

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

**APLICAÇÃO DO ÍNDICE DE ANCESTRALIDADE AFRICANA PARA O
CONTROLE DE ESTRATIFICAÇÃO POPULACIONAL EM ESTUDOS DE
ASSOCIAÇÃO**

Verônica Marques Zembrzuski

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da Universidade Federal do Rio Grande do Sul como requisito parcial para obtenção do grau de Mestre.

Orientadora: Profa. Dra. Mara Helena Hutz

Co-Orientadora: Profa. Sidia M. Callegari-Jacques

Porto Alegre, março de 2005

INSTITUIÇÕES E FONTES FINANCIADORAS

As pesquisas realizadas no Laboratório de DNA do Departamento de Genética da Universidade Federal do Rio Grande do Sul foram subvencionadas pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e pelo Programa de Apoio a Núcleos de Excelência (PRONEX).

A aluna recebeu bolsa de estudos concedida pelo CNPq.

SUMÁRIO

RESUMO	07
ABSTRACT	08
CAPÍTULO I – INTRODUÇÃO	09
1. Considerações gerais	10
2. Estudos de associação do tipo caso-controle	10
3. Estrutura populacional	11
4. Alternativas aos estudos de associação do tipo caso-controle	14
4.1 Método de controle genômico	15
4.2 Marcadores população-específicos	16
5. Raça	20
6. Características da população	24
CAPÍTULO II – JUSTIFICATIVA E OBJETIVOS	26
CAPÍTULO III – APPLICATION OF AN AFRICAN ANCESTRY INDEX AS A GENOMIC CONTROL APPROACH IN A BRAZILIAN POPULATION	29
CAPÍTULO IV – DISCUSSÃO	48
REFERÊNCIAS BIBLIOGRÁFICAS	53

RESUMO

A população brasileira possui um considerável grau de miscigenação. O objetivo deste trabalho foi verificar a ocorrência de estratificação populacional em um estudo do tipo caso-controle em uma amostra da população de Porto Alegre considerada fenotipicamente de ancestralidade européia. Foram genotipados, por métodos baseados em PCR, 10 marcadores informativos de ancestralidade africana. A comparação das freqüências alélicas não mostrou diferenças significativas em casos e controles, sugerindo ausência de estratificação populacional, embora as estimativas de mistura racial tenham indicado uma baixa introgressão (2 e 6%, respectivamente) de genes africanos nessas amostras. A utilização de um Índice de Ancestralidade Africana (IAA) também indicou que não havia diferenças significantes entre casos e controles. Como esse índice permite a avaliação da mistura racial ao nível individual, foi possível, retirando os indivíduos com valores de IAA maiores do que o descrito na literatura para os portugueses, homogeneizar ainda mais as amostras. Portanto, no presente trabalho propomos a utilização do Índice de Ancestralidade Africana como Método de Controle Genômico em estudos de associação, evitando associações espúrias devido à estratificação populacional.

ABSTRACT

The Brazilian population has a high degree of admixture. The objective of this investigation was to verify the occurrence of population stratification in a case-control study in a sample from the Porto Alegre population considered phenotypically as of European ancestry. Ten African ancestry informative markers were genotyped by PCR based methods. The allele frequency comparisons did not show statistically significant differences between cases and controls, suggesting an absence of population stratification, although racial admixture estimates showed a low introgression (2 and 6% respectively) of African genes in these samples. The use of an African Ancestry Index (AAI) also did not show significant differences between cases and controls. As this index allow to estimate admixture at the individual level, it was possible to select those individuals with AAI values higher than those described in the literature for Portuguese and drop them from the samples, making cases and controls more homogeneous. Therefore, in the present study we propose the use of the African Ancestry Index as a Genomic Control Method for association studies, avoiding spurious associations due to population stratification.

CAPÍTULO I
INTRODUÇÃO

1. Considerações gerais

Os estudos de associação são amplamente aplicados para identificar a base genética de características quantitativas, como a suscetibilidade às doenças complexas. Esses estudos possibilitam a detecção de efeitos genéticos pequenos a moderados e possibilitam testar interações com os diversos fatores ambientais envolvidos (Deng, 2001).

Dependendo da característica sob investigação, os estudos de associação podem ser do tipo caso-controle ou baseados em famílias. Os estudos do tipo caso-controle examinam as diferenças das frequências alélicas de genes candidatos entre indivíduos afetados e controles, não-afetados e não-relacionados (Risch, 2000). Já os estudos baseados em famílias levam em conta a transmissão da característica dos pais para um filho afetado, ou seja, os alelos não-transmitidos são usados como controle (Cardon e Palmer, 2003).

2. Estudos de associação do tipo caso-controle

Ao contrário do que acontece nos estudos baseados em famílias, em estudos de associação do tipo caso-controle, há o risco de se encontrar

associações espúrias devido às diferenças no *background* genético dos indivíduos casos e dos controles. Pritchard e Rosenberg (1999) utilizaram o termo “associação espúria” para descrever uma associação entre um fenótipo e um loco marcador, quando esse marcador não está relacionado a quaisquer locos envolvidos com o fenótipo sob investigação.

Em estudos do tipo caso-controle, diferenças em fatores como idade e sexo, por exemplo, podem levar a um excesso de resultados falso-positivos e/ou falso negativos (Risch, 2000). Estrutura populacional e mistura recente também podem levar a tais resultados (Pritchard e Rosenberg, 1999; Devlin e cols., 2001; Ardlie e cols., 2002; Kittles e cols., 2002; Hoggart e cols., 2003; Ziv e Burchard, 2003). Apesar disso, há poucos exemplos claros de estudos em que a estrutura da população levou a associações espúrias (Thomas e Witte, 2002; Cardon e Palmer, 2003).

3. Estrutura populacional

Quando há diferenças entre grupos de indivíduos em uma população, diz-se que ela é estratificada ou subdividida. Entretanto, essas diferenças podem, muitas vezes, não ser detectadas. Assim, a não-detecção de estratificação populacional pode levar a associações espúrias entre um marcador candidato e um determinado fenótipo (Lander e Schork, 1994; Pritchard e Rosenberg, 1999; Pritchard e cols., 2000; Risch, 2000; Pritchard e Donnelly, 2001; Cardon e Palmer, 2003; Hoggart e cols., 2003; Hinds e cols., 2004).

A detecção da diferença entre casos e controles depende, basicamente, da prevalência da doença entre as sub-populações e suas diferenças genéticas. Se uma mutação aumenta a susceptibilidade a uma doença, então se espera que ela seja mais freqüente entre os indivíduos afetados (casos) do que entre os indivíduos não-afetados (controles) (Pritchard e Donnelly, 2001). A possibilidade de associações falsas deve-se, principalmente, à população amostrada, à característica em estudo e/ou ao marcador que está sendo testado (Ziv e Burchard, 2003). Talvez o fator mais importante que leva a tais associações está relacionado às características da população, principalmente a sua etnia (Risch, 2000; Santos e cols., 2002). A idéia é que membros de uma “raça” pré-estabelecida compartilham uma ancestralidade comum que pode incluir fatores genéticos de risco e predisposição a doenças (Kittles e Weiss, 2003). Diferenças sistemáticas na ancestralidade entre casos e controles podem ser encontradas quando subgrupos populacionais distintos geneticamente têm uma prevalência diferente do fenótipo alvo (Hinds e cols., 2004).

É importante, entretanto, distinguir entre heterogeneidade e estrutura dentro de uma amostra caso-controle. Se a heterogeneidade é equivalente em casos e controles (isto é, os dois grupos têm a mesma mistura de subgrupos étnicos/genéticos) a estratificação não ocorrerá (Ardlie e cols., 2002). Se a freqüência alélica de um loco candidato é equivalente entre os diferentes subgrupos populacionais (por exemplo, para populações miscigenadas), associações espúrias não se originariam. Similarmente, se a prevalência da doença é a mesma entre os subgrupos, as associações espúrias também não se originariam, mesmo que a freqüência alélica do loco candidato variasse por

subgrupo. Já no caso de doenças genéticas complexas com múltiplos fatores de risco ambientais e genéticos, é possível que em alguns casos o grupo de alto risco tenha uma baixa prevalência do alelo de alto risco, se outras variantes genéticas e fatores de risco ambientais forem mais altos naquele grupo (Ziv e Burchard, 2003).

Um dos exemplos mais conhecidos de associação espúria devido à estratificação populacional foi descrito por Knowler e cols. (1988), em um estudo de associação com diabetes não insulino-dependente nas tribos Pima e Papago (nativos americanos). Os autores verificaram que muitos dos indivíduos amostrados tinham ancestralidade europeia recente, e, ainda, que a proporção média da mesma nos controles era mais alta que nos indivíduos afetados. Os europeus possuem baixa suscetibilidade ao diabetes, enquanto que os indivíduos dessas tribos possuem uma alta frequência de diabetes. Os dados indicaram uma forte associação negativa entre diabetes e um haplótipo do loco da imunoglobulina G. Os resultados podem ser espúrios porque indivíduos Pima com um alto grau de ancestralidade caucasóide têm baixa suscetibilidade ao diabetes. Então, locos marcadores que estavam em frequências mais altas nos Pima do que nos caucasóides foram “associados” com a doença (Pritchard e Rosenberg, 1999).

É interessante salientar que Ardlie e cols. (2002) sugeriram que a estratificação populacional não pode ser responsável pelo excesso de resultados falso-positivos nos estudos de associação, e que um cuidadoso pareamento étnico entre casos e controles seria suficiente para evitar esse tipo de problema. Mas como seria feito um cuidadoso pareamento étnico?

4. Alternativas aos estudos de associação do tipo caso-controle

Uma solução encontrada para os problemas citados acima é o uso dos testes de associação baseados em famílias (ou teste de ligação em famílias), pois esses não sofrem do problema de resultados falso-positivos devido à estrutura populacional. Entretanto, esses estudos podem ser caros e, algumas vezes, impraticáveis, principalmente para doenças de início tardio (Pritchard e Donnelly, 2001; McGinnis e cols., 2002), pois se torna difícil coletar amostras do indivíduo afetado e de pelo menos um dos seus genitores, visto que os mesmos já podem ter falecido.

Outra solução seria o uso de um modelo de caso-controle “correspondente”, no qual os controles são etnicamente equiparados aos casos (Pritchard e Donnelly, 2001; Hinds e cols., 2004). Esse modelo poderia ser aplicado às populações homogêneas e com casamentos ao acaso, apesar das mesmas não representarem a realidade. Entretanto, estudos de associação realizados na Finlândia (população considerada relativamente homogênea) são comumente menos sujeitos a fatores de “confusão” do que os realizados em populações heterogêneas da América do Norte (Risch, 2000). Essa heterogeneidade é ainda maior entre os subgrupos latinos que possuem uma proporção de ancestralidade africana variável: porto-riquenhos têm uma ancestralidade africana estimada em 30%, enquanto mexicanos têm aproximadamente 9% (Hanis e cols., 1991).

O que se pode fazer nesses casos é realizar, após a coleta de dados da amostra, a “estratificação” das análises de acordo com a etnia relatada (Pritchard e Donnelly, 2001; Hinds e cols., 2004).

4.1. Método de controle genômico

Muitos métodos têm sido desenvolvidos para que se possa utilizar um grande número de marcadores, ou para identificar a estrutura básica da população ou ajustar o viés devido à estratificação populacional (Zhao e cols., 2003).

O método de controle genômico, inicialmente descrito por Devlin e Roeder (1999), utiliza locos de marcadores independentes para ajustar a distribuição de um teste estatístico padrão (Pritchard e Donnelly, 2001), examinando a distribuição da estatística de associação (χ^2) entre as variantes genéticas não-ligadas genotipadas em casos e controles (Freedman e cols., 2004). Assim, a detecção de estrutura nas sub-populações depende, além do tamanho amostral, principalmente da seleção dos marcadores genéticos analisados. Entretanto, o poder para detectar estrutura populacional é altamente dependente do número de locos randômicos utilizados (Turakulov e Easteal, 2003).

Pritchard e Rosenberg (1999) propuseram um método em que marcadores genéticos não-ligados, entre si e com a doença, são genotipados na amostra de casos e controles para detectar diferenças não esperadas nas frequências gênicas, o que evidenciaria a estratificação populacional. Nesse trabalho, os autores sugerem, com base nos dados de simulação, que de 15 a 20

microsatélites seriam suficientes para testar uma fraca estratificação com uma probabilidade de erro do tipo I menor que 5%. No entanto, se marcadores bialélicos forem utilizados, mais locos seriam necessários.

Apesar dos marcadores de DNA distinguirem bem os indivíduos (Petkovski e cols., 2003), a grande maioria deles tem pouco poder para detectar diferenças étnicas ou características de um indivíduo. No caso de populações miscigenadas, a escolha de marcadores com grandes diferenças de frequências alélicas nas populações parentais aumentaria o poder para detectar a estratificação, sendo poucos marcadores necessários (Shriver e cols., 1997).

4.2. Marcadores população-específicos

Marcadores informativos de ancestralidade (AIMs) podem ser aplicados para estimar ancestralidade biogeográfica, em nível populacional, de subgrupos (casos e controles) e individual (Kittles e Weiss, 2003; Shriver e cols., 2003). Um parâmetro muito utilizado neste tipo de estudo é a diferença de frequência alélica (δ), definida como o valor absoluto da diferença de frequência entre populações ancestrais (Dean e cols., 1994; Shriver e cols., 1997; Parra e cols., 1998; Pfaff e cols., 2001; Kittles e cols., 2002; Shriver e cols., 2003). Por possuírem uma alta informação sobre ancestralidade, esses marcadores aumentam o poder para detectar associações espúrias entre casos e controles.

Shriver e cols. (1997) identificaram um conjunto de marcadores genéticos população-específicos e demonstraram que esses marcadores seriam úteis, tanto para as estimativas de mistura interétnica (individual e populacional), como para o

mapeamento de genes pelo uso do desequilíbrio de ligação criado por fluxo gênico entre as populações, pois devido às grandes diferenças nas frequências alélicas, mostrariam um alto poder de discriminação nos testes estatísticos.

Parra e cols. (1998) sugeriram o uso de um conjunto de marcadores autossômicos que possuem uma alta diferença de frequência entre populações européias e africanas (δ maior que 45%) ou são população-específicos (Tabela 1). Nesse trabalho, os autores analisaram a contribuição genética européia em 10 populações de afro-americanos dos Estados Unidos e da Jamaica. O grau de mistura européia variou de 6,8% (Jamaica) a 22,5% (New Orleans, Estados Unidos). Uma associação, não ao acaso, significativa entre os marcadores nucleares FY-null e AT3 (separados por 22 cM), foi detectada provavelmente devido ao desequilíbrio de ligação criado durante a miscigenação das duas populações parentais (africana e européia). Lautenberger e cols. (2000) e Pfaff e cols. (2001), trabalhando com simulações de modelos de mistura no computador, também encontraram desequilíbrio de ligação significativo entre FY-null e AT3. No trabalho conduzido por Pfaff e cols. (2001) foram utilizados 11 marcadores, descritos por Shriver e cols. (1997) e Parra e cols. (1998), em amostras populacionais de afro-americanos de Jackson, MS (Mississippi), e da costa da Carolina do Sul, nos Estados Unidos. Ambas as amostras mostraram associação significativa entre pares de marcadores não-ligados (37% de pares não-ligados na amostra de Jackson e 20%, na amostra da Carolina do Sul). O número significativo de associações observadas entre os locos não-ligados indica a presença de estrutura genética nessas amostras, e poderia resultar em uma alta taxa de falso-positivos em estudos de mapeamento de mistura recente.

Tabela 1. Descrição dos marcadores autossômicos informativos de ancestralidade.

Marcador	Localização Cromossômica ^a	Frequências alélicas médias ^b	
		Africanos	Europeus
APO*1	11q23	0,441	0,927
AT3*1	1q23-q25	0,874	0,279
FY-NULL*1	1q22-q23	0,000	1,000
ICAM*1	19p13.3-p13.2	0,756	1,000
LPL*1	8p22	0,973	0,486
OCA2*1	15q11.2-q12	0,098	0,769
RB2300*1	13q14.3	0,920	0,333
Sb19.3*1	19	0,425	0,910
GC-1F	4q12-q13	0,824	0,156
GC-1S	4q12-q13	0,078	0,607

NOTA: o alelo 1 representa a presença das inserções *Alu* e a ausência dos sítios de restrição nos polimorfismos; para o loco GC, os alelos são nomeados conforme as designações alélicas dos ensaios baseados em proteínas.

^a Pfaff e cols. (2001).

^b Parra e cols. (1998).

Kittles e cols. (2002) também utilizaram em seus estudos um conjunto de marcadores moleculares (dez ao todo), com altas diferenças nas frequências alélicas entre as populações ancestrais africanas e européias, e encontraram uma diferença significativa no *background* genético de casos e controles em afro-americanos, mostrando, assim, uma forte evidência de estratificação populacional. Esse resultado é importante porque tanto as amostras de casos como as de controles foram correspondentes em termos da etnia relatada pelos próprios indivíduos.

Já em um estudo onde a pigmentação da pele (medida por reflectometria) foi utilizada como modelo fenotípico, Shriver e cols. (2003) utilizaram trinta e quatro marcadores informativos de ancestralidade (AIMs), com o objetivo de identificar genes associados com a cor de pele. Foram observadas correlações significantes entre as estimativas de ancestralidade individual e pigmentação da pele. Os resultados mostraram que OCA2 [um dos marcadores também utilizado por Parra e cols. (1998)] teria efeitos mensuráveis nas diferenças de pigmentação da pele entre populações parentais do oeste africano e do oeste da Europa. Bonilla e cols. (2004) também utilizaram AIMs (22 marcadores ao todo) para descrever o processo e a dinâmica de mistura na população hispânica de San Luis Valley (Colorado, Estados Unidos), obtendo uma alta proporção de ancestralidade nativa americana ($34,1 \pm 1,9\%$), comparada aos $62,7 \pm 2,1\%$ de ancestralidade européia e $3,2 \pm 1,5\%$ de africanos.

5. Raça

A raça, entre outras características, deve-se, principalmente, à adaptação populacional ao ambiente de cada nicho geográfico. Grupos étnicos definidos culturalmente não refletem adequadamente a estrutura básica da população (Pritchard e Donnelly, 2001).

Raça é um conceito difícil de ser definido rigorosamente. A definição clássica é que raças diferem nas frequências de, pelo menos, alguns genes (Kittles e Weiss, 2003). O termo tem uma conotação complexa, que reflete cultura, história, status político e sócio-econômico, bem como suas origens geográficas ancestrais. Cultura, linguagem, religião e etnia têm fortes componentes sócio-culturais e nem sempre são bons indicadores de ancestralidade compartilhada. E a origem geográfica pode não ser adequada para definir raça, devido às recentes migrações de pessoas, tanto históricas como pré-históricas (Tishkoff e Kidd, 2004). Em muitos casos, as origens ancestrais têm uma correlação, embora imprecisa, com raça ou etnia auto-classificada, apesar de não ser estritamente verdadeiro que raça ou etnia não tenha uma conexão biológica. Ainda, há muitas outras conotações não-genéticas de raça, bem como falta de limites definidos entre as populações; e muitos indivíduos têm ancestrais de múltiplas regiões do mundo (Collins, 2004).

Características morfológicas como cor da pele, textura do cabelo, formato e tamanho da cabeça não definem raça (Kittles e Weiss, 2003), porque provavelmente resultam da adaptação às condições ambientais, podem ter sofrido evolução convergente (Tishkoff e Kidd, 2004) e, principalmente, não descrevem a

maioria dos indivíduos com a complexidade requerida para entender e reconhecer nossa diversidade genética (Patrinos, 2004).

A maioria da variação vista hoje originou-se no período de tempo no qual os humanos distribuíram-se ao longo do Velho Mundo; o aumento do tamanho populacional ocorreu nos lugares onde a agricultura ou outros fatores permitiu. A quantidade de variação genética humana pode ser medida por π (diversidade nucleotídica), que entre as diversas populações humanas é quatro vezes menor do que a observada em chimpanzés. Esta diferença pode ser indicativa de uma ancestralidade recente dos humanos modernos (Kittles e Weiss, 2003). Aproximadamente, 90% do total da variação genética detectada ocorre entre indivíduos de um mesmo continente e apenas 10% entre populações de continentes diferentes. Portanto, o padrão de variação compartilhada tem implicações importantes para o nosso entendimento sobre as diferenças e similaridades entre as populações (Jorde e Wooding, 2004).

As populações nunca são “puras” em um sentido genético, e definir limites entre indivíduos ou populações pode ser inadequado, pois elas tendem a formar agrupamentos de acordo com a distância geográfica que há entre elas. As populações da África são as que mais divergem, e a maior distância genética é vista entre essas e as não-africanas. Por exemplo, a anemia falciforme é mais comum em africanos e populações do Mediterrâneo do que em populações do norte da Europa; o contrário é verdadeiro para a fibrose cística e a hemocromatose (Jorde e Wooding, 2004).

A variação genética baseada nas populações é muito usada para explicar diferenças na saúde entre grupos raciais e étnicos (Royal e Dunston, 2004).

Respostas a terapias médicas são freqüentemente comparadas entre populações que são classificadas de acordo com as divisões raciais tradicionais (Jorde e Wooding, 2004).

Tishkoff e Kidd (2004) mostraram que as classificações raciais descrevem inadequadamente a distribuição da variação genética em nossa espécie. Entretanto, como alguns genes que determinam algumas doenças podem ser geograficamente restritos devido à mutação, deriva genética, migração e seleção natural, o conhecimento da ancestralidade individual é importante para estudos biomédicos.

Com o seqüenciamento do genoma humano, e os recentes avanços na identificação e genotipagem da variação genética em centenas de locos e em centenas de indivíduos, é possível fornecer-se maiores detalhes sobre o entendimento do padrão global de variação genética. Através do processo de deriva genética, as populações se diferenciaram através do tempo, e o isolamento genético de grupos étnicos foi reforçado pelos cruzamentos aleatórios e a endogamia localizada baseada em fatores sócio-culturais, como linguagem e etnia. Devido à grande mobilidade que temos hoje, a distância não é uma barreira; barreiras sócio-culturais contra casamentos interétnicos foram reduzidas e a miscigenação tornou-se muito comum (Tishkoff e Kidd, 2004).

O entendimento de etnia ou raça (não apenas relacionado à ancestralidade geográfica) e testes estatísticos de subestrutura são importantes para o delineamento dos estudos de associação do tipo caso-controle e para a identificação de alelos predispondo doenças, que podem diferir entre os grupos étnicos (Tishkoff e Kidd, 2004).

Outro elemento central na maioria das discussões sobre raça e genética é a pigmentação da pele. Pesquisas baseadas na genética da variação populacional desse fenótipo são escassas (Parra e cols. 2004).

A pigmentação da pele é um dos mais variáveis fenótipos em humanos, mas muito pouco se conhece sobre a base genética e história evolutiva desta característica poligênica. Ainda, pode servir de base para o racismo. Diferenças na cor da pele entre as populações são comumente (e incorretamente) entendidas com uma indicação de grandes diferenças biológicas entre as populações. Pigmentação da pele é utilizada como um modelo fenotípico para ancestralidade e estudos de mapeamento de mistura. Entretanto, a pigmentação constitutiva, isto é, o conteúdo de melanina de áreas não-expostas da pele, é uma característica poligênica que é relativamente não-afetada por fatores ambientais (ver referências de Shriver e cols., 2003 e Parra e cols., 2004).

Parra e cols. (2004) estudaram a relação entre a pigmentação da pele e a ancestralidade em cinco populações, cultural e geograficamente definidas com diferentes proporções de ancestralidade. O principal objetivo desse estudo foi avaliar se a pigmentação constitutiva e a ancestralidade estão correlacionadas em populações miscigenadas. Para estimar as ancestralidades relativas em cada indivíduo, foram utilizados marcadores genéticos com grandes diferenças de frequência entre as populações parentais (do oeste africano, europeia e nativa americana) e pequenas diferenças dentro dessas regiões. Apesar da relação entre a cor da pele e ancestralidade ter sido variável, observou-se uma correlação significativa entre pigmentação constitutiva e ancestralidade individual em cada uma das cinco amostras populacionais. Os resultados desse trabalho

demonstraram a utilidade dos marcadores genéticos informativos de ancestralidade e métodos de mistura, e enfatizam a necessidade de cuidado ao usar a pigmentação como indicador de ancestralidade, ou então extrapolar os resultados de uma população miscigenada para outra.

6. Características da população

A análise de dados sobre raça no Brasil mostra um gradiente – do Norte para o Sul – nas proporções relativas das cores da pele: brancos totalizam 22,7% da população na região Norte e 83,3% na região Sul. Nota-se ainda que o Sudeste é a região em que as proporções mais se assemelham às do Brasil como um todo (IBGE, 1991).

No Rio Grande do Sul, alguns estudos já foram realizados visando avaliar a contribuição de cada etnia para o “pool” gênico da população.

Schneider e Salzano (1979), utilizando o sistema Gm de imunoglobulinas, estimaram em 53% a proporção de genes de ancestrais europeus em afro-brasileiros de Porto Alegre. Franco e cols. (1982), também em uma amostra de afro-brasileiros de Porto Alegre, encontraram 65% de contribuição européia utilizando doze sistemas genéticos. Bortolini e cols. (1997) estimaram a contribuição relativa de africanos no “pool” gênico da população afro-derivada de Porto Alegre e região metropolitana em 41%, enquanto que a contribuição de brancos foi de 59%. Nessa estimativa, 13 sistemas genéticos foram utilizados.

Dornelles e cols. (1999) investigaram, em 2708 indivíduos com descendência européia da população do Rio Grande do Sul, 17 sistemas

genéticos de proteínas. A amostra foi dividida em sete mesoregiões, e a análise da mistura interétnica na mesoregião metropolitana de Porto Alegre mostrou os seguintes resultados: contribuição europeia em 79,6%, africana em 8,1%, e contribuição ameríndia em 12,2%. Em estudos anteriores (Schneider e Salzano, 1979; Franco e cols., 1982) a contribuição africana encontrada em indivíduos de descendência europeia de Porto Alegre foi também de 8%.

Para verificar se o grau de aparência física de um indivíduo está relacionado à ancestralidade genômica africana, Parra e cols. (2003), utilizando os mesmos marcadores moleculares propostos por Parra e cols. (1998), criaram um índice de ancestralidade africana (IAA) individual. O IAA foi calculado como o logaritmo da razão da probabilidade de um dado genótipo ocorrer na população africana pela probabilidade do mesmo ocorrer na população europeia. Os resultados mostraram que brasileiros possuem valores intermediários de IAA entre europeus e africanos. Para a amostra da região sul do Brasil – 52 indivíduos auto-classificados como brancos –, foi encontrado o valor médio de IAA mais baixo, comparado às outras regiões em estudo. Entretanto, nesse trabalho os valores de IAA para a região Sul ainda foram significativamente mais altos quando comparados aos europeus. Esses dados sugeriram que a raça do indivíduo (ou mais popularmente, a sua “cor”) não seria um bom indicativo de ancestralidade genômica para a população brasileira.

CAPÍTULO II
JUSTIFICATIVA E OBJETIVOS

Os estudos de associação do tipo caso-controle são amplamente utilizados em trabalhos que visam entender a etiologia genética e ambiental de muitas doenças complexas em populações humanas. A estrutura da população, entretanto, torna-se um obstáculo para esses estudos, podendo levar a resultados falso-positivos ou falso-negativos, não permitindo detectar a real associação existente entre os genes candidatos e o fenótipo em questão. Além disso, poucos estudos de caso-controle são avaliados para a presença de estrutura populacional (Ardlie e cols., 2002).

No nosso grupo de pesquisa, alguns estudos do tipo caso-controle já foram realizados para verificar se a prevalência e a severidade de cardiopatia isquêmica aterosclerótica está associada com variantes em genes candidatos (ver, por exemplo, Almeida, 2001; De Andrade e cols., 2004; Rios e cols., 2003a,b). Este tipo de delineamento justifica a importância de caracterizar geneticamente a etnia das amostras estudadas, com o intuito de evitar falsas associações.

Este trabalho tem como objetivos específicos:

- 1) Estimar as frequências alélicas, para um conjunto de marcadores informativos de ancestralidade (AIMs) africana e europeia (APO, AT3-I/D, GC, FY-null, LPL, OCA2, RB2300, Sb19.3 e APOA183), na amostra de pacientes com cardiopatia e na amostra de controles da população de Porto Alegre.

- 2) Estimar o grau de mistura racial nas duas amostras.
- 3) A partir das frequências alélicas, calcular o índice de ancestralidade africana (IAA) para cada amostra e para cada indivíduo.
- 4) Verificar se existem diferenças nas medianas do IAA entre casos e controles.
- 5) Verificar a possibilidade de se selecionar casos e controles através do IAA para estudos de associação.

CAPÍTULO III

**APPLICATION OF AN AFRICAN ANCESTRY INDEX AS A GENOMIC
CONTROL APPROACH IN A BRAZILIAN POPULATION**

HUMAN GENETICS (EM PREPARAÇÃO)

Application of an African Ancestry Index as a Genomic Control approach in a Brazilian population

Verônica M. Zembrzuski¹, Sidia M. Callegari-Jacques^{1,2}, Mara H. Hutz¹

¹*Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.*

²*Departamento de Estatística, Instituto de Matemática, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.*

Keywords: case-control studies – spurious association – ancestry

Corresponding author:

Prof. Mara H.Hutz
Departamento de Genética
Instituto de Biociências, UFRGS
Caixa postal 15053
91501-970 Porto Alegre, RS, Brazil
Tel. 55 51 3316-6720
Fax. 55 51 3316-7311
E-mail: mara.hutz@ufrgs.br

V. M. Zembrzuski, S. M. Callegari-Jacques, M. H. Hutz (✉)
Departamento de Genética,
Universidade Federal do Rio Grande do Sul,
Caixa Postal 15053, 91501-970, Porto Alegre, RS, Brazil
Fax: 55 51 33167311
e-mail: mara.hutz@ufrgs.br

Abstract Ten ancestry informative markers were investigated in 101 coronary artery disease patients and 102 healthy controls from a Southern Brazilian population in order to determine if stratification occurs in this population. The degree of African admixture detected in this population was estimated as much as 6%, but no differences between cases and controls were observed. Using an African Ancestry Index (AAI) that estimates admixture at the individual level it was possible to remove from the samples those individuals with evidences of African admixture. Therefore we have shown that it is possible to control for population stratification by choosing individuals without the loss of statistical power as occurs with the use of other methods of genomic control.

Introduction

Association studies have been widely used to help understand the genetic basis of quantitative traits, such as the susceptibility to complex diseases. However there has been much debate about the impact of population stratification on case-control association studies, and the fraction of associations that is attributable to stratification is still unknown (Ardlie et al, 2002; Freedman et al, 2004).

Ethnic population stratification exists when the total population has been formed by admixture between subpopulations, and when admixture proportions vary between individuals (Hoggart et al, 2003). The most likely source of confounding in association studies is ethnicity, whereby allele frequencies vary according to ethnicity, and cases and controls are not adequately matched in terms of ethnicity (Pritchard and Rosenberg, 1999; Risch, 2000). As long as the heterogeneity is equivalent in case and control subjects (i.e., the two groups have the same mix of ethnic/genetic subgroups), stratification bias will not occur (Ardlie et al, 2002). Statistical methods, including use of genomic markers, have been frequently suggested to control for population stratification in genetic association studies (Reich and Goldstein, 2001; Chen et al, 2003).

In order to verify if the degree of physical appearance of an individual is related to his/her African genomic ancestry, Parra et al (2003), using the same set of ancestry informative markers (AIMs) described by Parra et al (1998), created an individual index of African ancestry that allowed to estimate African genomic ancestry at the individual level. Parra et al (2003) also showed that overall, based on self-classification, Brazilians have an intermediate African ancestry index (AAI) between Europeans and Africans.

As in Brazil there might be a strong cultural bias toward claiming European ancestry, we used morphological classification based on skin color and morphological traits instead of self-classification for ethnic ascertainment in association studies (Rios et al, 2003; Mattevi et al, 2004). Within a country of continental size like Brazil, population composition vary widely among regions. As a whole, the population is highly mixed as pointed out by Parra et al (2003), but the southern Brazilian population differs from this general pattern as has been shown by admixture quantification studies. This population is mainly of Portuguese descent, although Italians, Spaniards and Germans have also contributed to its gene pool. A review about the ethnic admixture in Brazilian and other Latin American populations was recently published by Salzano and Bortolini (2002). These authors also showed a good agreement between morphologic classification, based on skin color and morphological traits, and genetic estimates of admixture.

The current study was undertaken to examine if population stratification in this European derived population would be a confounding factor in association studies, and to test if the African ancestry index derived by Parra et al (2003) would be useful to control for population stratification in case-control association studies.

Material and methods

Population sample

The sample comprised 101 individuals with coronary artery disease (cases) and 102 healthy controls. All subjects were of European ancestry as ascertained by skin color and morphological characteristics. Clinical and demographic characteristics of this sample have

been described elsewhere (Rios et al, 2003). All subjects gave informed consent to participate in the study.

Genotyping

The markers studied in this study are those defined by Parra et al (1998) with the exception of ICAM1 which was replaced by APOA1*83 described by Shriver et al (2003). The polymorphisms investigated are listed in Table 1. They had also been used by Parra et al (2003). These genetic markers were chosen because they possess differences of frequency (δ) between the parental populations (European and African) higher than 45% or are population-specific. All markers except APOA1*83 were genotyped as previously described by Parra et al (1998) whereas the primers and protocols from Wang et al (1996) were used for APOA1*83.

Statistical analyses

The estimates of the allele frequencies, tests for Hardy-Weinberg equilibrium, and comparisons between cases and controls for these frequencies were performed using the GENEPOP v.3.1d program (Raymond and Rousset, 1995). The proportions of admixture in the samples were estimated using Long's (1991) weighted least squares (WLS) method (ADMIX program), and Chakraborty's, (1985) approach, which provides least-squares estimates using gene identities probabilities. The calculations for this method were performed by the Admix routine written by R. Chakraborty, modified and adapted for Windows by B. Bertoni (available at <http://www.genetica.fmed.edu.uy/software.htm>). The AAI was calculated for each individual using the allelic frequencies obtained from the literature, as described by Parra et al (2003). The AAI comparison between samples was

performed with the Mann-Whitney U test using SPSS[®] version 8.0. Additionally, the *Structure* program (Pritchard et al, 2000) was used to search for subpopulations of individuals genetically similar (Pritchard and Donnelly, 2001). This program was run with $k=2$ as the predefined setting for the number of populations, with 30,000 iterations for the burn-in period, and 100,000 additional iterations to obtain parameter estimates. The Mann-Whitney U test was also used to compare the individual probabilities of assignment to the inferred cluster 1 (Q values) between cases and controls.

Results

Allele frequencies for the ten polymorphisms investigated are shown in table 1. The genotype frequencies for all markers were in agreement with those expected under Hardy-Weinberg equilibrium (data not shown). No significant differences between cases and controls in the frequency of each marker were observed.

African admixture estimates with both methods showed low levels of admixture in this European derived population (table 2). The introgression of African genes was lower in cases (2%) than in controls (6%), but considering the standard errors of these estimations they were of the same order of magnitude.

Table 3 presents information about AAI in cases and controls. The AAI median for controls (-10.78) was somewhat higher than that estimated for cases (-10.89) but the difference was not statistically significant ($p=0.392$) (Figure 1A). The median values are more similar to that described for the Portuguese sample (-11.73) reported by Parra et al (2003) than to the median observed for southern Brazilians (-9.11) in the same study (Table 3). The highest AAI (-4.86) observed in that Portuguese sample was used as a cut-point value to define European ancestry in our samples. In order to test the use of this

index to select cases and controls in association studies, the AAI was recalculated removing the outliers, individuals with an index $>-4,86$ (9 control and 2 case individuals, values ranging from -4.77 to -0.78). The resulting medians for the controls (-10.92) and cases (-10.98) became practically identical ($p=0.999$) (Table 3; Figure 1B). It should be noticed that the lowest value described by Parra et al (2003) for Africans was +2.86, therefore no overlap with the African AAI was observed in this population.

To corroborate the absence of structure in this European-derived population the *Structure* program was used to identify clusters of individuals in the whole sample of 203 subjects. Again, no evidence of stratification in the total sample was observed (Figure 2), and no significant differences between cases and controls were detected when Q values were compared (cases: mean \pm SE: 0.507 ± 0.023 ; controls: 0.510 ± 0.024 ; Mann-Whitney *U* test, $p=0.903$).

Discussion

Admixed populations are an important resource that can be used to study the genetics of complex disorders. A prerequisite to this application is a better understanding of the admixture proportions and dynamics of the admixture process. Several investigations have been conducted using AIMs in order to describe the admixture process and population dynamics in American admixed populations (Parra et al, 1998; Shriver et al, 2003; Bonilla et al, 2004; Collins-Schramm et al, 2004). These studies provided evidence that AIMs have applicability in admixture mapping as well as to control for structure in association tests. Although these data provide support for these practical applications, they were only validated in few populations.

The emerging picture is that populations do, generally, cluster by broad geographic regions that correspond with common racial classifications, thus knowledge of ethnicity is important for proper design of case-control association studies, and for identifying disease predisposing alleles that may differ across ethnic groups. Several methods have been described to correct for population stratification (Devlin and Roeder, 1999; Pritchard and Rosenberg, 1999; Pritchard et al, 2000; Reich and Goldstein, 2001). These methods are quite general, and they correct mathematically for the degree of stratification at the sample level. The African ancestry index used in the present study can be used at the individual level, therefore it is possible to correct for stratification by removing some individuals from the sample without loss of statistical power due to statistical corrections.

In Brazil, many population studies have been carried out with the objective of assessing the degree of European, African, and Amerindian contributions to their gene pools using blood groups and protein markers (for a review see Salzano and Bortolini, 2002). Most of these studies estimated the African contribution to the southern European - derived population as 8%, which is a close figure to that estimated for our control sample using DNA SNPs (6%). Parra et al (2003) using the same set of SNPs as well as the same estimation method (Long, 1991) showed 13% of African admixture in southern Brazilians, which is more than twice the highest value observed in the present study. Although their sample is not strictly comparable to the one investigated herein because it comprises individuals from all three southern states, the main difference is that Parra et al (2003) used self-classification as the criterion to define European ancestry. As pointed out by Ziv and Burchard (2003) in populations from the United States and Latin America where admixture has been ongoing for several generations self-described ethnicity may be a less accurate predictor of genetic ancestry.

Considering the tri-ethnic nature of the Brazilian population as a whole it would be important to define a set of markers which would be able to identify the Amerindian contribution. Although some authors have identified markers that might be useful for this purpose (Shriver et al, 2003) they were not validated in South American natives.

In conclusion, we suggest that the use of markers that are particularly informative for ancestry to estimate an African ancestry index at the individual level would be a useful method for genomic control in association studies in admixed populations. More empirical data is needed to improve the feasibility of this approach in case-control association studies in admixed population.

Acknowledgements Thanks are due to our colleagues Fabiana M. de Andrade, Marilu Fiegenbaum, Silvana de Almeida and Domingos L. Souza Rios for help in sample collection. Financial support was provided by Programa de Apoio a Núcleos de Excelência (PRONEX, Brazil) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil).

References

- Ardlie KG, Lunetta KL, Seielstad M (2002) Testing for population subdivision and association in four case-control studies. *Am J Hum Genet* 71:304-311
- Bonilla C, Parra EJ, Pfaff CL, Dios S, Marshall JA, Hamman RF, Ferrell RE, Hoggart CL, McKeigue PM, Shriver MD (2004) Admixture in the Hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping. *Ann Hum Genet* 68:139-153
- Chakraborty R (1985) Gene identity in racial hybrids and estimation of admixture rates. In: Neel JV, Ahuja Y, editors. *Genetic microdifferentiation in man and other animals*. New Delhi: Indian Anthropological Association. p 171-180
- Chen HS, Zhu X, Zhao H, Zhang S (2003) Qualitative semi-parametric test for genetic associations in case-control designs under structured populations. *Ann Hum Genet* 67:250-264
- Collins-Schramm HE, Chima B, Morii T, Wah K, Figueroa Y, Criswell LA, Hanson RL, Knowler WC, Silva G, Belmont JW, Seldin MF (2004) Mexican American ancestry-informative markers: examination of population structure and marker characteristics in European Americans, Mexican Americans, Amerindians and Asians. *Hum Genet* 114:263-271
- Devlin B, Roeder K (1999) Genomic control for association studies. *Biometrics* 55:997-1004
- Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, Gabriel SB, Topol EJ, Smoller JW, Pato CN, Pato MT, Petryshen TL, Kolonel LN, Lander ES, Sklar P, Henderson B, Hirschhorn JN, Altshuler D (2004) Assessing the

- impact of population stratification on genetic association studies. *Nat Genet* 36:388-393
- Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG, McKeigue PM (2003) Control of confounding of genetic associations in stratified populations. *Am J Hum Genet* 72:1492-1504
- Long JC (1991) The genetic structure of admixed populations. *Genetics* 127:417-428
- Mattevi VS, Zembrzuski VM, Hutz MH (2004) A resistin gene polymorphism is associated with body mass index in women. *Hum Genet* 115:208-212
- Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, Shriver MD (1998) Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet* 63:1839-1851
- Parra FC, Amado RC, Lambertucci JR, Rocha J, Antunes CM, Pena SD (2003) Color and genomic ancestry in Brazilians. *Proc Natl Acad Sci USA* 100:177-182
- Pritchard JK, Donnelly P (2001) Case-control studies of association in structured or admixed populations. *Theor Popul Biol* 60:227-237
- Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 65:220-228
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959
- Raymond M, Rousset F (1995) <http://wbiomed.curtin.edu.au/genepop>
- Reich DE, Goldstein DB (2001) Detecting association in a case-control study while correcting for population stratification. *Genet Epidemiol* 20:4-16

- Rios DL, Vargas AF, Torres MR, Zago AJ, Callegari-Jacques SM, Hutz MH (2003) Interaction between SREBP-1a and APOB polymorphisms influences total and low-density lipoprotein cholesterol levels in patients with coronary artery disease. *Clin Genet* 63:380-385
- Risch NJ (2000) Searching for genetic determinants in the new millennium. *Nature* 405:847-856
- Salzano FM, Bortolini MC (2002) The evolution and genetics of Latin American populations. Cambridge University Press, Cambridge
- Shriver MD, Parra EJ, Dios S, Bonilla C, Norton H, Jovel C, Pfaff C, Jones C, Massac A, Cameron N, Baron A, Jackson T, Argyropoulos G, Jin L, Hoggart CJ, McKeigue PM, Kittles RA (2003) Skin pigmentation, biogeographical ancestry and admixture mapping. *Hum Genet* 112:387-399
- Wang XL, Badenhop R, Humphrey KE, Wilcken DE (1996) New MspI polymorphism at +83 bp of the human apolipoprotein AI gene: association with increased circulating high density lipoprotein cholesterol levels. *Genet Epidemiol* 13:1-10
- Ziv E, Burchard EG (2003) Human population structure and genetic association studies. *Pharmacogenomics* 4:431-441

Table 1 Allele frequencies of the 10 Ancestry Informative Markers analyzed in cases and controls

Group	APO ¹	APOAI83 ¹	AT3 ¹	FY-NULL ¹	LPL ¹	OCA2 ¹	RB2300 ¹	Sb19.3 ¹	GC-F	GC-S
Cases	0.942	0.947	0.306	0.990	0.505	0.699	0.257	0.854	0.296	0.534
Controls	0.951	0.956	0.235	0.936	0.500	0.681	0.368	0.819	0.275	0.549
Europeans ²	0.927	0.925	0.279	1.000	0.486	0.769	0.333	0.910	0.156	0.607
Africans ²	0.441	0.420	0.874	0.000	0.973	0.098	0.920	0.425	0.824	0.078

¹Presence of *Alu* insertions and absence of the polymorphic restriction sites.

²European and African frequencies were described by Parra et al (1998) and Shriver et al (2003) (APOAI83).

Table 2 African admixture estimated by two different methods based on the frequencies of 10 Ancestry Informative Markers in cases and controls

Group	WLS Method ^a	Chakraborty's Method ^b
Cases	0.022 ± 0.020	0.046 ± 0.001
Controls	0.060 ± 0.024	0.065 ± 0.001

^aLong, 1991.

^bChakraborty, 1985.

Table 3 Ancestry African Index (AAI) values in cases and controls

Group	With outliers ^a		Without outliers ^a	
	N	AAI Median (range)	N	AAI Median (range)
Cases	101	-10.89 (-17.66 to -4.38)	99	-10.98 (-17.66 to -5.65)
Controls	102	-10.78 (-17.66 to -0.78)	93	-10.92 (-17.66 to -5.25)
Southern Brazilians ^b	52	-9.11	-	-
Portuguese ^b	20	-11.73	-	-

^aMann-Whitney *U* test between cases and controls (with outliers: $z=0.855$, $p=0.392$;

without outliers: $z=0.001$, $p=0.999$).

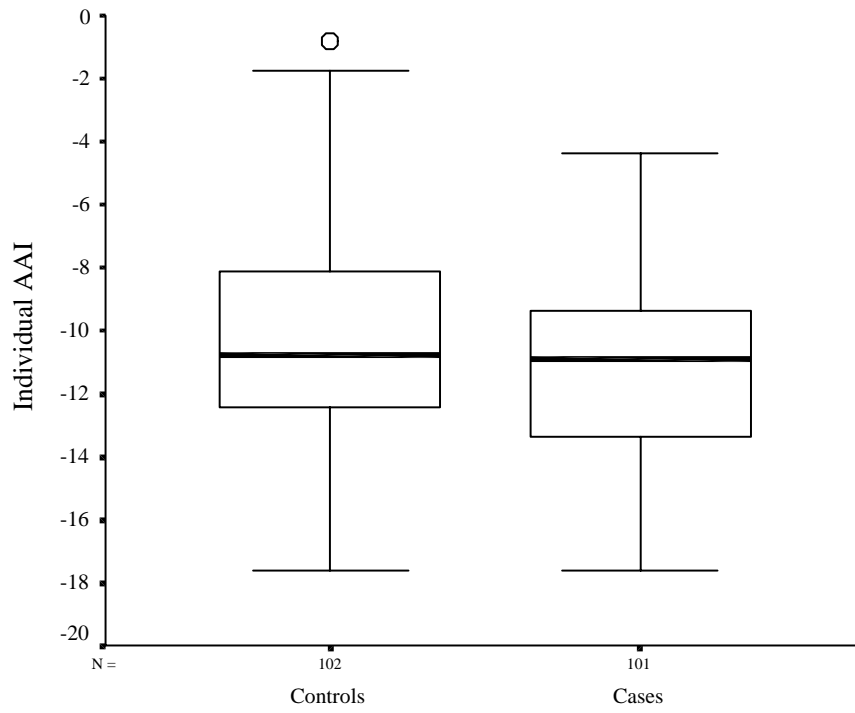
^bParra et al (2003).

Figure legends

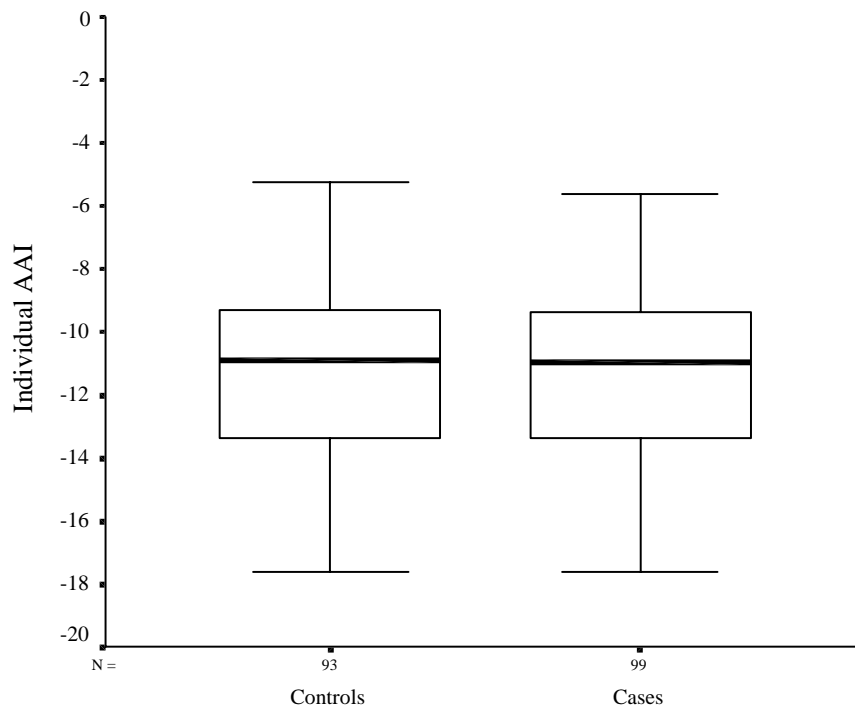
Fig. 1 Distribution of individual Ancestry African Index (AAI) values in the samples of cases and controls. Each group is represented as a box whose top and bottom are drawn at the lower and upper quartiles, with a line at the median. Thus, the box contains the middle half (50%) of the scores of the distribution. Vertical lines outside the box extend to the largest and the smallest observations within 1.5 interquartile ranges from the box. (A) Distribution of the ascertained samples (the open circle indicates an outlier) and (B) samples without outliers (individuals with $AAI < -4.86$).

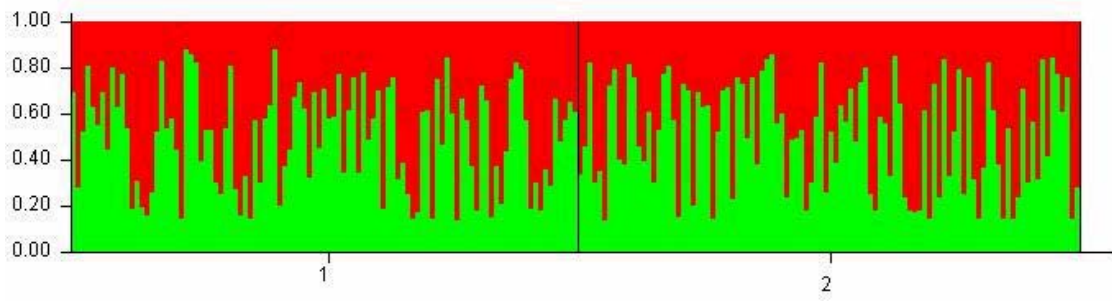
Fig. 2 Plot of estimates of Q (*Structure* program). Each individual is represented by a single vertical line broken into 2 colored segments, with lengths proportional to each of the inferred clusters. The numbers (1 and 2) correspond to the predefined populations (cases and controls).

A)



B)





CAPÍTULO IV
DISCUSSÃO

A discussão específica e completa dos resultados obtidos no presente trabalho encontra-se no artigo (Capítulo 3). Neste capítulo serão discutidos alguns pontos mais gerais como a questão da auto-classificação racial em estudos de associação, a busca de marcadores autossômicos para as estimativas de miscigenação e o novo tipo de abordagem para evitar a estratificação em populações miscigenadas, com o intuito de não abandonar este tipo de estudo (caso-controle) em favor de outros.

Conhecer a ancestralidade do indivíduo através de marcadores genéticos pode ser útil em muitas situações, como por exemplo, mapeamento por mistura, predição de riscos médicos, estudos de dispersão de populações, fluxo gênico e história evolutiva. Ainda, a pesquisa inovadora na área da farmacogenética mostra diferenças significativas, inter e intrapopulacionais, no metabolismo, eficácia e toxicidade de medicamentos (Kittles e Weiss, 2003).

A população brasileira tem um grande potencial de investigação, pois devido à miscigenação, a diversidade de combinações de variantes em diferentes locos possibilita o estudo de interações gene-gene, que em outras populações homogêneas não seria possível. Brasileiros formam uma das mais heterogêneas populações do mundo, e é interessante analisar como os três grupos - ameríndios, europeus e africanos – contribuíram para o *pool* gênico da população

brasileira atual (Carvalho-Silva e cols., 2001). Estudos com o DNA mitocondrial e o cromossomo Y no Brasil mostram um padrão de “cruzamentos” direcionais entre homens europeus e mulheres ameríndias e africanas, concordando com a história de povoamento do Brasil desde o ano de 1500 (Alves-Silva e cols., 2000; Carvalho-Silva e cols., 2001).

Em relação aos marcadores autossômicos, o grau de mistura observado a partir de nossos dados sugere uma pequena contribuição africana (aproximadamente 4% em casos e 6% em controles) (tabela 2, Capítulo 3). Outros estudos realizados em nossa população (Schneider e Salzano, 1979; Franco e cols., 1982; Dornelles e cols., 1999) para sistemas protéicos e outros tipos de análises, indicam uma proporção de mistura africana ligeiramente maior do que a obtida em nosso trabalho (aproximadamente 8%). Porém, esses resultados diferem das proporções obtidas por Parra e cols. (2003) (13 e 17%) para 52 indivíduos da região Sul auto-classificados como brancos. A auto-classificação, recentemente, foi sugerida como modelo fenotípico em alguns estudos. Parra e cols. (2003) e Tang e cols. (2005) mostram que a auto-classificação poderia ser utilizada em diferentes abordagens sem causar “confusão”. Já outros trabalhos (por exemplo, Parra e cols., 2004) pedem cautela quanto ao seu uso em estudos que envolvam populações miscigenadas. Nossos resultados corroboram isso, devido ao fato que nossa amostra foi escolhida fenotipicamente com base na observação das características físicas dos indivíduos. Portanto, deve-se ter cuidado em considerar a auto-classificação, principalmente no Brasil, que tem uma população altamente miscigenada e onde há uma tendência social ao branqueamento.

Os dados obtidos por Freedman e cols. (2004), sobre o impacto da estratificação populacional nos estudos de associação genética, mostram que a estratificação não pode ser excluída em estudos de associação caso-controles, mas que não há necessidade de abandonar esses estudos em favor dos estudos baseados em famílias (como os testes de desequilíbrio de transmissão). Evidências teóricas e práticas sugerem que os estudos baseados em populações com controles não-relacionados, se bem delineados e conduzidos, e analisados e interpretados corretamente, são eficientes para evitar “desvios” devido à estratificação populacional (Cardon e Palmer, 2003).

Levando isso em conta, este novo tipo de abordagem proposto no presente trabalho – utilização do Índice de Ancestralidade Africana –, em que se realiza uma análise pontuada, pode ser utilizado para evitar a estratificação populacional nos estudos de associação do tipo caso-controle, pois os indivíduos que não possuam a proporção de mistura “exigida” para determinado estudo podem ser retirados da amostra. Assim, se permite um controle maior do estudo em relação a esse fator de confusão. Um exemplo disso pode ser observado na figura 1 do capítulo 3; quando os indivíduos com IAA maior que $-4,86$ foram retirados da análise, as amostras de casos e controles ficaram mais homogêneas (o valor das medianas se aproximou).

Um outro ponto a ser questionado é que, em nossa análise, não foram incluídos marcadores genéticos informativos para ameríndios. Para o Brasil ainda não se tem muitos estudos com marcadores autossômicos; a maioria dos trabalhos se refere ao cromossomo Y e ao DNA mitocondrial (por exemplo, Bortolini e cols., 2003 e Dornelles e cols., 2004). A perspectiva é que possamos

validar mais marcadores para africanos e europeus em nossa população e, principalmente, descobrir e validar marcadores autossômicos informativos de ancestralidade ameríndia para nativos sul-americanos. Pode-se assim obter uma coleção de controles populacionais, “tipados” com os AIMS, para serem utilizados em múltiplos estudos de caso-controles.

Trabalhos recentes mostram que o número de AIMS para negros e brancos pode ser bem maior, o que seria uma vantagem, uma vez que um conjunto maior de marcadores proporcionaria traçar um perfil mais adequado relativo à ancestralidade da população ou do indivíduo em estudo.

Os resultados deste trabalho sugerem que, apesar de existir o risco de associação espúria devido à estratificação populacional em estudos do tipo caso-controle, o surgimento de novas ferramentas (AIMs e análises estatísticas diferenciadas) permite que os riscos de se encontrar falso-positivos e/ou falso-negativos sejam minimizados, possibilitando, assim, a continuação destes estudos.

REFERÊNCIAS BIBLIOGRÁFICAS

Almeida S (2001) Análise da influência de polimorfismos no gene do receptor de estrógenos sobre os níveis lipídicos e aterosclerose na população de Porto Alegre. Tese de Mestrado, Universidade Federal do Rio Grande do Sul, Porto Alegre.

Alves-Silva J, da Silva Santos M, Guimaraes PE, Ferreira AC, Bandelt HJ, Pena SD, Prado VF (2000) The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet* 67:444-461.

Ardlie KG, Lunetta KL e Seielstad M (2002) Testing for population subdivision and association in four case-control studies. *Am J Hum Genet* 71:304-311.

Bonilla C, Parra EJ, Pfaff CL, Dios S, Marshall JA, Hamman RF, Ferrell RE, Hoggart CL, McKeigue PM, Shriver MD (2004) Admixture in the Hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping. *Ann Hum Genet* 68:139-153.

Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SP, Bau CH, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, Bradman N, Labuda D, Ruiz-Linares A (2003) Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet* 73:524-539.

Bortolini MC, Weimer TA, Salzano FM, Moura LB e Silva MCBO (1997) Genetic structure of two urban Afro-Brazilian populations. *Int J Anthropol* 12:5-16.

Cardon LR e Palmer LJ (2003) Population stratification and spurious allelic association. *Lancet* 361:598-604.

Carvalho-Silva DR, Santos FR, Rocha J, Pena SD (2001) The phylogeography of Brazilian Y-chromosome lineages. *Am J Hum Genet* 68:281-286.

Collins FS (2004) What we do and don't know about 'race', 'ethnicity', genetics and health at the dawn of the genome era. *Nat Genet* 36:13-15.

De Andrade FM, Silveira FR, Arsand M, Antunes AL, Torres MR, Zago AJ, Callegari-Jaques SM e Hutz MH (2004) Association between -250G/A polymorphism of the hepatic lipase gene promoter and coronary artery disease and HDL-C levels in a Southern Brazilian population. *Clin Genet* 65:390-395.

Dean M, Stephens JC, Winkler C, Lomb DA, Ramsburg M, Boaze R, Stewart C, Charbonneau L, Goldman D, Albaugh BJ, Goedert JJ, Beasley RP, Hwang L-Y, Buchbinder S, Weedon M, Johnson P, Eichelberger M e O'Brien SJ (1994) Polymorphic admixture typing in human ethnic populations. *Am J Hum Genet* 55: 788-808.

Deng HW (2001) Population admixture may appear to mask, change or reverse genetic effects of genes underlying complex traits. *Genetics* 159:1319-1323.

Devlin B e Roeder K (1999) Genomic control for association studies. *Biometrics* 55:997-1004.

Devlin B, Roeder K e Wasserman L (2001) Genomic control, a new approach to genetic-based association studies. *Theor Popul Biol* 60:155-166.

Dornelles CL, Battilana J, Fagundes NJ, Freitas LB, Bonatto SL, Salzano FM (2004) Mitochondrial DNA and *Alu* insertions in a genetically peculiar population: the Ayoreo Indians of Bolivia and Paraguay. *Am J Hum Biol* 16:479-488.

Dornelles CL, Callegari-Jacques SM, Robinson WM, Weimer TA, Franco MHL, Hickmann AC, Geiger CJ e Salzano FM (1999) Genetics, surnames, grandparents' nationalities, and ethnic admixture in Southern Brazil – Do the patterns of variation coincide? *Genet Mol Biol* 22:151-161.

Franco MHL, Weimer TA e Salzano FM (1982) Blood polymorphisms and racial admixture in two Brazilian populations. *Am J Phys Anthropol* 58:127-132.

Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, Gabriel SB, Topol EJ, Smoller JW, Pato CN, Pato MT, Petryshen TL, Kolonel LN, Lander ES, Sklar P, Henderson B, Hirschhorn JN, Altshuler D (2004) Assessing

the impact of population stratification on genetic association studies. *Nat Genet* 36:388-393.

Hanis CL, Hewett-Emmett D, Bertin TK e Schull WJ (1991) Origins of U.S. Hispanics. Implications for diabetes. *Diabetes Care* 14:618-627.

Hinds DA, Stokowski RP, Patil N, Konvicka K, Kershnerobich D, Cox DR e Ballinger DG (2004) Matching strategies for genetic association studies in structured populations. *Am J Hum Genet* 74:317-325.

Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG e McKeigue PM (2003) Control of confounding of genetic associations in stratified populations. *Am J Hum Genet* 72:1492-1504.

IBGE (1991) <http://www.ibge.gov.br>

Jorde LB e Wooding SP (2004) Genetic variation, classification and 'race'. *Nat Genet* 36:28-33.

Kittles RA, Chen W, Panguluri RK, Ahaghotu C, Jackson A, Adebamowo CA, Griffin R, Williams T, Ukoli F, Adams-Campbell L, Kwagyan J, Isaacs W, Freeman V e Dunston GM (2002) CYP3A4-V and prostate cancer in African Americans: causal or confounding association because of population stratification? *Hum Genet* 110:553-560.

Kittles RA e Weiss KM (2003) Race, ancestry, and genes: implications for defining disease risk. *Annu Rev Genomics Hum Genet* 4:33-67.

Knowler WC, Williams RC, Pettitt DJ e Steinberg AG (1988) Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture. *Am J Hum Genet* 43:520-526.

Lander ES e Schork NJ (1994) Genetic dissection of complex traits. *Science* 265:2037-2048.

Lautenberger JA, Stephens JC, O'Brien SJ e Smith MW (2000) Significant admixture linkage disequilibrium across 30 cM around the FY locus in African Americans. *Am J Hum Genet* 66:969-978.

McGinnis R, Shifman S, Darvasi A (2002) Power and efficiency of the TDT and case-control design for association scans. *Behav Genet* 32:135-144.

Parra EJ, Kittles RA e Shriver MD (2004) Implications of correlations between skin color and genetic ancestry for biomedical research. *Nat Genet* 36:54-60.

Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE e Shriver MD (1998) Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet* 63:1839-1851.

Parra FC, Amado RC, Lambertucci JR, Rocha J, Antunes CM e Pena SD (2003) Color and genomic ancestry in Brazilians. *Proc Natl Acad Sci USA* 100:177-182.

Patrinos A (2004) 'Race' and the human genome. *Nat Genet* 36:1-2.

Pfaff CL, Parra EJ, Bonilla C, Hiester K, McKeigue PM, Kamboh MI, Hutchinson RG, Ferrell RE, Boerwinkle E e Shriver MD (2001) Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. *Am J Hum Genet* 68:198-207.

Pritchard JK e Donnelly P (2001) Case-control studies of association in structured or admixed populations. *Theor Popul Biol* 60:227-237.

Pritchard JK e Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 65:220-228.

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

Rios DL, Vargas AF, Ewald GM, Torres MR, Zago AJ, Callegari-Jacques SM e Hutz MH (2003a) Common variants in the lipoprotein lipase gene in Brazil: association with lipids and angiographically assessed coronary atherosclerosis. *Clin Chem Lab Med* 41:1351-1356.

Rios DL, Vargas AF, Torres MR, Zago AJ, Callegari-Jacques SM e Hutz MH (2003b) Interaction between SREBP-1a and APOB polymorphisms influences total and low-density lipoprotein cholesterol levels in patients with coronary artery disease. Clin Genet 63:380-385.

Risch NJ (2000) Searching for genetic determinants in the new millennium. Nature 405:847-856.

Royal CD e Dunston GM (2004) Changing the paradigm from 'race' to human genome variation. Nat Genet 36:5-7.

Santos JL, Perez F, Carrasco E e Albala C (2002) Use of case-parents trio for epidemiological studies of association between genetic polymorphisms and complex diseases. Rev Med Chil 130: 1307-1315.

Schneider H e Salzano FM (1979) Gm allotypes and racial admixture in two Brazilian populations. Hum Genet 53:101-105.

Shriver MD, Parra EJ, Dios S, Bonilla C, Norton H, Jovel C, Pfaff C, Jones C, Massac A, Cameron N, Baron A, Jackson T, Argyropoulos G, Jin L, Hoggart CJ, McKeigue PM e Kittles RA (2003) Skin pigmentation, biogeographical ancestry and admixture mapping. Hum Genet 112:387-399.

Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R e Ferrell RE (1997) Ethnic-affiliation estimation by use of population-specific DNA markers. *Am J Hum Genet* 60:957-964.

Tang H, Quertermous T, Rodriguez B, Kardia SL, Zhu X, Brown A, Pankow JS, Province MA, Hunt SC, Boerwinkle E, Schork NJ, Risch NJ (2005) Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies. *Am J Hum Genet* 76:268-275.

Thomas DC e Witte JS (2002) Point: population stratification: a problem for case-control studies of candidate-gene associations? *Cancer Epidemiol Biomarkers Prev* 11:505-512.

Tishkoff AS e Kidd KK (2004) Implications of biogeography of human populations for 'race' and medicine. *Nat Genet* 36:21-27.

Turakulov R e Easteal S (2003) Number of SNPS loci needed to detect population structure. *Hum Hered* 55:37-45.

Zhao H, Pfeiffer R, Gail MH (2003) Haplotype analysis in population genetics and association studies. *Pharmacogenomics* 4:171-178.

Ziv E e Burchard EG (2003) Human population structure and genetic association studies. *Pharmacogenomics* 4:431-441.