Dissecting the role of Amerindian genetic ancestry and *ApoE* ε4 allele on Alzheimer disease in an admixed Peruvian population

M. Cornejo-Olivas[1,2], F. Rajabli[3], V. Marca[1], P.G. Whitehead[3], N. Hofmann[3], O. Ortega[1], K. Milla-Neyra[1], D. Veliz-Otani[1], N. Custodio[4], R. Montesinos[4], S. Castro-Suarez[5,6], M. Meza[5,6], L.D. Adams[3], P.R. Mena[3], R. Isasi[3,7], M.L. Cuccaro[3,7], J.M. Vance[3,7], G.W. Beecham[3,7], P. Mazzetti[1,6], M.A. Pericak-Vance[3,7*]

1) Neurogenetics Research Center, Instituto Nacional de Ciencias Neurológicas, 2) Center for Global Health, Universidad Peruana Cayetano Heredia, Lima, Peru 3) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, Florida, USA 4) Instituto Peruano de Neurociencias 5) CBI en Demencias y enfermedades desmielinizantes del sistema nervioso, Instituto Nacional de Ciencias Neurológicas 6) School of Medicine, Universidad Nacional Mayor de San Marcos 7) Dr. John Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, Florida, USA

[*]Corresponding Author:

Margaret A. Pericak-Vance, PhD

Director, John P. Hussman Institute for Human Genomics

Dr. John T Macdonald Foundation Professor of Human Genetics

Executive Vice Chair, Dr. John T Macdonald Foundation Department of Human Genetics

John P. Hussman Institute for Human Genomics

University of Miami Miller School of Medicine

1501 NW 10th Avenue, BRB 319

Miami, FL 33136

mpericak@med.miami.edu

Abstract

Alzheimer disease (AD) is the leading cause of dementia in the elderly and occurs in all ethnic and racial groups. *ApoE ε4* is the most significant genetic risk factor for late-onset AD and shows the strongest effect among East Asian populations followed by non-Hispanic White populations and has a relatively lower effect in African descent populations. Admixture analysis in the African American and Puerto Rican populations showed that the variation in *ε4* risk is correlated with the genetic ancestral background local to the *ApoE* gene. Native American populations are substantially underrepresented in AD genetic studies. The Peruvian population with up to ~80 of Amerindian ancestry provides a unique opportunity to assess the role of Amerindian ancestry in Alzheimer disease. In this study we assess the effect of the *ApoE* ε4 allele on AD in the Peruvian population.

A total of 78 AD cases and 128 unrelated cognitive healthy controls were included in the study. Genome-wide genotyping was performed using the Illumina Global screening array. Global ancestry and local ancestry analyses were assessed. The effect of the *ApoE ε4* allele on Alzheimer disease was tested using a logistic regression model by adjusting for age, gender, and population substructure (first three principal components). Logistic regression results showed that *ApoE ε4* allele is significantly associated with AD in Peruvian population with the high-risk effect (OR = 5.02, CI: 2.3-12.5, p-value = 2e-4). The average values of the local ancestries surrounding the *ApoE* gene (chr19:44Mb-46Mb) have the highest proportion of Amerindian (60.6%), followed by European (33.9%) and African (5.5%) ancestral backgrounds.

Our results showed that the risk for AD from *ApoE ε4* in Peruvians is higher than we have observed in non-Hispanic White populations. Given the high admixture of Amerindian ancestry in the Peruvian population, it suggests that the Amerindian local ancestry is contributing to a strong risk for AD in *ApoE* ε4 carriers. Our data also support the findings of an interaction between the genetic risk allele *ApoE ε4* and the ancestral backgrounds located around the genomic region of *ApoE* gene.

Background

Alzheimer disease (AD) is a neurodegenerative disease accounting for over 70% of dementia cases in individuals ≥ 70 years of age[1]. AD has a multifactorial etiology, with both genetic and non-genetic risk factors, with liability-scale heritability estimates based on twin studies ranging between 0.58 and 0.79 with over 25 genetic risk factors contributing to AS risk[2,3].

The apolipoprotein E (*ApoE*) gene (19q13.32) is the strongest known genetic risk factor for AD explaining up to 6% of the liablity-scale phenotypic variance[4,5]. *ApoE* codes for a protein that transports cholesterol through the interaction with cell surface receptors[6]. There are three *ApoE* alleles, ε2, ε3 and ε4, defined by two polymorphisms rs429358 and rs7412, that code for three protein isoforms ApoE2 (Cys130, Cys176), ApoE3 *(*Cys130, Arg176) and ApoE4 (Arg130, Arg176)[7].

The association of *ApoE* with AD risk differs between populations and is not clearly established in groups of Amerindian (AI) descent. The strongest association of *ApoE* and AD risk has been observed in East Asian (EA) populations (ε3/ε4 odds ratio OR: 3.1–5.6; ε4/ε4 OR: 11.8–33.1) followed by non-Hispanic White (NHW) populations (ε3/ε4 OR: 3.2; ε4/ε4 OR: 14.9) [8,9]. Its effect is weaker in African-descent and Hispanic populations (ε3/ε4 OR:1.1–2.2; ε4/ε4 OR: 2.2–5.7)[10-14].Genetic studies examining the interaction of genetic ancestry and risk effect of the *ApoE* in Caribbean Hispanic populations (Puerto Rican and Dominican Republic) showed that the effect of the ε4 is correlated with the ancestral background around *ApoE* with the attenuated effect on African-originated haplotypes[15,16].

Peruvian population exhibits ~83% Amerindian ancestral background, higher than other Latin American populations, such as Mexico (50%), Chile (40%), Colombia (28%), Argentina

(28%) and Puerto Rico (16%)[17-19]. Peruvian Native American inhabitants show ancestry of three ancestral groups that originated by the split of an ancient group that migrated down the Americas after diverging from the East Asians and crossing the Bering Strait[20]. By admixing together and with non-Native inhabitants that arrived after Peru's Spanish colonization, these AI groups gave rise to the current Peruvian mestizo population, resulting from admixture with European (EU), Asian and small African (AF) component[18-20]. Distribution of *ApoE* alleles in a sample of Mestizo Peruvian population from northern Lima suggest large contribution of *ApoE* ε3 genotypes, with approximately 9.5 % of cases harboring one or two *ApoE ε4* alleles[21]. No previous published studies have been addresses association of *ApoE* and AD in Peruvian population.

The heterogeneous ancestral make-up of Peruvians provides a unique opportunity to study the effect of global and local Amerindian ancestry on the effect of the ε4 allele over the risk of AD. If the discrepancies seen in the effect of the *ε4* allele across populations is caused by social factors, global, but not local to *ApoE*, ancestry is expected to also be associated to AD risk. Our goal is to use data from the Peruvian population to assess the role of AI genetic ancestry and the *ApoE* gene on AD.

Methods

Study samples and ascertainment

Unrelated cases and controls were ascertained from the Instituto Nacional de Ciencias Neurologicas in Peru as part of a genetics study in AD**.** All cases were assessed by trained neurologists following NINCDS-ADRDA criteria for probable AD [22]. Cognitively intact controls were screened using the Clock drawing test, and the Pfeffer functional activities questionnaire [23,24]. Controls were defined as individuals with no evidence of cognitive problems and age of exam

(AOE) higher than 65 years of age. The dataset contained 79 AD cases (67.0% female, mean age at onset (AAO) = 72.3 years [SD=8.4]) and 128 cognitively healthy controls (59.1 % female, mean AOE = 75.0 years [SD =6.6]). This study was approved by the Ethical Committee of Instituto Nacional de Ciencias Neurologicas of Lima and the IRB of the University of Miami. Miller School of Medicine.

Genotyping and quality control procedures

Genome-wide genotyping have been performed using llumina Global Screening Array. Quality control (QC) analyses were performed using software PLINK v.2[25]. Variants with the call score less than 95%, minor allele frequency less than 0.01, or not in Hardy-Weinberg equilibrium (HWE) ($p<1.e-6$) were eliminated. The concordance between reported sex and genotype-inferred sex was checked using X-chromosome data. The relatedness among the individuals were assessed using "identical by descent" (IBD) allele sharing. *ApoE* genotyping was performed as in Saunders et al.[26]

Assessment of Genetic Ancestry

Global ancestry was evaluated using GENESIS software program that is robust to known and cryptic relatedness[27]. Firstly, the KING-Robust kinship coefficient estimator was used to calculate the KING matrix that includes pairwise relatedness and measures of pairwise ancestry divergence[28]. PC-AiR method was then applied to calculate "preliminary" principal components (PC) by using KING matrix. Default kinship and divergence threshold values have been used. The PC-Relate method that uses "preliminary" PCs to account for the samples ancestry variation and calculate the genetic relationship matrix (GRM) that is robust for the population structure, admixture, and departure from HWE was applied to the data. PC-AiR method was once more applied to the data by using the robust kinship estimates (GRM) and calculated PCs that accurately

capture population structure. PCs were calculated with and without population reference datasets. Four reference populations were used including AI, EU, AF, and EA from Human Genome Diversity Project (HGDP) data for the reference populations[29].

To estimate the admixture proportion, a model-based clustering algorithm was performed as implemented in the ADMIXTURE software[30]. Supervised ADMIXTURE analysis was used at K = 4 by including the same four reference populations from HGDP reference panel we used in PC-AiR approach.

To assess the local ancestry, HGDP reference panel was combined with the Peruvian data using the PLINK v2 software including approximately the same number of individuals from three reference populations EU, AF and AI (~ 100). Then, all individuals in combined dataset were phased using the SHAPEIT tool ver. 2 with default settings and 1000 Genomes Phase 3 reference panel[31,32]. Finally, RFMix was performed using the discriminative modeling approach, to infer the local ancestry at each loci across the genome. We ran RFMix with the PopPhased option and a minimum node size of 5[33].

The heterogenous risk effect of the *ApoE* gene across the populations is suggested to be correlated with the ancestral background local to the *ApoE* gene. Thus, to examine the ancestral background in our dataset we calculated the average ancestry proportions at the *ApoE* by taking the average of the local ancestry estimates around the *ApoE* gene (from 44 Mb to 46 Mb on chromosome 19)[15]. The pipeline to calculate the global and local ancestries was developed by our group using R and Python scripts.

Statistical Analysis

To assess the effect of the *ApoE ε4* allele in Peruvian population we performed logistic regression approach. In this model, the association was tested between the affection status and gene dose of the *ApoE ε4* allele by adjusting for age, gender, and populations substructure (PC1, PC2, and PC3). Statistical analysis was performed using the "GLM2" package available in R computing environment[34].

Results

The supervised ADMIXTURE analysis showed that Peruvians are four-way admixed population with the 63.6% AI, 29.7% EU, 3.8% AF and 2.9% EA ancestral background. Figure 1A shows the box-plot of the average ancestry across the all individuals in the dataset. The ancestral proportion of each individual is illustrated in the bar-plot Figure 1B, where each column reflects the admixture structure of a single individual as the proportion of different colors. Admixture analysis results confirm the recent genetic studies showing a four-way admixture (AI, EU, AF, and EA) structure in Peruvians.

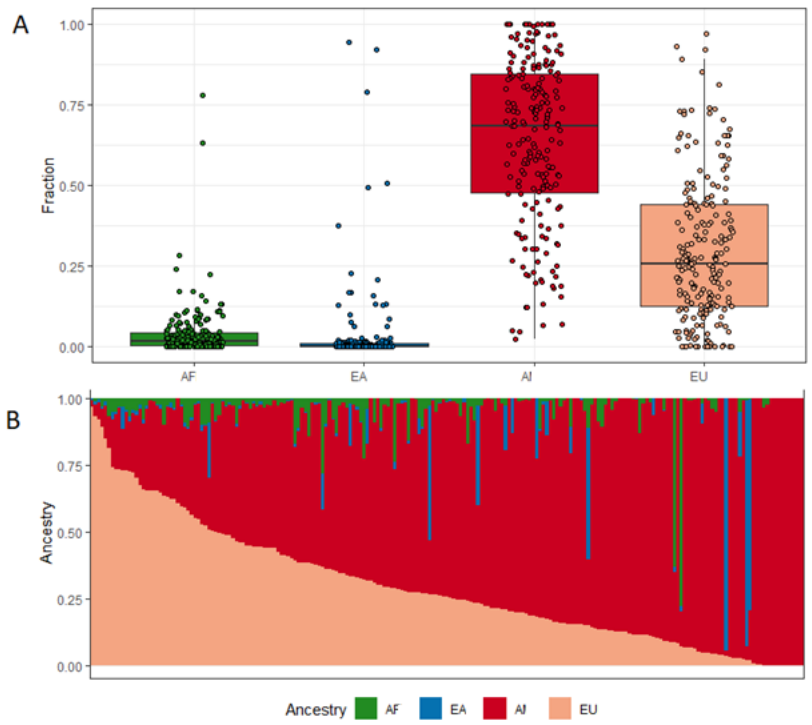

Figure 1 A. The box-plot of the four parental ancestries in Peruvian dataset B. Bar-plot of four-way admixed Peruvian individuals estimated using ADMIXTURE software at K = 4

The allele frequency distribution of the *ApoE* alleles are illustrated in Table 1. The affected individuals have higher frequency of *ApoE ε4* allele (9.2%) than individuals (4.6%) that are cognitively normal. Logistic regression results showed that the *ApoE ε4* allele is significantly associated with AD in Peruvian population with the high-risk effect (OR = 5.02, CI: 2.3-12.5, p-value = 2e-4). The average of the local ancestries around the *ApoE* gene showed that the distribution of the parental ancestries local to the *ApoE* gene is the similar to the average ancestry across the genome with the highest proportion of AI (60.6%), followed by EU (33.9%) and AF (5.5%) ancestral backgrounds.

	Cases (%)	**Controls (%)**
	Genotypes	
ε2ε3	3(1.4)	8(3.9)

ε3ε3	43(20.8)	102(49.3)
ε3ε4	28(13.5)	17(8.2)
ε4ε4	5(2.4)	1(0.5)
	Alleles	
ε2	3(0.7)	8(1.9)
ε3	117(28.3)	229(55.3)
ε4	38(9.2)	19(4.6)
Total	**79**	**128**

Table 1 ApoE genotype and allele frequencies in cases and controls

Discussion

The *ApoE* ε4 allele is the most significant genetic risk factor for late-onset AD with the differences in effect size among the populations. Our results showed that the risk for AD from *ApoE ε4* allele in Peruvians is higher than we have observed in NHW populations. Given the high admixture of AI in the Peruvian population, it suggests that the AI local ancestry is contributing to a strong risk for AD in *ApoE* ε4 carriers. This would align with the current believed migration pattern of AI from East Asia, where *ApoE* ε4 carriers have the highest *ApoE* ε4 risk for AD.

The prevalence of AD varies among the diverse populations. Moreover, AD genetic studies in different ethnic groups have shown variation in both risk effect size and variants (e.g. *ApoE, ABCA7, SORL1, etc*.)[8,35-37]. This heterogeneity suggests that distinct genetic architecture can lead to differing disease susceptibility. Thus, studying the diverse populations is critical to the

understanding of the molecular mechanism underlying the disease pathogenesis and the success of precision medicine. However, diverse populations and especially populations with the AI ancestry are substantially underrepresented in AD genetic studies. The Peruvian population with a large proportion of AI ancestry provides a unique opportunity to assess the role of AI ancestry in AD. This study by confirming the correlation of the genetic ancestry with the risk effect in ε4 allele shows the importance of studying different populations to evaluate the ancestry-specific genetic modifiers correlated with ancestry. Ultimately, studying diverse populations is essential to understand the genetic factors initiating AD pathogenesis that may contribute to health disparities and ultimately the development of effective therapies.

References

1. Alzheimer's Association 2011. Alzheimer's disease facts and figures. Alzheimer's Dementia 7: 20–21
2. Gatz M, Reynolds CA, Fratiglioni L, Johansson B, Mortimer JA, Berg S, Fiske A, Pedersen NL. Role of genes and environments for explaining Alzheimer disease. Arch Gen Psychiatry. 2006;63(2):168-74.
3. Kunkle WB, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. Nature Genetics. 2019; 51:414–430.
4. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science. 1993;261:921–923.
5. Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC Jr, et al. Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. Nat Genet. 1994;7(2):180–4.
6. Holtzman DM, Herz J, and Bu G. Apolipoprotein E and Apolipoprotein E Receptors: Normal Biology and Roles in Alzheimer Disease. Cold Spring Harb Perspect Med. 2012; 2(3): a006312.
7. Aleshkov S, Abraham CR, Zannis VI. 1997. Interaction of nascent apoE2, apoE3, and apoE4 isoforms expressed in mammalian cells with amyloid peptide b(1-40). Relevance to Alzheimer's disease. Biochemistry 36: 10571– 10580.
8. Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: A meta-analysis. JAMA. 1997;278(16):1349–1356.
9. Liu M, Bian C, Zhang J, Wen F. Apolipoprotein E gene polymorphism and Alzheimer's disease in Chinese population: a meta-analysis. Sci Rep. 2014;4:4383.
10. Tang MX, Stern Y, Marder K, Bell K, Gurland B, Lantigua R, et al. The APOE-epsilon4 allele and the risk of Alzheimer disease among African Americans, whites, and Hispanics. JAMA. 1998;279:751–755.

11. Tang MX, Cross P, Andrews H, Jacobs DM, Small S, Bell K, et al. Incidence of AD in African-Americans, Caribbean Hispanics, and Caucasians in northern Manhattan. Neurology. 2001;56(1) 49–56.
12. Tang MX, Maestre G, Tsai WY, Liu XH, Feng L, Chung WY, et al. Relative risk of Alzheimer disease and age-at-onset distributions, based on APOE genotypes among elderly African Americans, Caucasians, and Hispanics in New York City. Am J Hum Genet. 1996;58(3): 574–584.
13. Sahota A, Yang M, Gao S, Hui SL, Baiyewu O, Gureje O, et al. Apolipoprotein E-associated risk for Alzheimer's disease in the African-American population is genotype dependent. Ann Neurol. 1997;42:659–661.
14. Hendrie HC, Murrell J, Baiyewu O, Lane KA, Purnell C, Ogunniyi A, et al. APOE ε4 and the risk for Alzheimer disease and cognitive decline in African Americans and Yoruba. Int Psychogeriatr. 2014;26(9)977–985.
15. Rajabli F, Feliciano BE, Celis K, et al. Ancestral origin of ApoE ε4 Alzheimer disease risk in Puerto Rican and African American populations. PLoS Genet 2018; 14: e1007791.
16. Blue, E. E., Horimoto, A. R., Mukherjee, S., Wijsman, E. M., & Thornton, T. A. Local ancestry at APOE modifies Alzheimer's disease risk in Caribbean Hispanics. Alzheimer's & Dementia. 2019; doi: 10.1016/j.jalz.2019.07.016.
17. Norris ET, Rishishwar L, Wang L, Conley AB, Chande AT, Dabrowski AM, Valderrama-Aguirre A and Jordan IK. Assortative Mating on Ancestry-Variant Traits in Admixed Latin American Populations. Front Genet, 2019; https://doi.org/10.3389/fgene.2019.00359
18. Homburger, J. R. et al. Genomic insights into the ancestry and demographic history of South America. PLoS Genet. 2015; http://dx.doi.org/10.1371/journal.pgen.1005602
19. Norris ET, Wang L, Conley AB, Rishishwar L, Mariño-Ramírez L, Valderrama-Aguirre A, Jordan IK. Genetic ancestry, admixture and health determinants in Latin America. BMC Genomics. 2018; 19(8):75-87.
20. Harris DN, et al., Evolutionary genomic dynamics of Peruvians before, during, and after the Inca Empire. PNAS. 2018; 115(28) E6526-E6535.
21. Marca V, Acosta O, Cornejo-Olivas M, Ortega O, Huerta D, Mazzetti P. Genetic polymorphism of apolipoprotein E in a Peruvian population. Rev Peru Med Exp Salud Publica. 2011; 28(4):589-94.

22. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology. 1984;34:939–44.
23. Manos PJ, Wu R. The ten point clock test: a quick screen and grading method for cognitive impairment in medical and surgical patients. Int J Psychiatry Med. 1994;24:229–244.
24. Pfeffer R. I., Kurosaki T. T., Harrah C. H., Jr., Chance J. M., Filos S. (1982). Measurement of functional activities in older adults in the community. J. Gerontol.37, 323–329 10.1093/geronj/37.3.323
25. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet.2007; 81(3):559-75.
26. Saunders AM, Hulette C, Welsh-Bohmer KA, Schmechel DE, Crain B, Burke JR, et al. Specificity, sensitivity, and predictive value of apolipoprotein-E genotyping for sporadic Alzheimer's disease. Lancet.1996;348(9020):90–93. pmid:8676723
27. Conomos MP, Gogarten SM, Brown L, Chen H, Rice K, Sofer T, Thornton T, Yu C (2019). GENESIS: GENetic EStimation and Inference in Structured samples (GENESIS): Statistical methods for analyzing genetic data from samples with population structure and/or relatedness. R package version 2.14.3, https://github.com/UW-GAC/GENESIS.
28. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM (2010) Robust relationship inference in genome-wide association studies. Bioinformatics 26(22):2867-2873
29. Cavalli-Sforza LL (2007) Human evolution and its relevance for genetic epidemiology. Annu Rev Genomics Hum Genet 8: 1–15
30. Alexander DH, Novembre J, and Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Research, 19:1655–1664, 2009.
31. Delaneau O, Marchini J. The 1000 Genomes Project Consortium. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. Nature Communications.2014;5:3934. pmid:25653097.
32. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature.2015;526: 68–74 pmid:26432245.

33. Maples BK, Gravel S, Kenny EE, Bustamante CD. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am. J. Hum. Genet.2013;93:278–288 pmid:23910464
34. Marschner I. glm2: Fitting Generalized Linear Models. R package version 1.1.2.2014;https://CRAN.R-project.org/package=glm2
35. Reitz C, Jun G, Naj A, Vardarajan BN, Wang LS, Valladares O, et al. Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E 4, and the risk of late-onset Alzheimer disease in African Americans. JAMA.
36. Cukier HN, Kunkle BW, Vardarajan BN, Rolati S, Hamilton-Nelson KL, Kohli MA, et al. ABCA7 frameshift deletion with Alzheimer disease in African Americans. Neurol Genet.2016;2(3):e79.
37. Liu F, et al. A study of the SORL1 gene in Alzheimer's disease and cognitive function. J Alzheimers Dis. 2009;18(1):51-64.