Ancestral Heterogeneity in a Bi-ethnic Stroke Population

Lynda D Lisabeth[1,2], **Lewis B Morgenstern**[1,2], **David T Burke**[3], **Yan V Sun**[1], and **Jeffrey C Long**[4]

[1] Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI

[2] Stroke Program, University of Michigan, Ann Arbor, MI

[3] Department of Human Genetics, University of Michigan, Ann Arbor, MI

[4] Department of Anthropology, University of New Mexico, Albuquerque, NM

SUMMARY

To test for and characterize heterogeneity in ancestral contributions to individuals among a population of Mexican American (MA) and non-Hispanic white (NHW) stroke/TIA cases, data from a community-based stroke surveillance study in south Texas were used. Strokes/TIA cases were identified (2004–2006) with a random sample asked to provide blood. Race-ethnicity was self-reported. Thirty-three ancestry informative markers (AIMs) were genotyped and individual genetic admixture estimated using maximum likelihood methods. Three hypotheses were tested for each MA using likelihood ratio tests: 1) H0: $\mu_i=0$ (100% Native American), 2) H0: $\mu_i=1.00$ (100% European), 3) H0: $\mu_i=0.59$ (average European). Among 154 self-identified MAs, estimated European ancestry varied from 0.26–0.98, with an average of 0.59(se=0.014). We rejected hypothesis 1 for every MA and rejected hypothesis 2 for all but two MAs. We rejected hypothesis 3 for 40 MAs (20<59%, 20>59%). Among 84 self-identified NHWs, the estimated fraction of European ancestry ranged from 0.83–1.0, with an average of 0.97 (se=0.014). Self-identified MAs, and to a lesser extent NHWs, from an established bi-ethnic community were heterogeneous with respect to genetic admixture. Researchers should not use simple race-ethnic categories as proxies for homogeneous genetic populations when conducting gene mapping and disease association studies in multi-ethnic populations.

Keywords

stroke; ethnicity; ancestry

INTRODUCTION

Mexican Americans (MA) are the largest subgroup of the largest minority group in the US. Several health disparities have been identified for the MA population, including increased risk of complex neurologic diseases such as ischemic stroke, compared with non-Hispanic whites (NHW).(Morgenstern *et al.*, 2004) Reasons for these health disparities are largely unknown but are likely multi-factorial with environmental, social, and genetic underpinnings.

Characterization of health disparities among MAs has historically relied on self-reported race and ethnicity. Complicating an understanding of the observed health disparities in this

Address for correspondence: Lynda Lisabeth, PhD, 1415 Washington Heights, Ann Arbor, MI 48109, Telephone: 734-936-9649, Fax: 734-764-3192, llisabet@umich.edu.

population is an incomplete knowledge of what self-reported MA race-ethnicity represents from a genetic perspective. Recent advances in technology allow researchers to quantify race-ethnicity at the molecular level using ancestry informative genomic DNA markers (AIMs). Ancestry informative marker alleles provide quantitative estimates of the proportional contributions of African, European, and Native American ancestors to MA individuals and to the current MA population as a whole. Recent studies utilizing AIMs have reported that Native American ancestors contributed on average 35–52% of the genome to MA individuals.(Salari *et al.*, 2005, Kosoy *et al.*, 2009, Tang *et al.*, 2006, Basu *et al.*, 2008, Shtir *et al.*, 2009, Bonilla *et al.*, 2004a)

Although it is possible to quantify ancestry independently of an individual's self-reported information using genetic markers, large-scale epidemiology studies are likely to continue to use self-reported race-ethnicity for several reasons. First, DNA is expensive to collect and genotype relative to acquiring self-report information. Second, self reported race-ethnicity might indicate disease risk better than genetic ancestry alone because it is a proxy for lifestyle and other social factors as well as genetic inheritance. Third, we are uncertain how well ancestry from different populations serves as a proxy for disease risk, although recent studies have demonstrated associations of ancestry with subclinical cardiovascular disease (Wassel *et al.*, 2009) and complex neurologic diseases such as multiple sclerosis.(Reich *et al.*, 2005)

An understanding of ancestry at the molecular level in MAs would aid researchers trying to identify reasons for health disparities in this population through epidemiologic research by informing the degree to which self-reported race-ethnicity approximates genetic admixture. The objective of this study was to use previously identified AIMS to characterize and test for the heterogeneity in ancestral contributions to individuals among a population of self-identified MA and NHW stroke or transient ischemic attack (TIA) cases in southeast Texas.

METHODS

Participants in this study consist of $n = 154$ MAs and $n = 84$ NHWs from the Brain Attack Surveillance in Corpus Christi (BASIC) Project, a population-based stroke surveillance study in Nueces County, Texas. Detailed methods for this project have been published. (Smith *et al.*, 2004, Morgenstern *et al.*, 2004) Nueces County is located in south Texas on the Gulf Coast, and has a population size of roughly 300,000. MAs comprise the majority of residents, at 56% of the population based on the 2000 US Census. NHWs comprise 38% of the population, and other race-ethnicities comprise the remaining 6%. MAs in this county are primarily second and third generation US citizens. We previously reported that 87% of MAs and 93% of NHWs were born in the US. Mexico was the reported origin of all MA subjects not born in the US. On average, these individuals had been living in the US for 60 years (range 19–86 years).(Smith *et al.*, 2003) Stroke/TIA cases were identified among individuals ≥45 years seen at one of seven area hospitals located within Nueces County between June 2004 and June 2006. Cases were also identified through neurologists practicing in Nueces County. Cerebrovascular events were validated by board certified neurologists based on published criteria and blinded to subjects' ethnicity and age.(Asplund *et al.*, 1988) A random sample was asked to participate in an in-person interview and to provide a blood sample. The response rate for the blood draw was 70% with no ethnic difference (MA: 73%, NHW: 65%; $p = 0.07$). All study participants signed an informed consent document and the study was approved by the Institutional Review Boards at the University of Michigan and all local hospitals.

Peripheral venous blood samples were collected by venipuncture from each participant by a trained phlebotomist. Clinical blood samples were sent to the NINDS Human Genetics

Resource Center DNA and Cell Line Repository (http://ccr.coriell.org/ninds). According to established protocols, genomic DNA was extracted from the whole blood or lymphocyte cell pellets using the Qiagen Autopure method. Briefly, cells are lysed by addition of anionic detergent containing RNase and EDTA. After mixing, a salt solution is added and the insoluble cell debris is removed by centrifugation. An equal volume of isopropanol is added to the supernatant and the resulting DNA precipitate is collected by centrifugation. Following a brief rinse with 70% ethanol to remove residual salt the DNA pellet is solubilized overnight in TE buffer (0.01 M Tris, pH 8.0/0.001 M EDTA). After extraction, the DNA proceeds through several processing steps and must meet specific criteria: 260/280 nm absorbance ratio is between 1.65 and 1.95, concentration is at least 0.1 mg/ml, sample contains less than 0.1 μg protein per μg of DNA, and restriction enzyme digestion yields a broad size distribution of DNA fragments. Amplification by PCR with microsatellite and amelogenin gene-specific primers must also produce amplicon sizes that bin into expected allele sizes, and give fragment peak heights that are at least 3-fold above background. The amplified product allele peak heights are within 70% of each other, and there are not more than 2 allele peaks observed for each microsatellite locus.

Race-ethnicity was self-reported and collected as in the US Census. MA ethnicity was defined as self-reported ethnicity "of Hispanic origin", either with race of "white" or with race "refused". Refused is included as it is common among this population to consider "Hispanic" or "Mexican American" as a race. NHW was defined by a self-reported race of "white" and ethnicity of "not of Hispanic origin". Individuals who reported a race-ethnicity other than MA or NHW were excluded due to small numbers (n = 30).

Ancestry Informative Markers: We analyzed genotypes from 33 genomic single nucleotide polymorphisms (SNPs) dispersed across 17 chromosomes. The nearest physical distance between markers on the same chromosome was >1 million base pairs. This set of markers has been previously identified as being AIMs for estimating European and Native American contributions to admixed populations in the Americas.(Tian et al., 2007, Seldin et al., 2007) The absolute value of the difference in allele frequency between two ancestral populations, δ, is a simple measure of the effectiveness of a marker for estimating ancestry. Previous reports have used δ >0.3 as the threshold for declaring a SNP as being "ancestry informative".(Mao et al., 2007, Shtir et al., 2009, Bonilla et al., 2004a) All markers used in this study (table 1) had δ between Europeans and Native Americans ≥ 0.5 (median=0.8). For European and Native American parental population allele frequencies, we used published values.(Seldin *et al.*, 2007, Tian *et al.*, 2007) The AIMs employed in this study are useful for analysis of Native American and European ancestral contributions because they show high allele frequency differences between indigenous populations from the Americas and Europe, and low allele frequency differences among local populations on the same continent.

Genotyping Methods

The 33 AIMs were genotyped using oligonucleotide ligation(Barany, 1991) followed by electrophoresis using four main steps. First, we performed multiplex PCR amplification in batches of approximately 10 loci. Each locus was amplified using locus-specific primers. Second, to enrich the amplicon concentrations for all loci, we re-amplified the products of step 1 using primers that are complementary to a universal tag sequence incorporated into the initial locus-specific PCR primer pairs. Third, we ligated fluorescently labeled oligonucleotides specific to SNP alleles to the PCR amplification products. The nucleotide lengths of the ligation oligonucleotide products yield size classes that allow unambiguous separation by gel electrophoresis. Finally, electrophoretic separation and detection of the ligated products occurred using a capillary DNA sequencer.

Statistical Analysis

We tested deviations from Hardy-Weinberg equilibrium using likelihood ratio statistics, and measured the degree of departure from equilibrium using the within locus intraclass allelic correlation, F_1, as defined by Risch et al.(Risch *et al.*, 2009) We tested for deviations from 'linkage equilibrium' between all pairs of loci using chi-squared tests based on the r^2 statistic.(Weir, 1996) We estimated individual genetic admixture for each participant using the method of maximum likelihood (Chakraborty, 1986) based on two parental populations, European Americans and Native Americans. For each person we evaluated the likelihood function $L(\mu_i)$, where μ_i represents the fraction of ancestors of that person who were of European origin. By this method, the estimate of individual ancestry is the value $\hat{\mu}_i$ that maximizes the likelihood function. For each estimate, $\hat{\mu}_i$, we estimated the standard error of the estimate $s_{\hat{\mu}_i}$ from Fisher's information criterion $I_{\hat{\mu}_i} = -(d^2/d\mu^2)\ln[L(\mu_i)]$ using the formula

$s_{\widehat{\mu}_i} = \sqrt{1/I_{\widehat{\mu}_i}}$.(Edwards, 1992) We estimated average admixture for participants in each race-ethnic group using two methods: 1) the average of individual estimates described above, and 2) the method of weighted least squares as implemented in ADMIX.(Long, 1991)

Individual ancestry estimates from genetic markers have high standard errors that lead to wide confidence intervals. This presents two challenges: 1) showing that an individual deviates statistically from a predetermined reference point, such as 100% ancestry from either, or both, of the putative parental populations, and 2) showing that individuals in a sample are heterogeneous, with respect to their true proportions of ancestry from the putative parental populations. We used the following likelihood ratio statistic to address these problems,

$$G = -2[(\ln L(\mu_1) - \ln L(\widehat{\mu}_i))]$$

where μ_1 is a specified fraction of European ancestry and $\hat{\mu}_i$ is the ancestry fraction that maximizes the likelihood function for the ith individual. The null hypothesis is H_0: $\mu_i = \mu_1$. G is distributed asymptotically as a χ^2 random variable with degrees of freedom one less than the number of parental populations.(Edwards, 1992)

Finally, we compared the proportion of European ancestry with age at stroke onset and having a high school education using correlation coefficients and t-tests separately among MAs and NHWs.

RESULTS

Among the 238 stroke/TIA cases, mean age was 69 years (σ=13) and 49% were female. MAs were younger ($p < 0.0001$) and less likely to have a high school education ($p < 0.0001$) than NHWs (table 2). Among the 154 participants of self-reported MA race-ethnicity, the range of estimated fraction of European ancestry was 0.259–0.975 (table 3). The average of individual European ancestry estimates was 0.591±0.014. Using weighted least squares method, we estimated the fraction of European ancestry for the group to be 0.589±0.011, which agrees well with the average of individual estimates. Among the 84 participants of self-reported NHW race-ethnicity, the estimated fraction of European ancestry ranged from 0.827–1.00. The average of individual European ancestry was 0.968±0.014. Using the weighted least squares method, we estimated the fraction of European ancestry for the group to be 0.963±0.014, which also agrees well with the average of individual estimates for NHWs.

The next step was to document heterogeneity in ancestral contributions to the individual MA and NHW participants. We tested the following three hypotheses for each MA case, $H_0 : \mu_i = 0$, $H_0 : \mu_i = 1.00$, and $H_0 : \mu_i = 0.591$. The first hypothesis establishes whether a MA participant differs significantly from a person who has 100% Native American ancestry. The second hypothesis establishes whether a MA participant differs significantly from a person who has 100% European Ancestry. The third hypothesis establishes whether a MA participant differs significantly from the average European ancestry for the group as a whole (i.e., 59% European). We rejected the "100% Native American ancestry" hypothesis for every MA case, and we rejected the "100% European ancestry" hypothesis for all but two MA cases. We rejected the hypothesis that a self-reported MA had ancestry consistent with the MA population (59% European ancestry) for 40 of the 154 MA cases. Twenty individuals were significantly higher and 20 significantly lower than the mean European ancestry (figure 1).

Given that there is significant heterogeneity in ancestral contributions to MA individuals, we expect to see an excess of homozygosity within loci, and linkage disequilibrium among pairs of loci (even unlinked loci). Our results confirm these expectations. At $\alpha = 0.05$ or less, we found a significant excess of homozygotes at 6 (18%) loci. The mean intraclass correlation, F_1, was 0.035. At $\alpha = 0.05$ or less, we found significant linkage disequilibrium between 84 (16%) locus pairs.

To establish the extent of ancestral heterogeneity in the NHW sample, we tested the hypothesis that each person has 100% European ancestry, i.e., $H_0 : \mu_i = 1.00$. We rejected this hypothesis for 15 cases (figure 1). To follow-up on this result, we also tested the following two hypotheses: $H_0 : \mu_i = 0.94$ and $H_0 : \mu_i = 0.591$. The first hypothesis establishes whether a NHW participant differs significantly from a person who has the equivalent of one Native American great-great-grandparent. The second hypothesis establishes whether a NHW participant differs significantly from the average European ancestry for the MA sample. We rejected the $\mu_i = 0.94$ hypothesis for three NHWs, each of whom was estimated to have 100% European ancestry. We rejected the $\mu_i = 0.591$ hypothesis for all NHWs.

Given the small degree of heterogeneity in ancestral contributions to NHW individuals, we tested for excess of homozygosity within loci and linkage disequilibrium among pairs of loci. At $\alpha = 0.05$ or less, we found a significant excess of homozygotes at three loci (9%). The mean intraclass correlation was 0.012. At $\alpha = 0.05$ or less, we found significant linkage disequilibrium between 42 locus pairs (8%) in NHWs.

In further analysis, European ancestry was not associated with age at stroke among MAs ($p = 0.93$) or NHWs ($p = 0.16$). European ancestry was also not associated with having a high school education, a proxy for socio-economic status, among MAs ($p = 0.48$) or NHWs ($p = 0.93$)

DISCUSSION

Today many distinct populations live in the Americas with ancestry mixed between people who lived in Africa, Europe, or the Americas before the colonial era. Although many people refer collectively to mixed populations in the Americas as Hispanic, Latino, or Mestizo, geneticists recognize that these groups have distinct gene pools that trace different proportions of ancestors to each of the three continental regions. Self-identified MAs living in different cities typically have 35–50% Native American ancestry and a trace component, 4–6%, of African ancestry, whereas the self-identified Puerto Rican population as a whole has 15–18% Native American ancestry but a more substantial component of African ancestry (~20%).(Basu *et al.*, 2008, Collins-Schramm *et al.*, 2004, Risch *et al.*, 2009, Tseng

et al., 1998, Bonilla *et al.*, 2004a, Salari *et al.*, 2005) In this study, where we compared the degree to which a sample of self-identified MAs approximated a random mating population in genetic equilibrium, we found that MAs were a heterogeneous group regarding genetic ancestry, with individual estimates of Native American ancestry ranging from 2–74%. While our estimated average of 41% Native American ancestry is consistent with recently reported estimates,(Basu *et al.*, 2008, Shtir *et al.*, 2009, Kosoy *et al.*, 2009, Salari *et al.*, 2005, Tang *et al.*, 2006, Risch *et al.*, 2009, Bonilla *et al.*, 2004a) we found that more than a quarter of the MA cases were significantly different from the average European ancestry in the MA population as a whole.

We also found some heterogeneity in the ancestry of NHWs, with individual estimates of Native American ancestry ranging from 0–17%. While we found that 18% of NHWs had significantly less than 100% European ancestry, the average Native American ancestry in the NHW sample was roughly equivalent to one Native American great-great-grandparent. Thus, we should recognize that people who consider themselves ethnically NHW may have ancestors who were Native American. Direct unions between Native Americans and NHWs may have introduced this ancestry, but it is also likely that unions between MA and NHWs introduced this Native American ancestry indirectly. Genetic marker analysis cannot resolve this issue, but questionnaires could provide some information about the patterns of gene flow. No matter the origin of Native American ancestry in the NHW sample, it is likely the cause of departures from Hardy-Weinberg equilibrium and linkage equilibrium in the sample. Our results confirm that individuals within both Hispanic and non-Hispanic white US Census categories are heterogeneous with respect to European and Native American ancestry. While neither Census group in southeastern Texas constituted a genetic population, the NHW group was far more ancestrally homogeneous than the MA group.

Heterogeneity of population ancestry in other Hispanic communities has been reported including populations in New York,(Bonilla *et al.*, 2004b) southern Colorado,(Bonilla *et al.*, 2004a) the state of Guerrero, Mexico,(Bonilla *et al.*, 2005) Mexico City and San Francisco. (Risch *et al.*, 2009) Population genetic principles show that random mating causes variation in ancestry to decrease from one generation to the next due to segregation and recombination. On this basis we expect that individuals in well-established admixed populations will be homogenous with respect to the composition of ancestral populations. Differences in ancestral contributions to individuals demonstrate some departure from random mating. Risch and colleagues recently found evidence that Mexicans in Mexico City and MAs in San Francisco prefer mates with similar ancestries.(Risch *et al.*, 2009) This assortative mating is one mechanism that can maintain inter-individual heterogeneity in contributions from ancestral populations. Although Risch et al did not formally test for heterogeneity, they observed a wide spread of Native American ancestry in their population with a mean of 0.44 (σ=0.14), similar to our results. Also paralleling our finding of 15% of unlinked locus pairs in linkage disequilibrium in our MA population, they reported that 10–16% of unlinked locus pairs in their samples were in linkage disequilibrium. In San Francisco, the mean correlation of alleles within loci was 0.015, whereas we observed 0.035. Together the results of these studies suggest that assortative mating may partially explain the observed inter-individual heterogeneity in ancestry estimates demonstrated in MAs.

One-way gene flow from NHWs to MAs is another mechanism that could maintain Hardy-Weinberg and linkage disequilibrium in the MA population. In addition, recent migrants from Mexico may consist of individuals with lower European ancestry than the second- and third-generation United States Citizens that make up 87% of our Nueces County MA sample. Finally, differences in socioeconomic status may partially explain the observed heterogeniety if individuals of lower socioeconomic status have higher Native American ancestry. However, our analysis considering the association between ancestry and having a

high school education, a proxy for socioeconomic status, did not support this hypothesis. We note that the mechanisms that can maintain heterogeneity in ancestry are not exclusive mutually. We currently lack the necessary data that could further distinguish among these possibilities.

Our findings are important to epidemiological studies because they show that researchers cannot use simple race-ethnic categories designed for the US Census and other government purposes as proxies for homogeneous genetic populations when conducting gene mapping and disease association studies. Heterogeneity in ancestral contributions to individuals creates correlations among alleles within loci and among loci. Both sources of correlation can strengthen the association between linked genetic markers and complex diseases such as ischemic stroke, but correlations that owe to non-random mating in populations are not always beneficial because they can create spurious associations between genetic markers and disease. Special care is needed to sort out the true nature of correlations in complex populations such as we have demonstrated for MAs in Texas.

The population studied consisted of stroke/TIA cases. Previous research in the study community has shown that stroke disproportionately affects MAs, especially at younger ages.(Morgenstern *et al.*, 2004) We have also demonstrated that having a first degree relative with stroke increases one's risk of stroke particularly in MAs.(Lisabeth *et al.*, 2008) Siblings of MA ischemic stroke/TIA cases have roughly double the stroke risk compared to what would be expected based on national estimates of stroke prevalence in MAs. These findings together with the current finding of ancestral heterogeneity in the MA population suggest that ischemic stroke may be a suitable phenotype for admixture disequilibrium mapping to identify stroke susceptibility genes in this population.

Limitations of this work warrant discussion. Individual ancestry estimates were derived from 33 AIMs, which is a smaller set than other recent reports which have characterized ancestry in MAs. This may have led to somewhat larger standard errors around our estimates. However, the AIMs used for the current study were chosen such that the δ between Europeans and Native Americans was ≥ 0.5. This criterion is stricter than most previous reports. More importantly, these 33 AIMs provided enough information for us to reject our null hypotheses of ancestry homogeneity in both the MA and NHW samples, and to show evidence for admixture-related departures from Hardy-Weinberg equilibrium and linkage equilibrium. Thus, they provide sufficient information to achieve the study's goals.

Our model for genetic admixture constructs MA ancestry in Nueces County, Texas using two parental populations, Europeans and Native Americans. However, several reports indicate that most MA populations also harbor a small proportion of African ancestry (~5%). We decided against a three population model because the fraction of African ancestry is likely to be low and our AIMs are powerful only for distinguishing Native American ancestry from European ancestry. This is consistent with the typical marker selection strategy for admixture analyses in MAs on the two populations contributing the most ancestry to MAs.(Tian *et al.*, 2007) In addition, African ancestry is unlikely to change our main findings that neither MAs nor NHWs in Nueces, County Texas homogeneous populations with respect to their ancestral compositions.

The study population was limited to individuals with stroke/TIA. It is possible that this disease outcome influenced the estimates of genetic admixture, but it is unlikely. If a major gene contributes to stroke/TIA in the MA and NHW population, then it can influence admixture estimates through linkage disequilibrium with our AIMs, but our AIMs are unlinked and this would prevent linkage disequilibrium with a gene for stroke from having a large influence on ancestry estimates. Moreover, since all of our participants in this study

are stroke/TIA patients, the disease outcome cannot account for our major finding, *i.e.,* both the MA and NHW samples are heterogeneous with respect to Native American and European ancestry.

Summary

We observed that self-identified MAs from a bi-ethnic US community were heterogeneous with respect to genetic admixture, with estimates of Native American ancestry ranging considerably among individuals. Our findings suggest that researchers should not use simple self-reported race-ethnic categories as proxies for homogeneous genetic populations when conducting gene mapping and disease association studies in this growing segment of the population. However, self-reported race-ethnicity is a proxy for lifestyle and social factors and thus retains importance in the study of complex diseases such as stroke.

Acknowledgments

ND11106	ND11487	ND11653	ND11806	ND12197	ND12207
ND12208	ND12415	ND12536	ND12538	ND12539	ND12624
ND12715	ND12784	ND13111	ND13128	ND13388	ND13389
ND13430	ND13431	ND13465	ND13518	ND13600	ND13885
ND13886	ND13887	ND13994	ND13995	ND13996	ND14146
ND14229	ND14279	ND14376	ND14377	ND14378	ND14632
ND14809	ND14810	ND14812	ND14849	ND14894	ND14895
ND14896	ND14897	ND14898	ND14930	ND15032	ND15133
ND15190	ND15223	ND15335	ND15522	ND15524	ND15525
ND15601	ND15602	ND15603	ND15628	ND15629	ND15630
ND15631	ND15649	ND15728	ND15757	ND15758	ND15786
ND15787	ND15792	ND15793	ND15810	ND15849	ND15850
ND15851	ND15988	ND16030	ND16076	ND16077	ND16079
ND16081	ND16092	ND16237	ND16238	ND16239	ND16240
ND16242	ND16243	ND16245	ND16310	ND16311	ND16350
ND16351	ND16353	ND16354	ND16355	ND16404	ND16535
ND16536	ND16537	ND16538	ND16564	ND16565	ND16566
ND16599	ND16600	ND16602	ND16635	ND16636	ND16637
ND16669	ND19236	ND19237	ND19238	ND19240	ND19241
ND19242	ND19288	ND19330	ND19487	ND19488	ND19489
ND19491	ND19532	ND19533	ND19603	ND19632	ND19740
ND19792	ND19793	ND19797	ND19854	ND19856	ND19858
ND19861	ND19922	ND19981	ND19985	ND19986	ND20030
ND20099	ND20148	ND20195	ND20196	ND20197	ND20394
ND20396	ND20397	ND20398	ND20399	ND20400	ND20401
ND20402	ND20469	ND20514	ND20515	ND11828	ND12045
ND12414	ND12535	ND12537	ND12716	ND12717	ND12847

ND12978	ND12979	ND13108	ND13209	ND13599	ND13766
ND13858	ND13859	ND14033	ND14145	ND14147	ND14424
ND14633	ND14687	ND14688	ND14689	ND14746	ND14807
ND15132	ND15224	ND15521	ND15523	ND15604	ND15648
ND15675	ND15676	ND15677	ND15719	ND15756	ND15784
ND15791	ND16029	ND16032	ND16033	ND16080	ND16093
ND16235	ND16236	ND16244	ND16309	ND16312	ND16329
ND16331	ND16352	ND16403	ND16598	ND16667	ND16668
ND19179	ND19235	ND19239	ND19283	ND19284	ND19285
ND19286	ND19328	ND19406	ND19407	ND19423	ND19486
ND19490	ND19602	ND19791	ND19853	ND19855	ND19857
ND19919	ND19920	ND19921	ND19987	ND20033	ND20145
ND20147	ND20245	ND20440	ND20516		

References

Asplund K, Tuomilehto J, Stegmayr B, Wester PO, Tunstall-Pedoe H. Diagnostic criteria and quality control of the registration of stroke events in the MONICA project. Acta medica Scandinavica. 1988; 728:26–39. [PubMed: 3202029]

Barany F. The ligase chain reaction in a PCR world. PCR Methods Appl. 1991; 1:5–16. [PubMed: 1842922]

Basu A, Tang H, Zhu X, Gu CC, Hanis C, Boerwinkle E, Risch N. Genome-wide distribution of ancestry in Mexican Americans. Human genetics. 2008; 124:207–14. [PubMed: 18752003]

Bonilla C, Gutierrez G, Parra EJ, Kline C, Shriver MD. Admixture analysis of a rural population of the state of Guerrero, Mexico. American journal of physical anthropology. 2005; 128:861–9. [PubMed: 16118786]

Bonilla C, Parra EJ, Pfaff CL, Dios S, Marshall JA, Hamman RF, Ferrell RE, Hoggart CL, Mckeigue PM, Shriver MD. Admixture in the Hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping. Annals of human genetics. 2004a; 68:139–53. [PubMed: 15008793]

Bonilla C, Shriver MD, Parra EJ, Jones A, Fernandez JR. Ancestral proportions and their association with skin pigmentation and bone mineral density in Puerto Rican women from New York city. Human genetics. 2004b; 115:57–68. [PubMed: 15118905]

Chakraborty R. Gene admixture in human populations - Models and predictions. Yearbook of physical anthropology. 1986; 29:1–43.

Collins-Schramm HE, Chima B, Morii T, Wah K, Figueroa Y, Criswell LA, Hanson RL, Knowler WC, Silva G, Belmont JW, Seldin MF. Mexican American ancestry-informative markers: examination of population structure and marker characteristics in European Americans, Mexican Americans, Amerindians and Asians. Human genetics. 2004; 114:263–71. [PubMed: 14628215]

Edwards, A. Likelihood. Baltimore: Johns Hopkins University Press; 1992.

Kosoy R, Nassir R, Tian C, White PA, Butler LM, Silva G, Kittles R, Alarcon-Riquelme ME, Gregersen PK, Belmont JW, De La Vega FM, Seldin MF. Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. Human mutation. 2009; 30:69–78. [PubMed: 18683858]

Lisabeth LD, Peyser PA, Long JC, Majerisk JJ, Smith MA, Morgenstern LB. Stroke among siblings in a biethnic community. Neuroepidemiology. 2008; 31:33–8. [PubMed: 18535398]

Long JC. The genetic structure of admixed populations. Genetics. 1991; 127:417–28. [PubMed: 2004712]

Mao X, Bigham AW, Mei R, Gutierrez G, Weiss KM, Brutsaert TD, Leon-Velarde F, Moore LG, Vargas E, Mckeigue PM, Shriver MD, Parra EJ. A genomewide admixture mapping panel for

Hispanic/Latino populations. American journal of human genetics. 2007; 80:1171–8. [PubMed: 17503334]

Morgenstern LB, Smith MA, Lisabeth LD, Risser JM, Uchino K, Garcia N, Longwell PJ, Mcfarling DA, Akuwumi O, Al-Wabil A, Al-Senani F, Brown DL, Moye LA. Excess stroke in Mexican Americans compared with non-Hispanic Whites: the Brain Attack Surveillance in Corpus Christi Project. American journal of epidemiology. 2004; 160:376–83. [PubMed: 15286023]

Reich D, Patterson N, De Jager PL, Mcdonald GJ, Waliszewska A, Tandon A, Lincoln RR, Deloa C, Fruhan SA, Cabre P, Bera O, Semana G, Kelly MA, Francis DA, Ardlie K, Khan O, Cree BA, Hauser SL, Oksenberg JR, Hafler DA. A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. Nature genetics. 2005; 37:1113–8. [PubMed: 16186815]

Risch N, Choudhry S, Via M, Basu A, Sebro R, Eng C, Beckman K, Thyne S, Chapela R, Rodriguez-Santana JR, Rodriguez-Cintron W, Avila PC, Ziv E, Gonzalez Burchard E. Ancestry-related assortative mating in Latino populations. Genome biology. 2009; 10:R132. [PubMed: 19930545]

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. Genetic epidemiology. 2005; 29:76–86. [PubMed: 15918156]

Seldin MF, Tian C, Shigeta R, Scherbarth HR, Silva G, Belmont JW, Kittles R, Gamron S, Allevi A, Palatnik SA, Alvarellos A, Paira S, Caprarulo C, Guilleron C, Catoggio LJ, Prigione C, Berbotto GA, Garcia MA, Perandones CE, Pons-Estel BA, Alarcon-Riquelme ME. Argentine population genetic structure: large variance in Amerindian contribution. Am J Phys Anthropol. 2007; 132:455–62. [PubMed: 17177183]

Shtir CJ, Marjoram P, Azen S, Conti DV, Le Marchand L, Haiman CA, Varma R. Variation in genetic admixture and population structure among Latinos: the Los Angeles Latino eye study (LALES). BMC genetics. 2009; 10:71. [PubMed: 19903357]

Smith MA, Risser JM, Lisabeth LD, Moye LA, Morgenstern LB. Access to care, acculturation, and risk factors for stroke in Mexican Americans: the Brain Attack Surveillance in Corpus Christi (BASIC) project. Stroke; a journal of cerebral circulation. 2003; 34:2671–5.

Smith MA, Risser JM, Moye LA, Garcia N, Akiwumi O, Uchino K, Morgenstern LB. Designing multi-ethnic stroke studies: the Brain Attack Surveillance in Corpus Christi (BASIC) project. Ethnicity & disease. 2004; 14:520–6. [PubMed: 15724771]

Tang H, Jorgenson E, Gadde M, Kardia SL, Rao DC, Zhu X, Schork NJ, Hanis CL, Risch N. Racial admixture and its impact on BMI and blood pressure in African and Mexican Americans. Human genetics. 2006; 119:624–33. [PubMed: 16738946]

Tian C, Hinds DA, Shigeta R, Adler SG, Lee A, Pahl MV, Silva G, Belmont JW, Hanson RL, Knowler WC, Gregersen PK, Ballinger DG, Seldin MF. A genomewide single-nucleotide-polymorphism panel for Mexican American admixture mapping. Am J Hum Genet. 2007; 80:1014–23. [PubMed: 17557415]

Tseng M, Williams RC, Maurer KR, Schanfield MS, Knowler WC, Everhart JE. Genetic admixture and gallbladder disease in Mexican Americans. American journal of physical anthropology. 1998; 106:361–71. [PubMed: 9696151]

Wassel CL, Pankow JS, Peralta CA, Choudhry S, Seldin MF, Arnett DK. Genetic ancestry is associated with subclinical cardiovascular disease in African-Americans and Hispanics from the multi-ethnic study of atherosclerosis. Circulation. 2009; 2:629–36. [PubMed: 20031644]

Weir, B. Genetic Data Analysis II: Methods for Discrete Population Genetic Data. Sinauer Associates, Incorporated; 1996.
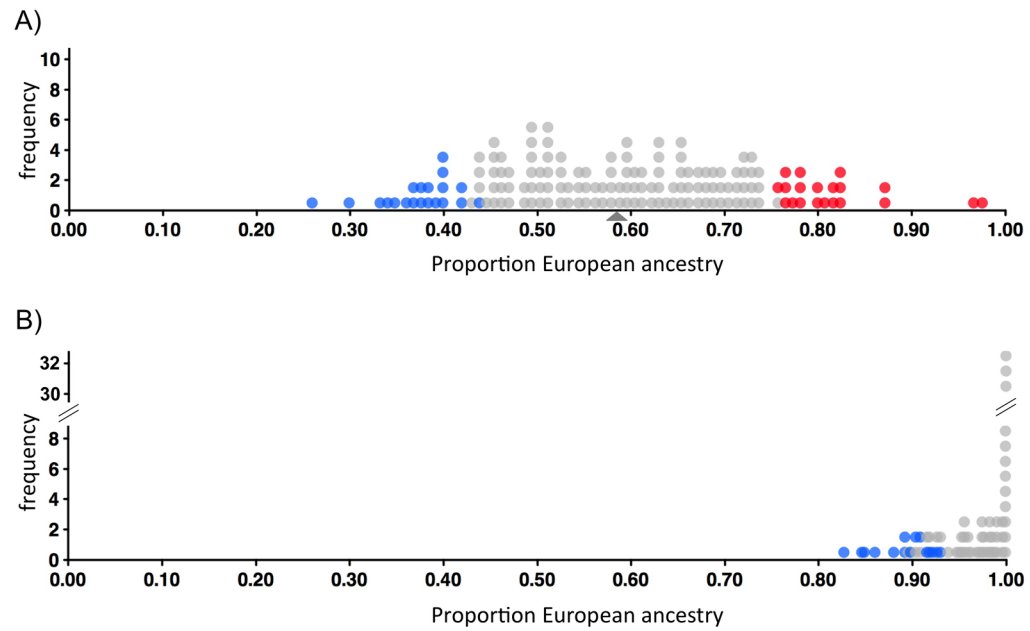
Fig. 1.
(A) Proportion of European ancestry in the Mexican American sample (n=154). Black triangle indicates mean European Ancestry for the Mexican American sample (59%). For Mexican Americans, blue dots represent individuals that had statistically lower European ancestry than the mean value (n = 20). Red dots represent individuals that had statistically greater European ancestry than the mean value (n = 20)
(B) Proportion of European ancestry in the non-Hispanic white sample (n = 84). For non-Hispanic whites, blue dots represent individuals that had statistically lower European ancestry than 100% (n = 15). In total, 32 non-Hispanic whites had genotype results with 100% European ancestry (32/84 = 38%)

Table 1

Ancestry informative markers (AIMs) and absolute value of the difference in allele frequency(δ) between ancestral populations (European and Native American).

Locus	Allele	European	Native American	δ	Location
rs1951936	A	0.85	0.06	0.79	10p12
rs11256014	A	0.05	0.55	0.5	10p14
rs1638567	C	0.06	0.64	0.58	11q13
rs11169154	A	0.95	0.15	0.8	12q13
rs7995033	C	0.85	0.19	0.66	13q12
rs9319336	C	0.04	0.89	0.85	13q12
rs2324596	C	0.05	0.92	0.87	13q13
rs1540979	A	0.89	0.21	0.68	13q31
rs12102256	A	0.91	0.05	0.86	15q14
rs1426654	A	1	0.05	0.95	15q21
rs1950030	A	0.93	0.1	0.83	15q21
rs6587216	C	0.8	0.2	0.6	17p11.2
rs17638989	C	0.56	0.01	0.55	19p13.2
rs1931059	A	0.19	0.77	0.58	1p35
rs7504	A	0.22	0.95	0.73	1p36.1
rs1407434	C	0.92	0.08	0.84	1q25
rs6086473	C	0.22	0.84	0.62	20p12
rs293553	A	0.67	0.02	0.65	20q11.2
rs3755095	A	0.92	0.05	0.87	2p12
rs3907854	C	0.99	0.21	0.78	2p13
rs3827760	C	0.02	0.96	0.94	2q12
rs7432238	A	0.93	0.1	0.83	3p24
rs2700394	C	0.99	0.13	0.86	3q21
rs2165139	A	0.88	0.04	0.84	3q22
rs11725412	A	0.06	0.99	0.93	4p14
rs12501010	C	0.06	0.93	0.87	4q26
rs262838	A	0.92	0.21	0.71	5q36
rs12662498	A	0.94	0.04	0.9	6p12

Locus	Allele	European	Native American	δ	Location
rs9369677	C	0.07	0.88	0.81	6p12
rs2439522	A	0.88	0.26	0.62	8q22
rs4478653	C	0.36	1	0.64	9p21
rs10809782	A	0.08	0.88	0.8	9p23
rs7863917	A	0.01	0.8	0.79	9q31

Table 2

Socio-demographic characteristics by self-reported race-ethnicity, Mexican American and non-Hispanic white (n = 238).

Variable	Mexican American (n = 154)	Non-Hispanic White (n = 84)
Mean Age (sd)	66.3 (12.8)	73.4 (12.6)
% Female (n)	50.0 (77)	47.6 (40)
% High School Education (n)	45.5 (70)	79.8 (67)

Table 3

Average contributions of European and Native American ancestry by self-reported race-ethnicity, Mexican American and non-Hispanic white (n = 238).

	Mexican American (n = 154)			Non-Hispanic White (n = 84)		
	European	Native American	se	European	Native American	se
WLS	0.589	0.411	0.011	0.968	0.032	0.014
average μ_i	0.591	0.409	0.014	0.963	0.037	0.005

WLS = weighted least squares, se = standard error