Effect of race, genetic population structure, and genetic models in two-locus association studies: clustering of functional renin-angiotensin system gene variants in hypertension association studies

A.C. Pereira[1], G.A. Mota[1],
I. Benseñor[2], P.A. Lotufo[2]
and J.E. Krieger[1,2]

[1]Laboratório de Genética e Cardiologia Molecular, Instituto do Coração,
Faculdade de Medicina, Universidade de São Paulo, São Paulo, SP, Brasil
[2]Departamento de Clínica Médica, Faculdade de Medicina,
Universidade de São Paulo, São Paulo, SP, Brasil

Abstract

Correspondence
J.E. Krieger
Laboratório de Genética e
Cardiologia Molecular
InCor, HC, FM, USP
Av. Dr. Eneas C. Aguiar, 44
05403-000 São Paulo, SP
Brasil
Fax: +55-11-3069-5068
E-mail: krieger@incor.usp.br

Previous genetic association studies have overlooked the potential for biased results when analyzing different population structures in ethnically diverse populations. The purpose of the present study was to quantify this bias in two-locus association studies conducted on an admixtured urban population. We studied the genetic structure distribution of angiotensin-converting enzyme insertion/deletion (ACE I/D) and angiotensinogen methionine/threonine (M/T) polymorphisms in 382 subjects from three subgroups in a highly admixtured urban population. Group I included 150 white subjects; group II, 142 mulatto subjects, and group III, 90 black subjects. We conducted sample size simulation studies using these data in different genetic models of gene action and interaction and used genetic distance calculation algorithms to help determine the population structure for the studied loci. Our results showed a statistically different population structure distribution of both ACE I/D (P = 0.02, OR = 1.56, 95% CI = 1.05-2.33 for the D allele, white versus black subgroup) and angiotensinogen M/T polymorphism (P = 0.007, OR = 1.71, 95% CI = 1.14-2.58 for the T allele, white versus black subgroup). Different sample sizes are predicted to be determinant of the power to detect a given genotypic association with a particular phenotype when conducting two-locus association studies in admixtured populations. In addition, the postulated genetic model is also a major determinant of the power to detect any association in a given sample size. The present simulation study helped to demonstrate the complex interrelation among ethnicity, power of the association, and the postulated genetic model of action of a particular allele in the context of clustering studies. This information is essential for the correct planning and interpretation of future association studies conducted on this population.

Introduction

Previous genetic association studies have overlooked the potential for biased results when analyzing different population structures in ethnically diverse populations. Particularly, studies with putative genetic risk factors for hypertension have been criticized by many investigators because their results have not been consistent (1).

Variants of the renin-angiotensin system (RAS) genes have the potential to reveal certain aspects of the genesis of some cardiovascular diseases. Two of these allelic variants have been undoubtedly associated with particular intermediate phenotypes that could help explain the RAS role in some of the pathophysiological derangements observed in cardiovascular homeostasis in a number of diseases (2,3).

A diallelic polymorphism in the angiotensin-converting enzyme (ACE) gene, characterized by a deletion (D) or insertion (I) in the 16th intron of the gene, has been unequivocally associated with differences in plasma ACE levels (4). However, its role in predicting hypertension and other cardiovascular diseases (e.g., myocardial infarction, cardiac hypertrophy) (4-10) remains unclear.

Additionally, a methionine→threonine variant at position 235 (M235T) of the angiotensinogen (ATG) gene has been considered a marker for primary hypertension and elevated plasma ATG concentrations (3). Other studies, however, have found no association between the M235T ATG variant and a number of cardiovascular phenotypes (11).

The hypothesis that a synergistic effect of both RAS genetic variants present in the same individual would increase cardiovascular risk has also been contemplated in a small number of studies and, again, no consensus has been obtained (12,13).

Finally, few studies have contemplated the relationship between the statistical power of association studies and sample size for different genetic models. Much criticism has been directed at the real value of not only positive, but also negative results (1). These facts have raised great concern about the real role of these variants in predicting the occurrence, or the outcomes, of cardiovascular disease in a particular population. Different ethnic allele frequencies and genetic structures among the studied populations, and the proposed genetic model of action of these genes in the genesis of a particular phenotype have been considered as biases in many of the studies to date (14,15).

No study to date has quantified this bias. In addition, the distribution of the RAS genetic polymorphisms in relation to population structure has been overlooked in previous investigations (1). To quantify this bias, we studied the distribution of RAS genetic polymorphisms and population genetic structure in three subgroups from a highly admixtured urban population. Sample size simulations were conducted using these data in different genetic models of gene action and interaction. These data can have important applications in the planning and interpretation of genetic association studies in our area and other similar urban areas worldwide.

Material and Methods

Study population

The study population comprised 382 consecutive unrelated men and women admitted to the Blood Donation Center of the Fundação Pró-Sangue, São Paulo University Medical School. The population was divided into three groups according to ethnic morphological criteria. Group I included 150 white subjects; group II, 142 mulatto subjects, and group III, 90 black subjects. Mulatto, in this study, was defined as an individual with a known family history of admixture between black and white populations. The ethnic morphological subgroup classification was

based on phenotype pigmentation of the abdomen, hair color, type and conformation of the nose and lips, and family history, as determined and agreed upon by two examining physicians (16). It should be noted that our subgroup stratification was based on morphological criteria and not on a true ethnic group stratification. Ethnic group stratification is defined by a group of genetic, linguistic and cultural characteristics. In particular, it has been shown that in the Brazilian population morphological criteria are not highly concordant with ancestry (17,18). Venous blood was obtained for genomic DNA extraction.

Assessment of the angiotensin-converting enzyme gene polymorphism genotype

Five-milliliter blood samples were drawn into tubes containing EDTA. The ACE I/D polymorphism was determined using a three-primer system (19), which minimizes the mistyping (20) that can occur with a two-primer system. The polymerase chain reaction (PCR) products were visualized by electrophoresis on 3% agarose gel with ethidium bromide and stored in digital form.

Assessment of the angiotensinogen gene polymorphism genotype

The M235T variant of the ATG gene was detected by a PCR method. Primers were selected according to Russ et al. (21). The PCR parameters were described by Kiema et al. (22). Three microliters of unpurified PCR product was diluted to 10 µl in recommended restriction buffer containing 5 U of *Tth* 111 I and digested at 65°C overnight. The PCR products were visualized by electrophoresis on 3% agarose gel with ethidium bromide and stored in digital form.

Statistical analysis

Data were analyzed by the chi-square test

and ANOVA using the statistical program GraphPad Prism (version 2.0, GraphPad Software Inc., San Diego, CA, USA). The Hardy-Weinberg equilibrium for the distribution of genotypes was estimated using the chi-square test. Allele and genotype frequencies were compared using the chi-square test and the chi-square test for linear trends available in EpiInfo (version 6.0). Genetic structure analyses, including population pairwise F_{ST} values, sample heterozygosity and the population pairwise differentiation test, were conducted using the statistical program Arlequin (version 1.1, Genetics and Biometry Laboratory, University of Geneva, Switzerland) for population genetic data analysis (23). Sample size calculations were conducted using EpiInfo (version 6.0). Modeling different mechanisms of action of a particular allele was conducted by grouping individuals with one or two particular genotypes regarding the chosen model (i.e., dominant models contemplate as being at increased risk of developing the studied phenotype both the homozygous and heterozygous individuals harboring allele conferring increased risk; similarly, in the recessive model of action only the homozygous individuals for the allele conferring increased risk were considered to be at increased risk of developing the studied phenotype). Values of P<0.05 were considered to be statistically significant.

Results

Population demographics

Population demographics are summarized in Table 1. The mean age for each subgroup was 32.4 ± 0.8 years for the white population, 30.6 ± 0.7 for the mulatto population, and 30.3 ± 0.8 for the black population. ANOVA did not reveal any relationship between the parameters studied and age in any of the populations. While there was approximately the same number of men in each

Table 1. Population demographics.

Group	N	Male (female)	Mean age (years)
White	150	75 (75)	32.4
Mulatto	142	73 (69)	30.6
Black	90	73 (17)	30.3

There were no significant differences between groups (one-way ANOVA).

Table 2. Allele and genotype frequency distribution of the ACE I/D and ATG M/T polymorphisms among the studied groups.

A. ACE I/D polymorphism (N = 382).

	White	Mulatto	Black	P value (overall)	P value (white vs black)
Allele					
D allele	0.54	0.57	0.65	0.068	0.02
I allele	0.46	0.43	0.35		
OR (D allele)	1.0	1.12	1.56	0.027	P value for trend
Genotype					
DD	0.28	0.34	0.43	0.14	
DI	0.53	0.45	1.43		
II	0.19	0.20	0.13		
OR (DD genotype)	1.0	1.31	1.97	0.016	P value for trend

B. ATG M/T polymorphism (N = 382).

	White	Mulatto	Black	P value (overall)	P value (white vs black)
Allele					
M allele	0.43	0.44	0.31	0.008	0.007
T allele	0.57	0.56	0.69		
OR (D allele)	1.0	0.96	1.71	0.016	P value for trend
Genotype					
MM	0.23	0.23	0.07	0.027	
MT	0.39	0.41	0.45		
TT	0.37	0.35	0.47		
OR (TT genotype)	1.0	0.91	1.61	0.12	P value for trend

ACE I/D = angiotensin-converting enzyme insertion/deletion polymorphism; ATG M/T = angiotensinogen methionine/threonine polymorphism; DD, DI, II = ACE I/D gene variants; MM, MT, TT = ATG M/T gene variants; OR = odds ratio. P value (overall) refers to the comparison among the three subgroups. P value for trend refers to all groups (presented OR in relation to white DD genotype frequency) (chi-square test for linear trends).

group, the black population was primarily male, with 81% (73/90) being men. Groups I and II had approximately equal numbers of men and women, with 50% (75/150) and 51% (73/142) males, respectively.

Allele, genotype and association frequencies

Table 2 displays the allele and genotype frequencies of the ACE I/D and the ATG M/T gene variants. The genotype frequencies were consistent with Hardy-Weinberg equilibrium.

There were no sex differences in allele or genotype frequencies in any of the three groups. Therefore, male and female data were pooled and analyzed together as the same group.

Allele frequencies for the ACE gene variants differed significantly between groups: 54.3% in the Caucasian population (group I), 65.0% in the black population (group III), and 57.0% in the mulatto population (group II) for the D allele (P = 0.02, odds ratio (OR) = 1.56, for white versus black groups). ATG gene variants also showed a statistically different distribution among ethnic groups: 57.0% in group I, 56.0% in group II, and 69.4% in group III for the T allele (P = 0.007, OR = 1.71, for white versus black groups). When clustering both genotype frequencies a higher frequency of a postulated worse genotypic association (DD + TT) was evident in group III, the black population (11.3% in group I as compared to 21.1% in group III, P = 0.04, chi-square = 4.22) (Figure 1).

Figure 1. Genotypic distribution of the genetic variants among the subgroups. As shown, the clustering of "worse" variants (DD/TT) was more prominent in the black than in both the white and mulatto populations. DD, DI, II = angiotensin-converting enzyme I/D gene variants. MM, MT, TT = angiotensinogen M/T gene variants.
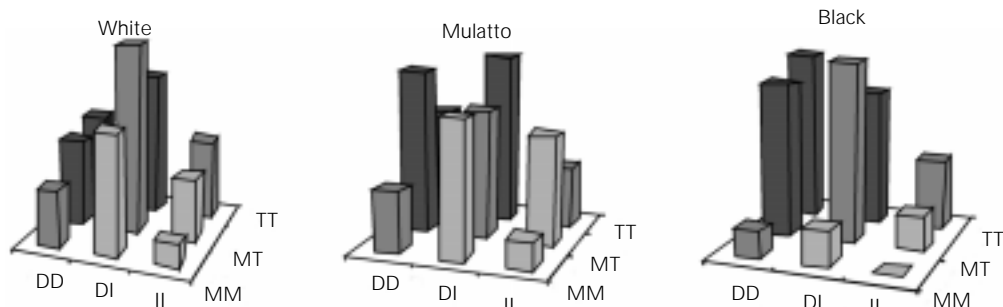
Sample heterozygosity

Sample heterozygosity, the proportion of heterozygote individuals for at least one of the studied polymorphisms, was 0.70 in group I, 0.73 in group II and 0.67 in group III.

Genetic structure analysis

Population pairwise F_{ST} values were significantly different only between the white (group I) and black (group III) subgroups (P = 0.03 for group I versus group III, P = 0.83 for group I versus group II, and P = 0.07 for group III versus group II).

Finally, genotype frequency analysis by a population pairwise differentiation test using 10,000 Markov chain steps showed significant difference between all populations in the sample (probability (P) of nondifferentiation = 0.048). When pairs of populations were tested there was a significant difference between the mulatto and black populations (P = 0.03) and the black and Caucasian populations (P = 0.03). No difference was observed when comparing the Caucasian and the mulatto populations. These data imply that the populations studied differ in their genotypic components.

Sample size simulations

To quantify the possible role of ethnicity and population structure in association studies using these genes we conducted sample size simulations for both genes. Additionally, we used different genetic models (dominant/recessive) to study the influence of a particular model on the sample size and the design of the study. These data are shown in Tables 2 and 3. As illustrated in Figures 2 and 3, size was strongly dependent on relative risk of the variant for the development of the studied phenotype, ethnicity and genetic model of action, particularly when two-locus studies are considered.

Table 3. Sample size calculations considering the effect of ethnicity, odds ratio and genetic model for both genes individually (A,B) and when clustering studies are performed (C).

A. Dominant models

	Odds ratio		
	1.5	2.0	3.0
1. ACE gene			
White	735	289	142
Mulatto	704	277	136
Black	1017	407	203
2. ATG gene			
White	632	247	120
Mulatto	634	248	120
Black	1674	678	345

B. Recessive models

	Odds ratio		
	1.5	2.0	3.0
1. ACE gene			
White	461	157	63
Mulatto	423	147	61
Black	403	144	62
2. ATG gene			
White	413	145	61
Mulatto	420	146	61
Black	403	145	63

C. Clustering models

	Odds ratio		
	1.5	2.0	3.0
1. ACE dominant x ATG dominant			
White	525	201	96
Mulatto	437	163	75
Black	685	269	131
2. ACE dominant x ATG recessive			
White	461	157	63
Mulatto	454	155	63
Black	412	144	61
3. ACE recessive x ATG dominant			
White	539	179	70
Mulatto	460	156	63
Black	407	144	61
4. ACE recessive x ATG recessive			
White	862	278	103
Mulatto	866	279	103
Black	542	180	70

Sample size calculations used confidence intervals of 95%, power of 80% and a 1:1 case-control ratio. ACE, angiotensin-converting enzyme; ATG, angiotensinogen.

Discussion

Association studies are increasingly being used in the genetic dissection of complex diseases. This approach has become an alternative to linkage studies, which have limited power to detect genes of modest effect (24). Despite the greater power of genetic association studies, there are several limitations to their application in the study of complex traits and diseases.

One of the major concerns is the little knowledge generally available about the allelic distribution among populations of different ethnic origins. It has been argued that
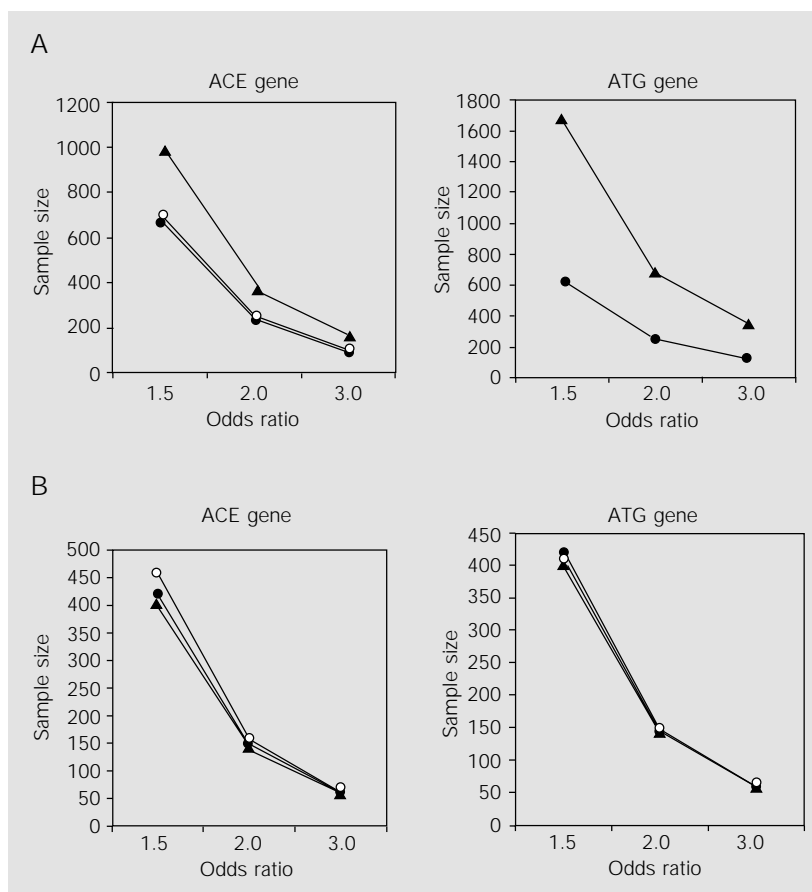

Figure 2. Effect of ethnicity and odds ratio on the recommended sample size of an association study conducted on this population using dominant (A) or recessive (B) models of action of the genetic variant studied (i.e., increased risk of the DD or the TT genotype in a recessive model, or the presence of the D or T allele in a dominant model). ACE = angiotensin-converting enzyme; ATG = angiotensinogen. Open circles = white subjects; closed circles = mulatto subjects; triangles = black subjects.

differences among these populations could explain the contradictory findings of some of these studies (14). Indeed, the ACE I/D polymorphism is one of the best-known examples.

Another major concern about the validity of genetic association studies is related to the small amount of knowledge about the molecular biology of the variants studied. It is especially noteworthy that analysis and interpretation remain primarily based on a particular assumption about the mode of action (i.e., dominant or recessive) of the genetic polymorphism being studied.

The data presented here establish the distribution of RAS genetic polymorphisms and the population genetic structure in three ethnic subgroups from a highly admixtured urban population. On this basis, it is interesting to note the different genotypic distribution of these markers among the studied groups (Figure 1). Whereas the observed distributions could be responsible for possible selection bias in association studies conducted on this population, the clustering of "worse" variants was more prominent in the black than in both the Caucasian and the mulatto populations. These findings may be related to a higher frequency and a worse outcome of hypertension and cardiovascular disease observed in black populations (25). The hypothesis of an association between these variants and the development or the severity of hypertension or another cardiovascular phenotype in this group should be tested in large, population-based studies. In addition, the use of population genetic algorithms as presented here can help investigators to determine the genetic differences between two populations, especially when one is working with a multi-locus model.

The use of sample size calculation models is of help in the quantification of this selection bias. Based on the data from this study, for example, to detect a 2.0 OR association of the ATG T allele with a particular phenotype in a dominant model of action it

would be necessary to test approximately 250 subjects with the phenotype (and the same number of control individuals) if one is studying white subjects. However, if studying black subjects, it would be necessary to study 680 subjects with the same phenotype. Conversely, if one is working with the hypothesis of an association of the ACE D allele in a recessive model of action, to identify a 1.5 OR relationship with a particular phenotype, 461 subjects would be necessary in a white population, but only 403 in a black population. Thus, ethnicity should be considered a potential bias in unmatched studies, especially when one is working with the hypothesis of a dominant model of action.

Most important, however, is the use of this approach when working with two-locus models. As shown in this study, sample size is highly dependent on both ethnicity of the studied population and genetic model of action of the genes involved in the analysis. For instance, 685 black subjects would be required in a study in which the ACE D allele and the ATG T allele both have a dominant effect on the phenotype studied. Only 525 subjects would be required in a white population. This number would be completely different if one were working with a model in which the ACE D allele has a recessive action and the ATG T allele has a dominant one. In this case, for the same 1.5 OR, one would need 539 Caucasian subjects as opposed to only 407 black subjects with the phenotype of interest. It should also be noted that epistatic effects, which can substantially add complexity to the model, were not considered in these simulations.

Matched control groups (e.g., sib-pair studies) remain the best choice to avoid the ethnicity bias (24). However, it should be noted that the proposed genetic model of action of a particular combination of alleles can add a much stronger bias than
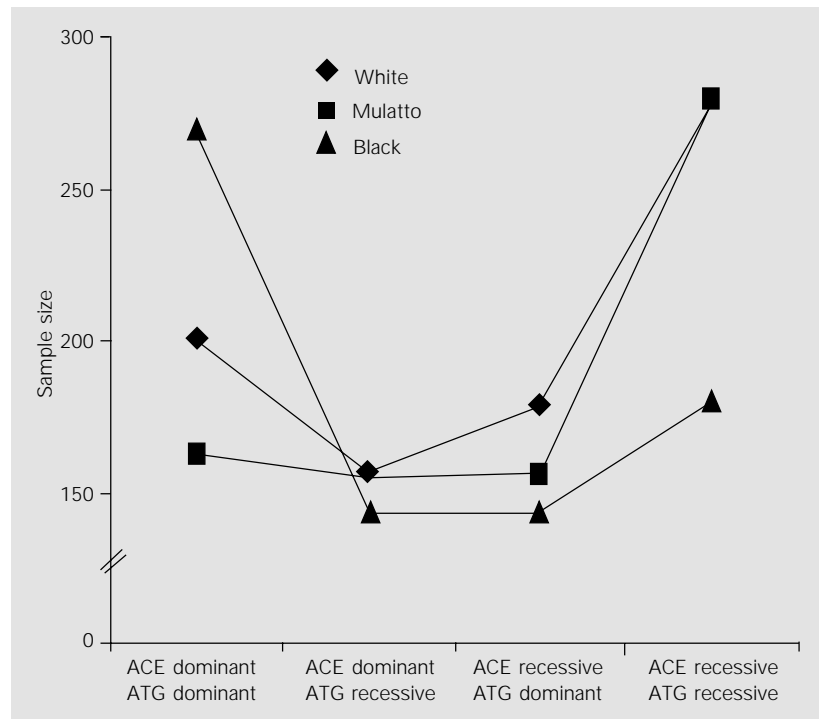

Figure 3. The effect of ethnicity and genetic model on the recommended sample size for a fixed odds ratio (2.0) for the respective genotypic association. As shown, sample size is strongly dependent on both relative risk of the variant for the development of the studied phenotype, ethnicity and genetic model of action, particularly when clustering studies are taken. ACE = angiotensin-converting enzyme; ATG = angiotensinogen.

ethnicity itself.

Our data highlight the importance of elements that should be taken into consideration in association studies in admixtured urban populations. The increasing knowledge about population genetics and genetic variance among different populations will add important elements to the analysis of complex diseases and traits. Similarly, knowledge of gene action and interaction should be exercised whenever interpreting and analyzing such studies.

Acknowledgments

References

1. Corvol P, Persu A, Gimenez-Roqueplo AP & Jeunemaitre X (1999). Seven lessons from two candidate genes in human essential hypertension: angiotensinogen and epithelial sodium channel. Hypertension, 33: 1324-1331.

2. Tiret L, Rigat B, Visvikis S, Breda C, Corvol P, Cambien F & Soubrier F (1992). Evidence from combined segregation and linkage analysis that a variant of the angiotensin I-converting enzyme (ACE) gene controls plasma ACE levels. American Journal of Human Genetics, 51: 197-205.

3. Corvol P & Jeunemaitre X (1997). Molecular genetics of human hypertension: role of angiotensinogen. Endocrine Reviews, 18: 662-677.

4. Jeunemaitre X, Lifton RP, Hunt SC, Williams RR & Lalouel JM (1992). Absence of linkage between the angiotensin converting enzyme locus and human essential hypertension. Nature Genetics, 1: 72-75.

5. O'Donnell CJ, Lindpaintner K, Larson MG, Rao VS, Ordovas JM, Schaefer EJ, Myers RH & Levy D (1998). Evidence for association and genetic linkage of the angiotensin-converting enzyme locus with hypertension and blood pressure in men but not women in the Framingham Heart Study. Circulation, 97: 1766-1772.

6. Zee RY, Ridker PM, Stampfer MJ, Hennekens CH & Lindpaintner K (1999). Prospective evaluation of the angiotensin-converting enzyme insertion/deletion polymorphism and the risk of stroke. Circulation, 99: 340-343.

7. Gired X, Hanon O, Mourad JJ, Boutouyrie P, Laurent S & Jeunemaitre X (1998). Lack of association between renin-angiotensin system gene polymorphism and wall thickness of the radial and carotid arteries. Hypertension, 32: 579-583.

8. Agerholm-Larsen B, Nordestgaard BG, Steffensen R, Sorensen TI, Jensen G & Tybjaerg-Hansen A (1997). ACE gene polymorphism: Ischemic heart disease and longevity in 10,150 individuals. A case-referent and retrospective cohort study based on the Copenhagen City Heart Study. Circulation, 95: 2358-2367.

9. Lindpaintner K, Lee M, Larson MG, Rao VS, Pfeffer MA, Ordovas JM, Schaefer EJ, Wilson AF, Wilson PW, Vasan RS, Myers RH & Levy D (1996). Absence of association or genetic linkage between the angiotensin-converting-enzyme gene and left ventricular mass. New England Journal of Medicine, 334: 1023-1028.

10. Lechin M, Quinones MA, Omran A, Hill R, Yu QT, Rakowski H, Wigle D, Liew CC, Sole M, Roberts R & Ali JM (1995). Angiotensin-I converting enzyme genotypes and left ventricular hypertrophy in patients with hypertrophic cardiomyopathy. Circulation, 92: 1808-1812.

11. Brand E, Chatelain N, Keavney B, Caulfield M, Citterio L, Connell J, Grobbee D, Schmidt S, Schunkert H, Schuster H, Sharma AM & Soubrier F (1998). Evaluation of the angiotensinogen locus in human essential hypertension: a European study. Hypertension, 31: 725-729.

12. Vasku A, Soucek M, Znojil V, Rihacek I, Tschoplova S, Strelcova L, Cidl K, Blazkova M, Hajek D, Holla L & Vacha J (1998). Angiotensin I-converting enzyme and angiotensinogen gene interaction and prediction of essential hypertension. Kidney International, 53: 1479-1482.

13. Ko YL, Ko YS, Wang SM, Chu PH, Teng MS, Cheng NJ, Chen WJ, Hsu TS, Kuo CT, Chiang CW & Lee YS (1997). Angiotensinogen and angiotensin-I converting enzyme gene polymorphisms and the risk of coronary artery disease in Chinese. Human Genetics, 100: 210-214.

14. Tiret L & Cambien F (1995). Departure from Hardy-Weinberg equilibrium should be systematically tested in studies of association between genetic markers and disease. Circulation, 92: 3364-3365.

15. Barley J, Blackwood A, Carter ND, Crews DE, Cruickshank JK, Jeffery S, Ogunlesi AO & Sagnella GA (1994). Angiotensin converting enzyme insertion/deletion polymorphism: association with ethnic origin. Journal of Hypertension, 12: 955-957.

16. Krieger H, Morton NE, Mi MP, Azevedo E, Freire-Maia A & Yasuda N (1965). Racial admixture in north-eastern Brazil. Annals of Human Genetics, 29: 113-125.

17. Alves-Silva J, Santos MS, Guimarães PEM, Ferreira ACS, Bandelt HJ, Pena SDJ & Prado VF (2000). The ancestry of Brazilian mtDNA lineages. American Journal of Human Genetics, 67: 444-461.

18. Carvalho-Silva DR, Santos FR, Rocha J & Pena SDJ (2001). The phylogeography of Brazilian Y-chromosome lineages. American Journal of Human Genetics, 68: 281-286.

19. Evans AE, Poirier O, Kee F, Lecerf L, McCrum E, Falconer T, Crane J, O'Rourke DF & Cambien F (1994). Polymorphisms of the angiotensin-converting-enzyme gene in subjects who died from coronary heart disease. Quarterly Journal of Medicine, 87: 211-214.

20. Shanmugam V, Sell KW & Saha BK (1993). Mistyping ACE heterozygotes. PCR Methods and Applications, 3: 120-121.

21. Russ AP, Maerz W, Ruzicka V, Stein U & Gross W (1993). Rapid detection of the hypertension-associated Met235→Thr allele of the human angiotensinogen gene. Human Molecular Genetics, 2: 609-610.

22. Kiema TR, Kauma H, Rantala AO, Lilja M, Reunanen A, Kesaniemi YA & Savolainen MJ (1996). Variation at the angiotensin-converting enzyme gene and angiotensinogen gene loci in relation to blood pressure. Hypertension, 28: 1070-1075.

23. Scheneider S, Kueffer JM, Roessli D & Excoffier L (1997). Arlequin Version 1.1: A Software for Population Genetic Data Analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

24. Lander ES & Schork NJ (1994). Genetic dissection of complex traits. Science, 30: 2037-2048.

25. Cooper R & Rotimi C (1997). Hypertension in blacks. American Journal of Hypertension, 10: 804-812.