Population Stratification in a Case-Control Study of Brain Arteriovenous Malformation in Latinos

Helen Kim^{a,b,e} Pirro G. Hysi^a Ludmila Pawlikowska^{a,b} Shweta Choudhry^d
Esteban González Burchard^{b,d} Pui-Yan Kwok^{b,c} Stephen Sidney^h
Charles E. McCulloch^e William L. Young^{a,f,g}

^aCenter for Cerebrovascular Research, Department of Anesthesia and Perioperative Care, ^bInstitute for Human Genetics, ^cCardiovascular Research Institute, Departments of ^dMedicine, ^eEpidemiology and Biostatistics, ^fNeurological Surgery and ^gNeurology, University of California, San Francisco, San Francisco, Calif., and ^hDivision of Research, Kaiser Permanente Medical Care Program, Oakland, Calif., USA

Key Words

Arteriovenous malformations • Epidemiology • Genetics • Inflammation • Risk factors • Race/ethnicity

Abstract

Background: Genetic association studies conducted in admixed populations may be confounded by population stratification resulting in spurious associations. The purpose of this pilot study was to determine the presence and effect of population stratification in a case-control study of brain arteriovenous malformation (BAVM). **Methods:** We tested 83 ancestry informative markers in BAVM cases and healthy controls of self-reported Latino race/ethnicity (n = 294). Individual ancestry estimates (IAE) were obtained using the Structure program, assuming 3 underlying subpopulations. Summary χ^2 tests comparing genotype frequency of ancestry informative markers were used to detect stratification and IAE were included as covariates in logistic regression analysis to account for differences in genetic background. **Results:** Admixture estimates for Latinos (overall 47% native American, 45% European and 8% African ancestry) revealed heterogeneity between individuals within ancestral groups. The summary χ^2 test was significant (p = 0.005), suggesting

ancestral differences between cases and controls. Furthermore, genetic ancestry was associated with frequency differences in a promoter variant in the IL-6 gene (IL-6 –174G>C). On average, subjects with the IL-6 –174 GG genotype had 6% greater Native American ancestry (p = 0.023). Age- and sex-adjusted risk of BAVM associated with the IL-6 –174 GG genotype was 1.85 (95% CI 0.99–3.48, p = 0.055), and further adjustments for IAE yielded an OR of 1.96 (95% CI 1.03–3.72, p = 0.039). **Conclusion:** The IL-6 –174G>C polymorphism was associated with increased risk of BAVM among Latinos after accounting for differences in ancestral background. These results suggest subtle, negative confounding and illustrate the importance of addressing population stratification in case-control studies conducted in admixed populations.

Copyright © 2008 S. Karger AG, Basel

Introduction

Case-control studies in admixed populations, such as Latinos or African Americans, are susceptible to genetic confounding by population stratification, which can result in both false-positive or false-negative associations. If the risk of disease varies with ancestry, then

admixture will confound associations of disease with genotypes at any locus where allele frequencies also vary between ancestral populations [1, 2]. Several studies have demonstrated that confounding by population stratification can exist, even despite stringent recruitment of ethnically matched cases and controls from the same clinics [2–5].

We recently reported that compared to Caucasians, Latinos were at 2-fold increased risk of hemorrhage in the untreated course of 1,464 brain arteriovenous malformation (BAVM) patients [6]. The incidence of primary spontaneous intracerebral hemorrhage is also known to be higher among Latinos and African Americans [7]. Therefore, the purpose of this pilot study was to determine if population stratification exists in a case-control study of BAVM conducted among self-reported Latinos. We illustrate the effect of population stratification on a promoter polymorphism in the IL-6 gene (IL-6 –174G>C), which is known to vary by race/ethnicity [8] and has been associated with increased risk of subarachnoid hemorrhage [9] and intracerebral hemorrhage presentation in BAVM patients [10].

Methods

Study Population

BAVM cases were recruited from the University of California, San Francisco or the Northern California Kaiser Permanente Medical Care Program [6]. Healthy controls from 4 racial/ethnic groups (Chinese, African American, Mexican and Caucasian) were recruited for a pharmacogenetics study of membrane transporters at the University of California, San Francisco [11]. Self-reported race/ethnicity was recorded from the medical records (cases) or medical history questionnaire (controls). For this study, we focused on subjects who self-reported as Latino ($n = 294$), including 79 BAVM cases and 215 controls.

Marker Selection and Genotyping

We selected 96 ancestry informative markers (AIMs) with large absolute allele frequency differences ($\delta > 0.5$) between African, European or Native American populations (online suppl. table, www.karger.com/doi/10.1159/000160215) that were not known to be associated with BAVM. The unlinked genetic markers were selected from a large pool of AIMs identified by genotyping DNA from 3 ancestral populations (42 European Caucasians, 30 Native Americans from Mexico and 37 West Africans) using the Affymetrix GeneChip Human Mapping 100K array [12]: 8,242 AIMs for African-Native American, 5,781 AIMs for African-European and 5,040 AIMs for European-Native American populations were identified on the 100K array with $\delta > 0.5$. Selected AIMs were genotyped using a template-directed dye-terminator incorporation assay on the Beckman Coulter SNPstream 48-plex platform [13].

Statistical Analysis

The Cochran-Armitage χ^2 trend test was used to evaluate differences in genotype frequency distributions between BAVM cases and controls for each AIM. To test for population stratification, a global χ^2 test of association was performed by summing the χ^2 values ($\Sigma\chi^2$) and degrees of freedom (d.f.) over all tests [14]. $p < 0.05$ implies that cases and controls have different genetic backgrounds. Quantile-quantile plots were used to compare the observed χ^2 distribution for unlinked AIMs to the expected distribution with 1 d.f. under the null hypothesis of no stratification.

Individual admixture estimates (IAE) were obtained using a Bayesian Markov Chain-Monte Carlo method implemented in the Structure 2.1 program [14] under an admixture model, assuming 3 underlying subpopulations ($K = 3$) and using prior ancestral population genotype information (described above). The Markov Chain-Monte Carlo scheme was run for 10,000 iterations to obtain proportions of Native American, European and African ancestry for each individual; individual proportions summed to 1.0. IAE values can then be included as covariates in the regression analysis to adjust for differences in genetic background.

To determine the effect of population stratification in our study, we investigated the association between an IL-6 –174G>C promoter polymorphism and risk of BAVM among Latinos. A total of 14 subjects with AIM data were missing IL-6 –174G>C genotypes for a total of 280 Latinos (75 cases and 205 controls) available for analysis. Odds ratios (OR) and 95% confidence intervals (CI) were calculated using logistic regression analysis, including age, sex and IAE to account for differences in genetic background.

Results

Of 96 AIMs, 7 failed 48-plex genotyping and 6 were monomorphic (online suppl. table), leaving a total of 83 AIMs available for association analysis in 79 BAVM cases and 215 healthy controls, all of self-reported Latino race/ethnicity. Descriptive characteristics of BAVM cases and controls are summarized in table 1. Mean age at diagnosis for cases and at study enrollment for controls was similar; however, a greater percentage of cases were male (57%). Among cases, 41% presented with hemorrhage, 18% had deep-only venous drainage and 64% had BAVM in eloquent location.

Group and individual admixture estimates are shown in table 2 and figure 1, respectively. The majority of Latinos in our study population are primarily of European and Native American ancestry (fig. 1a, b). However, there is dramatic heterogeneity in admixture levels, with some individuals being almost of 100% European ancestry while others are of almost 100% Native American ancestry. Overall, Latinos in our study population were of 47% Native American, 45% European and 8% African ancestry. These admixture estimates are consistent with previ-

Fig. 1. Proportion of individual admixture estimates from *Structure* program assuming three subpopulations, for Latino BAVM cases (**a**; $n = 79$) and Latino controls (**b**; $n = 215$). Individual subjects are sorted along the x-axis by increasing proportion of Eu-

ropean ancestry. Each subject has 3 stacked bars representing the relative proportion of ancestry (y-axis): Native American, African and European Caucasian.

Table 1. Descriptive characteristics of Latino BAVM cases and controls

	Cases ($n = 79$)	Controls ($n = 215$)
Age, years	30.5 ± 15.6	28.1 ± 7.1
Gender, %		
Male	57.0	35.4
Female	43.0	64.7
Initial presentation, %		
Hemorrhage	40.5	NA
Seizure	25.3	NA
Headache	13.9	NA
Other/incidental	20.3	NA
Venous drainage, %		
Superficial	44.1	NA
Deep	17.7	NA
Both	38.2	NA
Eloquent location, %		
Yes	63.9	NA
No	36.1	NA
BAVM size, cm	3.3 ± 1.8	NA

NA = Not applicable.

Table 2. Average group admixture ancestry proportions estimated from *Structure* 2.1 software; rows sum to 100%

Population	Sample size	African	European Caucasian	Native American
Latino cases	79	8.3%	46.7%	45.0%
Latino control	215	7.4%	44.8%	47.8%
Latino total	294	7.7%	45.3%	47.1%

ous studies conducted among Latinos recruited based on stringent ethnicity criteria [3, 4].

Next, we compared the genotype frequency of AIMs between BAVM cases and controls. The mean and me-

dian χ^2 values for Latinos were 1.45 and 0.55, respectively. The summary χ^2 test was significant ($\Sigma\chi^2 = 119.99$, d.f. = 83, $p = 0.005$), suggesting differences in ancestral background between Latino cases and controls. In figure 2, we plotted quantiles of the observed against the expected χ^2 distribution with 1 d.f. under the null hypothesis of no stratification. The observed distribution clearly deviates from the expected distribution. That is, a greater number of observed χ^2 tests had higher values than expected for unlinked AIMs.

Next, we investigated the effect of population stratification on a promoter polymorphism in the IL-6 gene ($-174G>C$) and BAVM. In our Latino cohort, genetic ancestry proportions were associated with the IL-6 $-174G>C$ polymorphism; subjects with the IL-6 -174 GG genotype had, on average, 5.5% higher Native American ancestry ($p = 0.023$).

Table 3. IL-6 -174G>C polymorphism and risk of BAVM among Latinos

IL-6 -174G>C genotype	Cases		Controls		OR	95% CI	p
	n	%	n	%			
CC	2	2.7	5	2.4	0.93	0.18–4.94	0.933
GC	15	20.0	65	31.7	0.54	0.28–1.02	0.057
GG	58	77.3	135	65.9	reference		
CC+GC	17	22.7	70	34.1	reference		
GG ¹	58	77.3	135	65.9	1.77	0.96–3.26	0.068
GG vs. CC+GC (adjusted for age and sex)					1.85	0.99–3.48	0.055
GG vs. CC+GC (adjusted for age, sex and IAE)					1.96	1.03–3.72	0.039

¹ Reference group has changed to any C allele. Keeping the GG genotype as the reference group yields an OR of $1/1.77 = 0.56$.

Fig. 2. Quantile-quantile plots of observed χ^2 values (dots) to the expected distribution with 1 d.f. (diagonal line) under the null hypothesis of no stratification for Latino BAVM cases and controls.

The genotype frequency and association between IL-6 -174G>C promoter polymorphism and risk of BAVM among Latinos is summarized in table 3. The univariate OR for BAVM associated with the IL-6 -174 GG versus any C reference genotype was 1.77 (95% CI 0.96–3.26, $p = 0.068$). This estimate increased after adjusting for the effects of age and sex (OR = 1.85, 95% CI 0.99–3.48, $p = 0.055$). Further adjustments for IAE resulted in the highest OR (1.96, 95% CI 1.03–3.72, $p = 0.039$) for the IL-6 -174 GG genotype.

Discussion

In summary, we found evidence for population stratification in BAVM cases and controls of self-reported Latino race/ethnicity. After accounting for differences in genetic background, a common variant in the promoter region of the IL-6 gene (-174G>C) was associated with an approximately 2-fold increased risk of BAVM among Latinos ($p = 0.039$).

The proinflammatory cytokine IL-6 may play a key role in intracranial vessel dysplasia and rupture. We previously reported an almost 3-fold increased risk of hemorrhagic presentation in BAVM patients with the IL-6 -174 GG genotype [10], which further correlated with the highest IL-6 mRNA and protein levels in BAVM tissue [15]. The same polymorphism has also been associated with increased risk of subarachnoid hemorrhage among Caucasians [9], as well as a number of other neurological diseases. Additionally, the distribution of IL-6 -174 alleles is known to vary between populations [8]: the -174 G and C alleles are equally common in Caucasians (approximately 50:50), whereas the C allele is extremely rare (present at <5% frequency) among Africans and Asians genotyped in the International HapMap project (dbSNP build 128). These findings suggest that underlying genetic differences may play a role in BAVM pathogenesis and progression.

Our admixture estimates obtained from the Structure program are similar to those reported in other genetic association studies conducted among Latinos [3, 4], which used both Bayesian and maximum likelihood methods to infer individual ancestry proportions. Previous work has

suggested that approximately 50–100 AIMs are sufficient to accurately infer individual ancestry for Latinos [5]. In our study, genetic ancestry was associated with both case-control status (outcome) and IL-6 –174G>C genotype (predictor) among Latinos, meeting the classic epidemiological definition for a potential confounder. The increased OR after accounting for IAE suggests subtle negative confounding here. However, the adverse effects of even small amounts of admixture will increase markedly with larger sample sizes, and illustrates the importance of addressing population stratification.

Conclusion

In conclusion, the IL-6 –174G>C polymorphism may be a risk factor for BAVM among Latinos. This association was observed after accounting for differences in ge-

netic ancestry, and illustrates the importance of addressing population stratification in genetic association studies conducted in admixed populations. Further work in population stratification methods will allow for more accurate risk estimates and a better understanding of the influence of race/ethnicity in genetic association studies.

Acknowledgments

This work was supported by grants to H.K. (AHA Scientist Development Grant, 0735242N), W.L.Y. (NIH NS34949, NS41877), E.G.B. (NIH HL078885, American Lung Association of California, RWJ Amos Medical Faculty Development Award, NCMHD Health Disparities Scholar, Extramural Clinical Research Loan Repayment Program for Individuals from Disadvantaged Backgrounds, 2001–2003), S.C. (ATS BOLD Award, ATS-05-078) and K. Giacomini (NIH GM61390).

References

- Cardon LR, Palmer LJ: Population stratification and spurious allelic association. *Lancet* 2003;361:598–604.
- Ziv E, Burchard EG: Human population structure and genetic association studies. *Pharmacogenomics* 2003;4:431–441.
- Choudhry S, Coyle NE, Tang H, Salari K, Lind D, Clark SL, Tsai HJ, Naqvi M, Phong A, Ung N, Matallana H, Avila PC, Casal J, Torres A, Nazario S, Castro R, Battle NC, Perez-Stable EJ, Kwok PY, Sheppard D, Shriver MD, Rodriguez-Cintron W, Risch N, Ziv E, Burchard EG: Population stratification confounds genetic association studies among Latinos. *Hum Genet* 2006;118:652–664.
- Salari K, Choudhry S, Tang H, Naqvi M, Lind D, Avila PC, Coyle NE, Ung N, Nazario S, Casal J, Torres-Palacios A, Clark S, Phong A, Gomez I, Matallana H, Perez-Stable EJ, Shriver MD, Kwok PY, Sheppard D, Rodriguez-Cintron W, Risch NJ, Burchard EG, Ziv E: Genetic admixture and asthma related phenotypes in Mexican American and Puerto Rican asthmatics. *Genet Epidemiol* 2005;29:76–86.
- Tsai HJ, Choudhry S, Naqvi M, Rodriguez-Cintron W, Burchard EG, Ziv E: Comparison of three methods to estimate genetic ancestry and control for stratification in genetic association studies among admixed populations. *Hum Genet* 2005;118:424–433.
- Kim H, Sidney S, McCulloch CE, Poon KY, Singh V, Johnston SC, Ko NU, Achrol AS, Lawton MT, Higashida RT, Young WL: Racial/ethnic differences in longitudinal risk of intracranial hemorrhage in brain arteriovenous malformation patients. *Stroke* 2007;38:2430–2437.
- Labovitz DL, Halim A, Boden-Albala B, Hauser WA, Sacco RL: The incidence of deep and lobar intracerebral hemorrhage in whites, blacks, and Hispanics. *Neurology* 2005;65:518–522.
- Ness RB, Haggerty CL, Harger G, Ferrell R: Differential distribution of allelic variants in cytokine genes among African Americans and White Americans. *Am J Epidemiol* 2004;160:1033–1038.
- Morgan L, Cooper J, Montgomery H, Kitchen N, Humphries SE: The interleukin-6 gene –174G>C and –572G>C promoter polymorphisms are related to cerebral aneurysms. *J Neurol Neurosurg Psychiatry* 2006;77:915–917.
- Pawlikowska L, Tran MN, Achrol AS, McCulloch CE, Ha C, Lind DL, Hashimoto T, Zaroff J, Lawton MT, Marchuk DA, Kwok PY, Young WL: Polymorphisms in genes involved in inflammatory and angiogenic pathways and the risk of hemorrhagic presentation of brain arteriovenous malformations. *Stroke* 2004;35:2294–2300.
- Shu Y, Brown C, Castro R, Shi R, Lin E, Owen R, Sheardown S, Yue L, Burchard E, Brett C, Giacomini K: Effect of genetic variation in the organic cation transporter 1, OCT1, on metformin pharmacokinetics. *Mol Ther* 2008;18:273–280.
- Choudhry S, Taub M, Mei R, Rodriguez-Santana J, Rodriguez-Cintron W, Shriver MD, Ziv E, Risch NJ, Burchard EG: Genome-wide screen for asthma in Puerto Ricans: evidence for association with 5q23 region. *Hum Genet* 2008;123:455–468.
- Hsu TM, Kwok PY: Homogeneous primer extension assay with fluorescence polarization detection. *Methods Mol Biol* 2003;212:177–187.
- Pritchard JK, Rosenberg NA: Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 1999;65:220–228.
- Chen Y, Pawlikowska L, Yao JS, Shen F, Zhai W, Achrol AS, Lawton MT, Kwok PY, Yang GY, Young WL: Interleukin-6 involvement in brain arteriovenous malformations. *Ann Neurol* 2006;59:72–80.