986.

KAUL, D. K.; FABRY, M. E.; NAGEL, R. L. Microvascular sites and characteristics of sickle cell adhesion to vascular endothelium in shear flow conditions: pathophysiological implications. **Proceedings of the National Academy of Sciences of the United States of America**, v. 86, n. 9, p. 3356–60, maio 1989.

KAUL, D. K.; FABRY, M. E.; NAGEL, R. L. The pathophysiology of vascular obstruction in the sickle syndromes. **Blood Reviews**, v. 10, n. 1, p. 29–44, 1996.

KEHDY, F. S. G. et al. Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. **Proceedings of the National Academy of Sciences**, v. 112, n. 28, p. 8696–8701, 2015.

KELLER, M. P.; CHANCE, P. F. Inherited peripheral neuropathy. **Seminars in neurology**, v. 19, n. 4, p. 353–62, 1999.

KINNEY, T. R. et al. Silent Cerebral Infarcts in Sickle Cell Anemia: A Risk Factor Analysis. **Pediatrics**, v. 103, n. 3, p. 640–645, 1999.

KIOUSSIS, D. et al. Beta-globin gene inactivation by DNA translocation in gamma beta-thalassaemia. **Nature**, v. 306, n. 5944, p. 662–6, [s.d.].

KNIGHT, J. C. Regulatory polymorphisms underlying complex disease traits. **Journal of molecular medicine Berlin Germany**, v. 83, n. 2, p. 97–109, 2005.

KOMURA, D. et al. Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. **Genome Research**, v. 16, n. 12, p. 1575–1584, 2006.

KORBER, E. Uber Differenzen des Blufarstoffs. **Inaugural dissertation. Universität Dorpat**, 1866.

LABIE, D. et al. Common haplotype dependency of high G gamma-globin gene expression and high Hb F levels in beta-thalassemia and sickle cell anemia patients. **Proceedings of the National Academy of Sciences of the United States of America**, v. 82, n. 7, p. 2111–4, abr. 1985.

LANDER, E. S.; SCHORK, N. J. Genetic dissection of complex traits. **Science**, v. 265, n. 5181, p. 2037–2048, 1994.

LAPOUMÉROULIE, C. et al. A novel sickle cell mutation of yet another origin in Africa: the Cameroon type. **Human Genetics**, v. 89, n. 3, p. 333–337, 1992.

LARSEN, M. H. et al. CCL3L gene copy number and survival in an HIV-1 infected Zimbabwean population. **Infection, Genetics and Evolution**, v. 12, n. 5, p. 1087–1093, 2012.

LAWSON, D. J. et al. Inference of population structure using dense haplotype data. **PLoS genetics**, v. 8, n. 1, p. e1002453, jan. 2012.

LEE, E. Y. et al. The impact of CCL3L1 copy number in an HIV-1-infected white population. **AIDS (London, England)**, v. 24, n. 10, p. 1589–91, 19 jun. 2010.

LEE, J. A.; CARVALHO, C. M. B.; LUPSKI, J. R. A DNA Replication Mechanism for Generating Nonrecurrent Rearrangements Associated with Genomic Disorders. **Cell**, v. 131, n. 7, p. 1235–1247, 2007.

LEONARDO, F. C. et al. Reduced rate of sickle-related complications in Brazilian patients carrying HbF-promoting alleles at the BCL11A and HMIP-2 loci. **British Journal of Haematology**, v. 173, n. 3, p. 456–460, 2016.

LEW, V. L. et al. Compartmentalization of sickle-cell calcium in endocytic inside-out vesicles. **Nature**, v. 315, n. 6020, p. 586–589, 1985.

LIMA-COSTA, M. F. et al. Genomic ancestry and ethnoracial self-classification based on 5,871 community-dwelling Brazilians (The Epigen Initiative). **Scientific reports**, v. 5, p. 9812, 2015.

LIPOWSKY, H. H.; KOVALCHECK, S.; ZWEIFACH, B. W. The distribution of blood rheological parameters in the microvasculature of cat mesentery. **Circulation Research**, v. 43, n. 5, p. 738–749, 1978.

LISKAY, R. M.; LETSOU, A.; STACHELEK, J. L. Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. **Genetics**, v. 115, n. 1, p. 161–7, jan. 1987.

LIU, P. et al. Mechanisms for recurrent and complex human genomic rearrangements. **Current opinion in genetics & development**, v. 22, n. 3, p. 211–20, jun. 2012.

LUPSKI, J. R. et al. DNA duplication associated with Charcot-Marie-Tooth disease type 1A. **Cell**, v. 66, n. 2, p. 219–232, 1991.

LUPSKI, J. R. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. **Trends in genetics : TIG**, v. 14, n. 10, p. 417–22, out. 1998.

LUPSKI, J. R. et al. Clan genomics and the complex architecture of human disease. **Cell**, v. 147, n. 1, p. 32–43, 30 set. 2011.

MALHOTRA, D. et al. High frequencies of de novo CNVs in bipolar disorder and schizophrenia. **Neuron**, v. 72, n. 6, p. 951–63, 22 dez. 2011.

MANFREDINI, V.; CASTRO, S.; WAGNER, S. A fisiopatologia da anemia falciforme. **Hemoglobin**, p. 4–7, 2007.

MANOLIO, T. A. et al. Finding the missing heritability of complex diseases. **Nature**, v. 461, n. 7265, p. 747–53, 8 out. 2009.

MARTH, G. T. et al. The functional spectrum of low-frequency coding variation. **Genome biology**, v. 12, n. 9, p. R84, 2011.

MAVILIO, F. et al. Molecular mechanisms of human hemoglobin switching: selective undermethylation and expression of globin genes in embryonic, fetal, and adult erythroblasts. **Proceedings of the National Academy of Sciences of the United States of America**, v. 80, n. 22, p. 6907–11, nov. 1983.

MAY, A. Martin H. Steinberg, Bernard G. Forget, Douglas R. Higgs, David J. Weatherall: Disorders of Hemoglobin: Genetics, Pathophysiology, and Clinical Management. **Human Genetics**, v. 128, p. 231–232, 2010.

MAYNARD, T. M. et al. 22q11 DS: genomic mechanisms and gene function in DiGeorge/velocardiofacial syndrome. **International journal of developmental neuroscience: the official journal of the International Society for Developmental**

Neuroscience, v. 20, n. 3–5, p. 407–19, [s.d.].

MCCARROLL, S. A et al. Integrated detection and population-genetic analysis of SNPs and copy number variation. **Nature genetics**, v. 40, n. 10, p. 1166–74, out. 2008.

MCCARTHY, M. I. et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. **Nature Reviews Genetics**, v. 9, n. 5, p. 356–69, 2008.

MCCLINTOCK, B. Chromosome organization and genic expression. **Cold Spring Harbor symposia on quantitative biology**, v. 16, p. 13–47, 1951.

MERLA, G. et al. Submicroscopic deletion in patients with Williams-Beuren syndrome influences expression levels of the nonhemizygous flanking genes. **American journal of human genetics**, v. 79, n. 2, p. 332–41, ago. 2006.

MILLER, S. T. et al. Silent infarction as a risk factor for overt stroke in children with sickle cell anemia: a report from the Cooperative Study of Sickle Cell Disease. **The Journal of pediatrics**, v. 139, n. 3, p. 385–390, 2001.

MOHANDAS, N.; ROSSI, M. E.; CLARK, M. R. Association between morphologic distortion of sickle cells and deoxygenation-induced cation permeability increase. **Blood**, v. 68, n. 2, p. 450–454, 1986.

MONTALEMBERT, M. et al. Cerebrovascular accidents in sickle cell disease. Risk factors and blood transfusion influence. French Study Group on Sickle Cell Disease. **European Journal of Pediatrics**, v. 152, n. 3, p. 201–204, 1993.

MONTINARO, F. et al. Unravelling the hidden ancestry of American admixed populations. **Nature Communications**, v. 6, p. 6596, 2015.

MOREIRA NETO, F. et al. The clinical impact of MTHFR polymorphism on the vascular complications of sickle cell disease. **Brazilian journal of medical and biological research = Revista brasileira de pesquisas médicas e biológicas / Sociedade Brasileira de Biofísica ... [et al.]**, v. 39, n. 10, p. 1291–5, out. 2006.

MUSUMECI, S. et al. The genetics and molecular basis of alpha thalassaemia in association with Hb S in Jamaican Negroes. **British Journal of Haematology**, v. 43, n. 3, p. 43–56, 1981.

NAGEL, R. L. et al. Hematologically and Genetically Distinct Forms of Sickle Cell Anemia in Africa. **New England Journal of Medicine**, v. 312, n. 14, p. 880–884, 4 abr. 1985.

NAGEL, R. L. Origins and dispersion of the sickle gene. In: EMBURY, S. H. (Ed.). . **Sickle Cell Disease Basic Principles and Clinical Practice**. [s.l.] Raven Press, 1994. p. 353–380.

NAGEL, R. L. Pleiotropic and epistatic effects in sickle cell anemia. **Current opinion in hematology**, v. 8, n. 2, p. 105–10, mar. 2001.

NAGEL, R. L.; LABIE, D. DNA haplotypes and the beta s globin gene. **Progress in clinical and biological research**, v. 316B, p. 371–93, 1989.

NEEL, J. V. The Inheritance of Sickle Cell Anemia. **Science**, v. 110, n. 2846, p. 64–66, 1949.

NOLAN, V. G. et al. Association of single nucleotide polymorphisms in klotho with priapism in sickle cell anaemia. **British journal of haematology**, v. 128, n. 2, p. 266–72, jan. 2005.

NUSSBAUM, R. L.; MCINNES, R. R.; WILLARD, H. F. **Thompson & Thompson. Genética en medicina**. [s.l.] Elsevier, 2008.

OHENE-FREMPOG, K. et al. Cerebrovascular accidents in sickle cell disease: rates and risk factors. **Blood**, v. 91, n. 1, p. 288–294, 1998.

PAIXÃO, M. C. et al. Reliability of isoelectrofocusing for the detection of Hb S, Hb C, and HB D in a pioneering population-based program of newborn screening in Brazil. **Hemoglobin**, v. 25, n. 3, p. 297–303, 2001.

PASANIUC, B. et al. Enhanced Statistical Tests for GWAS in Admixed Populations: Assessment using African Americans from CARE and a Breast Cancer Consortium. **PLoS Genetics**, v. 7, n. 4, p. e1001371, 21 abr. 2011.

PATTERSON, N.; PRICE, A. L.; REICH, D. Population structure and eigenanalysis. **PLoS genetics**, v. 2, n. 12, p. e190, dez. 2006.

PAULING, L.; ITANO, H. A. Sickle cell anemia a molecular disease. **Science**, v. 109, n. 2835, p. 543–548, 1949.

PAWLOSKI, J. R.; HESS, D. T.; STAMLER, J. S. Impaired vasodilation by red blood cells in

sickle cell disease. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 7, p. 2531–6, 15 fev. 2005.

PAYEN, C. et al. Segmental Duplications Arise from Pol32-Dependent Repair of Broken Forks through Two Alternative Replication-Based Mechanisms. **PLoS Genetics**, v. 4, n. 9, p. e1000175, 5 set. 2008.

PEARSON, K. On lines and planes of closest fit to systems of points in space. **The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science**, v. 2, n. 1, p. 559–572, 1901.

PEGELow, C. H. et al. Natural history of blood pressure in sickle cell disease: risks for stroke and death associated with relative hypertension in sickle cell anemia. **The American Journal of Medicine**, v. 102, n. 2, p. 171–177, 1997.

PEGELow, C. H. et al. **Silent infarcts in children with sickle cell anemia and abnormal cerebral artery velocity.** **Archives of Neurology**. [s.l: s.n.]. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/11735775>>.

PEGELow, C. H. et al. Longitudinal changes in brain magnetic resonance imaging findings in children with sickle cell disease. **Blood**, v. 99, n. 8, p. 3014–3018, 2002.

PENA, S. D. J. 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. 9, 2011.

PENA, S. D. J.; BORTOLINI, M. C. Pode a genética definir quem deve se beneficiar das cotas universitárias e demais ações afirmativas? **Estudos Avançados**, v. 18, n. 50, p. 31–50, abr. 2004.

PERUTZ, M. F. **Stereochemistry of cooperative effects in hemoglobin. Heme-heme interaction and the problem of allostery** **Nature**, 1970.

PINTO, D. et al. Functional impact of global rare copy number variation in autism spectrum disorders. **Nature**, v. 466, n. 7304, p. 368–372, 15 jul. 2010.

PLATT, O. S. et al. Mortality in sickle cell disease. Life expectancy and risk factors for early death. **The New England Journal of Medicine**, v. 330, n. 23, p. 1639–1644, 1994.

POLLEX, R. L.; HEGELE, R. A. Copy number variation in the human genome and its implications for cardiovascular disease. **Circulation**, v. 115, n. 24, p. 3130–3138, 2007.

POOL, J. E.; NIELSEN, R. Inference of historical changes in migration rate from the lengths of migrant tracts. **Genetics**, v. 181, n. 2, p. 711–9, fev. 2009.

PRICE, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. **Nature genetics**, v. 38, n. 8, p. 904–9, ago. 2006.

PRICE, A. L. et al. Long-range LD can confound genome scans in admixed populations. **American journal of human genetics**, v. 83, n. 1, p. 132–5–9, jul. 2008.

PRITCHARD, J. K. et al. Association mapping in structured populations. **American journal of human genetics**, v. 67, n. 1, p. 170–81, jul. 2000.

PRITCHARD, J. K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, v. 155, n. 2, p. 945–959, 2000.

PUGACH, I. et al. Dating the age of admixture via wavelet transform analysis of genome-wide data. **Genome Biology**, v. 12, n. 2, p. R19, 2011.

PURCELL, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. **American journal of human genetics**, v. 81, n. 3, p. 559–75, set. 2007.

RATHORE, A. et al. Association of CCR5-59029 A/G and CCL3L1 copy number polymorphism with HIV type 1 transmission/progression among HIV type 1-seropositive and repeatedly sexually exposed HIV type 1-seronegative North Indians. **AIDS research and human retroviruses**, v. 25, n. 11, p. 1149–56, nov. 2009.

REDON, R. et al. Global variation in copy number in the human genome. **Nature**, v. 444, n. 7118, p. 444–454, 2006.

REICH, D. et al. A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. **Nature Genetics**, v. 37, n. 10, p. 1113–8, 2005.

REICH, D. et al. Reconstructing Indian population history. **Nature**, v. 461, n. 7263, p. 489–494, 24 set. 2009.

REICH, D. E.; GOLDSTEIN, D. B. Detecting association in a case-control study while correcting for population stratification. **Genetic epidemiology**, v. 20, n. 1, p. 4–16, jan. 2001.

REICH, D. E.; LANDER, E. S. On the allelic spectrum of human disease. **Trends in Genetics**, v. 17, n. 9, p. 502–510, 2001.

REITER, L. T. et al. Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients. **American journal of human genetics**, v. 62, n. 5, p. 1023–33, maio 1998.

RICARD, G. et al. Phenotypic Consequences of Copy Number Variation: Insights from Smith-Magenis and Potocki-Lupski Syndrome Mouse Models. **PLoS Biology**, v. 8, n. 11, p. e1000543, 23 nov. 2010.

ROSENBERG, N. A. et al. Genetic structure of human populations. **Science (New York, N.Y.)**, v. 298, n. 5602, p. 2381–5, 20 dez. 2002.

SALZANO, F. M. Incidence, effects, and management of sickle cell disease in Brazil. **The American journal of pediatric hematologyoncology**, v. 7, n. 3, p. 240–244, 1985.

SANDERS, S. J. et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. **Neuron**, v. 70, n. 5, p. 863–85, 9 jun. 2011.

SANTOS, H. C. et al. A minimum set of ancestry informative markers for determining admixture proportions in a mixed American population: the Brazilian set. **European journal of human genetics : EJHG**, n. July, p. 1–7, 2015.

SANTOS, S. E. et al. Multiple founder haplotypes of mitochondrial DNA in Amerindians revealed by RFLP and sequencing. **Annals of human genetics**, v. 60, n. Pt 4, p. 305–19, jul. 1996.

SARNAIK, S. A.; BALLAS, S. K. Molecular characteristics of pediatric patients with sickle cell anemia and stroke. **American Journal of Hematology**, v. 67, n. 3, p. 179–182, 2001.

SATTEN, G. A.; FLANDERS, W. D.; YANG, Q. Accounting for unmeasured population substructure in case-control studies of genetic association using a novel latent-class model.

American journal of human genetics, v. 68, n. 2, p. 466–77, fev. 2001.

SCHWANZER-PFEIFFER, D. et al. Characterization of SVEP1, KIAA, and SRPX2 in an in vitro cell culture model of endotoxemia. **Cellular Immunology**, v. 263, n. 1, p. 65–70, 2010.

SEBASTIANI, P. et al. Genetic dissection and prognostic modeling of overt stroke in sickle cell anemia. v. 37, n. 4, p. 435–440, 2005a.

SEBASTIANI, P. et al. Genetic dissection and prognostic modeling of overt stroke in sickle cell anemia. **Nature genetics**, v. 37, n. 4, p. 435–40, abr. 2005b.

SEBASTIANI, P. et al. Genetic modifiers of the severity of sickle cell anemia identified through a genome-wide association study. **American journal of hematology**, v. 85, n. 1, p. 29–35, jan. 2010.

SEBAT, J. et al. Large-scale copy number polymorphism in the human genome. **Science**, v. 305, n. 5683, p. 525–8, 2004.

SETAKIS, E.; STIRNADEL, H.; BALDING, D. J. Logistic regression protects against population structure in genetic association studies. **Genome research**, v. 16, n. 2, p. 290–6, fev. 2006.

SHOSTAKOVICH-KORETSKAYA, L. et al. Combinatorial content of CCL3L and CCL4L gene copy numbers influence HIV-AIDS susceptibility in Ukrainian children. **AIDS (London, England)**, v. 23, n. 6, p. 679–88, 27 mar. 2009.

SLACK, A. et al. On the Mechanism of Gene Amplification Induced under Stress in *Escherichia coli*. **PLoS Genetics**, v. 2, n. 4, p. e48, 2006.

SOLOVIEFF, N. et al. Fetal hemoglobin in sickle cell anemia: genome-wide association studies suggest a regulatory region in the 5' olfactory receptor gene cluster. **Blood**, v. 115, n. 9, p. 1815–22, 4 mar. 2010a.

SOLOVIEFF, N. et al. Fetal hemoglobin in sickle cell anemia: genome-wide association studies suggest a regulatory region in the 5' olfactory receptor gene cluster. **Blood**, v. 115, n. 9, p. 1815–22, 4 mar. 2010b.

STAM, P. The distribution of the fraction of the genome identical by descent in finite random mating populations. **Genetical Research**, v. 35, n. 2, p. 131, 14 abr. 1980.

STANKIEWICZ, P.; LUPSKI, J. R. Genome architecture, rearrangements and genomic disorders. **Trends in genetics : TIG**, v. 18, n. 2, p. 74–82, fev. 2002.

STANKIEWICZ, P.; LUPSKI, J. R. The genomic basis of disease, mechanisms and assays for genomic disorders. **Genome dynamics**, v. 1, p. 1–16, 2006.

STANKIEWICZ, P.; LUPSKI, J. R. Structural variation in the human genome and its role in disease. **Annual review of medicine**, v. 61, p. 437–55, 2010.

STEINBERG, M. H. et al. Gender and haplotype effects upon hematological manifestations of adult sickle cell anemia. **American journal of hematology**, v. 48, n. 3, p. 175–81, mar. 1995.

STEINBERG, M. H. Pathophysiology of sickle cell disease. **Baillière's clinical haematology**, v. 11, n. 1, p. 163–84, mar. 1998.

STEINBERG, M. H.; SEBASTIANI, P. Genetic modifiers of sickle cell disease. **American Journal of Hematology**, v. 87, n. 8, p. 795–803, ago. 2012.

STEVEN, A. et al. Neurologic and Head and Neck Manifestations of Sickle Cell Disease. **Hematology/Oncology Clinics of North America**, v. 30, n. 4, p. 779–798, 2016.

STEVENS, M. C. et al. Fetal hemoglobin and clinical severity of homozygous sickle cell disease in early childhood. **The Journal of pediatrics**, v. 98, n. 1, p. 37–41, jan. 1981.

STRANGER, B. E. et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. **Science**, v. 315, n. 5813, p. 848–853, 2007.

STROUSE, J. J. et al. Primary hemorrhagic stroke in children with sickle cell disease is associated with recent transfusion and use of corticosteroids. **Pediatrics**, v. 118, n. 5, p. 1916–24, nov. 2006.

SWITZER, J. A. et al. Pathophysiology and treatment of stroke in sickle-cell disease: present and future. **The Lancet**, v. 5, n. 6, p. 501–512, 2006.

TANAKA, H. et al. Large DNA palindromes as a common form of structural chromosome

aberrations in human cancers. **Human Cell**, v. 19, n. 1, p. 17–23, fev. 2006.

TANG, D. C. et al. Polymorphisms within the angiotensinogen gene (GT-repeat) and the risk of stroke in pediatric patients with sickle cell disease: a case-control study. **American Journal of Hematology**, v. 68, n. 3, p. 164–169, 2001.

TANG, H. et al. Reconstructing genetic ancestry blocks in admixed individuals. **American journal of human genetics**, v. 79, n. 1, p. 1–12, jul. 2006.

TANG, H. et al. Recent genetic selection in the ancestral admixture of Puerto Ricans. **American journal of human genetics**, v. 81, n. 3, p. 626–33, set. 2007.

THEIN, S. L. et al. Detection of a major gene for heterocellular hereditary persistence of fetal hemoglobin after accounting for genetic modifiers. **American journal of human genetics**, v. 54, n. 2, p. 214–28, fev. 1994.

TOSTESON, D. C.; SHEA, E.; DARLING, R. C. Potassium and sodium of red blood cells in sickle cell anemia. **Journal of Clinical Investigation**, v. 31, n. 4, p. 406–411, 1952.

VISSERS, L. E. L. M. et al. Complex chromosome 17p rearrangements associated with low-copy repeats in two patients with congenital anomalies. **Human genetics**, v. 121, n. 6, p. 697–709, jul. 2007.

WANG, K. et al. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. **Genome Research**, v. 17, n. 11, p. 1665–1674, 1 nov. 2007.

WANG, W. C. et al. Abnormalities of the central nervous system in very young children with sickle cell anemia. **The Journal of pediatrics**, v. 132, n. 6, p. 994–998, 1998.

WANG, W. C. The pathophysiology, prevention, and treatment of stroke in sickle cell disease. **Current Opinion in Hematology**, v. 14, n. 3, p. 191–197, 2007.

WARD, R. H. et al. Mitochondrial DNA polymorphism in three Brazilian Indian tribes. **American Journal of Human Biology**, v. 8, n. 3, p. 317–323, 1996.

WARE, R. E. et al. Prevention of secondary stroke and resolution of transfusional iron overload

in children with sickle cell anemia using hydroxyurea and phlebotomy. **The Journal of pediatrics**, v. 145, n. 3, p. 346–52, set. 2004.

WEISCHENFELDT, J. et al. Phenotypic impact of genomic structural variation: insights from and for human disease. **Nature Reviews Genetics**, v. 14, n. 2, p. 125–138, 18 jan. 2013.

WHEELER, J. T.; KREVANS, J. R. The homozygous state of persistent fetal hemoglobin and the interaction of persistent fetal hemoglobin with thalassemia. **Bulletin of the Johns Hopkins Hospital**, v. 109, p. 217–33, nov. 1961.

WIERENGA, K. J.; SERJEANT, B. E.; SERJEANT, G. R. **Cerebrovascular complications and parvovirus infection in homozygous sickle cell disease.****The Journal of pediatrics**, 2001. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/11562626>>

WILBER, A.; NIENHUIS, A. W.; PERSONS, D. A. Review article Transcriptional regulation of fetal to adult hemoglobin switching : new therapeutic opportunities. v. 117, n. 15, p. 3945–3954, 2014.

WINKLER, C. A.; NELSON, G. W.; SMITH, M. W. Admixture mapping comes of age. **Annual review of genomics and human genetics**, v. 11, p. 65–89, 22 set. 2010.

WONG, K. K. K. et al. A Comprehensive Analysis of Common Copy-Number Variations in the Human Genom. **The American Journal of Human Genetics**, v. 80, n. 1, p. 91–104, 2007.

WOOD, K. C.; HEBBEL, R. P.; GRANGER, D. N. Endothelial cell NADPH oxidase mediates the cerebral microvascular dysfunction in sickle cell transgenic mice. **The FASEB journal official publication of the Federation of American Societies for Experimental Biology**, v. 19, n. 8, p. 989–991, 2005.

WUN, T. et al. Platelet activation and platelet-erythrocyte aggregates in patients with sickle cell anemia. **The Journal of laboratory and clinical medicine**, v. 129, n. 5, p. 507–16, maio 1997.

ZAGO, M. A.; FALCÃO, R.P.; PASQUINI, R. Anemia Falciforme. In: **Tratado de Hematologia**. 1ª ed. [s.l.] Atheneu, 2013. p. 205–223.

ZARBOCK, A.; POLANOWSKA-GRABOWSKA, R. K.; LEY, K. Platelet-neutrophil-interactions: linking hemostasis and inflammation. **Blood reviews**, v. 21, n. 2, p. 99–111, mar.

2007.

ZHANG, D. et al. Neutrophils , platelets , and in fl ammatory pathways at the nexus of sickle cell disease pathophysiology. v. 127, n. 7, p. 801–810, 2016.

ZHANG, S. X. et al. Pigment epithelium-derived factor (PEDF) is an endogenous antiinflammatory factor. **FASEB journal: official publication of the Federation of American Societies for Experimental Biology**, v. 20, n. 2, p. 323–5, fev. 2006.

ZHU, X.; TANG, H.; RISCH, N. Admixture Mapping and the Role of Population Structure for Localizing Disease Genes. **Advances in Genetics**, v. 60, n. 7, p. 547–569, 2008.

ZUK, O. et al. The mystery of missing heritability: Genetic interactions create phantom heritability. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 4, p. 1193–8, 24 jan. 2012.

ANEXO I**FORMULÁRIO DE CONSENTIMENTO LIVRE E ESCLARECIDO
(Pacientes)**

FACULDADE DE CIÊNCIAS MÉDICAS

COMITÊ DE ÉTICA EM PESQUISA

Caixa Postal 6111, 13083-970 Campinas, SP

(0_19) 3521-8936

FAX (0_19) 3521-7187

www.fcm.unicamp.br/pesquisa/etica/index.htmlcep@fcm.unicamp.br**TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO**

Você está sendo convidado(a) a participar, como voluntário, em uma pesquisa. Após ser esclarecido(a) sobre as informações a seguir, no caso de aceitar fazer parte do estudo, assine ao final deste documento, que está em duas vias. Uma delas é sua e a outra do pesquisador responsável. Em caso de recusa você não será penalizado(a) de forma alguma.

Informações sobre a pesquisa:

Título do projeto: **AVALIAÇÃO DAS VARIAÇÕES NO NÚMERO DE CÓPIAS (CNVs) NA SUSCETIBILIDADE AO ACIDENTE VASCULAR CEREBRAL EM PACIENTES COM ANEMIA FALCIFORME.**

Pesquisador responsável: Mônica Barbosa de Melo

Telefone para contato: 019 3521-1143

As doenças falciformes podem levar a uma série de complicações vasculares, dentre elas, o acidente vascular cerebral ou AVC. O objetivo geral do estudo é identificar regiões do genoma humano associadas ao risco de desenvolvimento de acidente vascular cerebral (derrame) nos referidos pacientes. Esses estudos poderão levantar dados que possibilitarão um conhecimento mais aprofundado sobre o quadro clínico da anemia falciforme.

Caso você concorde em participar desse estudo, os pesquisadores participantes farão perguntas a respeito dos seus antecedentes médicos e familiares. Você será submetido a um exame físico e neurológico para confirmação do estado clínico. Além disso, poderá ser submetido a exames laboratoriais adicionais e/ou de imagem (como tomografia computadorizada ou ressonância magnética de crânio). Para participar do estudo é necessária sua permissão para colhermos uma amostra de sangue (20 a 30 ml, o equivalente a duas colheres de sopa). O exame é muito seguro e fora o desconforto da

injeção não costuma causar efeitos colaterais. Hospitalização não será necessária. Os procedimentos mencionados acima serão realizados dentro do primeiro ano após o consentimento em participar no estudo e, com exceção dos exames de imagem, fazem parte dos cuidados médicos de rotina para um paciente com anemia falciforme.

Será garantida, aos pacientes e responsáveis, respostas a quaisquer perguntas que possam ocorrer, em qualquer momento da pesquisa, e esclarecimento de qualquer dúvida acerca dos assuntos relacionados com esta pesquisa.

O participante do estudo não receberá qualquer valor em dinheiro. Os resultados serão publicados e sua identidade será preservada.

O DNA, extraído a partir do sangue de cada paciente, será preservado conforme o Regimento do Laboratório de Genética Molecular Humana do CBMEG (UNICAMP) se o paciente assim o consentir.

Liberdade de abandonar a pesquisa sem prejuízo:

Eu, paciente ou responsável, entendi que a participação neste projeto de pesquisa é voluntária e que eu posso recusar ou retirar meu consentimento, a qualquer momento (incluindo sobre a amostra de sangue), sem comprometer os cuidados médicos que recebo ou receberei no futuro.

Fui informado que poderei autorizar que esse material genético seja guardado no Banco de Material Biológico do Laboratório de Genética Molecular Humana CBMEG/UNICAMP. Caso este material venha a ser útil para outros estudos aprovados pelo Comitê de Ética, estou ciente que serei informado para que autorize ou não o uso da(s) amostra(s).

- Inclui o armazenamento do material para pesquisa genéticas futuras.
- Determina que o material deva ser descartado após o término da pesquisa.

Nome e RG do sujeito da pesquisa

Consentimento

Eu, _____ RG _____
 ___abaixo assinado, estou de acordo em participar do estudo proposto acima, como sujeito. Fui devidamente informado(a) e esclarecido pelo pesquisador _____ sobre a pesquisa e os procedimentos nela envolvidos. Foi-me garantido que posso retirar meu consentimento a qualquer instante, sem prejuízo em meu acompanhamento ou tratamento. Declaro que recebi cópia do presente Termo de Consentimento.

Local e data: _____

Nome do paciente: _____

Nome e assinatura do responsável legal: _____

ANEXO II

FORMULÁRIO DE CONSENTIMENTO LIVRE E ESCLARECIDO (Controles)

PARTICIPAÇÃO DE INDIVÍDUOS CONTROLE

Eu estou ciente de que a minha participação no estudo intitulado “**AVALIAÇÃO DAS VARIACIONES NO NÚMERO DE CÓPIAS (CNVs) NA SUSCETIBILIDADE AO ACIDENTE VASCULAR CEREBRAL EM PACIENTES COM ANEMIA FALCIFORME.**” será como um indivíduo controle em testes de biologia molecular.

Entendi que todas as informações constadas neste termo quanto a procedimentos, sigilo, resultados e contatos para resultados (se eu desejar saber), serão iguais aos demais indivíduos não-controles deste estudo. Fui informado também que poderei autorizar que o meu material genético seja guardado no Banco de Material Biológico do Laboratório de Genética Molecular Humana CBMEG/UNICAMP. Caso este material venha a ser útil para outros estudos aprovados pelo Comitê de Ética, estou ciente que serei informado para que autorize ou não o uso da (s) amostra (s).

Nome e RG do indivíduo controle

Assinatura do sujeito da pesquisa ou responsável legal

Data

RESPONSABILIDADE DO PESQUISADOR

Eu expliquei a _____ o objetivo do estudo, os procedimentos requeridos e os possíveis riscos que poderão advir do estudo, usando o melhor do meu conhecimento. Eu me comprometo a fornecer uma cópia desse formulário de consentimento ao participante ou responsável.

Nome e RG do pesquisador

Assinatura do pesquisador

Data

Telefone para contato: 19-3521-1143 (Mônica B. Melo)

Telefone do Comitê de Ética em Pesquisa – FCM – UNICAMP para dúvidas éticas e/ou reclamações:
19-3521-8936

ANEXO III**FORMULÁRIO DE CONSENTIMENTO LIVRE E ESCLARECIDO
(Exame de ressonância magnética)****TERMO DE CONSENTIMENTO INFORMADO LIVRE E ESCLARECIDO**

Projeto: Aplicação da Técnica De Microarray de Alta Densidade de Marcadores na Avaliação da Suscetibilidade ao Acidente Vascular Cerebral em Pacientes com Anemia Falciforme

Pesquisadores Responsáveis:

Profa. Dra. Mônica Barbosa de Melo (CBMEG/UNICAMP)

Pedro Rodrigues Sousa da Cruz (CBMEG/UNICAMP)

Identificação do paciente:

Nome do paciente: _____

Registro Hospitalar: _____

Endereço: _____

Fone: _____

Responsável: _____

RG: _____

Grau de parentesco: _____

Justificativa e objetivos da pesquisa:

Este projeto tem por objetivo o estudo dos genes envolvidos no acidente vascular cerebral (AVC) em indivíduos com anemia falciforme. Os indivíduos serão selecionados de acordo com as características da idade, gênero e status da anemia falciforme e de AVC, se afetado ou não, segundo histórico clínico.

Risco e desconforto:

As imagens serão obtidas por meio de exame de Ressonância Magnética, um procedimento diagnóstico no qual o(a) paciente é colocado(a) em um forte campo magnético, onde ondas de rádio-frequência são enviadas aos tecidos no interior do corpo. Os "ecos" retornantes são então transformados em imagens das estruturas internas do corpo através de processamento computadorizado. O protocolo adotado nesta pesquisa não inclui uso de meio de contraste.

Se você está grávida, ou imagina que possa estar grávida, deve relatar disso antes da realização do exame de Ressonância Magnética. A Ressonância Magnética não poderá ser realizada em pacientes com marca-passo cardíaco ou defibriladores, alguns tipos de clips de aneurisma intra-craniano, corpo estranho metálico intra-ocular, e alguns outros tipos de implantes ou próteses metálicas. Excetuando-se os casos acima citados e possível desconforto decorrente do tempo em decúbito no equipamento de imagem, não há maiores riscos para o indivíduo submetido ao exame.

Benefícios esperados:

Existe a necessidade de se associar os testes genéticos aos já existentes programas de prevenção e diagnóstico precoce da AVC, a fim de se conhecer a magnitude deste problema e dar um passo de alerta à população, aos profissionais da saúde e às autoridades locais. Isto gera apenas benefícios à população afetada e a seus familiares, pois a precocidade do diagnóstico e reabilitação gera grandes reduções no orçamento público.

Nos casos em que for detectado segmento gênico associado ao quadro, o diagnóstico etiológico poderá ser elucidado, instituindo-se a reabilitação precoce.

Garantia de resposta às perguntas:

Será garantida, aos pacientes e responsáveis, respostas a quaisquer perguntas que possam ocorrer, em qualquer momento da pesquisa, e esclarecimento de qualquer dúvida acerca dos assuntos relacionados com esta pesquisa.

Garantia de privacidade:

Será mantido o sigilo e o caráter confidencial das informações, e a identificação dos pacientes não será exposta nas conclusões ou publicações.

Compromisso com informações atualizadas do estudo:

Durante todo o decorrer da pesquisa será mantido o compromisso de se proporcionar informações atualizadas referentes aos resultados dos procedimentos realizados.

Liberdade de abandonar a pesquisa sem prejuízo para si:

Eu, paciente ou responsável, aceito que a participação neste projeto de pesquisa é voluntária e que eu posso recusar ou retirar meu consentimento, a qualquer momento, sem comprometer os cuidados médicos que recebo ou receberei no futuro.

Declaro que recebi cópia do presente Termo de Consentimento

Nome do Paciente ou Responsável

Assinatura do Responsável pela Pesquisa

ANEXO IV

Folha de Rosto do CONEP



MINISTÉRIO DA SAÚDE
Conselho Nacional de Saúde
Comissão Nacional de Ética em Pesquisa - CONEP

FOLHA DE ROSTO PARA PESQUISA ENVOLVENDO SERES HUMANOS				FR - 267422	
Projeto de Pesquisa AVALIAÇÃO DAS VARIACÕES NO NÚMERO DE CÓPIAS (CNVs) NA SUSCETIBILIDADE AO ACIDENTE VASCULAR CEREBRAL EM PACIENTES COM ANEMIA FALCIFORME					
Área de Conhecimento 4.00 - Ciências da Saúde - 4.01 - Medicina - Nenhum				Grupo Grupo II	Nível Não se aplica
Área(s) Temática(s) Especial(s) Genética Humana,				Fase Não se Aplica	
Unitermos anemia falciforme, acidente vascular cerebral, CNVs					
Sujeitos na Pesquisa					
Nº de Sujeitos no Centro 240	Total Brasil 240	Nº de Sujeitos Total 240	Grupos Especiais Criança e ou menores de 18 anos.		
Placebo NAO	Medicamentos HIV / AIDS NÃO	Wash-out NÃO	Sem Tratamento Especifico NÃO	Banco de Materiais Biológicos SIM	
Pesquisador Responsável					
Pesquisador Responsável Mônica Barbosa de Melo			CPF 357.258.771-91	Identidade 319873	
Área de Especialização GENÉTICA			Maiores Titulação DOUTORADO	Nacionalidade BRASILEIRA	
Endereço CBMEG - UNICAMP - AV. CÂNDIDO RONDON, 400			Bairro BARÃO GERALDO	Cidade CAMPINAS - SP	
Código Postal 13083-875	Telefone 19-3521-1143 / 19-9193-1130	Fax 19-3521-1089	Email melomb@uol.com.br		
Termo de Compromisso					
Declaro que conheço e cumprirei os requisitos da Res. CNS 196/96 e suas complementares. Comprometo-me a utilizar os materiais e dados coletados exclusivamente para os fins previstos no protocolo e publicar os resultados sejam eles favoráveis ou não.					
Aceito as responsabilidades pela condução científica do projeto acima.					
Data: 03 / 06 / 2009.			Assinatura 		
Instituição Onde Será Realizado					
Nome UNICAMP/Faculdade de Ciências Médicas - SP		CNPJ 04.606.842/0001-33	Nacional/Internacional Nacional		
Unidade/Órgão Hemocentro		Participação Estrangeira NÃO	Projeto Multicêntrico NÃO		
Endereço Rua Tessália Vieira de Camargo 126		Bairro Barão Geraldo	Cidade Campinas - SP		
Código Postal 13084970	Telefone 19 35218936	Fax 19 35218936	Email cep@fcm.unicamp.br		
Termo de Compromisso					
Declaro que conheço e cumprirei os requisitos da Res. CNS 196/96 e suas complementares e como esta instituição tem condições para o desenvolvimento deste projeto, autorizo sua execução.					
Nome: <u>Carmino De Souza</u>			Assinatura 		
Data: <u>01 / 06 / 09</u>			Assinatura 		
Patrocinadora					
Nome Fundação de Amparo a Pesquisa do Estado de São Paulo - FAPESP			CNPJ 50.644.053/0001-10		
Endereço Rua Pio XI 1500		Bairro Alto da Lapa	Cidade São Paulo - SP		
Código Postal 05468901	Telefone 3645-2383	Fax (11) 3838-4000	Email		
Responsável Mônica Barbosa de Melo			Cargo / Função Pesquisadora		

ANEXO V

Parecer do Comitê de Ética da FCM



FACULDADE DE CIÊNCIAS MÉDICAS
COMITÊ DE ÉTICA EM PESQUISA

www.fcm.unicamp.br/pesquisa/etica/index.html

CEP, 29/07/09.
(Grupo III)

PARECER CEP: Nº 516/2009 (Este nº deve ser citado nas correspondências referente a este projeto)
CAAE: 0413.0.146.000-09

I - IDENTIFICAÇÃO:

PROJETO: “AVALIAÇÃO DAS VARIAÇÕES NO NÚMERO DE CÓPIAS (CNVS) NA SUSCETIBILIDADE AO ACIDENTE VASCULAR CEREBRAL EM PACIENTES COM ANEMIA FALCIFORME”.

PESQUISADOR RESPONSÁVEL: Mônica Barbosa de Melo

INSTITUIÇÃO: Hemocentro/UNICAMP

APRESENTAÇÃO AO CEP: 09/06/2009

APRESENTAR RELATÓRIO EM: 23/06/10 (O formulário encontra-se no *site* acima)

II - OBJETIVOS

Identificar região do genoma envolvido com o aumento do risco para AVC em pacientes com anemia falciforme através do estudo de CNVs por meio de *microarrays* de alta densidade.

III - SUMÁRIO

Estudo com coleta de 20-30 ml de sangue de 200 pacientes portadores de anemia falciforme (100 com AVC e 100 sem AVC). Também serão excluídos 40 pacientes normais. Extração de ADN. Copy Number Variation (CNVs) – Análise por meio de software.

IV - COMENTÁRIOS DOS RELATORES

Trata-se de um projeto de pós-doutorado e vinculado a um projeto temático de anemia falciforme. Após respostas às pendências tanto o projeto como o Termo de Consentimento Livre e Esclarecido estão adequados.

V - PARECER DO CEP

O Comitê de Ética em Pesquisa da Faculdade de Ciências Médicas da UNICAMP, após acatar os pareceres dos membros-relatores previamente designados para o presente caso e atendendo todos os dispositivos das Resoluções 196/96 e complementares, resolve aprovar sem restrições o Protocolo de Pesquisa, o Termo do Consentimento Livre e Esclarecido, bem como todos os anexos incluídos na pesquisa supracitada.

O conteúdo e as conclusões aqui apresentados são de responsabilidade exclusiva do CEP/FCM/UNICAMP e não representam a opinião da Universidade Estadual de Campinas nem a comprometem.



**FACULDADE DE CIÊNCIAS MÉDICAS
COMITÊ DE ÉTICA EM PESQUISA**

www.fcm.unicamp.br/pesquisa/etica/index.html

VI - INFORMAÇÕES COMPLEMENTARES

O sujeito da pesquisa tem a liberdade de recusar-se a participar ou de retirar seu consentimento em qualquer fase da pesquisa, sem penalização alguma e sem prejuízo ao seu cuidado (Res. CNS 196/96 – Item IV.1.f) e deve receber uma cópia do Termo de Consentimento Livre e Esclarecido, na íntegra, por ele assinado (Item IV.2.d).

Pesquisador deve desenvolver a pesquisa conforme delineada no protocolo aprovado e descontinuar o estudo somente após análise das razões da descontinuidade pelo CEP que o aprovou (Res. CNS Item III.1.z), exceto quando perceber risco ou dano não previsto ao sujeito participante ou quando constatar a superioridade do regime oferecido a um dos grupos de pesquisa (Item V.3.).

O CEP deve ser informado de todos os efeitos adversos ou fatos relevantes que alterem o curso normal do estudo (Res. CNS Item V.4.). É papel do pesquisador assegurar medidas imediatas adequadas frente a evento adverso grave ocorrido (mesmo que tenha sido em outro centro) e enviar notificação ao CEP e à Agência Nacional de Vigilância Sanitária – ANVISA – junto com seu posicionamento.

Eventuais modificações ou emendas ao protocolo devem ser apresentadas ao CEP de forma clara e sucinta, identificando a parte do protocolo a ser modificada e suas justificativas. Em caso de projeto do Grupo I ou II apresentados anteriormente à ANVISA, o pesquisador ou patrocinador deve enviá-las também à mesma junto com o parecer aprovatório do CEP, para serem juntadas ao protocolo inicial (Res. 251/97, Item III.2.e)

Relatórios parciais e final devem ser apresentados ao CEP, de acordo com os prazos estabelecidos na Resolução CNS-MS 196/96.

III – DATA DA REUNIÃO

Homologado na VI Reunião Ordinária do CEP/FCM, em 23 de junho de 2009.

Prof. Dr. Carlos Eduardo Steiner
PRESIDENTE DO COMITÊ DE ÉTICA EM PESQUISA
FCM/UNICAMP

ANEXO VI

Profa. Dra. Rachel Meneguello
Presidente
Comissão Central de Pós-Graduação
Declaração

As cópias de artigos de minha autoria ou de minha co-autoria, já publicados ou submetidos para publicação em revistas científicas ou anais de congressos sujeitos a arbitragem, que constam da minha Dissertação/Tese de Mestrado/Doutorado, intitulada **AVALIAÇÃO DAS VARIACÕES NO NÚMERO DE CÓPIAS (CNVs) NA SUSCETIBILIDADE AO ACIDENTE VASCULAR CEREBRAL EM PACIENTES COM ANEMIA FALCIFORME**, não infringem os dispositivos da Lei n.º 9.610/98, nem o direito autoral de qualquer editora.

Campinas, 23 de Jul. de 16

Assinatura : 
Nome do(a) autor(a): **Pedro Rodrigues Sousa da Cruz**
RG n.º 42117090-6

Assinatura : 
Nome do(a) orientador(a): **Mônica Barbosa de Melo**
RG n.º 319.873
