HGG Advances

ARTICLE

Admixture mapping implicates 13q33.3 as ancestry-of-origin locus for Alzheimer disease in Hispanic and Latino populations

Andrea R.V.R. Horimoto,[1,20,*] Lisa A. Boyken,[1,21] Elizabeth E. Blue,[2,3] Kelsey E. Grinde,[1,4] Rafael A. Nafikov,[1,2,22] Harkirat K. Sohi,[2,5] Alejandro Q. Nato, Jr.,[2,6] Joshua C. Bis,[7] Luis I. Brusco,[8] Laura Morelli,[9] Alfredo Ramirez,[10,11,12,13,14,15] Maria Carolina Dalmasso,[10,16] Seth Temple,[17] Claudia Satizabal,[15,18,19] Sharon R. Browning,[1] Sudha Seshadri,[19] Ellen M. Wijsman,[1,2,23,*] and Timothy A. Thornton[1,17]

Summary

Alzheimer disease (AD) is the most common form of senile dementia, with high incidence late in life in many populations including Caribbean Hispanic (CH) populations. Such admixed populations, descended from more than one ancestral population, can present challenges for genetic studies, including limited sample sizes and unique analytical constraints. Therefore, CH populations and other admixed populations have not been well represented in studies of AD, and much of the genetic variation contributing to AD risk in these populations remains unknown. Here, we conduct genome-wide analysis of AD in multiplex CH families from the Alzheimer Disease Sequencing Project (ADSP). We developed, validated, and applied an implementation of a logistic mixed model for admixture mapping with binary traits that leverages genetic ancestry to identify ancestry-of-origin loci contributing to AD. We identified three loci on chromosome 13q33.3 associated with reduced risk of AD, where associations were driven by Native American (NAM) ancestry. This AD admixture mapping signal spans the *FAM155A*, *ABHD13*, *TNFSF13B*, *LIG4,* and *MYO16* genes and was supported by evidence for association in an independent sample from the Alzheimer's Genetics in Argentina—Alzheimer Argentina consortium (AGA-ALZAR) study with considerable NAM ancestry. We also provide evidence of NAM haplotypes and key variants within 13q33.3 that segregate with AD in the ADSP whole-genome sequencing data. Interestingly, the widely used genome-wide association study approach failed to identify associations in this region. Our findings underscore the potential of leveraging genetic ancestry diversity in recently admixed populations to improve genetic mapping, in this case for AD-relevant loci.

Introduction

Alzheimer disease (AD) is a neurodegenerative disorder that progress slowly from mild cognitive impairment to severe dementia. It is the most common form of dementia and the sixth leading cause of death in the United States, affecting 5.8 million Americans age ≥ 65.[1] However, the estimated risk of AD can vary in different ethnic groups, particularly if these groups have diverse genetic ancestries; for example, Caribbean Hispanic (CH) individuals have higher prevalence and incidence rates of AD and other dementias than non-Hispanic White individuals.[2,3] Despite increased AD risk and significant economic impacts of AD and need for long-term care,[1] CH individuals remain underrepresented in medical genetic research.[4–6] While recent research is beginning to uncover the genetic factors affecting AD among Hispanic populations,[7–11] more studies are necessary to better understand the pathogenesis of AD in Hispanic and other admixed populations.

[1]Department of Biostatistics, University of Washington, Seattle, WA 98195, USA; [2]Division of Medical Genetics/Department of Medicine, University of Washington, Seattle, WA 98195, USA; [3]Brotman Baty Institute for Precision Medicine, Seattle, WA 98195, USA; [4]Department of Mathematics, Statistics and Computer Science, Macalester College, Saint Paul, MN 55105, USA; [5]Biomedical and Health Informatics Program, University of Washington, Seattle, WA 98195, USA; [6]Department of Biomedical Sciences, Joan C. Edwards School of Medicine, Marshall University, Huntington, WV 25755, USA; [7]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA 98101, USA; [8]CENECON - Center of Behavioural Neurology and Neuropsychiatry, School of Medicine, University of Buenos Aires, C1121A6B Buenos Aires, Argentina; [9]Laboratory of Brain Aging and Neurodegeneration-Fundación Instituto Leloir-IIBBA- National Scientific and Technical Research Council (CONICET), C1405BWE Ciudad Autónoma de Buenos Aires, Argentina; [10]Division of Neurogenetics and Molecular Psychiatry, Department of Psychiatry and Psychotherapy, University of Cologne, Medical Faculty, 50937 Cologne, Germany; [11]Department of Neurodegeneration and Gerontopsychiatry, University of Bonn, 53127 Bonn, Germany; [12]German Center for Neurodegenerative Diseases (DZNE), 53127 Bonn, Germany; [13]Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD) University of Cologne, 50674 Cologne, Germany; [14]Department of Psychiatry, UT Health San Antonio, San Antonio, TX 78229, USA; [15]Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, UT Health San Antonio, San Antonio, TX 78229, USA; [16]Neurosciences and Complex Systems Unit (EnyS), CONICET, Hospital El Cruce, National University A. Jauretche (UNAJ), B1888AAE Florencio Varela, Argentina; [17]Department of Statistics, University of Washington, Seattle, WA 98195, USA; [18]Department of Population Health Sciences, University of Texas, San Antonio, TX 78229, USA; [19]Department of Neurology, University of Texas, San Antonio, TX 78229, USA
[20]Present address: Division of Aging, Brigham and Women's Hospital, Boston, MA, USA
[21]Present address: Zymeworks Biopharmaceuticals, Inc., Seattle, WA 98101, USA
[22]Present address: La Jolla Labs Inc., La Jolla, CA 92037, USA
[23]Lead contact
*Correspondence: andrearh@uw.edu (A.R.V.R.H.), wijsman@uw.edu (E.M.W.)
https://doi.org/10.1016/j.xhgg.2023.100207.

Admixture mapping is a powerful approach for complex trait mapping in multi-ethnic, recently admixed populations. The rationale behind admixture mapping is that causative genetic variants may be more frequent on haplotypes derived from parental populations with a higher incidence of the disease, and leveraging genetic ancestry diversity of the admixed individuals can improve the discovery of genomic regions harboring variants associated positively or negatively with the disease. In this way, genome-wide admixture and association testing can yield complementary as well as differing results.[12–14] Although the concept has existed for >20 years,[15] applications of admixture mapping were limited in the early years because of insufficient number of genetic markers that both blanketed the genome and had substantially different allele frequencies across ancestral populations. The availability of dense SNPs from genotyping arrays and sequencing data has therefore relatively recently facilitated the identification of segments originating from particular ancestral populations genome-wide.

Admixture mapping faces similar statistical analysis issues as genome-wide association studies (GWASs). One common issue of concern is spurious association caused by population structure and relatedness in the sample if not properly accounted for. Linear mixed models (LMMs) have become a method of choice for genetic association testing of continuous outcomes because they have been shown to provide valid associations in the presence of sample structure.[16,17] Although LMMs have also been used to analyze categorical outcomes, simply treated as continuous, the assumption of homoscedasticity can be violated in the presence of population structure and covariates, incurring an inflated type I error. The generalized LMM association test (GMMAT) method[18] uses a logistic model framework for valid genome-wide association testing of dichotomous outcomes in the presence of sample structure.

Although LMMs have been previously used in admixture mapping studies to accommodate population structure and relatedness in the sample,[19–22] to our knowledge there have not been any methods proposed that implement a logistic mixed model for admixture mapping with proper analysis of binary traits that also allow for relatedness among sampled individuals. Here, we introduce the general framework of a linear or logistic admixture mapping analysis (LLAMA) for samples with relatedness and population structure. The approach is an extension of the GMMAT logistic-mixed-model framework that provides valid admixture mapping of outcomes (dichotomous or continuous) in samples derived from an arbitrary number of ancestral populations. We apply our LLAMA approach to data from CH families from the Alzheimer Disease Sequencing Project (ADSP), where we identify a region with evidence of association between late-onset AD and ancestry of origin on chromosome 13q33.3. This region is supported by association results in an independent sample from the Alzheimer's Genetics in Argentina - Alzheimer Argentina consortium (AGA-ALZAR) with more Native American (NAM) ancestry than the ADSP sample, followed by haplotype reconstructed using whole genome sequencing (WGS) variants in the ADSP sample that further narrows the variants of greatest interest.

Material and methods

Samples and phenotypes

The ADSP CH data (ADSP Hispanic; dbGaP: phs000572.v7.p4) comprised 545 subjects of the ADSP discovery set, distributed among 68 families. The genotype data were obtained using Illumina HumanOmniExpress12.v1.18, Human650Y.v2 and Human Omni1-Quad.v1.0.H SNP arrays and combined using PLINK,[23] keeping markers with a genotyping rate of >90% across panels. A subset of 526 individuals had *APOE* genotypes. Details about the study design and sample selection are described elsewhere[24]; sequencing and quality control analyses are also described elsewhere.[25] A total of 541 subjects had available phenotype data (Table 1). For analysis purposes, individuals were considered affected if their AD status was coded as possible or probable AD. Definitions of AD are detailed elsewhere.[26] There were no individuals with autopsy data to provide the information needed to provide a diagnosis of definite AD. Individuals with AD reported by family or who showed a mixed diagnosis were treated as unknown status. Age was the age at onset for affected and age at last examination for unaffected subjects and was available for all subjects with AD status. The study was carried out under Human Subjects Approval number STUDY00001230 from the University of Washington.

We used SNP genotype data from publicly available population reference samples for our principal component and local ancestry analyses. We selected 165 European (EUR) (CEU, Utah residents) and 203 African (AFR) (YRI, Yoruba) samples from HapMap phase 3 (https://ftp.ncbi.nlm.nih.gov/hapmap/)[27] and 63 NAM (Colombian, Pima, Maya, Karitiana, Surui) samples from the Human Genome Diversity Project (HGDP: https://hagsc.org/hgdp/files.html).[28,29] We additionally used 1000 Genomes phase 3 reference samples (https://www.internationalgenome.org/data)[30] for phasing and to obtain population-specific allele frequencies of variants within NAM haplotypes.

Two independent samples were used for the validation study (Table 1). The Columbia University Study of Caribbean Hispanics with Familial and Sporadic Late Onset Alzheimer Disease dataset (CU Hispanic; dbGaP: phs000496.v1.p1)[10,31] had 3,056 subjects with phenotype data and SNP genotyping performed on the Illumina HumanOmni1-Quad_v1-0_B SNP array. The AGA-ALZAR samples were recruited in the Buenos Aires, Jujuy, Córdoba, and Mendoza provinces. The study (protocol CBFIL#22) was approved by the ethics committee (HHS IRB#00007572, IORG#006295, FWA00020769), and all participants and/or family members gave their informed consent.[32] The diagnosis of AD followed diagnostic criteria from the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer disease and Related Disorders Association.[33] The samples were genotyped using the Illumina Infinium Global Screening Array (GSA) v.1.0 combined to a GSA shared custom content. Only GWAS summary statistics were available to conduct the validation study in the AGA-ALZAR samples.

We used WGS data from the ADSP for haplotype analysis of sequence variation in the region of interest on chromosome 13. The samples used for admixture mapping were sequenced in the early ADSP discovery-extension sequencing, but then called again

Table 1. Descriptive characteristics of the subjects across datasets

Trait	ADSP Hispanic (n = 541)	CU Hispanic (n = 3056)	AGA-ALZAR (n = 962)
Males (%)	42.1	33.6	34.7
Age[a] (years)	71.9 ± 10.0	73.9 ± 9.2	74.6 ± 7.0
ε2[b] (%)	3.8	5.6	4.2
ε3[b] (%)	78.9	73.8	77.8
ε4[b] (%)	17.3	20.6	18.0
Unaffected			
Number	173	1659	490
Age[a] (years)	67.7 ± 9.7	73.0 ± 8.9	73.4 ± 7.3
Affected			
Number	368	1397	388
Age[a] (years)	73.8 ± 9.6	74.9 ± 9.4	76.4 ± 6.4
Ancestry proportions[c]			
AFR	0.27 ± 0.14 (0.22–0.89)	0.33 ± 0.19 (0.00–0.99)	0.05 ± 0.03 (0.00–0.20)
EUR	0.64 ± 0.13 (0.06–0.76)	0.58 ± 0.17 (0.01–0.99)	0.71 ± 0.26 (0.00–1.00)
NAM	0.09 ± 0.03 (0.01–0.27)	0.09 ± 0.08 (0.00–0.96)	0.25 ± 0.27 (0.00–0.99)

[a]Age: mean ± standard deviation.
[b]Frequency of *APOE* ε2/ε3/ε4 alleles.
[c]Ancestry proportions: mean ± standard deviation (minimum – maximum).

together with additional samples (a total of 16,905 subjects), to Human Genome sequence build GRCh38. For the analysis, here, we used data from both accession number NIAGADS: fsa000003 (release NG00067.v2 – the 5K sample), which provided the original WGS variant calls for 4,788 subjects, and NNIAGADS: fsa000006 (release NG0067v7 – the 17K sample) on 16,905 subjects, which included the earlier 5K sample. Both are available from NIAGADS, and the NG00067.v2 release is also available from dbGaP. For consistency with the other analyses performed on the earlier data release here, we focused haplotype analyses only on the same subjects. The later, larger release was used only to provide the larger sample of reference sequences needed for accurate phasing the WGS data within the single jointly called and quality controlled sample; the use of external WGS is contraindicated in this situation because of strong batch effects introduced by different sequence-calling pipelines. To maintain consistency of presentation for purposes of this paper, all variant positions are provided as original GRCh37 sequence build coordinates.

Local ancestry inference

HapMap phase 3 and HGDP population samples, described above, were used here as reference samples. We merged reference datasets with PLINK,[23] keeping 603,611 SNPs with an overall genotyping rate of 0.998. We then merged the reference and ADSP Hispanic datasets, leaving 273,523 common autosomal SNPs with a genotype missing rate of <7% (overall genotyping rate of 0.996). For CU Hispanic analysis, we updated the physical positions of the reference data to build NCBI37/hg19 using LiftOver[34] to match the CU Hispanic data, and randomly removed CEU and YRI samples to keep the reference populations balanced. Merging of the reference and CU Hispanic datasets left 294,252 autosomal markers with a genotype missing rate of <5% and a total genotyping rate of 0.993. We phased reference and inference samples jointly. The phasing of the ADSP Hispanics was performed using Beagle version 3.3.2,[35] while the CU Hispanics were phased using Shapeit version 2[36] with 1000 Genomes Phase 3 samples[30] as a reference. Both programs impute sporadic missing genotypes in the dataset during the phasing. We obtained high-confidence local ancestry calls with RFMix version 1.5.4.[37] Proportions of AFR, EUR, and NAM global ancestries for ADSP Hispanics and CU Hispanics were estimated by averaging the local ancestry estimates across all chromosomes.

Principal components and genetic relatedness matrix

Principal components (PCs) and the genetic relatedness matrix (GRM) were estimated in a recursive manner using the PC-AiR[38] and PC-Relate[39] methods implemented in the GENESIS R package.[40] PC-AiR uses measures of ancestry divergence estimated using the KING-Robust algorithm[41] to partition samples into related and unrelated ancestry representative sets. The population reference dataset used for local ancestry inference was included here to improve inference of the population structure. Standard PCs were estimated on the unrelated set and PC values were projected for the related samples. PC-Relate then uses these ancestry representative PCs to estimate pairwise kinship coefficients adjusted for population structure. A second round of PC-AiR and PC-Relate analyses were performed starting with the GRM obtained in the previous step to obtain PCs robust to relatedness and an ancestry adjusted GRM.

Admixture mapping

We extended the GMMAT logistic mixed model framework for association testing to admixture mapping. Our admixture mapping logistic mixed model[22] is implemented in LLAMA, available in the GENESIS R package.[40] We performed a joint admixture mapping test, in which all AFR, EUR, and NAM ancestries are tested simultaneously,[42] using genotype data from the ADSP and CU Hispanic

samples. To test the association between ancestry at each locus, j, and AD, we first fit the model under the null hypothesis of no SNP effect, using the top four PCs as fixed effect covariates and the GRM as random effect. GENESIS uses a penalized quasi-likelihood approximation to the generalized LMM, implemented as the GMMAT.[18] The full admixture mapping logistic mixed model is described by:

$$logit(\pi) = X\alpha + A_j\beta_j + g,$$

where $\pi = P(y = 1|X, A_j, g)$ represents the $Nx1$ column vector of probabilities of being affected for the N individuals conditional on covariate, local ancestry calls, and random effects; X is the vector of covariates; and α is the vector of fixed covariate effects including an intercept. We assume that $g \sim N(0, \sigma_a^2\Phi)$ is a vector $g = (g_1, \ldots, g_N)$ of random effects for the N subjects, where σ_a^2 is the additive genetic variance and Φ is the GRM. Letting the third ancestral population be the reference population, A_j represents a $Nx(K-1)$ matrix of the local ancestry dosages at the locus j for the $K-1$ parental populations, with the corresponding effect size vector β_j of length $K-1$. The null hypothesis $\beta_j = 0$ was assessed via a multivariate score test. Single-ancestry admixture mapping analyses were also performed to identify which population was driving the signal. The significance threshold ($p < 4.5 \times 10^{-5}$) was obtained by simulating a null phenotype,[42] and accounts for the number of independent tests performed under historical population structure and time depth of this population.

Evaluation of the admixture mapping logistic approach

The performance of our admixture mapping logistic approach was evaluated by (a) conducting a simulation study, and (b) comparing how much better it fits to the data in relation to an admixture mapping LMM when analyzing a binary trait.

Simulation study

We conducted a simulation study using simulated binary phenotypes and real genetic data. This framework permits admixture mapping signals that are more realistic than those observed using simulated genotypes. Binary phenotypes for 545 individuals were simulated in R[43] using a logistic approach, considering a similar complex correlation structure as that observed in the ADSP data. We selected a subset of 1,027 causal SNPs from a set of 5,134 SNPs in linkage equilibrium across the genome (selecting one out of each set of five SNPs to guarantee the correct representation for each chromosome). Risk was assigned according to the number of EUR alleles, considering increasing effect sizes for the causal SNPs from 0.5 to 1.5 in increments of 0.25. In total, we performed 5,135 simulations. The vector of probabilities of being affected for the N individuals was estimated from:

$$logit(\pi) = \beta_0 + A_j\beta_j + g,$$

where β_0 was set to -2, representing a probability of approximately 1% of being affected under no other predictor or random effects, β_j is the effect size of the causal locus j, A_j is the vector of EUR allelic dosages for the N individuals at locus j, and $g \sim N(0, \sigma_a^2\Phi)$ is the vector of random effects estimated from a multivariate normal distribution, with σ_a^2 set to 5 and a semi-definite positive kinship matrix based on the real ADSP kinship matrix. The affected and unaffected AD status was randomly set based on the uniform distribution.

Genome-wide admixture mapping was conducted on real genotype data, adjusting the logistic mixed model for population structure and dependence effects, as described above. Power and type I error were assessed as performance measures of our method. Power was recorded as the proportion of simulations that capture a significant p value for the causal SNP at different effect sizes and significance levels. Type I error of the admixture mapping logistic approach was assessed as the mean proportion of false-positive results across the genome at different significance levels.

Comparing admixture mapping logistic and LMMs

The full admixture mapping LMM is described by:

$$y = X\alpha + A_j\beta_j + g + \varepsilon,$$

where y is the vector of binary outcomes for the N individuals, X is the vector of covariates, α is the vector of fixed covariate effects including an intercept, and A_j represents a $Nx(K-1)$ matrix of the local ancestry dosages at locus j for the $K-1$ parental populations, with the corresponding effect size vector β_j of length $K-1$. We assume that $g \sim N(0, \sigma_a^2\Phi)$ is a vector $g = (g_1, \ldots, g_N)$ of random effects for the N subjects, where σ_a^2 is the additive genetic variance and Φ is the GRM, and $\varepsilon \sim N(0, \sigma_e^2 I)$ is a vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)$ of residual effects, where σ_e^2 represents the residual variance and I is an identity matrix. GENESIS uses the average information restricted maximum likelihood procedure to estimate the variance components of the random effects under the null model. Generalized least squares are used to fit this LMM to test the null hypothesis $H_0 : \beta_j = 0$. Admixture mapping for AD using the LMM was conducted using the steps described above for the logistic mixed model, including the same adjustments for the fixed and random effects.

Haplotype analysis

We investigated the existence of potential variants of interest in the ADSP Hispanic subjects with NAM ancestry via haplotype analysis, using the ADSP WGS data. This was the only analysis performed on genotypes derived from WGS; all other analyses used SNP array genotype data. The approach was predicated on the assumption that a previously unknown admixture mapping signal that is driven by NAM local ancestry implies a low-frequency or rare variant that is more common on NAM chromosomes than on chromosomes of EUR or AFR ancestry. This context also suggests the possibility of a unique haplotypic context of such variants. For these analyses, therefore, the initial exploratory analysis limited attention to a subset of genotypes extracted from the post-quality control discovery phase vcf files.[25] This was followed by a deeper analysis of haplotypes that were further informed by the larger approximately 17K data release. Both parts of the analysis made use of the same set of individual admixture proportions estimated for SNPs at the associated GWAS-marker loci at 13q33.3 for the group of 356 subjects noted previously. For these purposes, each subject was classified into one of six groups, defined by being homozygous or heterozygous for EUR, AFR, or NAM ancestral chromosome(s).

Exploratory analyses consisted of ancestry-based variant prioritization of 2363 polymorphic variants in the three genomic loci defined by admixture mapping. Allele frequencies were computed by direct counting from the initial data release of this 356-member sample for each ancestral chromosome group. We prioritized variants for which alleles were uniquely present or absent in individuals with NAM ancestry, or were close to this ideal (i.e., the minor allele existed in at most two families out of all subjects who did not have estimated NAM ancestry at that variant position). We used the absence of an alternative allele in all individuals from the ADSP non-Hispanic White families from fsa000003 as further evidence to justify the selection of variants. The existence of similar allele frequencies among variants identified through this approach was used as an initial prediction of the presence of a shared

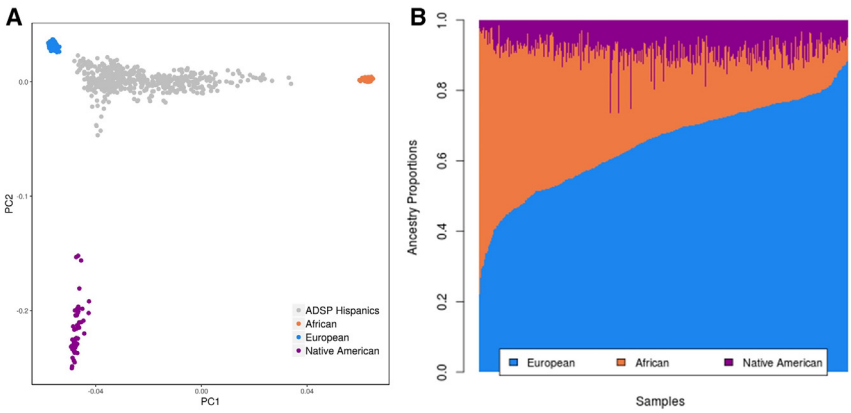

Figure 1. Population structure and admixture proportions of ADSP Hispanic samples
(A) Population structure captured by the two first PCs (PC1 and PC2); AFR, EUR, and NAM populations are represented by YRI and CEU samples from HapMap phase 3, and NAM samples from HGDP, respectively.
(B) Vertical bars represent the ancestry proportions of each individual, arranged in order of increasing estimated EUR ancestry.

haplotype among NAM carriers. Finally, for the variants that were prioritized through these procedures, we additionally referred to and report allele frequencies in reference samples (1000 Genomes phase 3 and HGDP) containing independent representative major geographic populations, as well as other populations with NAM ancestry. For further variant prioritization, we considered a combined annotation dependent depletion (CADD) scaled score[44] cutoff of >10 as suggestive of potential deleteriousness.

Variants suggested by this initial analysis were further evaluated by more formal haplotype analyses. We used the same general approach for initial steps here as we used for the initial local ancestry estimation: we used BEAGLE (v. 5.4)[35] to perform population-based phasing of the WGS data from the full 17K multi-ethnic sample to provide reference genomes. We used the following variant inclusion filters: minor allele count >2, missingness <5%, Phred score >40, FILTER = PASS, and read depth >10. We did not use an HWE filter because of the existence of related individuals in the multi-ethnic sample. After the phasing, we tabulated the haplotypes from our most highly prioritized region with reference to the inferred haplotypes within families for the whole 356-member sample, together with the comparable ancestry labels.

Linkage disequilibrium plots were prepared using LDmatrix, a tool of the LDlink suite (https://ldlink.nci.nih.gov). All 1000 Genomes AFR, EUR, and NAM (admixed American) population samples were used as reference.

GWAS

The association analyses were performed on the ADSP Hispanic using a set of 504,628 autosomal SNPs, which passed on filters for a minor-allele frequency of >0.01 and a genotype missing rate of <5%. For comparison, we fit a logistic mixed model to the data applying the same fixed and random effects of the admixture mapping logistic mixed model. The analyses were conducted using the GENESIS R package.[40] The full statistical model is described by:

$$logit(\pi) = X\alpha + G_j\beta_j + g,$$

where $\pi = P(y = 1|X, G_j, g$ represents the $Nx1$ column vector of probabilities of being affected for the N individuals conditional to covariates, allelic dosages; and random effects; X, α, and g are defined as above; G_j is a vector with the allelic dosages (0, 1, or 2 copies of the reference allele) at the locus j; and β_j is its corresponding effect size. The null hypothesis of $\beta_j = 0$ was assessed using a multivariate score test.

Samples and markers of the AGA-ALZAR dataset were submitted to quality control procedures using PLINK version 1.9.[45] The remaining samples (n = 962) had <3% of missing genotypes and passed sex-check and identity-by-state filters. The remaining SNPs had >97% call rate, a minor-allele frequency of >1% and were in Hardy-Weinberg equilibrium ($p > 10^{-6}$). No differences in call rate were observed between cases and controls ($p < 1 \times 10^{-5}$). We estimated the ancestry proportions of the samples using ADMIXTURE[46] from a panel of 446 ancestry informative markers developed elsewhere[47] for Latin American populations. The GWAS was performed fitting a simple logistic regression model, which included an adjustment for sex, age, and the top four PCs.

Results

Sample description

We analyzed data from CH samples from the family-based ADSP Discovery and Extension study. The multiplex ADSP Hispanic families were selected with the goal of improving the detection of genomic regions harboring extremely rare causal variants. The potential enrichment of multiple copies of rare causal variants segregating in pedigrees is an advantage for discovery purposes; however, their effects can be very difficult to detect in unrelated cases and controls from the general population. We additionally analyzed GWAS data from Columbia University (CU) Hispanics, and Argentinian samples from the AGA-ALZAR study in our validation studies (Table 1). The datasets differ in sample size, with the CU Hispanic sample being much bigger than the other two samples. The CU Hispanic and AGA-ALZAR samples are more similar in terms of proportion of affected subjects (45.7% and 40.3%, respectively) and sex distribution, compared with ADSP Hispanic (68% affected individuals, 42.1% males). The frequency of *APOE* alleles is comparable across studies (ε2, 3.8%–5.6%; ε3, 73.8%–78.9%; ε4, 17.3%–20.6%), with lower ε2 and ε4 and higher ε3 frequencies observed in the ADSP Hispanic cohort, likely because of the ascertainment for inclusion in the ADSP.[24] The subjects also have equivalent mean age (72–75 years), but unaffected subjects are slightly younger in the ADSP (68 years vs. 73–74 years).

The population structure of ADSP Hispanic, captured by the two first PCs, showed clustering primarily toward the AFR and EUR reference samples (Figure 1A). The genome-wide global ancestry proportions for each individual, calculated by averaging the local ancestry calls, confirmed

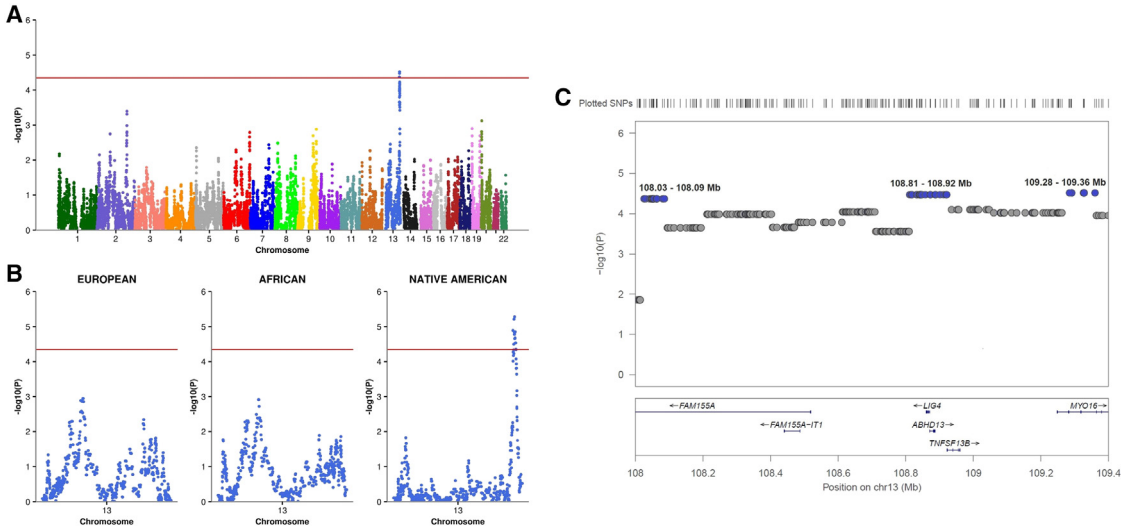

Figure 2. Admixture mapping for AD in ADSP Hispanic samples

(A) Admixture mapping joint test, in which all AFR, EUR, and NAM ancestries are tested simultaneously for association.

(B) Single ancestry admixture mapping analyses on chromosome 13 testing the ancestry noted against the combined remaining ancestries.

(C) Regional association plot on 13q33.3 admixture mapping associated region, with blue dots denoting association results that are genome-wide significant. Red lines in (A) and (B) represent the genome-wide significance threshold ($p < 4.5 \times 10^{-5}$).

the ancestry background of the ADSP Hispanic samples (Figure 1B; Table 1). As expected, the CU Hispanic and AGA-ALZAR samples showed different ancestry composition and population structure (Figure S1) that reflects their different geographic origins within the Americas and corresponding colonization histories. On average, CU Hispanic individuals had a smaller proportion of EUR ancestry and greater AFR proportion, whereas AGA-ALZAR showed a considerably greater NAM proportion (Table 1).

Admixture mapping and GWAS in the ADSP Hispanic sample

Admixture mapping for AD in the ADSP Hispanic sample identified genome-wide significant association between AD and local ancestry at 13q33.3 ($p < 4.5 \times 10^{-5}$) (Figure 2; Table 2). Figure 2A shows the results for the joint admixture mapping test, in which all ancestries were tested simultaneously for association with AD. Further single-ancestry admixture mapping analyses were conducted and showed that the association signal on chromosome 13 is driven by NAM ancestry (Figure 2B).

The 13q33.3 AD association signal represents three loci in the region (108.0–109.4Mb, $p \leq 4.2 \times 10^{-5}$), in which the associations were driven by the NAM background and suggested an a protective effect against AD risk (odds ratio [OR], 0.37–0.41; $p < 4.1 \times 10^{-5}$) (Table 2). This region spans five protein-coding genes (*FAM155A* [MIM: 619899], *MYO16* [MIM: 615479], *ABHD13*, *TNFSF13B* [MIM: 603969], and *LIG4* [MIM: 601837]) and includes variants with regulatory features (Figure 2C; Table S1). Additional evidence of the NAM contribution at 13q33.3 was provided by the linkage disequilibrium (LD) structure in the region. We observed three well-defined LD blocks exhibiting higher correlation between SNPs within each locus when NAM samples were used as the reference, suggesting the existence of NAM haplotypes in the region (Figure 3). Conditional admixture mapping analyses including the lead SNP of each associated locus, either jointly or separately as covariates, removed the signal at 13q33.3 (Figure S2), suggesting that the admixture-LD blocks are equally contributing to the association signal.

A standard GWAS that used PCs to adjust for population structure did not identify any genome-wide significant associations between SNPs in the 13q33.3 region and AD (using $p < 5 \times 10^{-8}$) (Figure S3). We performed the GWAS using a similar logistic mixed model as used for the admixture mapping to avoid any bias caused by adjustments. There are 95 SNPs within the 13q33.3 region defined by the three loci. None of these 95 SNPs were found to be significant in the standard GWAS. However, we identified significant association between AD and a single SNP at 10q26.13 (Figure S3A; $p = 4.5 \times 10^{-8}$, rs7082865, chr10:124,066,652:G:A, ADSP Hispanic minor-allele frequency = 0.44). SNP rs7082865 is an intronic variant in *BTBD16*, a gene that has previously shown a suggestive pleiotropic effect on AD,[48] and suggestive or significant associations with related traits superior parietal cortex volume,[49] and type 2 diabetes,[48] as well as with bipolar disorder.[50] However, it is worth noting that the genomic region surrounding this gene contains extensive structural variation, which can easily lead to spurious conclusions.

Evaluation of the admixture mapping logistic approach

To investigate the performance of the admixture mapping logistic mixed model, we applied simulations that used simulated binary outcome and real genetic data to measure the power and type I error; additionally, we performed

Table 2. Admixture mapping results for AD in the 13q33.3 region in ADSP Hispanic samples

Locus	SNPs[a]	Physical position[b]	Lead SNP	AM joint p value[c]	OR (95% CI)	Ancestry background	
						Ancestry	p value
1	14	108025668–108085618	rs4444189	4.2×10^{-5}	0.41 (0.27–0.63)	NAM	1.3×10^{-5}
2	19	108811734–108921373	rs16972067	3.3×10^{-5}	0.37 (0.23–0.59)	NAM	6.1×10^{-6}
3	11	109282805–109359834	rs16972815	3.0×10^{-5}	0.38 (0.24–0.59)	NAM	5.3×10^{-6}

[a]SNPs: number of SNPs included in the admixed-LD block.
[b]Physical position in genome build GRCh37.p13.
[c]AM Joint p value: p value for the admixture mapping joint test, in which AFR, EUR, and NAM ancestries are tested simultaneously.

admixture mapping of AD in ADSP Hispanic samples using a LMM to examined how much better the logistic approach fits to the data.

As expected, the power of the logistic approach increased with increasing effect sizes at all significance levels (Figure 4A). At a significance level (α) of 10^{-5}, the same magnitude of the nominal significance level applied to the ADSP Hispanic data, the true-positive rate increases from 1.5% to 91.1% when the effect size goes from 0.5 to 1.5. The type I error rates, measured as the mean proportion of false positive results across the genome at different significance levels, were contained within the respective confidence intervals (CIs): CI($\alpha = 1 \times 10^{-4}$)[$-5.2 \times 10^{-4}$; 7.2×10^{-4}], and CI($\alpha = 1 \times 10^{-5}$)[$-1.8 \times 10^{-4}$; 2.1×10^{-4}]) (Table 3).

A comparison of admixture mapping for AD using logistic and LMMs is described in Table 4, and Figures 4B and 4C. Both logistic and LMMs identified the NAM ancestry-derived region at 13q33.3 associated with AD. However, the logistic approach yielded slightly stronger p values in the extremes, conformed more closely to the null distribution than the LMM over most of the distribution (Figures 4B and 4C), and identified an additional AD associated locus (108.02–108.08 Mb) in this region (Table 4).

Validation analyses in CU Hispanic and AGA-ALZAR samples

We performed a validation study in two independent samples. An admixture mapping validation analysis in the CU Hispanic samples failed to provide evidence for association between local ancestry in the 13q33.3 loci and AD (Figure S4). The smallest p value within the ADSP significant loci was observed in a variant that is also part of the most significant AD locus but failed to reach statistical significance (rs16972815; chr13:109,282,805:G:T; $p = 0.132$). There are several factors that may explain the lack of validation in CU Hispanic samples: (1) the average proportion of NAM ancestry across all SNPs within 13q33.3 was 37.5% higher in ADSP Hispanic individuals (Table S2), suggesting that CU Hispanic samples may have an insufficient number of NAM alleles to capture the association in this region. We analyzed the ADSP and CU local ancestry calls to compare the frequency of NAM alleles within the 13q33.3 loci. The number of SNPs in this region was similar (44 for ADSP Hispanic vs. 46 for CU Hispanic), but the frequency of the NAM alleles differed across datasets (Figure S5). ADSP Hispanic samples showed a higher frequency in cases and more balanced frequency between AD cases and controls compared to CU Hispanic individuals (mean frequency, 0.05 [ADSP cases] vs. mean frequency, 0.03 [CU cases] and mean frequency, 0.05 [ADSP controls] vs. mean frequency, 0.05 [CU controls]. (2) ADSP and CU samples have a different ancestry composition and population structure (Figures 1 and S1), which implies that different LD structures may determine the occurrence of distinct haplotypes at 13q33.3. (3) ADSP Hispanic is a family-based cohort, where AD genetic variants should be enriched, and CU Hispanic is a case-control sample. The effect size may not be the same in both samples, since the method needs to deal with the complexity of family relationships in the ADSP Hispanic but not in CU Hispanic. (4) The occurrence of NAM substructure across Central and South American native populations have been previously discussed.[51,52] ADSP and CU Hispanic samples may have a different NAM substructure that also influences the differentiation of NAM haplotypes in this region.

The AD 13q33.3 admixture mapping signal was supported by results from GWAS summary statistics from the AGA-ALZAR sample (Table S3). This sample was used because of its greater NAM ancestry, with the hope that there would be sufficient power for validation of the genes implicated by our original results. From a total of 1,080 variants within the region of interest (chr13:108,025,668–109,359,834), 34 SNPs showed suggestive association under a nominal p value of <0.05 (Table S3). Most of the variants are intronic in the *FAM155A* and *MYO16* genes, and the results suggest the occurrence of at least three NAM haplotypes in this region. Only one SNP (rs9284245; $p = 0.05$), in a *MYO16* intron, was also included in the ADSP admixture mapping analysis.

Haplotype analysis

We further investigated the apparent protective effect of the NAM ancestry against AD by using the WGS data to identify haplotype(s) with elevated frequency on only the NAM ancestral background. Initial investigation identified 18 haplotype-defining variants within two of the three admixture mapping loci located at *LIG4* and *MYO16*. These variants contained alternative alleles for AD that were specific to ADSP Hispanic with NAM ancestry in the region, were absent

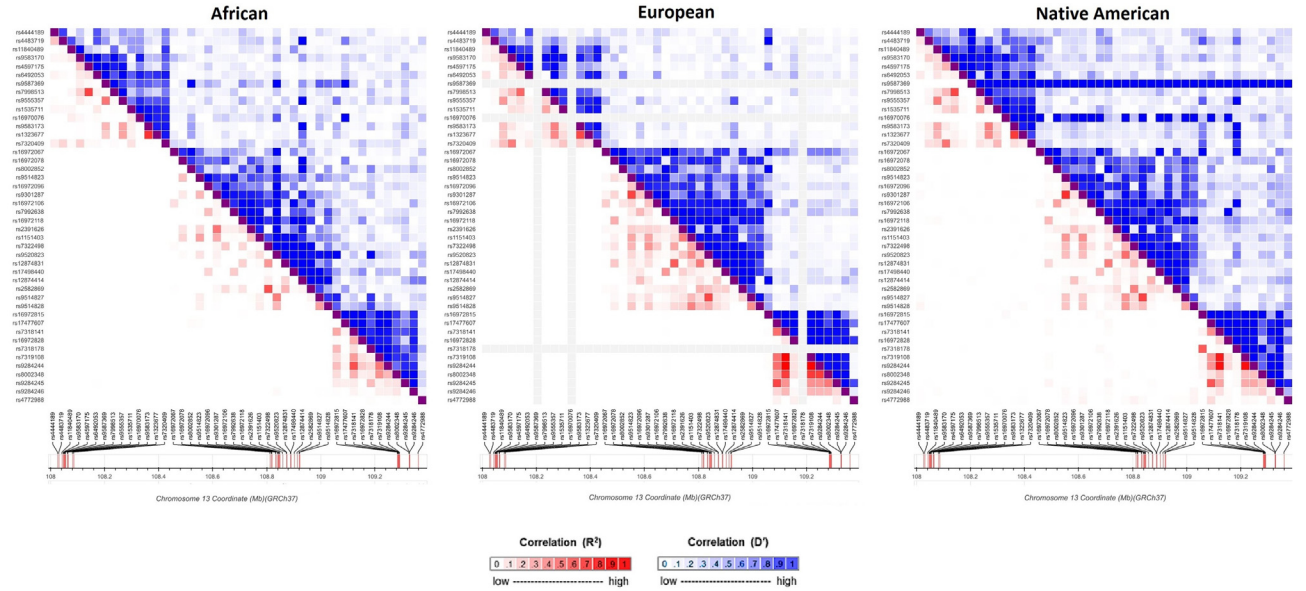

Figure 3. LD within the 13q33.3 region in 1000 Genomes AFR, EUR, and NAM reference samples

in ADSP homozygotes for both AFR or EUR ancestry in the region, and were present only in heterozygotes with these ancestries. These variant allele frequencies were also constant across variants on the NAM ancestral background (Table S4). An additional check of the allele frequencies in the 1000G and HGDP reference samples (Table S5) showed near constancy across the SNPs of the alternate allele frequencies within our data and across members from populations with diverse ancestral backgrounds. This supports NAM ancestry as the major source of the haplotypes. There is evidence of the haplotype, at low to moderate frequency, in East Asian reference samples, consistent with current understanding of the origin of the origins of the peoples of the Americas.[53] In addition, there is evidence for at least two haplotypes (one at each locus). These haplotypes are virtually non-existent on a EUR background in both regions, while the AFR ancestry might be the only other minor source for some of these alternative alleles in the *MYO16* gene region.

Population-based phasing strongly supports the existence of a NAM-specific haplotype in the *LIG4* gene region, while potentially slightly shrinking the most likely boundaries of the critical region. Fifteen of the 21 families with NAM ancestry had at least one copy of the haplotype. Thirty-one of the 62 chromosomes with NAM ancestry represent the full 11-SNP haplotype, and 4 more (from a single family) represent a 1-SNP shorter haplotype formed by dropping rs144294080 (Table S6). While there were only three individuals in three separate families where both chromosomes had NAM ancestry, two of these individuals were homozygous for the haplotype. None of the 148 chromosomes of AFR ancestry carried any of the ALT alleles that define the 11-SNP haplotype. Of the 502 chromosomes with EUR ancestral background, 3 chromosomes from 2 families carried an abbreviated 8-SNP haplotype.

Further evaluation of the haplotype and possible risk-modifying variants comes from available functional annotation. It is worth noting that one of these variants, rs12585114, has a high CADD[44] score of 14, suggesting a potential deleterious/pathogenic effect in the top fourth percentile of variants. In addition, rs12428235 and rs12430274 are regulatory region variants in the same enhancer and are nominally associated with *LIG4* expression. The region of interest also includes several promoter-like and enhancer-like elements active in neurons (http://screen.encodeproject.org),[54] again pointing to *LIG4* (Table S5).

Discussion

We implemented and applied an admixture mapping logistic mixed model approach, LLAMA, which is suitable for binary outcomes in recently admixed populations. This both permits proper control of population structure and relatedness in the data by inclusion of multiple fixed and random effects and leverages ancestry to permit admixture mapping. We do not expect that the broad differences in the AFR, EUR, and NAM ancestry proportions among admixed individuals have any influences in either the model performance or the admixture mapping results. Assessed on real data from the ADSP and on simulated data, our admixture mapping logistic framework demonstrated power to identify true-positive associations under different significance levels with suitable control of the type I error. Analysis of the ADSP CH family sample detected an associated region by admixture analysis at 13q33.3 (13:108,025,668–108,085,618) that was not identified using the standard LMM. Therefore, the proposed logistic mixed model approach can improve gene discovery when conducting admixture mapping for binary traits.

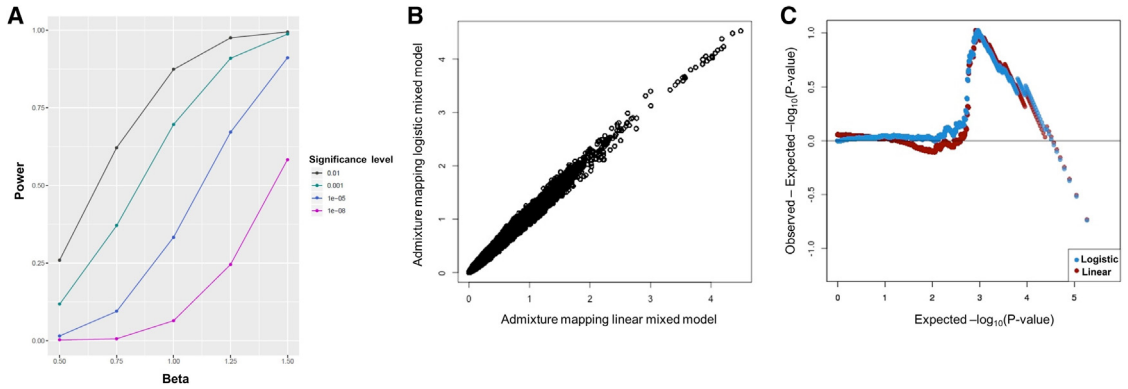

Figure 4. Evaluation of the admixture mapping logistic approach
(A) Power, assessed as the proportion of simulations capturing a significant p value for the causal SNP at different effect sizes (beta) and significance levels.
(B) Scatterplot of the -log(p value).
(C) Quantile difference QQ-plot (difference of -log[p value]) for the admixture mapping performed using logistic and LMMs.

The admixture mapping logistic mixed model identified a NAM ancestry-derived region at 13q33.3 associated with AD in ADSP Hispanic samples ($p < 4.5 \times 10^{-5}$). This included three loci with potential AD protective variants (OR, 0.37–0.41). A protective effect of the NAM ancestry against AD has been previously reported and may be partially related to a lower frequency of the *APOE* ε4 allele among NAM populations.[55–57] Despite not being able to validate the ADSP admixture mapping association at a Bonferroni-corrected value to 13q33.3 in the CU Hispanic cohort, there is suggestive evidence using GWAS summary statistics from the AGA-ALZAR study. Additionally, the implication of the NAM ancestry background with AD was supported by evidence of a uniquely NAM haplotype segregating in this region. The copies of a partial haplotype found in two families with local EUR ancestry are so rare that existing GWAS are considerably underpowered to detect association with the haplotype in Eurocentric populations. This sub-haplotype therefore does not contribute to fine mapping of the region. Our validation and haplotype analyses prioritized the *LIG4*, *MYO16*, and *FAM155A* genes within 13q33.3, with a NAM-specific haplotype identified for *LIG4*. However, further analyses should be conducted to more thoroughly explore and understand the implications of these genes on AD risk.

LIG4 encodes a DNA ligase, which is a core component of the non-homologous end-joining pathway to repair the DNA double-strand breaks (DSBs) in the mature nervous system.[58] DSBs are highly detrimental for neurons. The accumulation of DSBs coupled with a defective mechanism of DNA repair plays an important role in the onset and/or progression of AD, contributing to neuronal damage and impaired learning and memory, as documented in models and postmortem AD brains.[59,60] *LIG4* shows evidence of differential expression in AD (https://agora.ampadportal.org/) and its gene expression pattern is similar to two of the therapeutic targets for AD that were nominated by the Accelerating Medicines Partnership in Alzheimer Disease consortium (https://adknowledgeportal.synapse.org/).

MYO16 encodes a class of myosin proteins involved in morphogenesis of neuronal cells. It acts through the phosphoinositide 3-kinase signaling pathway,[61,62] with differential expression in excitatory and inhibitory neurons of aged brains.[63] It has been previously associated with different neurological disorders,[64–66] including slower AD cognitive decline,[67] AD age-at-onset delay,[68] and an AD phenotype combining affection status and age at onset.[69] *MYO16* is overexpressed in the prefrontal cortex of individuals with major depressive disorder through an enrichment of 5-hydroxymethylcytosine (5hmC).[70] One of the epigenetic markers that is altered in AD is 5hmC. This epigenetic mark is involved in the regulation of learning, synaptic plasticity, and memory processes. Its levels are significantly increased in AD middle frontal gyrus and middle temporal gyrus (MTG) brain regions, two areas that are vulnerable to AD, and positively correlated with neurofibrillary tangles, β-amyloid peptides, and ubiquitin levels in the MTG region. This suggests that 5hmC epigenetic marks play an important role in the determination and progression of AD.[71] A gene expression study in the offspring of male rats exposed to prenatal stress has also suggested an overrepresentation of the nicotine acetylcholine receptor signaling pathway, in which *MYO16* is one of the participant genes.[72] Nicotine acetylcholine receptors are linked to different neuroprotection pathways, which control the cell death and secretion of β-amyloid peptides.[73]

The third gene implicated, *FAM155A*, also known as *NLF-1*, is an auxiliary subunit of the sodium leak channel (NALCN). *FAM155A* is a critical component for neuronal excitability; it regulates NALCN ion permeation by protecting the ion-selectivity filter against neurotoxin attack, and is conserved across species.[74,75] It is highly expressed in the brain and, as other genes of the NALCN complex, and has been previously implicated in several neurological and cognitive-related disorders.[76] Associations with AD in a trans-ethnic meta-analysis that included samples of Japanese and EUR ancestries,[77] and with AD cerebrospinal fluid Aβ42/Aβ40 ratio biomarker in EUR samples,[31] suggests that *FAM155A* is a gene with a trans-ethnic effect on AD.

Table 3. Mean proportion of false positive associations (type I error) across the genome observed for the admixture mapping logistic mixed model in a simulation study

	Significance level	
Effect size	1×10^{-4}	1×10^{-5}
0.50	5.3×10^{-4}	8.1×10^{-5}
0.75	5.4×10^{-4}	8.2×10^{-5}
1.00	5.0×10^{-4}	7.7×10^{-5}
1.25	5.0×10^{-4}	7.7×10^{-5}
1.50	4.3×10^{-4}	6.4×10^{-5}

The identification of haplotypes that are specific to NAMs within the three genetic ancestries lends further support for the presence of AD risk modifiers in the 13q33.3 region. These haplotypes are supported by well-defined LD blocks in 1000 Genomes NAM populations and by haplotype analysis using ADSP WGS data. Admixture mapping leverages the information from coarse blocks of different ancestries that occur in recently admixed populations. An identifiable haplotype that is heavily enriched in one diagnostic group in one genetic ancestry, such as the haplotype identified in *LIG4*, as seems to be the case here, suggests the existence of a single risk-altering event, which should improve the prognosis of downstream analyses.

A challenge and limitation of our study and others that include admixed samples is the lack of a genetically similar population to validate the findings. Although Hispanic and Latino populations share the AFR, EUR, and NAM ancestry components, the ancestry proportions and population structure are very diverse across admixed populations from different geographic regions as a result of their historical process and migration.[78–80] Along with the fact that these populations remain underrepresented in genetic studies, genetic heterogeneity complicates replication. Here, we used two independent samples to validate the NAM-derived loci associated with AD. The CU Hispanic sample had a similar NAM global ancestry to ADSP Hispanic, but the smallest number of NAM alleles at 13q33.3. Our results suggest that validation may depend on both global and local (at the region of interest) ancestries, which brings an additional problem to find the right sample for replication. The AGA-ALZAR Argentinian sample had a greater NAM component, which was an advantage for validating the NAM-derived region, but only GWAS summary statistics were available. Unlike GWAS, admixture mapping captures ancestry-specific regions associated with an outcome and has no resolution to detect the underlying risk or protective genetic variants. Although we validated the 13q33.3 region with a suggestive threshold in the AGA-ALZAR samples, distinct variants could be adjacent to the admixture mapping NAM loci because of the different genetic architectures influencing AD in the ADSP Hispanic and AGA-ALZAR samples. In addition, the occurrence of population substructure in Hispanic and Latino populations have been discussed,[81] including substructure across Central and South American native populations.[51,52] ADSP Hispanic and AGA-ALZAR samples may have a NAM substructure that was not accounted for in our study. Working with NAM populations is really challenging. There are no publicly available datasets to broadly represent the diversity of NAM populations, and, for the few available datasets, the number of samples is small. Like ours, other studies in such admixed populations have struggled with replication,[13,82,83] which highlights the needs of discussion about the replication requirements for genetic studies involving admixed populations.

In summary, we implemented an admixture mapping logistic mixed model approach, LLAMA, to analyze binary outcomes that correctly considered the sources of structure and dependence of the data and showed adequate statistical power to detect true positive associations while controlling for type I error. By applying this logistic framework to ADSP Hispanic samples, we identified an association between three NAM ancestry-derived loci at 13q33.3 and AD that is not detected by the traditional GWAS. While only 34

Table 4. Admixture mapping for AD in ADSP Hispanic data using logistic vs. LMMs

				Logistic mixed model			LMM		
					Ancestry background			Ancestry background	
Locus	SNPs[a]	Physical position[b]	Lead SNP	AM joint p value[c]	Ancestry	p value	AM joint p value[c]	Ancestry	p value
1	14	108025668–108085618	rs4444189	$\mathbf{4.2 \times 10^{-5}}$	NAM	1.3×10^{-5}	–	–	–
2	19	108811734–108921373	rs16972067	$\mathbf{3.3 \times 10^{-5}}$	NAM	6.1×10^{-6}	4.4×10^{-5}	NAM	8.4×10^{-6}
3	11	109282805–109359834	rs16972815	$\mathbf{3.0 \times 10^{-5}}$	NAM	5.3×10^{-6}	$\mathbf{3.2 \times 10^{-5}}$	NAM	5.9×10^{-6}

[a]SNPs: number of SNPs included in each associated locus.
[b]Physical position in genome build GRCh37.p13.
[c]AM Joint p value: p value for the admixture mapping joint test, in which all AFR, EUR, and NAM ancestries are tested simultaneously.

SNPs from the 13q33.3 association were nominally significant in the AGA-ALZAR samples, the occurrence of a unique AD-protective NAM variant in this region was suggested by haplotype analysis using the ADSP whole-genome sequencing data. Genes in this region (*LIG4*, *MYO16*, and *FAM155A*) have been previously implicated in AD and other neurological and cognitive disorders. Our study adds to the understanding of the genetic contributions to AD, providing new insights into ancestry-specific genetic regions influencing AD in Hispanic and Latino populations.

Data and code availability

The Alzheimer Disease Sequencing Project (ADSP Hispanic) data supporting the findings of this study is available in the dbGaP repository at dbGaP: https://www.ncbi.nlm.nih.gov/gap/, under the accession number dbGaP: phs000572.v7.p4 [GWAS data], and dbGap: fsa000003 [WGS data; NG00067.v2 release]. The ADSP WGS data is also available in the NIAGADS repository (https://www.niagads.org/), under the accession number NIAGADS: fsa000003 (release NG00067.v2; 5K sample), and NIAGADS: fsa000006 (release NG0067.v7; 17K sample). The CU Hispanic dataset is available in the dbGaP repository (accession number dbGaP: phs000496.v1.p1). The AGA-ALZAR AD GWAS results are currently under publication process; once published, the GWAS summary statistics will be available upon request. Reference samples used in the PCs and local ancestry analyses are publicly available at NCBI: https://ftp.ncbi.nlm.nih.gov/hapmap/ (HapMap phase 3), HGDP: https://hagsc.org/hgdp/files.html (HGDP), and IGSR: https://www.internationalgenome.org/data (1000 Genomes phase 3). The LLAMA framework is implemented in the GENESIS R package freely available at http://bioconductor.org/packages/release/bioc/html/GENESIS.html.

Supplemental information

Supplemental information includes five figures (Figures S1-S5), six tables (Tables S1-S6), and supplementary acknowledgements, and can be found online at https://doi.org/10.1016/j.xhgg.2023.100207.

Acknowledgments

The Alzheimer Disease Sequencing Project (ADSP) comprises two Alzheimer Disease (AD) genetics consortia and four National Human Genome Research Institute (NHGRI) funded Large Scale Sequencing and Analysis Centers (LSAC). The two AD genetics consortia are the Alzheimer Disease Genetics Consortium (ADGC) funded by NIA (U01 AG032984) and the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) funded by the NIA (R01 AG033193), the National Heart, Lung, and Blood Institute (NHLBI), other National Institutes of Health (NIH) institutes, and other foreign governmental and non-governmental organizations. The Discovery Phase analysis of sequence data is supported through UF1AG047133 (to Drs. Schellenberg, Farrer, Pericak-Vance, Mayeux, and Haines); U01AG049505 to Dr. Seshadri; U01AG049506 to Dr. Boerwinkle; U01AG049507 to Dr. Wijsman; and U01AG049508 to Dr. Goate and the Discovery Extension Phase analysis is supported through U01AG052411 to Dr. Goate, U01AG052410 to Dr. Pericak-Vance and U01 AG052409 to Drs. Seshadri and Fornage. The four LSACs are the Human Genome Sequencing Center at the Baylor College of Medicine (U54 HG003273), the Broad Institute Genome Center (U54HG003067), The American Genome Center at the Uniformed Services University of the Health Sciences (U01AG057659), and the Washington University Genome Institute (U54HG 003079). Data generation and harmonization is supported by U54AG052427 (to Drs. Schellenberg and Wang). Genotyping of the Argentinian sample was supported by a grant (European Alzheimer DNA BioBank, EADB) from the EU Joint Program – Neurodegenerative Disease Research (JPND) to Alfredo Ramirez (German Federal Ministry of Education and Research (BMBF) grant: 01ED1619A). Dr. Dalmasso was supported by the Alexander von Humboldt Foundation.

Author contributions

A.R.V.R.H.: formal analysis, visualization, writing - original draft; E.E.B.: conceptualization, resources, review, and editing; L.A.B.: conceptualization, methodology; K.E.G.: methodology, software; H.K.S.: data curation, software; A.Q.N.: software; J.C.B.: conceptualization; L.I.B., L.M., A.R., and C.D.: resources, formal analysis; R.N.: formal analysis, software; R.M.: resources, supervision, conceptualization, funding acquisition, project administration; S.R.B.: methodology; E.M.W.: resources, conceptualization, funding acquisition, project administration, supervision, data analysis, writing - original draft and review and editing; T.A.T.: conceptualization, methodology, resources, funding acquisition, project administration, writing - original draft and review and editing.

Declaration of interests

The authors declare no competing interests.

Received: November 28, 2022
Accepted: May 16, 2023

Web resources

AD Knowledge Portal, https://adknowledgeportal.synapse.org/.

ADMIXTURE, https://dalexander.github.io/admixture/.

AGORA, https://agora.adknowledgeportal.org/.

Beagle, http://faculty.washington.edu/browning/beagle/.

CADD, https://cadd.gs.washington.edu/.

GENESIS, https://bioconductor.org/packages/release/bioc/html/GENESIS.html.

KING, https://www.kingrelatedness.com/.

LDMatrix, https://ldlink.nci.nih.gov/?tab=ldmatrix.

LiftOver, https://genome.ucsc.edu/cgi-bin/hgLiftOver.

OMIM, https://www.omim.org/.

PLINK 1.0.7, https://zzz.bwh.harvard.edu/plink/.

PLINK 1.9, https://www.cog-genomics.org/plink/.

R, https://www.r-project.org/.

RFMix 1.5.4, https://sites.google.com/site/rfmixlocalancestryinference/.

SCREEN, http://screen.encodeproject.org/.

Shapeit 2, https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html.

References

1. (2020). 2020 Alzheimer's disease facts and figures. Alzheimer's Dement. *16*, 391–460. https://doi.org/10.1002/alz.12068.
2. Matthews, K.A., Xu, W., Gaglioti, A.H., Holt, J.B., Croft, J.B., Mack, D., and McGuire, L.C. (2019). Racial and ethnic estimates of Alzheimer's disease and related dementias in the United States (2015–2060) in adults aged ≥65 years. Alzheimer's Dement. *15*, 17–24. https://doi.org/10.1016/j.jalz.2018.06.3063.
3. Mehta, K.M., and Yeo, G.W. (2017). Systematic review of dementia prevalence and incidence in United States race/ethnic populations. Alzheimers Dement. *13*, 72–83. https://doi.org/10.1016/j.jalz.2016.06.2360.
4. Buniello, A., Macarthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. *47*, D1005–D1012. https://doi.org/10.1093/nar/gky1120.
5. Need, A.C., and Goldstein, D.B. (2009). Next generation disparities in human genomics: concerns and remedies. Trends Genet. *25*, 489–494. https://doi.org/10.1016/j.tig.2009.09.012.
6. Popejoy, A.B., and Fullerton, S.M. (2016). Genomics is failing on diversity. Nature *538*, 161–164. https://doi.org/10.1038/538161a.
7. Tang, M.X., Maestre, G., Tsai, W.Y., Liu, X.H., Feng, L., Chung, W.Y., Chun, M., Schofield, P., Stern, Y., Tycko, B., and Mayeux, R. (1996). 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. *58*, 574–584.
8. Tang, M.X., Stern, Y., Marder, K., Bell, K., Gurland, B., Lantigua, R., Andrews, H., Feng, L., Tycko, B., and Mayeux, R. (1998). The *APOE*-epsilon4 allele and the risk of Alzheimer disease among African Americans, whites, and Hispanics. JAMA *279*, 751–755. https://doi.org/10.1001/jama.279.10.751.
9. Rogaeva, E., Meng, Y., Lee, J.H., Gu, Y., Kawarai, T., Zou, F., Katayama, T., Baldwin, C.T., Cheng, R., Hasegawa, H., et al. (2007). The neuronal sortilin-related receptor *SORL1* is genetically associated with Alzheimer disease. Nat. Genet. *39*, 168–177. https://doi.org/10.1038/ng1943.
10. Lee, J.H., Cheng, R., Barral, S., Reitz, C., Medrano, M., Lantigua, R., Jiménez-Velazquez, I.Z., Rogaeva, E., St George-Hyslop, P.H., and Mayeux, R. (2011). Identification of novel loci for Alzheimer disease and replication of *CLU, PICALM,* and *BIN1* in Caribbean Hispanic individuals. Arch. Neurol. *68*, 320–328. https://doi.org/10.1001/archneurol.2010.292.
11. Blue, E.E., Horimoto, A.R.V.R., Mukherjee, S., Wijsman, E.M., and Thornton, T.A. (2019). Local ancestry at *APOE* modifies Alzheimer's disease risk in Caribbean Hispanics. Alzheimers Dement. *15*, 1524–1532. https://doi.org/10.1016/j.jalz.2019.07.016.
12. Horimoto, A.R.V.R., Xue, D., Thornton, T.A., and Blue, E.E. (2021). Admixture mapping reveals the association between Native American ancestry at 3q13.11 and reduced risk of Alzheimer's disease in Caribbean Hispanics. Alzheimer's Res. Ther. *13*, 122. https://doi.org/10.1186/S13195-021-00866-9.
13. Horimoto, A.R.V.R., Xue, D., Cai, J., Lash, J.P., Daviglus, M.L., Franceschini, N., and Thornton, T.A. (2022). Genome-wide admixture mapping of estimated glomerular filtration rate and chronic kidney disease identifies European and African ancestry-of-origin loci in hispanic and Latino individuals in the United States. J. Am. Soc. Nephrol. *33*, 77–87. https://doi.org/10.1681/ASN.2021050617.
14. Loesch, D.P., V R Horimoto, A.R., Heilbron, K., Irem Sarihan, E., Inca-Martinez, M., Mason, E., Cornejo-Olivas, M., Torres, L., Mazzetti, P., Cosentino, C., et al. Characterizing the genetic architecture of Parkinson's disease in Latinos. Ann. Neurol. 10.1002/ana.26153.
15. Stephens, J.C., Briscoe, D., and O'Brien, S.J. (1994). Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. Am. J. Hum. Genet. *55*, 809–824.
16. Sul, J.H., Martinid, L.S., and Eskin, E. (2018). Population structure in genetic studies: confounding factors and mixed models. PLoS Genet. https://doi.org/10.1371/journal.pgen.1007309.
17. Li, G., Rivas, P., Bedolla, R., Thapa, D., Reddick, R.L., Ghosh, R., and Kumar, A.P. (2013). Genetic studies: the linear mixed models in genome-wide association studies. Cancer Prev. Res. *6*, 27–39.
18. Chen, H., Wang, C., Conomos, M.P., Stilp, A.M., Li, Z., Sofer, T., Szpiro, A.A., Chen, W., Brehm, J.M., Celedón, J.C., et al. (2016). Control for population structure and relatedness for binary traits in genetic association studies via logistic mixed models. Am. J. Hum. Genet. *98*, 653–666. https://doi.org/10.1016/j.ajhg.2016.02.012.
19. Ziyatdinov, A., Parker, M.M., Vaysse, A., Beaty, T.H., Kraft, P., Cho, M.H., and Aschard, H. (2020). Mixed-model admixture mapping identifies smoking-dependent loci of lung function in African Americans. Eur. J. Hum. Genet. *28*, 656–668. https://doi.org/10.1038/s41431-019-0545-8.
20. Wang, H., Cade, B.E., Sofer, T., Sands, S.A., Chen, H., Browning, S.R., Stilp, A.M., Louie, T.L., Thornton, T.A., Johnson, W.C., et al. (2019). Admixture mapping identifies novel loci for obstructive sleep apnea in Hispanic/Latino Americans. Hum. Mol. Genet. *28*, 675–687. https://doi.org/10.1093/hmg/ddy387.
21. Sofer, T., Baier, L.J., Browning, S.R., Thornton, T.A., Talavera, G.A., Wassertheil-Smoller, S., Daviglus, M.L., Hanson, R., Kobes, S., Cooper, R.S., et al. (2017). Admixture mapping in the Hispanic Community Health Study/Study of Latinos reveals regions of genetic associations with blood pressure traits. PLoS One. https://doi.org/10.1371/journal.pone.0188400.
22. Brown, L.A., Sofer, T., Stilp, A.M., Baier, L.J., Kramer, H.J., Masindova, I., Levy, D., Hanson, R.L., Moncrieft, A.E., Redline, S., et al. (2017). Admixture mapping identifies an amerindian ancestry locus associated with albuminuria in hispanics in the United States. J. Am. Soc. Nephrol. *28*, 2211–2220. https://doi.org/10.1681/ASN.2016091010.
23. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575. https://doi.org/10.1086/519795.
24. Beecham, G.W., Bis, J.C., Martin, E.R., Choi, S.H., DeStefano, A.L., Van Duijn, C.M., Fornage, M., Gabriel, S.B., Koboldt, D.C., Larson, D.E., et al. (2017). The Alzheimer's disease sequencing

project: study design and sample selection. Neurol. Genet. *3*, e194. https://doi.org/10.1212/NXG.0000000000000194.
25. Naj, A.C., Lin, H., Vardarajan, B.N., White, S., Lancour, D., Ma, Y., Schmidt, M., Sun, F., Butkiewicz, M., Bush, W.S., et al. (2019). Quality control and integration of genotypes from two calling pipelines for whole genome sequence data in the Alzheimer's disease sequencing project. Genomics *111*, 808–818. https://doi.org/10.1016/j.ygeno.2018.05.004.
26. Nafikov, R.A., Nato, A.Q., Sohi, H., Wang, B., Brown, L., Horimoto, A.R., Vardarajan, B.N., Barral, S.M., Tosto, G., Mayeux, R.P., et al. (2018). Analysis of pedigree data in populations with multiple ancestries: strategies for dealing with admixture in Caribbean Hispanic families from the ADSP. Genet. Epidemiol. *42*, 500–515. https://doi.org/10.1002/gepi.22133.
27. International HapMap Consortium (2003). The International HapMap project. Nature *426*, 789–796. https://doi.org/10.1038/nature02168.
28. Cavalli-Sforza, L.L., Wilson, A.C., Cantor, C.R., Cook-Deegan, R.M., and King, M.C. (1991). Call for a worldwide survey of human genetic diversity: a vanishing opportunity for the Human Genome Project. Genomics *11*, 490–491. https://doi.org/10.1016/0888-7543(91)90169-F.
29. Cann, H.M., de Toma, C., Cazes, L., Legrand, M.F., Morel, V., Piouffre, L., Bodmer, J., Bodmer, W.F., Bonne-Tamir, B., Cambon-Thomsen, A., et al. (2002). A human genome diversity cell line panel. Science *296*, 261–262. https://doi.org/10.1126/science.296.5566.261b.
30. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74. https://doi.org/10.1038/nature15393.
31. Hong, S., Prokopenko, D., Dobricic, V., Kilpert, F., Bos, I., Vos, S.J.B., Tijms, B.M., Andreasson, U., Blennow, K., Vandenberghe, R., et al. (2020). Genome-wide association study of Alzheimer's disease CSF biomarkers in the EMIF-AD Multimodal Biomarker Discovery dataset. Transl. Psychiatry *10*, 403. https://doi.org/10.1038/s41398-020-01074-z.
32. Dalmasso, M.C., Brusco, L.I., Olivar, N., Muchnik, C., Hanses, C., Milz, E., Becker, J., Heilmann-Heimbach, S., Hoffmann, P., Prestia, F.A., et al. (2019). Transethnic meta-analysis of rare coding variants in *PLCG2, ABI3,* and *TREM2* supports their general contribution to Alzheimer's disease. Transl. Psychiatry *9*, 55. https://doi.org/10.1038/s41398-019-0394-9.
33. McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack, C.R., Kawas, C.H., Klunk, W.E., Koroshetz, W.J., Manly, J.J., Mayeux, R., et al. (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's Dement. *7*, 263–269. https://doi.org/10.1016/j.jalz.2011.03.005.
34. Kuhn, R.M., Haussler, D., and Kent, W.J. (2013). The UCSC genome browser and associated tools. Brief. Bioinform. *14*, 144–161. https://doi.org/10.1093/bib/bbs038.
35. Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. *81*, 1084–1097. https://doi.org/10.1086/521987.
36. Delaneau, O., Marchini, J., and Zagury, J.F. (2011). A linear complexity phasing method for thousands of genomes. Nat. Methods *9*, 179–181. https://doi.org/10.1038/NMETH.1785.
37. Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. Am. J. Hum. Genet. *93*, 278–288. https://doi.org/10.1016/j.ajhg.2013.06.020.
38. Conomos, M.P., Miller, M.B., and Thornton, T.A. (2015). Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. Genet. Epidemiol. *39*, 276–293. https://doi.org/10.1002/gepi.21896.
39. Conomos, M.P., Reiner, A.P., Weir, B.S., and Thornton, T.A. (2016). Model-free estimation of recent genetic relatedness. Am. J. Hum. Genet. *98*, 127–148. https://doi.org/10.1016/j.ajhg.2015.11.022.
40. Gogarten, S.M., Sofer, T., Chen, H., Yu, C., Brody, J.A., Thornton, T.A., Rice, K.M., and Conomos, M.P. (2019). Genetic association testing using the GENESIS R/Bioconductor package. Bioinformatics *35*, 5346–5348. https://doi.org/10.1093/bioinformatics/btz567.
41. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867–2873. https://doi.org/10.1093/bioinformatics/btq559.
42. Brown, L.A. (2016). Statistical Methods in Admixture Mapping: Mixed Model Based Testing and Genome-wide Significance Thresholds (University of Washington).
43. R Core Team (2021). R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing).
44. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. *47*, D886–D894. https://doi.org/10.1093/NAR/GKY1016.
45. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, 7. https://doi.org/10.1186/S13742-015-0047-8.
46. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109.
47. Galanter, J.M., Fernandez-Lopez, J.C., Gignoux, C.R., Barnholtz-Sloan, J., Fernandez-Rozadilla, C., Via, M., Hidalgo-Miranda, A., Contreras, A.V., Figueroa, L.U., Raska, P., et al. (2012). Development of a panel of genome-wide ancestry informative markers to study admixture throughout the americas. PLoS Genet. *8*, e1002554. https://doi.org/10.1371/journal.pgen.1002554.
48. Wang, X.F., Lin, X., Li, D.Y., Zhou, R., Greenbaum, J., Chen, Y.C., Zeng, C.P., Peng, L.P., Wu, K.H., Ao, Z.X., et al. (2017). Linking Alzheimer's disease and type 2 diabetes: novel shared susceptibility genes detected by cFDR approach. J. Neurol. Sci. *380*, 262–272. https://doi.org/10.1016/J.JNS.2017.07.044.
49. Alliey-Rodriguez, N., Grey, T.A., Shafee, R., Asif, H., Lutz, O., Bolo, N.R., Padmanabhan, J., Tandon, N., Klinger, M., Reis, K., et al. (2019). *NRXN1* is associated with enlargement of the temporal horns of the lateral ventricles in psychosis. Transl. Psychiatry *9*, 230. https://doi.org/10.1038/s41398-019-0564-9.
50. Smith, E.N., Bloss, C.S., Badner, J.A., Barrett, T., Belmonte, P.L., Berrettini, W., Byerley, W., Coryell, W., Craig, D., Edenberg, H.J., et al. (2009). Genome-wide association study of bipolar disorder in European American and African American individuals. Mol. Psychiatry *14*, 755–763. https://doi.org/10.1038/mp.2009.43.

51. Ara ujo Castro Silva, M., Ferraz, T., Couto-Silva, M., Lemes, R.B., Nunes, K., Comas, D., and abita Hünemeier, T. Population histories and genomic diversity of South American Natives. Mol. Biol. Evol. 10.1093/molbev/msab339.
52. Skoglund, P., Mallick, S., Cátira Bortolini, M., Chennagiri, N., Hünemeier, T., Luiza Petzl-Erler, M., Salzano, F.M., Patterson, N., and Reich, D. Genetic evidence for two founding populations of the Americas. Nature 10.1038/nature14895.
53. Willerslev, E., and Meltzer, D.J. (2021). Peopling of the Americas as inferred from ancient genomics. Nature *594*, 356–364. https://doi.org/10.1038/s41586-021-03499-y.
54. Abascal, F., Acosta, R., Addleman, N.J., Adrian, J., Afzal, V., Ai, R., Aken, B., Akiyama, J.A., Jammal, O.A., Amrhein, H., et al. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature *583*, 699–710. https://doi.org/10.1038/s41586-020-2493-4.
55. Benedet, A.L., Moraes, C.F., Camargos, E.F., Oliveira, L.F., Souza, V.C., Lins, T.C., Henriques, A.D., Carmo, D.G.S., Machado-Silva, W., Araújo, C.N., et al. (2012). Amerindian genetic ancestry protects against Alzheimer's disease. Dement. Geriatr. Cogn. Disord *33*, 311–317. https://doi.org/10.1159/000339672.
56. Weiner, M.F., Hynan, L.S., Beekly, D., Koepsell, T.D., and Kukull, W.A. (2007). Comparison of Alzheimer's disease in American Indians, whites, and African Americans. Alzheimer's Dement. *3*, 211–216. https://doi.org/10.1016/j.jalz.2007.04.376.
57. Henderson, J.N., Crook, R., Crook, J., Hardy, J., Onstead, L., Carson-Henderson, L., Mayer, P., Parker, B., Petersen, R., and Williams, B. (2002). Apolipoprotein E4 and tau allele frequencies among Choctaw Indians. Neurosci. Lett. *324*, 77–79. https://doi.org/10.1016/s0304-3940(02)00150-7.
58. McKinnon, P.J. (2009). DNA repair deficiency and neurological disease. Nat. Rev. Neurosci. *10*, 100–112. https://doi.org/10.1038/nrn2559.
59. Thadathil, N., Delotterie, D.F., Xiao, J., Hori, R., McDonald, M.P., and Khan, M.M. (2021). DNA double-strand break accumulation in Alzheimer's disease: evidence from experimental models and postmortem human brains. Mol. Neurobiol. *58*, 118–131. https://doi.org/10.1007/s12035-020-02109-8.
60. Madabhushi, R., Pan, L., and Tsai, L.H. (2014). DNA damage and its links to neurodegeneration. Neuron *83*, 266–282. https://doi.org/10.1016/j.neuron.2014.06.034.
61. Patel, K.G., Liu, C., Cameron, P.L., and Cameron, R.S. (2001). Myr 8, A novel unconventional myosin expressed during brain development associates with the protein phosphatase catalytic subunits 1α and 1γ1. J. Neurosci. *21*, 7954–7968. https://doi.org/10.1523/jneurosci.21-20-07954.2001.
62. Yokoyama, K., Tezuka, T., Kotani, M., Nakazawa, T., Hoshina, N., Shimoda, Y., Kakuta, S., Sudo, K., Watanabe, K., Iwakura, Y., and Yamamoto, T. (2011). NYAP: a phosphoprotein family that links PI3K to WAVE1 signalling in neurons. EMBO J. *30*, 4739–4754. https://doi.org/10.1038/emboj.2011.348.
63. Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J.Z., Menon, M., He, L., Abdurrob, F., Jiang, X., et al. (2019). Single-cell transcriptomic analysis of Alzheimer's disease. Nature *570*, 332–337. https://doi.org/10.1038/s41586-019-1195-2.
64. Kao, C.-F., Chen, H.-W., Chen, H.-C., Yang, J.-H., Huang, M.-C., Chiu, Y.-H., Lin, S.-K., Lee, Y.-C., Liu, C.-M., Chuang, L.-C., et al. (2016). Identification of susceptible loci and enriched pathways for bipolar II disorder using genome-wide association studies. Int. J. Neuropsychopharmacol. *19*, pyw064. https://doi.org/10.1093/ijnp/pyw064.
65. Liu, Y.F., Sowell, S.M., Luo, Y., Chaubey, A., Cameron, R.S., Kim, H.-G., and Srivastava, A.K. (2015). Autism and intellectual disability-associated *KIRREL3* interacts with neuronal proteins *MAP1B* and *MYO16* with potential roles in neurodevelopment. PLoS One *10*, e0123106. https://doi.org/10.1371/journal.pone.0123106.
66. Rodriguez-Murillo, L., Xu, B., Roos, J.L., Abecasis, G.R., Gogos, J.A., and Karayiorgou, M. (2014). Fine mapping on chromosome 13q32-34 and brain expression analysis implicates *MYO16* in schizophrenia. Neuropsychopharmacology *39*, 934–943. https://doi.org/10.1038/npp.2013.293.
67. Sherva, R., Tripodis, Y., Bennett, D.A., Chibnik, L.B., Crane, P.K., De Jager, P.L., Farrer, L.A., Saykin, A.J., Shulman, J.M., Naj, A., et al. (2014). Genome-wide association study of the rate of cognitive decline in Alzheimer's disease. Alzheimer's Dement. *10*, 45–52. https://doi.org/10.1016/j.jalz.2013.01.008.
68. Naj, A.C., Jun, G., Reitz, C., Kunkle, B.W., Perry, W., Park, Y.S., Beecham, G.W., Rajbhandary, R.A., Hamilton-Nelson, K.L., Wang, L.S., et al. (2014). Effects of multiple genetic loci on age at onset in late-onset Alzheimer disease: a genome-wide association study. JAMA Neurol. *71*, 1394–1404. https://doi.org/10.1001/jamaneurol.2014.1491.
69. Herold, C., Hooli, B.V., Mullin, K., Liu, T., Roehr, J.T., Mattheisen, M., Parrado, A.R., Bertram, L., Lange, C., and Tanzi, R.E. (2016). Family-based association analyses of imputed genotypes reveal genome-wide significant association of Alzheimer's disease with *OSBPL6, PTPRG,* and *PDCL3*. Mol. Psychiatry *21*, 1608–1612. https://doi.org/10.1038/mp.2015.218.
70. Gross, J.A., Pacis, A., Chen, G.G., Drupals, M., Lutz, P.-E., Barreiro, L.B., and Turecki, G. (2017). Gene-body 5-hydroxymethylation is associated with gene expression changes in the prefrontal cortex of depressed individuals. Transl. Psychiatry *7*, e1119. https://doi.org/10.1038/tp.2017.93.
71. Coppieters, N., Dieriks, B.V., Lill, C., Faull, R.L.M., Curtis, M.A., and Dragunow, M. (2014). Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. Neurobiol. Aging *35*, 1334–1344. https://doi.org/10.1016/j.neurobiolaging.2013.11.031.
72. Mychasiuk, R., Gibb, R., and Kolb, B. (2011). Prenatal stress produces sexually dimorphic and regionally specific changes in gene expression in hippocampus and frontal cortex of developing rat offspring. Dev. Neurosci. *33*, 531–538. https://doi.org/10.1159/000335524.
73. Buckingham, S.D., Jones, A.K., Brown, L.A., and Sattelle, D.B. (2009). Nicotinic acetylcholine receptor signalling: roles in Alzheimer's disease and amyloid neuroprotection. Pharmacol. Rev. *61*, 39–61. https://doi.org/10.1124/pr.108.000562.
74. Kschonsak, M., Chua, H.C., Noland, C.L., Weidling, C., Clairfeuille, T., Bahlke, O.Ø., Ameen, A.O., Li, Z.R., Arthur, C.P., Ciferri, C., et al. (2020). Structure of the human sodium leak channel NALCN. Nature *587*, 313–318. https://doi.org/10.1038/s41586-020-2570-8.
75. Xie, J., Ke, M., Xu, L., Lin, S., Huang, J., Zhang, J., Yang, F., Wu, J., and Yan, Z. (2020). Structure of the human sodium leak channel NALCN in complex with *FAM155A*. Nat. Commun. *11*, 5831. https://doi.org/10.1038/s41467-020-19667-z.
76. Cochet-Bissuel, M., Lory, P., Monteil, A., Altier, C., Feng, Z.-P., Bosman, L., and Mc, E. (2014). The sodium leak channel, NALCN, in health and disease. Front. Cell. Neurosci. *8*, 132. https://doi.org/10.3389/fncel.2014.00132.

77. Shigemizu, D., Mitsumori, R., Akiyama, S., Miyashita, A., Morizono, T., Higaki, S., Asanomi, Y., Hara, N., Tamiya, G., Kinoshita, K., et al. (2021). Ethnic and trans-ethnic genome-wide association studies identify new loci influencing Japanese Alzheimer's disease risk. Transl. Psychiatry *11*, 151. https://doi.org/10.1038/s41398-021-01272-3.
78. Conomos, M.P., Laurie, C.A., Stilp, A.M., Gogarten, S.M., McHugh, C.P., Nelson, S.C., Sofer, T., Fernández-Rhodes, L., Justice, A.E., Graff, M., et al. (2016). Genetic diversity and association studies in US hispanic/latino populations: applications in the hispanic community Health study/study of latinos. Am. J. Hum. Genet. *98*, 165–184. https://doi.org/10.1016/J.AJHG.2015.12.001.
79. Spear, M.L., Diaz-Papkovich, A., Ziv, E., Yracheta, J.M., Gravel, S., Torgerson, D.G., and Hernandez, R.D. (2020). Recent shifts in the genomic ancestry of Mexican Americans may alter the genetic architecture of biomedical traits. Elife *9*, e56029. https://doi.org/10.7554/ELIFE.56029.
80. Dai, C.L., Vazifeh, M.M., Yeang, C.H., Tachet, R., Wells, R.S., Vilar, M.G., Daly, M.J., Ratti, C., and Martin, A.R. (2020). Population histories of the United States revealed through fine-scale migration and haplotype analysis. Am. J. Hum. Genet. *106*, 371–388. https://doi.org/10.1016/J.AJHG.2020.02.002.
81. Healy, M.E., Hill, D., Berwick, M., Edgar, H., Gross, J., and Hunley, K. (2017). Social-group identity and population substructure in admixed populations in New Mexico and Latin America. PLoS One. https://doi.org/10.1371/journal.pone.0185503.
82. Kizil, C., Sariya, S., Kim, Y.A., Rajabli, F., Martin, E., Reyes-Dumeyer, D., Vardarajan, B., Maldonado, A., Haines, J.L., Mayeux, R., et al. (2022). Admixture mapping of Alzheimer's disease in Caribbean Hispanics identifies a new locus on 22q13.1. Mol. Psychiatry *27*, 2813–2820. https://doi.org/10.1038/s41380-022-01526-6.
83. Horimoto, A.R.V.R., Xue, D., Thornton, T.A., and Blue, E.E. Admixture mapping reveals the association between native American ancestry at 3q13.11 and reduced risk of Alzheimer's disease in Caribbean Hispanics.Alzheimers Res. Ther. 10.1186/s13195-021-00866-9.