

Beyond ‘Hispanic/Latino’: Self-Reported Race, Ethnicity, and Genetic Ancestry Variation in Latin American Cohorts of Participants with Neurodegeneration

Stefanie Danielle Piña Escudero^{1,2}, Jared Taylor³, Vianna Mashayekhi⁴, Marcelo Maito⁵, Guido Rocatti⁵, Maira Okada de Oliveira^{6,15}, Maria Eugenia Godoy^{2,5}, Shireen Javandel^{1,2}, José Alberto Ávila Funes⁷, María Isabel Behrens⁸, Martin Bruno⁹, Nilton Custodio¹⁰, David Aguillon¹¹, Diana Matallana Eslava¹², Andrea Slachevsky⁸, Leonel Tadao Takada⁶, Lina Maria Zapata-Restrepo¹³, Juliana Acosta-Uribe^{11, 14}, Mike A. Nalls Nalls¹⁵, J. Nicholas Cochran³, Agustín Ibáñez², Jennifer Yokoyama^{1,2}

¹Fein Memory and Aging Center, University of California San Francisco, San Francisco, CA, USA, ²Global Brain Health Institute, University of California San Francisco, San Francisco, USA and Trinity College Dublin, Ireland, ³HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA, ⁴University Highschool, San Francisco, CA, USA, ⁵Universidad de San Andrés, Buenos Aires, Argentina, ⁶Hospital Das Clinicas, University of Sao Paulo Medical School, São Paulo, Brazil, ⁷Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubiran, Tlalpan, Mexico, ⁸Universidad de Chile, Santiago, Chile, ⁹Universidad Católica de Cuyo, San Juan, Argentina, ¹⁰Unidad de Investigación y Docencia, Instituto Peruano de Neurociencias, Lima, Peru, ¹¹Universidad de Antioquia, Medellin, Colombia, ¹²Universidad Javeriana, Bogota, Colombia, ¹³Universidad del Valle de Llíli, Cali, Colombia, ¹⁴University of California, Santa Barbara, CA, USA, ¹⁵Data Technica, Maryland, MD, USA,

Funding Sources: UCSF Partnership Grand Program, National Institutes of Health, the National Institute on Aging (NIA) (R01 AG057234, R01 AG075775, R01 AG21051, R01 AG083799, CARDS-NIH 75N95022C00031), the Alzheimer's Association (SG-20-725707), the Rainwater Charitable Foundation – The Bluefield Project to Cure FTD, and the Global Brain Health Institute.

#6620

Applying only U.S. Census race/ethnicity categories to Latin American populations can obscure important variation in genetic ancestry. Accounting for both self-reported identity and genetic ancestry may improve the precision of neurodegeneration research across diverse Latin American populations.

Introduction

People living in Latin America and the Caribbean are often grouped into a single "Hispanic/Latino" race/ethnicity category in research and public health surveys. However, this classification may obscure substantial variation in genetic ancestry, which can influence disease risk and other biomedical outcomes. We examined global ancestry proportions across self-reported US Census race/ethnicity categories among participants from six Latin American countries (Mexico, Colombia, Peru, Chile, Argentina, and Brazil) enrolled in the ReDLat study. Participants included healthy controls (HC) and individuals diagnosed with Alzheimer's disease (AD), frontotemporal dementia (FTD), or other neurodegenerative diseases. We also analyzed the ancestry composition associated with local race/ethnicity self-identifications used in each country.

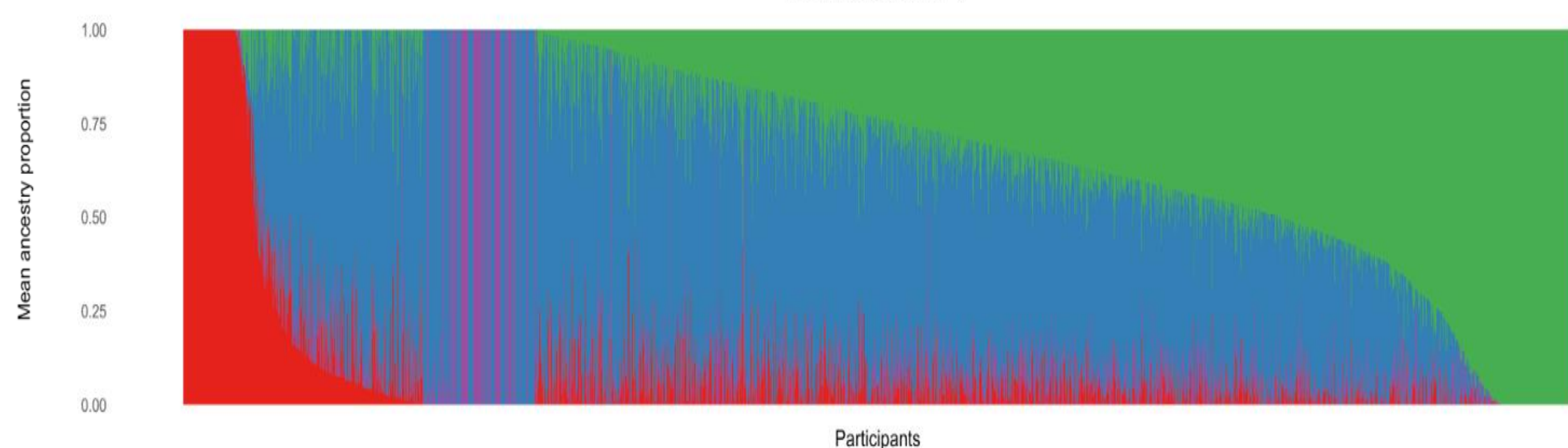
Methods

Blood samples were collected from all participants. DNA extraction and genotyping were performed at the HudsonAlpha Institute for Biotechnology, and genotype quality control was conducted at the University of California, Santa Barbara. Participants self-reported the US Census race/ethnicity category with which they most closely identified and also completed open-ended race and ethnicity fields to describe their identity using their own terms. Global genetic ancestry proportions were estimated using the 1000 Genomes Project reference populations. Descriptive statistics were used to characterize the study population. Linear regression models were fitted separately for each global ancestry component, using participants who self-reported as Hispanic/Latino as the reference group. Differences in ancestry proportions among participants who self-identified with other US Census race/ethnicity categories were estimated relative to the Hispanic/Latino reference group. Statistical significance was assessed using a two-sided significance threshold of $p < 0.05$.

Results

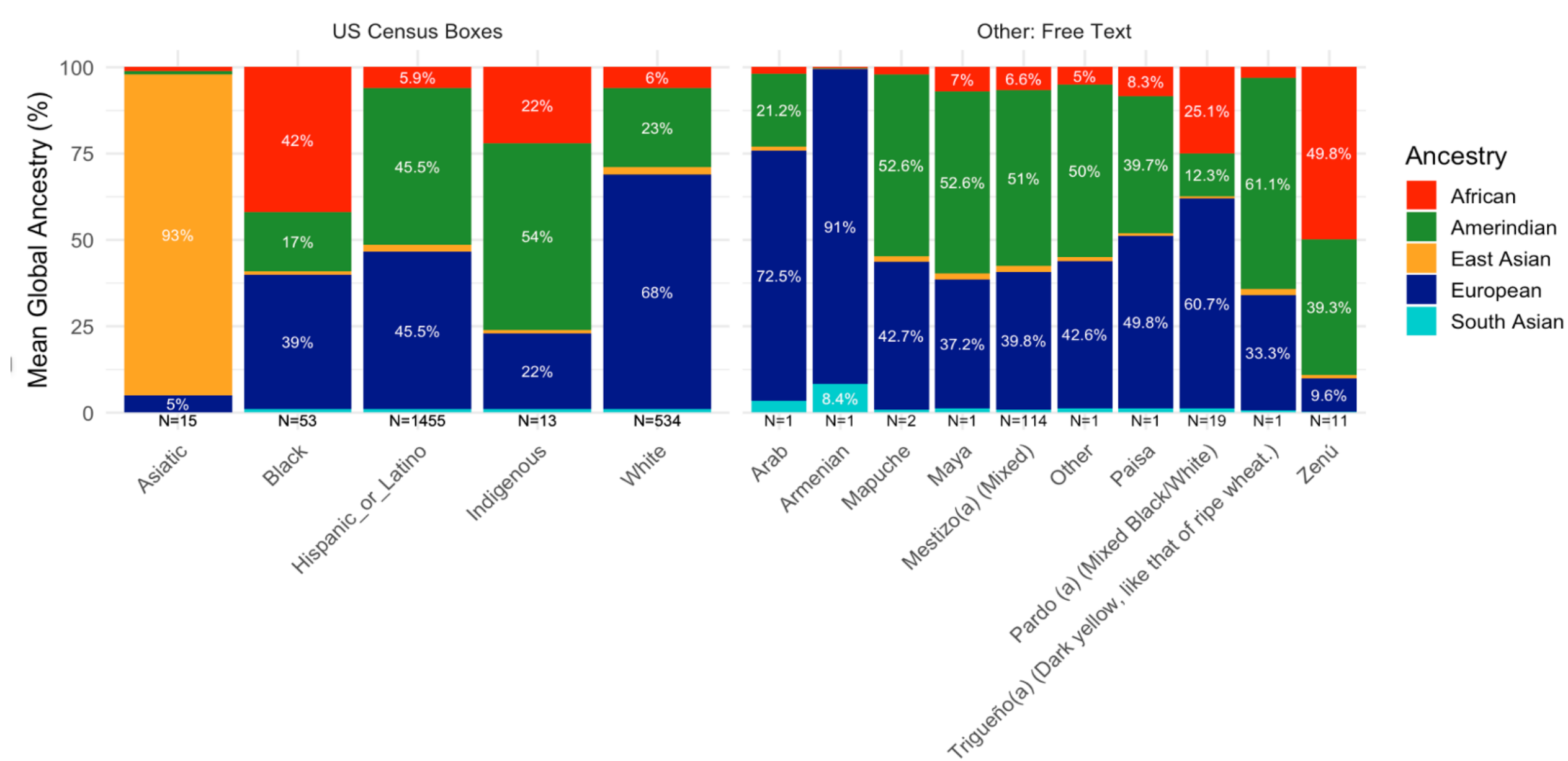
2,379 individuals were included in the study. The mean age was 64.6 years (SD, 11.3), 66.0% were women, 52.2% were healthy controls, 35.2% had AD, 11.3% had FTD, and 1.3% had other neurodegenerative diseases. 54 (2.3%) self-reported more than one US Census race/ethnicity category and were excluded from the analyses. The overall cohort had the following mean global ancestry proportions:

Individual ancestry proportions in the ReDLat cohort
ADMIXTURE K=4

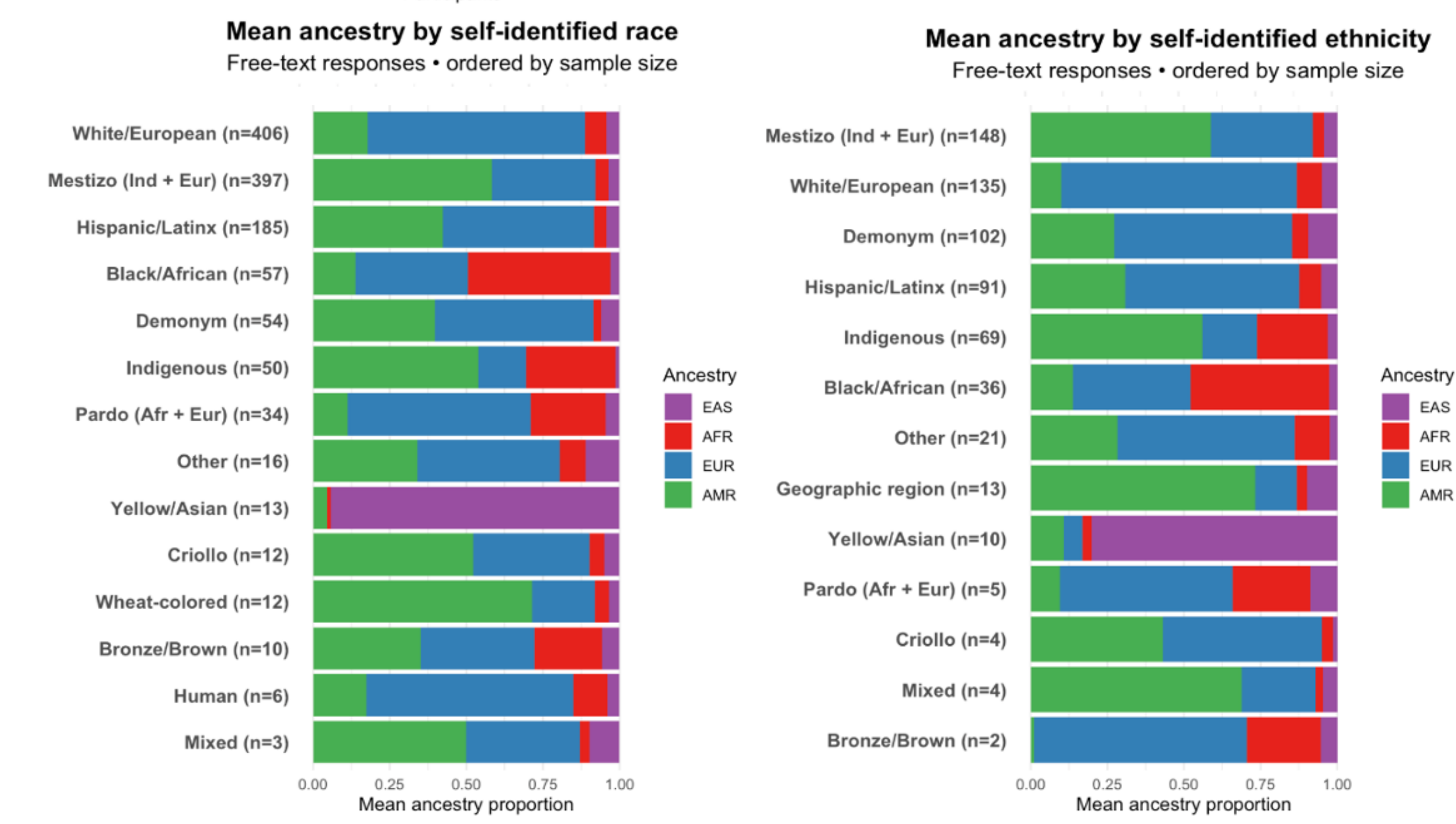


Results

Among the 2,222 participants who, under standard research practice, would have been classified as "Hispanic/Latino" based solely on their country of origin, only 65% self-identified as Hispanic/Latino when asked to report their race/ethnicity using the US Census categories. Participants who self-identified as Hispanic/Latino exhibited a highly admixed genetic ancestry profile, with mean ancestry proportions of 6% African, 46% Native American, 46% European, 2% East Asian, and 1% South Asian. Compared with participants who self-identified as Hispanic/Latino, those who selected other US Census race/ethnicity categories despite also originating from Latin America exhibited distinct ancestry profiles:



Participants were also asked to describe their race and ethnicity using open-ended response fields rather than predetermined categories. Responses were highly diverse, reflecting substantial variation in how race and ethnicity are conceptualized across Latin America.



Conclusions

Classifying Latin American populations into a single "Hispanic/Latino" category obscures substantial heterogeneity in both self-identified race/ethnicity and genetic ancestry. Although self-reported US Census race/ethnicity categories captured meaningful differences in ancestry, the relationship between identity and genetic ancestry was not uniform, highlighting the complexity of these constructs in admixed populations. The diversity observed in open-ended race and ethnicity responses further illustrates that these concepts are culturally contextual and not directly interchangeable across regions. Together, these findings suggest that combining genetic ancestry with locally relevant measures of self-identification may improve the characterization of Latin American populations in dementia research.