

Genetic Contributions to Alzheimer's Disease and Frontotemporal Dementia in Admixed Latin American Populations

Juliana Acosta-Urbe

University of California, Santa Barbara

Stefanie D. Piña Escudero

Global Brain Health Institute

J. Nicholas Cochran

HudsonAlpha Institute for Biotechnology

Jared W. Taylor

HudsonAlpha Institute for Biotechnology

Alejandra Castruita

University of California, San Francisco

Caroline Jonson

University of California, San Francisco

Erin A. Barinaga

HudsonAlpha Institute for Biotechnology

Kevin Roberts

HudsonAlpha Institute for Biotechnology

Alexandra R. Levine

University of California, San Francisco

Dawwod S. George

University of California, Santa Barbara

José Alberto AvilaFunes

Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán

María I. Behrens

University of Chile

Martin A. Bruno

Universidad Católica de Cuyo

Luis I. Brusco

National Scientific and Technical Research Council

Nilton Custodio

Instituto Peruano de Neurociencias

Claudia Duran-Aniotz

Adolfo Ibáñez University

Francisco Lopera

University of Antioquia

Diana L. Matallana

Pontificia Universidad Javeriana

Andrea Slachevsky

University of Chile

Leonel T. Takada

Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo

Lina M. Zapata-Restrepo

Fundación Valle del Lili

Dafne E. Durón-Reyes

Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán

Elisa de Paula França Resende

Universidade Federal de Minas Gerais

Nancy Gelvez

Instituto de Genética Humana

Luisa F. Gómez

Instituto de Genética Humana

Maria E. Godoy

Adolfo Ibáñez University

Marcelo A. Maito

Adolfo Ibáñez University

Shireen avandel

Global Brain Health Institute

Bruce L. Miller

Global Brain Health Institute

Mike A. Nalls

Data Tecnica International (United States)

Hampton Leonard

Data Tecnica International (United States)

Dan Vitale

Data Tecnica International (United States)

Sara Bandres-Ciga

Center for Alzheimer's and Related Dementias (CARD)

Mathew J. Koretsky

Data Tecnica International (United States)

Andrew B. Singleton

Center for Alzheimer's and Related Dementias (CARD)

Caroline B. Pantazis

Center for Alzheimer's and Related Dementias (CARD)

Victor Valcour

Global Brain Health Institute

Agustin Ibañez

Adolfo Ibáñez University

Kenneth S. Kosik

University of California, San Francisco

Jennifer S. Yokoyama

jennifer.yokoyama@ucsf.edu

Global Brain Health Institute

M. Beatriz Bistue

Universidad Católica de Cuyo

Article

Keywords:

Posted Date: May 6th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-5462510/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: Competing interest reported. M.A.N. owns stock from Character Bio Inc and Neuron23 Inc. K.S.K. and J.S.Y. are members of the scientific advisory board of the Epstein Family Alzheimer's Research Collaboration. K.S.K. is a member of the board of directors of the Tau Consortium and consults with Expansion Therapeutics and ADRx Pharma. J.S.Y. is Editor-in-Chief for npj Dementia. J.S.Y. was not involved in the journal's review of, or decisions related to, this manuscript.

Abstract

Latin America's diverse genetic landscape provides a unique opportunity to study Alzheimer's disease (AD) and frontotemporal dementia (FTD). The Multi-Partner Consortium to Expand Dementia Research in Latin America (ReDLat) recruited 2,162 participants with AD, FTD, or as healthy controls from six countries: Argentina, Brazil, Chile, Colombia, Mexico, and Peru. Participants underwent genomic sequencing and population structure analyses were conducted using Principal Component Analysis and ADMIXTURE. The study revealed a predominant mix of American, African, and European ancestries, with an additional East Asian component in Brazil. Variant curation identified 17 pathogenic variants, pathogenic *C9orf72* expansion, and 44 variants of uncertain significance. Seventy families showed autosomal dominant inheritance, with 48 affected by AD and 22 by FTD. This represents the first large-scale genetic study of AD and FTD in Latin America, highlighting the need to consider diverse ancestries, social determinants of health, and cultural factors when assessing genetic risk for neurodegenerative diseases.

Introduction

Latin America is home to a unique blend of Indigenous American, European, and African ancestries, thereby possessing a diverse admixed genetic landscape that emerged in the 1500s from the conquest of the Americas and the slave trade¹. Significant East Asian immigration that occurred during the early 20th century has further contributed to the continent's population diversity observed today^{2,3}. This admixture harbors a spectrum of novel genetic variants, including some that may modulate susceptibility to neurodegenerative diseases like Alzheimer's disease (AD) and frontotemporal dementia (FTD) or may result in a higher allelic frequency of known risk-conveying variants for neurodegeneration⁴. Studying these populations, with their complex genetic architecture, offers an invaluable resource for understanding neurodegenerative diseases.

Assessing admixed populations is particularly interesting because admixture introduces a rich heterogeneity of alleles, which can be crucial for understanding genetic risk. For instance, in the Colombian population, researchers from a single center in Medellín have identified 13 different pathogenic *PSEN1* variants from different ancestral backgrounds⁴. This contrasts with the nine independent variants described in a screening study from nine centers throughout the Iberian Peninsula⁵. The presumptive excess of *PSEN1* variants may have become fixed in a relatively small effective population due to the antimicrobial benefits of high beta-amyloid levels⁴. Additionally, in the Peruvian population, a recent genome-wide association study and functional analysis suggested that the *NFASC* gene, located on chromosome 1, is associated with AD. The *NFASC* locus showed significant contributions from both European and African ancestries⁶. This finding emphasizes the crucial role of relatively recent founder effects among diverse ancestral backgrounds in uncovering novel variation and understanding the complex genetic foundations of AD.

Due to cultural, religious, and historical factors, Latin American families have traditionally been large. Government policies during the 1960s and 1970s further encouraged population growth, resulting in an average family size of six children^{7,8}. Additionally, geographic and socioeconomic conditions often led to extended families residing in proximity^{7,9}. This unique demographic structure provides an exceptional opportunity to study founder effects in large families by tracing the lineage and impact of specific genetic variants across generations. For example, the *PSEN1* E280A (Glu280Ala) variant in Colombia can be traced back approximately 500 years to the time of the Spanish invasion. In the large families that descend from the original carrier, the variant became fixed and spread widely among numerous distant relatives. This historical and genetic tracing provides valuable insights into how genetic drift occurs over time in small populations, increasing the likelihood of identifying both risk and protective genetic variants^{10–12}.

The scientific community from the region recognized the immense potential of this approach and formed the Multi-Partner Consortium to Expand Dementia Research in Latin America (ReDLat)¹³. This consortium fosters collaboration among researchers, clinicians, and institutions across six Latin American countries and the United States to leverage the unique genetic diversity and demographic characteristics that influence AD and FTD in Latin American populations. ReDLat aims to highlight the unique genetic diversity and demographic features of these populations associated with neurodegenerative diseases of cognition¹⁴. Ultimately, the consortium's work seeks to refine diagnostic approaches, develop targeted therapeutic interventions, and significantly enhance our understanding of dementia across Latin America¹³.

This paper is the first genetic report from this consortium, focusing on AD and FTD in admixed Latin American participants, with a particular emphasis on families. Our research efforts aim to identify genetic variants associated with these neurodegenerative diseases in the region and provide insights into some of the clinical presentations observed within the studied families, thereby enhancing our understanding of AD and FTD across this vast and diverse population. By expanding the genomic dataset in the coming years, we aim to deepen our understanding of the genetic underpinnings and clinical presentations of these neurodegenerative diseases across diverse Latin American populations, enhancing the potential for targeted interventions and therapies.

Results

A. Characterization of the population

To investigate the genetic landscape of neurodegeneration in Latin America we recruited patients with mild to moderate AD or FTD, along with healthy controls from ten memory clinics across six Latin American countries: Argentina, Brazil, Chile, Colombia, Mexico, and Peru. Recruitment occurred in two phases due to the COVID-19 pandemic, which delayed the launch of ReDLat's prospective enrollment, originally planned for 2020. Despite the differences in timing, the inclusion criteria for both the retrospective and prospective cohorts were identical. As of August 17, 2024, a total of 5,834 participants had been enrolled in the study [Figure 1].

By the time of manuscript submission, PCR-free whole genome sequencing (WGS), whole exome sequencing (WES), and/or single nucleotide polymorphism (SNP) array genotyping had been performed in 2,254 participants from both cohorts. After thorough quality control, genomic data from 2,162 individuals were retained for analysis. The final dataset included 658 participants with SNP array data, of whom 174 also had WES; 1,495 participants with WGS; and 9 individuals with only WES [Figure 1]. All participants underwent medical and neuropsychological evaluation. Among those with high-quality genomic data, 999 were diagnosed with AD, 381 with FTD, and 755 were classified as cognitively healthy at the time of assessment. The dataset also included eight participants with mild cognitive impairment and 19 individuals with other neuropsychiatric diagnoses, such as Parkinson's disease, Lewy body dementia, atypical parkinsonism, neurodevelopmental disorders, cerebellar ataxia, brain tumor, cognitive impairment associated with non-brain cancer, vascular dementia, severe depressive disorder, bipolar disorder, obstructive sleep apnea, and chronic traumatic encephalopathy. These individuals were members of recruited families and had a relative enrolled in the study with AD- or FTD-related dementia.

As expected, a higher percentage of participants with FTD were under 65 years of age compared to those with AD (22.3% vs. 13.9%). Furthermore, the percentage of female participants was higher in the AD group (66.6%) compared to the FTD group (52.8%). Those in the AD group also showed a higher proportion of individuals who are heterozygous (41.5%) and homozygous (8.6%) for the *APOE* ϵ 4 allele. [Table 1] The distribution of *APOE* alleles of unrelated individuals varies per country as shown in Supplementary Table 1. Homozygous *APOE* ϵ 2 carriers were observed only in Colombia (0.2%) and Brazil (1%). Conversely, the highest numbers of *APOE* ϵ 4 alleles were found in Argentina and Colombia (39% in both), although Colombia has 1% more homozygous carriers than Argentina. It is worth noting that these numbers may not be fully representative due to the Colombian sample size being considerably larger. As larger numbers of samples from all regions are collected, allele frequency estimates will be further refined.

B. Genetic Ancestry

To assess genetic ancestry similarity among our samples, we initially generated a merged ReDLat dataset that included 2,153 participants with WGS or SNP array data that passed the concordance analysis. We then used WGS from the 1000 Genomes Project (1000GP) as reference populations to estimate the global ancestry of the participants, employing Principal Component Analysis (PCA) and ADMIXTURE software to estimate global ancestry. We used the 1000GP cohort to identify variants with allelic frequency >10% and in linkage equilibrium that were present in both the 1000GP and the merged ReDLat dataset, resulting in a total of 226,524 variants used for ancestry analyses.

PCA [Figure 2, and Supplementary Figure 1] reveals that the ReDLat dataset shows substantial overlap with the American populations (AMR) sampled by the 1000GP, which also included Colombian, Mexican, and Peruvian participants. Though a substantial number of participants overlap with European (EUR) cohorts, there were 21 individuals clustering with the East Asian (EAS) population, suggesting recent EAS descent subgroups within ReDLat.

When analyzing the data by country, participants from Peru, Mexico, Chile, and Argentina exhibit minimal variation along Principal Component 1 (PC1), which is associated with African ancestry. In contrast, there is greater variation along Principal Component 2 (PC2), which is associated with Amerindian ancestry. This variation is particularly pronounced among participants from Mexico and Peru, who have individuals with a majority of their ancestry being Amerindian. These findings suggest a predominant two-way ancestry pattern in these populations. There is clear overlap between Argentinian and Brazilian samples with the European (EUR) cohort; however, the Brazilian samples show significant variation along PC1, highlighting the African component of the sample. While most Brazilian samples are distributed primarily between African (AFR) and EUR populations, a small subset clusters with EAS, indicating a distinct ancestral subgroup within Brazil. Colombia displays a clear three-way admixture pattern, as evidenced by its wide distribution along both PC1 and PC2. Additionally, Peruvian and Mexican samples from ReDLat exhibit clear overlap with their 1000GP counterparts, while ReDLat Colombian samples showed greater diversity than those in 1000GP [Figure 2]. This increased diversity is likely due to ReDLat's broader sampling across multiple regions of Colombia. Overall, the PCA analysis confirms that the ReDLat cohort accurately represents the different historical admixture patterns previously described in the corresponding countries¹.

To calculate global ancestry proportions (Q-values), which are adjusted p-values accounting for multiple testing and controlling the false discovery rate, we projected the ReDLat samples onto the 1000GP dataset ADMIXTURE results at multiple clustering values. [Supplementary Figure 2]. At K=5, where K represents the number of ancestral populations in the clustering analysis, we observed a continental separation of ancestral origins and were able to differentiate the Amerindian component. [Figure 3] Peru is the only country where Amerindian ancestry exceeds European ancestry, followed by Mexico, where these two ancestries show similar distribution. In Argentina, Brazil, Colombia, and Chile, European ancestry is the most prevalent among the participants, with a median value of 86.9% (Mean value of 79.7%, Standard deviation 17.5) for Argentina and 84.2% (Mean value of 73.8%, Standard deviation 26.9) for Brazil. African ancestry is present in Colombia and Brazil at lower levels; we observe a continuum of this ancestry, with individuals having over 90% and 75% African descent, down to the mean levels for both countries (around 10%), suggesting ongoing admixture over generations. In contrast, this continuum is not observed in the EAS component of the Brazilian samples, suggesting a recent diaspora without intercontinental admixture. [Supplementary Figure 3]

C. Variant pathogenicity analysis

To identify Mendelian forms of neurodegeneration in our cohort, we analyzed data from 1,678 participants who had high-quality WGS or WES data to detect pathogenic variants. Following standard practices in complex systems analysis, we employed both “bottom-up” and “top-down” approaches.

Our bottom-up approach was a “gene-to-family” search, in which we initially assessed the genes most commonly associated with adult-onset neurodegeneration for pathogenic variants (see Methods). We identified a total of 17 pathogenic variants, a pathogenic *C9orf72* expansion, and 44 variants of uncertain significance (VUS). [Table 2 and Supplementary Table 2]. We note that the *C9orf72* expansions were identified using ExpansionHunter software from PCR-free WGS but were not explicitly confirmed by Southern blot, although this predictive tool is very accurate¹⁵.

In families with AD, we identified several previously described *PSEN1* variants: c.356C>T (p.Thr119Ile), c.415A>G (p.Met139Val) c.428T>C (p.Ile143Thr) and c.519G>T (p.Leu173Phe)^{4,16,17} [Supplementary Figure 4]. Notably, Thr119Ile and Ile143Thr are of European origin and exhibit identity by descent among Colombian carriers for Ile143Thr and between Colombian and Argentine carriers for Thr119Ile⁴. In contrast, the *PSEN1* c.415A>G (p.Met139Val) variant, identified in a large family from Argentina, was determined to be of Amerindian origin using the local ancestry inference software RFMix (see Methods). This haplotype was also present in two individuals from the 1000 Genomes Project, originating from Peru and Colombia [Supplementary Figure 5]. The *PSEN1* Met139Val variant has been reported across diverse ancestral backgrounds^{18–20} and is associated with either AD or atypical dementia, characterized by amnesic and behavioral symptoms. These symptoms may also include spastic paraparesis, psychosis, seizures, and myoclonus²¹. [See Supplementary Note 1 for the clinical description of this family].

In families with FTD, the gene most commonly associated with hereditary forms of the illness in our cohort was *GRN*, followed by *MAPT* [See Supplementary Notes 2 and 3 for the clinical description of these families]. Pathogenic *C9orf72* repeat expansions were observed in several families from geographically non-adjacent areas. The variants in *GRN*, *MAPT*, *C9orf72*, and *TARDBP* that were present in more than one individual were all of European descent [Table 3]. The pedigrees of families with pathogenic variants are shown in the Supplementary Figures 6–9.

After the initial analysis of these primary genes, we expanded our search to include secondary genes associated with adult-onset neurodegeneration. Utilizing the OMIM database, we identified a set of genes where single nucleotide variants or short insertions/deletions could be disease-causing [Supplementary Table 3]^{19,22–24}. In our analysis, we found four additional pathogenic variants. [Supplementary Table 4 and 5] Three of these variants are present in families with autosomal dominant disease and are located in the *PRNP* and *NOTCH3* genes. Most notably, we identified an FTD patient without motor symptoms carrying a pathogenic variant in *SOD1* c.388G>A (p.Phe21Leu) which was previously reported in another FTD patient from the same geographical region and is believed to have originated an Amerindian haplotype⁴.

In contrast, our top-to-bottom approach consisted of a “family-to-gene” search, where we analyzed the pedigrees of all recruited participants with WES or WGS data available (766 individuals grouped in 592 families). After excluding individuals recruited as “healthy”, we identified 426 independent families and classified them based on the presence of affected individuals [Figure 4]. In this cohort, 70 families exhibited autosomal dominant inheritance of neurodegenerative diseases, as evidenced by the presence of three affected individuals in two consecutive generations. The families were later classified according to the diagnosis of the proband, with 48 families identified as having AD and 22 as FTD. These families were subsequently categorized based on the age at disease onset: ‘late onset’ was assigned to families where all affected members presented dementia at ages older than 65 years; ‘early onset’ applied to those where all affected individuals were 65 years or younger at the time of dementia onset; and ‘mixed onset’ described families that included members with both early and late-onset

disease [Supplementary Table 6]. Though many of the carriers of pathogenic variants belonged to the ReDLat retrospective cohort, 14 of the recruited families were carriers of pathogenic variants, and the majority had a positive family history of neurodegeneration.

In a further analysis of families exhibiting autosomal dominant patterns, we assessed for the presence of at least one allele of *APOE* ϵ 4. Among the families diagnosed with AD, nine were negative for *APOE* ϵ 4 alleles, while 32 had had at least an *APOE* ϵ 4 carrier. We were able to determine the *APOE* ϵ 4 allele status for more than one participant in seven families, and only four families demonstrated a co-segregation of the *APOE* ϵ 4 allele with the illness. In contrast, when assessing the families with FTD, six were negative for *APOE* ϵ 4 alleles, and in the one family where we could determine the *APOE* ϵ 4 allele status for more than one participant, no co-segregation pattern was observed. [Figure 4]

Discussion

This initial release of genomic data from the ReDLat cohort provides early insights into the genetic underpinnings of neurodegeneration within a Latin American population, supported by genomic analyses of established variants associated with AD and FTD. Our genetic ancestry analysis, leveraging data from the 1000 Genomes Project, revealed tricontinental admixture patterns across most regions and an East Asian component in Brazil, reflecting historical migration and admixture events. The study notably identifies a significant prevalence of autosomal dominant inheritance patterns in AD and FTD, characterized by distinct age-of-onset categorizations, geographic distribution of genetic variants, and a stronger presence of the *APOE* ϵ 4 allele in AD families. These patterns also include newly discovered variants in the *PSEN1* and *APP* genes for AD, which play critical roles in the disease's pathogenesis and the recurrence of a *SOD1* variant presenting as FTD without motor symptoms, suggesting a novel disease phenotype association. Families with as-yet unidentified variation remain strong candidates for future novel gene discovery as additional family members are recruited for gene-mapping linkage studies.

Indeed, there is considerable potential for novel genetic discovery in diverse cohorts such as ReDLat, both in terms of risk for AD and related dementias (ADRD) and resilience against it, in both families and sporadic cases. Previous work in the region has unveiled more than 13 *PSEN1* pathogenic variants in Colombia, including the E280A kindred that spans more than 5000 descendants of a founder couple¹¹. Leveraging larger, diverse cohorts—as well as genetic families with substantial clinical heterogeneity—represents a unique opportunity for the discovery of resilience factors for ADRD, which may serve as strong targets for disease intervention.

As previously noted by Browning et. al. (2018)²⁵, historical factors like colonization, migration, and bottlenecks have significantly shaped the genetic landscape of Latin American populations. During and after the period of colonization, many Latin Americans lived in small, often isolated villages. This created a population structure characterized by multiple mini-bottlenecks as descendants of a small number of founders in each village tended to remain in the same location for many generations. As families expanded, the specific rare alleles in each place became common, even surpassing the allelic frequencies of the same variant in the ancestral population, as is the case for the *SQSTM1* FTD risk-conferring variant Pro392Leu (rs104893941)^{26–28}. This phenomenon resulted in a genetic map of the region that closely corresponds to the geographic map, where multiple individuals share the same deleterious variants and are identical by descent. The long stretches of identical haplotypes created by the bottlenecks and increased ancestral diversity within isolated populations are advantageous for researchers seeking to identify rare variants associated with diseases like AD and FTD or the interaction between genetic variation and ancestral haplotypes²⁹. These factors underscore the value of family studies in Latin America, offering unique insights into genetic patterns and the potential for discovering new genetic contributions to disease²⁵.

The first-wave study cohort reported here has several limitations. First, we have chosen not to conduct unbiased discovery efforts, such as genome-wide association studies and burden analyses, in this cohort due to the extensive family structure and relatively small sample size of the cross-sectional cohort collected to date. Second, despite the tangible advancement in global representation offered by this cohort, participants are still enriched for higher socioeconomic status due to the urban-centric recruitment. This drawback is being actively addressed through ongoing enrollments and community outreach efforts in more rural areas. Third, as with any clinic-based enrollment cohort, there is a possibility of ascertainment bias among recruited participants because the study recruits from clinical practices specializing in cognitive disorders, which may lead to an overrepresentation of more extreme clinical phenotypes. The findings to date represent significant advances in understanding the etiology of Alzheimer's and Frontotemporal dementia in this region. Continued enrollment in this project will provide additional valuable insights through future studies that map the genetic underpinnings of disease risk in large families, genetic risk burden in cases, and offer well-powered cohorts for case-control studies to identify common risk variants. Moreover, the robust family structure already observed in ReDLat provides a unique opportunity to map genetic modifiers and assess the impact of local genomic ancestry. As global population representation continues to expand, it will be critical to evaluate the generalizability of genetic risk factors for AD and FTD across diverse ancestral backgrounds, within the context of distinct social determinants of health, and accounting for modifiable risk factors that may influence disease risk and resilience across distinct cultures.

Methods

1. Participant recruitment

Clinical diagnosis was determined by site investigators through consensus conferences at each site, adhering to the current diagnostic criteria for AD and FTD^{30–32}. Healthy controls were recruited at the same locations meeting the following criteria: Clinical Dementia Rating (CDR)³³ of 0, a Mini-Mental State Examination (MMSE)³⁴ score greater than 25, or having been evaluated by a neuropsychologist who confirmed normal cognition in participants with few years of formal education. Family members of participants with AD or FTD, aged 18 years or older, were included if there were two or more individuals with neurodegenerative illnesses in the family, or were related to a study participant with a known dementia-associated genetic mutation and having undergone genetic counseling. All participants (diagnosed patients, healthy controls, and family members) demonstrated minimum fluency in the language of assessment (Spanish or Portuguese), had adequate vision and hearing for cognitive testing as determined by the investigator, and were required to have a study partner (informant) with at least six months of knowledge about their daily activities and cognitive/functional status. All participants (prospective and retrospective) had to be capable of providing informed consent or have a legally authorized representative.

Written informed consent was obtained from all participants or their legally authorized proxies for all evaluations and assessments conducted, following a detailed explanation of the procedures, associated risks, and potential benefits. This process adhered to the ethical guidelines of each participating country, the Code of Ethics of the World Medical Association, the Declaration of Helsinki, and the Belmont Report. Assent was also secured from participants themselves, ensuring their willingness to participate. The consent process included explicit permission to publish the findings. The study and informed consent procedures were approved by the Institutional Review Board of each participating medical institution.

Participating institutions and their respective Federalwide Assurance (FWA) numbers included the following: Argentina – INECO-Centro de Psicología Médica San Martín de Tours (FWA00028264); Brazil – Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (FWA00001035); Chile – Hospital Clínico Universidad de Chile (FWA00029089) and Universidad Adolfo Ibáñez (FWA00030846); Colombia – Comité de Bioética del Instituto de Investigaciones Médicas, Facultad de Medicina, Universidad de Antioquia (FWA00028864), Pontificia Universidad Javeriana - Hospital Universitario San Ignacio (FWA00001113), and Fundación Valle de Lili (FWA00029865); Mexico – Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán (FWA00014416); Peru – Hospital Nacional Docente Madre Niño San Bartolomé (FWA00010121); and the United States – University of California, San Francisco – Memory and Aging Center (FWA00000068).

2. Clinical characterization:

As part of the recruitment process, participants were interviewed about their family history of neurodegeneration. Information was collected via self-report from both patients and their study partners using the Genetic Pedigree Software-Progeny®³⁵. A positive family history was defined as having at least one first- or second-degree relative with dementia or another neurodegenerative disorder. Families with three or more affected individuals in two consecutive generations were then labeled as ‘strong family aggregation’. Medical history and a full neuropsychological examination were conducted as described in Ibañez et. al.(2021)¹³ Retrospective participants were assessed based on a re-evaluation of the available clinical data for each individual. The cognitive tests were harmonized as described in Maito et.al (2023)³⁶.

3. Genetic sequencing

Standardized phlebotomy with EDTA tubes was used for sample collection. Genomic DNA was extracted using Wizard® Genomic DNA Purification Kit (Promega), QIAamp® DNA Mini Kit (Qiagen), or similar salting-out methods. Samples were shipped quarterly from the various participating sites in Latin America to the United States. HudsonAlpha Institute for Biotechnology (Alabama, U.S.A) performed Single Nucleotide Polymorphism (SNP) Arrays, whole exome sequencing (WES), and/or whole genome sequencing (WGS) of the samples. Additional whole genome sequencing was performed at Psomagen, Inc. (Maryland, U.S.A)

Variants were genotyped using the NeuroBooster array from Illumina, designed to capture variants relevant to neurological conditions³⁷. Quality control (QC) of the SNP Array data was conducted using Genotools v1 default settings³⁸. Prior to imputation, the QC'ed output files from Genotools were processed with the *no_qc_imputation_prep.sh* script. This script ran the datasets through the Wrayner script to compare them against all TOPMed freeze 8 variants. Excluded variants were then flipped to rescue additional variants, after which the dataset was processed again through the Wrayner script to compare against PASS TOPMed freeze 8 variants. Data was subsequently imputed using the TOPMed Imputation Panel and Server v1.3.3, following a previously developed pipeline for multi ancestral sample sets as described in Vitale et. al. (2024)³⁸.

For WES, DNA was processed using Integrated DNA Technologies xGen Exome Hyb Panel v2, and sequenced on the NovaSeq 6000 platform using paired-end 100-base pair reads to a target depth of 100X.

WGS was performed at two institutions; Samples at Psomagen were prepared using the TruSeq DNA PCR-Free library prep method to avoid PCR amplification bias. Samples at HudsonAlpha (64 of the pass-QC genomes) underwent a custom PCR-free preparation involving Covaris shearing (fragmenting the DNA), end repair (preparing the DNA fragments for ligation), and adapter ligation, all without PCR amplification. All libraries from both sites were then normalized using KAPA qPCR and sequenced on the Illumina NovaSeq 6000 platform to a target depth of 30X. The sequencing was paired-end with a read length of 150 bp (Illumina 150bpPE).

4. Genetic data processing

The raw sequence data (fastq files) were aligned to the hg38 reference genome using the Sentieon v202112.05 implementation of the BWA MEM algorithm at HudsonAlpha. Sentieon v202112.05 utilities were used to sort the reads, mark duplicate sequences, and recalibrate the base quality scores. Variant calling, which identifies differences between the sample DNA and the reference genome, was performed using GATK4 tools implemented by Sentieon v202112.05. This step was conducted across all samples in a batch for exomes and one for genomes to maintain consistency. Finally, variant quality score recalibration (VQSR) was applied to filter out false positive variant calls, ensuring high-quality data. This comprehensive approach achieved an average recall rate of 99.22% when compared to the Genome in a Bottle high confidence truth sets, indicating a high level of accuracy in detecting genetic variants.

Variant Call Format (VCF) files were filtered according to established criteria to ensure high-quality data. For whole genome and exome sequences variants with genotype quality greater than 20 and read depth scores above 10 were retained. The filtered VCF was then annotated with gene names, variant types, and amino acid changes for all exonic variants using GRCh38.99 with SnpEff, dbSNP release 156, CADD 1.6, TOPMed Bravo Freeze 8 allele frequencies, and ClinVar through BCFtools and Annovar^{39–43}. Variants with genotyping rates below 95% by individual and 95% by variant were removed. Chromosomal sex was further validated via genetic data by splitting the pseudoautosomal regions of the X chromosome and analyzing the heterozygosity of X-chromosome, as well as the count of variants present on the Y chromosome. A detailed pipeline and scripts are available at <https://github.com/TauConsortium/redlat-genetics>

To combine the arrays and WGS data, all variant IDs in both datasets were first annotated with “chrom:pos:ref:alt”. The VCFs were then intersected using BCFtools v1.9 isec, producing a list of intersecting variants between the two datasets. Both VCFs were filtered to include only these intersecting variants and then merged using BCFtools v1.9 merge⁴⁵. After merging, the VCF was annotated with dbSNP 156⁴⁶.

To ensure concordance between the imputed arrays and WGS, concordance was checked for individuals with data from both methods. The imputed array data was filtered for varying allele frequencies (AF) using BCFtools v1.9⁴⁵. Concordance was assessed for each AF-filtered VCF using SnpSift concordance. The final concordance for each individual was determined by dividing the number of correct variant calls by the total number of intersecting variant calls. The complete script for the concordance check is available at the project's GitHub repository [see “Code Availability”]. Samples with concordance below 0.95 at sites with AF less than 0.0001 were excluded.

5. Genetic Relatedness

Family history was documented through elaboration of detailed pedigrees for each recruited participant. Disclosed relatedness was compared to expected genetic relatedness using KING⁴⁴. Individuals with cryptic relatedness (kinship coefficient <0.125 without a familial relationship documented on the pedigrees) or discrepancies between disclosed and genetic relatedness were removed.

6. Population stratification

To capture the ancestral diversity of the cohort and represent all participating countries, we merged the WGS data with imputed SNP array data for individuals lacking WGS data. The ReDLat dataset was then combined with high-depth WGS data from the 1000 Genomes Project (1000GP)⁴⁷ and filtered to include only biallelic variants with a genotyping rate of >95%. The resulting ReDLat-1000GP dataset was used for the following analyses:

Principal component analysis (PCA) was conducted using the smartpca package from EIGENSOFT (version 8.0.0)⁴⁸. PCA was performed on samples from 1000GP, and the ReDLat samples were subsequently projected onto the principal components.

Global ancestry was estimated using ADMIXTURE (version 1.3)⁴⁹. We performed an unsupervised ancestry analysis on the 1000GP data, modeling ancestry from two to eight populations (K). The ancestry fractions for the ReDLat samples were calculated using the allelic ancestry proportions derived from the 1000GP analysis.

7. Variant pathogenicity analysis

Data from WGS and WES was merged for a joint analysis for pathogenic variation. We used the Online Mendelian In Men (OMIM) database to search for genes associated with autosomal dominant, autosomal recessive, or X-linked forms of adult-onset dementia⁵⁰.

We manually curated protein-altering variants in the ten genes most commonly associated with adult-onset neurodegeneration: *APP*, *CHMP2B*, *FUS*, *GRN*, *MAPT*, *PSEN1*, *PSEN2*, *TARDBP*, *TBK1*, and *VCP*, as well as expansions in *C9orf72* (identified using ExpansionHunter v5.0.0)^{51,52}. Variants located in introns, the 3' untranslated region (3' UTR), the 5' untranslated region (5' UTR), and synonymous variants within exons were included if their in silico splice-predicting scores (dbSNV_RF_SCORE and dbSNV_ADA_SCORE) were both greater than 0.6, since these variants were considered likely to have an impact on splicing, making them relevant for further study⁵³. Exonic non-synonymous variants (missense, nonsense, and frameshift) were analyzed following guidelines from the American College of Medical Genetics and Genomics (ACMG)⁵² and the Guerreiro algorithm for *PSEN1* and *PSEN2* genes⁵. The variants identified in the remaining genes listed in Supplementary Table 3 were queried in ClinVar⁵⁴ databases and reported if previously classified as pathogenic or likely pathogenic. Variants identified through this process were then tested by hand for familial segregation to confirm their association with the disease within families.

8. Local ancestry of pathogenic variants

To determine the haplotypic origin of pathogenic variants, we first phased the ReDLat WGS using SHAPEIT5 (v5.1.1)⁵⁵. Pedigree information and 1000GP phased genomes were included as family and population references to improve phase accuracy. Since heterozygous variants in a single individual cannot be reliably assigned to a specific haplotype, we restricted our assessment of local ancestry to pathogenic variants present in two or more individuals.

After phasing, we estimated the local ancestry of pathogenic variants using RFMix (v2.03-r0)⁵⁶. We constructed the reference population panel by merging high-coverage sequence data from the Human Genome Diversity Project (HGDP)⁵⁷ and 1000GP. Due to the absence of Amerindian individuals in the 1000GP dataset, we used admixed American samples from 1000GP that were over 99.9% Amerindian in the ADMIXTURE analysis at K=3 as Amerindian reference samples. We then extracted representative individuals of African, Amerindian, European, and East Asian ancestry to build reference cohorts of similar size (100–150 individuals). Finally, we divided the ReDLat WGS into subgroups of individuals with similar global ancestry based on their ADMIXTURE results and ran RFMix with the following settings: terminal node size of five, five expectation-maximization iterations, and both with and without the `--reanalyze-reference` option.

Declarations

Data Availability

Data used in this publication has been uploaded to the FAIR Data Portal within the AD Workbench, which is supported by the Alzheimer's Disease Data Initiative, and can be accessed at <https://www.alzheimersdata.org/ad-workbench>.

Code Availability

All the scripts used for the study have been compiled into a single GitHub repository: <https://github.com/TauConsortium/redlat-genetics>

Acknowledgements

We dedicate this manuscript to Professor Francisco Lopera, a pioneer in the study of neurodegeneration and Alzheimer's disease in Latin America. We thank the individuals and the families who participated in this study. We are grateful to Dr. Sharon R. Browning for her assistance with determining local ancestry and to all the researchers in the participating centers, whose collaboration made ReDLat possible.

The Multi-Partner Consortium to Expand Dementia Research in Latin America (ReDLat) was supported in part by the National Institutes of Health; Intramural Research Program of the NIH; National Institute on Aging (NIA) [R01 AG057234, R01 AG075775, R01 AG021051, R01 AG083799]; NIH Department of Health and Human Services [Z01 AG000534 and AG000548]; National Institute of Neurological Disorders and Stroke (NINDS); Fogarty International Center; Alzheimer's Association (SG-20-725707); Rainwater Charitable foundation–Tau Consortium; Bluefield Project to Cure Frontotemporal Dementia; and the Global Brain Health Institute.

In addition to the grants reported above, A.S. is supported by ANID/FONDAP [15150012] and ANID/FONDECYT [Regular 1231839]. C.D.A. is supported by ANID/FONDECYT [Regular 1210622] and ANID/PIA/ANILLOS [ACT210096]. D.L.M. received funding from Bluefield Project to Cure Frontotemporal Dementia and Alector. A.I. is supported by ReDLat grants mentioned above and by ANID/FONDECYT Regular [1210195, 1210176 and 1220995]; ANID/FONDAP [15150012]; ANID/PIA/ANILLOS [ACT210096]; and FONDEF [ID20I10152 and ID22I10029]. J.SY. receives funding from NIH-NIA [R01AG062588, P30AG062422, P01AG019724, U19AG079774]; NIH-NINDS [U54NS123985]; NIH-NIDA [75N95022C00031]; AFTD Susan Marcus Memorial Fund; Larry L. Hillblom Foundation; Bluefield Project to Cure Frontotemporal Dementia; French Foundation; Mary Oakley Foundation; Alector and Transposon Therapeutics.

This work utilized the computational resources of the NIH STRIDES Initiative (<https://cloud.nih.gov>) through the Other Transaction agreement [Azure: OT2OD032100, Google Cloud Platform: OT2OD027060, Amazon Web Services: OT2OD027852] and the NIH high performance computing Biowulf cluster (<https://hpc.nih.gov>). M.A.N., H.L., C.J., D.V. and M.J.K.'s participation in this project was part of a competitive contract awarded to DataTecnica LLC by the National Institutes of Health to support open science research.

The contents of this publication are solely the responsibility of the authors and do not represent the official views of these institutions. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Authors' contributions

J.A-U, S.D.P-E, J.S.Y, and K.S.K. conceived and designed the project, with contributions from J.N.C., J.W.T., P.A.C., C.J., D.L.M., and M.A.N. A.R.L. and D.S.G. managed pedigree information and conducted the heritability assessment. E.A.B. handled sample processing, while K.R. generated SNP-array data at HudsonAlpha.

J.A-F, M.I.B., M.A.B., L.I.B., N.C., C.D-A., F.L., A.S., L.T.T., L.M.Z-R, D.E.D-R., E.P.F.R., L.F.G., N.G., M.B.B., M.E.G., M.A.M., S.J., B.L.M., C.B.P., A.B.S., V.V., and A.I. contributed to participant recruitment, evaluation, sample acquisition, genomic data generation, as well as project management and coordination. H.L., D.V., S.B-C., and M.J.K. assisted with genetic data analyses.

F.L. passed away on September 2024; although he was an active collaborator in data acquisition, analyses, and interpretation, he was not able to review the final version of the manuscript. All remaining authors read and approved the final manuscript.

J.A-U, S.D.P-E, contributed equally to this work. J.S.Y, and K.S.K. contributed equally to this work.

Competing interests

M.A.N. owns stock from Character Bio Inc and Neuron23 Inc. K.S.K. and J.S.Y. are members of the scientific advisory board of the Epstein Family Alzheimer's Research Collaboration. K.S.K. is a member of the board of directors of the Tau Consortium and consults with Expansion Therapeutics and ADRx Pharma. J.S.Y. is Editor-in-Chief for *npj Dementia*. J.S.Y. was not involved in the journal's review of, or decisions related to, this manuscript.

Author Information

Author Information for Multi-Partner Consortium to Expand Dementia Research in Latin America (ReDLat).

Juliana Acosta-Urbe^{1,2}

Stefanie D. Piña-Escudero^{3,4}

J. Nicholas Cochran⁵

Jared W. Taylor⁵

P. Alejandra Castruita⁴

Caroline Jonson^{4,6,7}, Erin A. Barinaga⁵

Kevin Roberts⁵

Alexandra R. Levine²

Dawwod S. George²

José Alberto Ávila-Funes^{8,9}

María I. Behrens^{10,11}

Martín A. Bruno^{12,13}

Luis I. Brusco^{14,15}

Nilton Custodio¹⁶

Claudia Duran-Aniotz¹⁷

Francisco Lopera^{1†}

Diana L. Matallana^{18,19,20}

Andrea Slachevsky^{10,21,22,23}

Leonel T. Takada²⁴

Lina M. Zapata-Restrepo^{25,26}

Dafne E. Durón-Reyes⁸

Elisa de Paula França Resende^{3,27}

Luisa F. Gómez²⁸

Nancy Gelvez²⁸

M. Beatriz Bistue^{12,13}

Maria E. Godoy¹⁷

Marcelo A. Maito¹⁷

Shireen Javandel^{3,4}

Bruce L. Miller^{3,4}

Mike A. Nalls^{6,7}

Hampton Leonard^{6,7}

Dan Vitale^{6,7}

Sara Bandres-Ciga⁷

Mathew J. Koretsky^{6,7}

Andrew B. Singleton^{7,29}

Caroline B. Pantazis⁷

Victor Valcour^{3,4,30}

Agustin Ibañez^{17,31}

Kenneth S. Kosik²

Jennifer S. Yokoyama^{3,4,32}

The Multi-Partner Consortium to Expand Dementia Research in Latin America (ReDLat).

References

1. Adhikari, K., Chacón-Duque, J. C., Mendoza-Revilla, J., Fuentes-Guajardo, M. & Ruiz-Linares, A. The Genetic Diversity of the Americas. *Annu Rev Genomics Hum Genet* **18**, 277–296 (2017).
2. Kunitomo, I. Japanese Migration to Latin America. in *Japan, the United States, and Latin America: Toward a Trilateral Relationship in the Western Hemisphere?* (eds. Stallings, B. & Székely, G.) 99–121 (Palgrave Macmillan UK, London, 1993). doi:10.1007/978-1-349-13128-

3. Hassan, W. S. Arabs and the Americas: A Multilingual and Multigenerational Legacy. *Review: Literature and Arts of the Americas* **52**, 166–169 (2019).
4. Acosta-Urbe, J. *et al.* A neurodegenerative disease landscape of rare mutations in Colombia due to founder effects. *Genome Medicine* **14**, 27 (2022).
5. Guerreiro, R. J. *et al.* Genetic screening of Alzheimer's disease genes in Iberian and African samples yields novel mutations in presenilins and APP. *Neurobiol Aging* **31**, 725–731 (2010).
6. Akgun, B. *et al.* A genome-wide association study in Peruvians suggests new risk loci for Alzheimer Disease. 2023.11.29.23299201 Preprint at <https://doi.org/10.1101/2023.11.29.23299201> (2023).
7. Arriagada, I. *Familias latinoamericanas: diagnóstico y políticas públicas en los inicios del nuevo siglo*. (Naciones Unidas, CEPAL, Div. de Desarrollo Social, Santiago de Chile, 2001).
8. Van Bavel, J. The world population explosion: causes, backgrounds and projections for the future. *Facts Views Vis Obgyn* **5**, 281–291 (2013).
9. Arriagada, I. Familias latinoamericanas: cambiantes, diversas y desiguales. *Papeles de población* **13**, 9–22 (2007).
10. Acosta-Baena, N. *et al.* Pre-dementia clinical stages in presenilin 1 E280A familial early-onset Alzheimer's disease: a retrospective cohort study. *Lancet Neurol* **10**, 213–220 (2011).
11. Lalli, M. A. *et al.* Origin of the PSEN1 E280A mutation causing early-onset Alzheimer's disease. *Alzheimers Dement* **10**, S277-S283.e10 (2014).
12. Cochran, J. N. *et al.* Genetic associations with age at dementia onset in the PSEN1 E280A Colombian kindred. *Alzheimers Dement* **19**, 3835–3847 (2023).
13. Ibanez, A. *et al.* The Multi-Partner Consortium to Expand Dementia Research in Latin America (ReDLat): Driving Multicentric Research and Implementation Science. *Front Neurol* **12**, 631722 (2021).
14. Moguilner, S. *et al.* Brain clocks capture diversity and disparities in aging and dementia across geographically diverse populations. *Nat Med* 1–12 (2024) doi:10.1038/s41591-024-03209-x.
15. Dolzhenko, E. *et al.* Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Research* **27**, 1895–1903 (2017).
16. Arango, D. *et al.* Systematic genetic study of Alzheimer disease in Latin America: Mutation frequencies of the amyloid β precursor protein and presenilin genes in Colombia. *American Journal of Medical Genetics* **103**, 138–143 (2001).
17. Itzcovich, T. *et al.* A novel mutation in PSEN1 (p.T119I) in an Argentine family with early- and late-onset Alzheimer's disease. *Neurobiology of Aging* **85**, 155.e9-155.e12 (2020).
18. Alzheimer's Disease Collaborative Group. The structure of the presenilin 1 (S182) gene and identification of six novel mutations in early onset AD families. *Nat Genet* **11**, 219–222 (1995).
19. Rippon, G. A. *et al.* Presenilin 1 Mutation in an African American Family Presenting With Atypical Alzheimer Dementia. *Archives of Neurology* **60**, 884–888 (2003).
20. Jia, L. *et al.* PSEN1, PSEN2, and APP mutations in 404 Chinese pedigrees with familial Alzheimer's disease. *Alzheimer's & Dementia* **16**, 178–191 (2020).
21. Ryan, N. S. *et al.* Clinical phenotype and genetic associations in autosomal dominant familial Alzheimer's disease: a case series. *The Lancet Neurology* **15**, 1326–1335 (2016).
22. Clark, R. F. *et al.* The structure of the presenilin 1 (S182) gene and identification of six novel mutations in early onset AD families. *Nat Genet* **11**, 219–222 (1995).
23. Sandbrink, R. *et al.* Missense mutations of the PS-1/S182 gene in german early-onset Alzheimer's disease patients. *Annals of Neurology* **40**, 265–266 (1996).
24. Islam, S. *et al.* Presenilin Is Essential for ApoE Secretion, a Novel Role of Presenilin Involved in Alzheimer's Disease Pathogenesis. *J. Neurosci.* **42**, 1574–1586 (2022).
25. Browning, S. R. *et al.* Ancestry-specific recent effective population size in the Americas. *PLOS Genetics* **14**, e1007385 (2018).
26. Karczewski, K. J. *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* 531210 (2019) doi:10.1101/531210.
27. van der Zee, J. *et al.* Rare mutations in SQSTM1 modify susceptibility to frontotemporal lobar degeneration. *Acta Neuropathol* **128**, 397–410 (2014).
28. gnomAD. Variant: 5-179836445-C-T. gnomAD (2024).

29. Blue, E. E., Horimoto, A. R. V. R., Mukherjee, S., Wijsman, E. M. & Thornton, T. A. Local ancestry at APOE modifies Alzheimer's disease risk in Caribbean Hispanics. *Alzheimers Dement* **15**, 1524–1532 (2019).
30. Jack Jr., C. R. *et al.* Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. *Alzheimer's & Dementia* **n/a**.
31. Rascofsky, K. *et al.* Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain* **134**, 2456–2477 (2011).
32. Gorno-Tempini, M. L. *et al.* Classification of primary progressive aphasia and its variants. *Neurology* **76**, 1006–1014 (2011).
33. Arnold, B. R., Cuellar, I. & Guzman, N. Statistical and clinical evaluation of the Mattis Dementia Rating Scale-Spanish adaptation: an initial investigation. *J Gerontol B Psychol Sci Soc Sci* **53**, P364-369 (1998).
34. Folstein, M. F., Folstein, S. E. & McHugh, P. R. 'Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189–198 (1975).
35. Progeny Genetics, L. *Pedigrees Were Constructed and Drawn Using the Genetic Data Management System, Progeny Clinical – Version N from Progeny Genetics*. (Copyright 2019. Reprinted with permission of Progeny Genetics LLC, Delray Beach, FL.).
36. Maito, M. A. *et al.* Classification of Alzheimer's disease and frontotemporal dementia using routine clinical and cognitive measures across multicentric underrepresented samples: A cross sectional observational study. *Lancet Reg Health Am* **17**, 100387 (2023).
37. Bandres-Ciga, S. *et al.* NeuroBooster Array: A Genome-Wide Genotyping Platform to Study Neurological Disorders Across Diverse Populations. *medRxiv* doi:10.1101/2023.11.06.23298176.
38. Vitale, D. *et al.* GenoTools: An Open-Source Python Package for Efficient Genotype Data Quality Control and Analysis. 2024.03.26.586362 Preprint at <https://doi.org/10.1101/2024.03.26.586362> (2024).
39. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
40. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
41. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
42. ClinVar. <https://www.ncbi.nlm.nih.gov/clinvar/>.
43. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research* **38**, e164 (2010).
44. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
45. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
46. Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308–311 (2001).
47. Byrsk-Bishop, M. *et al.* High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. *Cell* **185**, 3426-3440.e19 (2022).
48. Patterson, N., Price, A. L. & Reich, D. Population Structure and Eigenanalysis. *PLOS Genetics* **2**, e190 (2006).
49. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* **19**, 1655–1664 (2009).
50. Home - OMIM. <https://www.omim.org/>.
51. Dolzhenko, E. *et al.* ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. *Bioinformatics* **35**, 4754–4756 (2019).
52. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405–423 (2015).
53. Li, Q. & Wang, K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *Am J Hum Genet* **100**, 267–280 (2017).
54. ALZFORUM | NETWORKING FOR A CURE. <https://www.alzforum.org/>.
55. Hofmeister, R. J., Ribeiro, D. M., Rubinacci, S. & Delaneau, O. Accurate rare variant phasing of whole-genome and whole-exome sequencing data in the UK Biobank. *Nat Genet* **55**, 1243–1249 (2023).
56. Maples, B. K., Gravel, S., Kenny, E. E. & Bustamante, C. D. RFMix: A discriminative modeling approach for rapid and robust local-ancestry inference. *American Journal of Human Genetics* **93**, 278–288 (2013).
57. Bergström, A. *et al.* Insights into human genetic variation and population history from 929 diverse genomes. *Science (New York, N.Y.)* **367**, eaay5012 (2020).

Tables

Table 1. Clinical characteristics of included ReDLat participants

		AD n=999	FTD n=381	MCI n=8	OTHER n=19	HC n=755
Age<65years, n (%)		139 (13.9)	85 (22.3)	2 (25.0)	2 (10.5)	277 (36.7)
Female, n (%)		665 (66.6)	201 (52.8)	6 (75.0)	8 (42.1)	516 (68.3)
AD Phenotype, n (%)	Amnestic	521 (52.2)	-	-	-	-
	lvPPA	10 (1.0)	-	-	-	-
	PCA	6 (0.6)	-	-	-	-
	fvAD	1 (0.1)	-	-	-	-
	usAD	461 (46.1)	-	-	-	-
FTD Phenotype, n (%)	bvFTD	-	196 (51.4)	-	-	-
	nfPPA	-	25 (6.6)	-	-	-
	svPPA	-	37 (9.7)	-	-	-
	FTD-CBS	-	26 (6.8)	-	-	-
	FTD-MND	-	9 (2.4)	-	-	-
	FTD-PSP	-	9 (2.4)	-	-	-
	usFTD	-	72 (18.9)	-	-	-
	usPPA	-	7 (1.8)	-	-	-
APOEε4 carriers, n (%)	Heterozygous	408 (41.5)	96 (25.9)	1 (14.3)	2 (10.5)	153 (20.7)
	Homozygous	85 (8.6)	15 (4.0)	-	1 (5.3)	13 (1.8)
Country of Origin	Argentina	191 (19.1)	2 (0.5)	-	1 (5.3)	74 (9.8)
	Brazil	42 (4.2)	99 (26.0)	-	-	88 (11.7)
	Chile	115 (11.5)	45 (11.8)	6 (75.0)	11 (57.9)	115 (15.2)
	Colombia	536 (53.7)	193 (50.7)	1 (12.5)	6 (31.6)	289 (38.3)
	Mexico	59 (5.9)	17 (4.5)	-	-	133 (17.6)
	Peru	56 (5.6)	25 (6.6)	1 (12.5)	1 (5.3)	56 (7.4)
Genomic Data, n (%)	SNP Array	282 (28.2)	126 (33.1)	6 (75.0)	10 (52.6)	234 (31.0)
	WES	6 (0.6)	-	-	-	3 (0.4)
	WGS	711 (71.2)	255 (66.9)	2 (25.0)	9 (47.4)	518 (68.6)

AD: Alzheimer's disease, FTD: frontotemporal dementia, MCI: mild cognitive impairment, HC: healthy controls, n: number, lvPPA: logopenic variant primary progressive aphasia, PCA: posterior cortical atrophy, fvAD: frontal variant AD, usAD: unspecified Alzheimer's disease, bvFTD: behavioral variant frontotemporal dementia, nfPPA: nonfluent variant primary progressive aphasia, svPPA: semantic variant primary progressive aphasia, FTD-CBS: frontotemporal dementia-corticobasal syndrome, FTD-MND: frontotemporal dementia-motor neuron disease, FTD-PSP: frontotemporal dementia-progressive supranuclear paralysis, usFTD: unspecified frontotemporal dementia, usPPA: unspecified primary progressive aphasia. SNP: single nucleotide polymorphism, WES: whole exome sequencing, WGS: whole genome sequencing

Table 2. Pathogenic variants found in primary AD/FTD genes.

Gene	Coding Change	AA Change	ExAC	CADD	REVEL	Family history	ACMG Classification	Country of origin	Proband's Phenotype(s)
APP	c.2020G>C	Glu674Gln	< 0.00001	25.2	0.635	A-Dom	Likely pathogenic	Mexico	Amnestic AD
C9orf72	c.-45+163 GGGGCC [>24]	A-Dom	Pathogenic	Chile Brazil	bvFTD
GRN	c.21G>A	Trp7Ter	.	36	.	Sporadic	Pathogenic	Colombia	usAD
	c.58dupT	Cys20Leufs*45	.	.	.	A-Dom	Pathogenic	Chile	bvFTD
	c.328 C>T	Arg110Ter	< 0.00001	29.4	.	Positive	Pathogenic	Colombia	FTD-PSP
	c.462+1 G>A	.	.	26.3	.	Positive	Pathogenic	Peru	bvFTD
	c.767_768insCC	Gln257Profs*27	.	.	.	Unknown	Pathogenic	Brazil	bvFTD
	c.1098 T>A	Cys366Ter	.	36	.	Positive	Pathogenic	Colombia	svPPA
MAPT	c.796 C>G	Leu266Val	.	25.8	0.662	A-Dom	Pathogenic	Brazil	bvFTD
	c.915 T>C	Ser305Ser	.	.	.	A-Dom	Pathogenic	Colombia	bvFTD
	c.1280 C>T	Thr427Met	0.00001	33	0.508	Unknown	Pathogenic	Argentina	Amnestic AD
PSEN1	c.250 A>G	Met84Val	.	23.9	0.923	Unknown	VUS	Argentina	Amnestic AD
	c.280 G>A	Val94Met	< 0.00001	25.1	0.847	A-Dom	Pathogenic	Colombia	usAD
	c.356 C>T	Thr119Ile	.	24.4	0.795	A-Dom	Pathogenic	Colombia Argentina	bvFTD
	c.415 A>G	Met139Val	.	23.4	0.879	A-Dom	Pathogenic	Argentina	Amnestic AD
	c.428 T>C	Ile143Thr	.	26.8	0.985	A-Dom	Pathogenic	Colombia	Amnestic AD
	c.519 G>T	Leu173Phe	.	25.8	0.923	A-Dom	Pathogenic	Colombia	bvFTD
	c.1223 T>C	Ile408Thr	.	27.4	0.877	Unknown	VUS	Peru	Amnestic AD
VCP	c.283 C>T	Arg50Cys	.	27.1	0.606	A-Dom	Pathogenic for essential tremor	Colombia	Amnestic AD
TARDBP	c.1147 A>G	Ile383Val	< 0.00001	0.308	0.377	A-Dom	Pathogenic	Colombia	usFTD

AA: Amino acid. A-Dom: Autosomal dominant; three affected individuals in two generations. Positive family history; At least one first or second degree relative with neurodegeneration. usAD: Unspecified Alzheimer's disease, bvFTD: behavioral variant frontotemporal dementia, FTD-PSP: frontotemporal dementia-progressive supranuclear paralysis, svPPA: semantic variant primary progressive aphasia, AD: Alzheimer's disease, VUS: Variant of unknown significance. PVS1: The PVS1 rule is related to null variants which usually result with a loss of function effect. These variants include stop gain, frameshift insertions or deletions, splice donor or acceptor, a start loss

Table 3. Ancestral origin of pathogenic variants found in more than one carrier

Gene	Coding Change	AA Change	Number of independent families	Number of Carriers with WGS/WES	Ancestral Haplotype
<i>GRN</i>	c.767_768insCC	Gln257Profs*27	2	2	European
	c.1098 T>A	Cys366Ter	1	3	European
<i>MAPT</i>	c.915 T>C	Ser305Ser	1	3	European
<i>PSEN1</i>	c.356 C>T	Thr119Ile	4	16	European
	c.415 A>G	Met139Val	1	8	Amerindian
<i>TARDBP</i>	c.1147 A>G	Ile383Val	1	2	European

AA: Amino acid. WGS: whole genome sequencing, WES: whole exome sequencing. The 'number of independent families' clusters individuals into 'families' based on pedigree reconstruction from the participant interview.

Figures

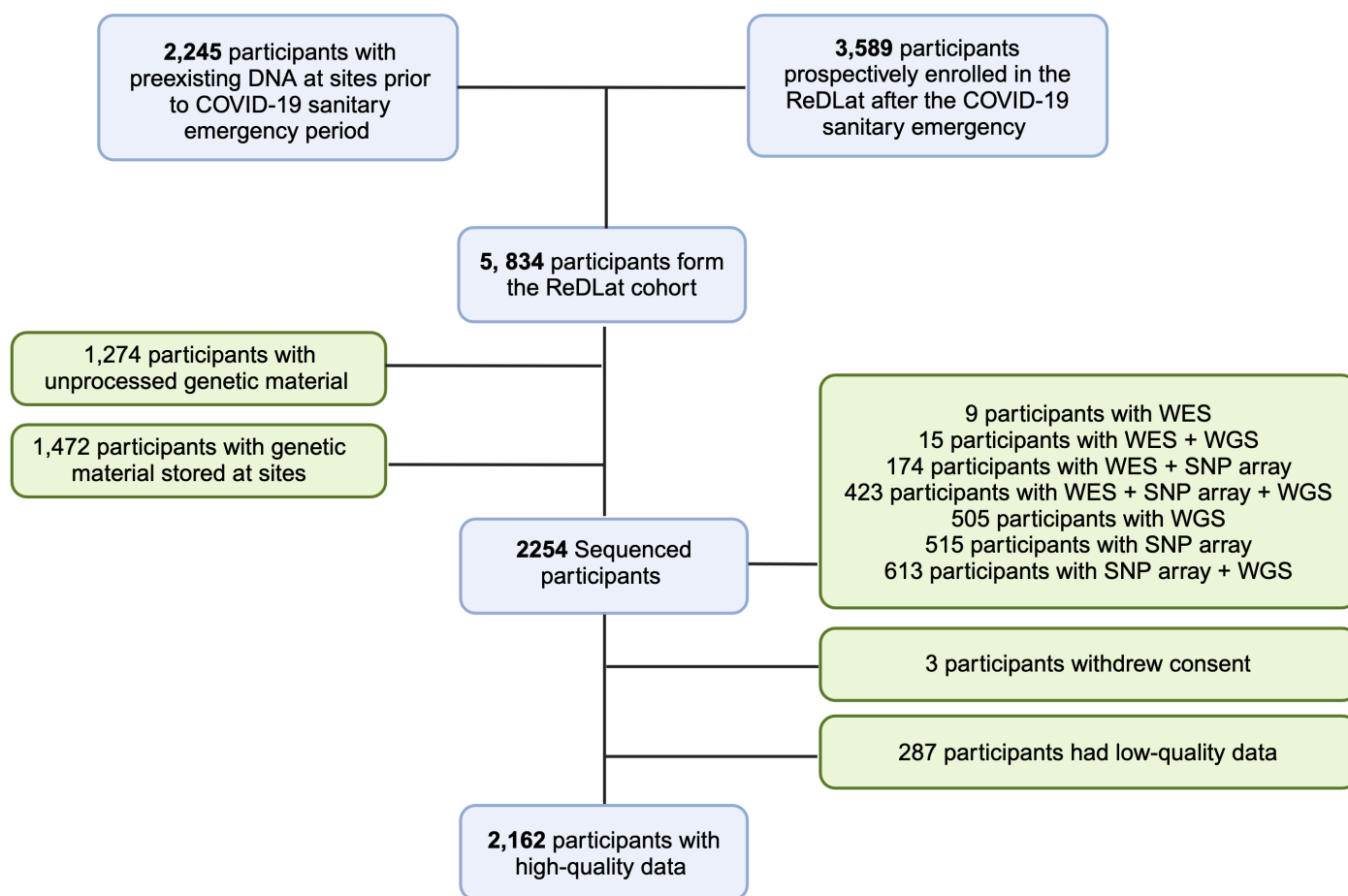


Figure 1

Assembly of ReDLat genomic dataset.

WES: whole exome sequencing. WGS: whole genome sequencing. SNP: single nucleotide polymorphism

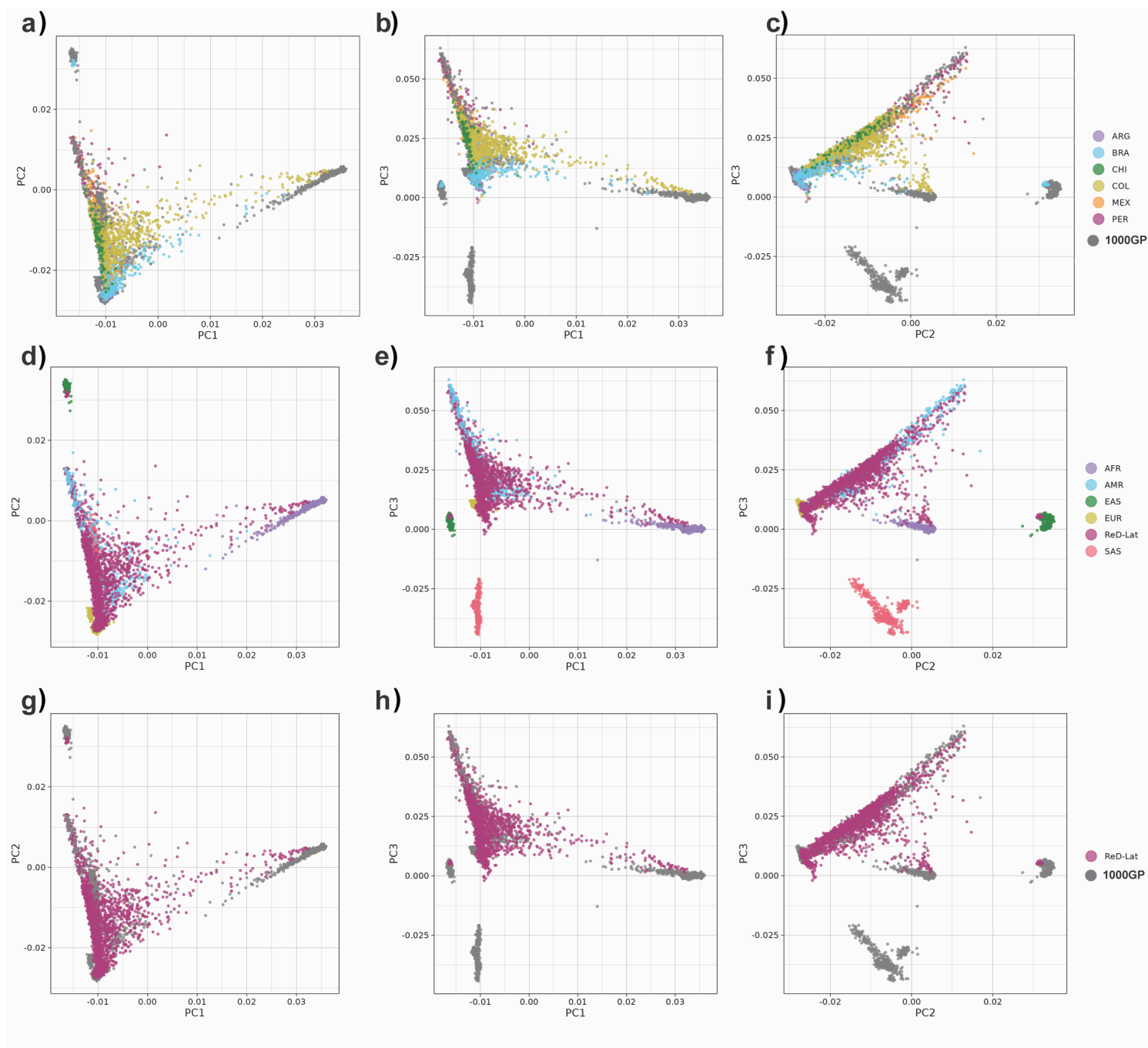


Figure 2

Principal component analysis of the ReDLat dataset and individuals from the 1000 Genome project (1000GP).

The 1000 Genomes project includes African (AFR), European (EUR), South Asian (SAS), East Asian (EAS) and Admixed American (AMR) individuals. ReDLat sub-cohorts include individuals from Argentina (ARG), Brazil (BRA), Chile (CHI), Colombia (COL), Mexico (MEX) and Peru (PER). PC: Principal component.

a) PC1 vs. PC2. Genomes from ReDLat are colored by sub-cohort, while all 1000GP genomes are shown in a single color. b) PC1 vs. PC3. Genomes from ReDLat are colored by sub-cohort, while all 1000GP genomes are shown in a single color. c) PC2 vs. PC3. Genomes from ReDLat are colored by sub-cohort, while all 1000GP genomes are shown in a single color. d) PC1 vs. PC2. Genomes from 1000GP are colored by sub-cohort, while all ReDLat genomes are shown in a single color. e) PC1 vs. PC3. Genomes from 1000GP are colored by sub-cohort, while all ReDLat genomes are shown in a single color. f) PC2 vs. PC3. Genomes from 1000GP are colored by sub-cohort, while all ReDLat genomes are shown in a single color. g) PC1 vs. PC2 of ReDLat and 1000GP genomes. h) PC1 vs. PC3 of ReDLat and 1000GP genomes. i) PC1 vs. PC2 of ReDLat and 1000GP genomes.

K = 5

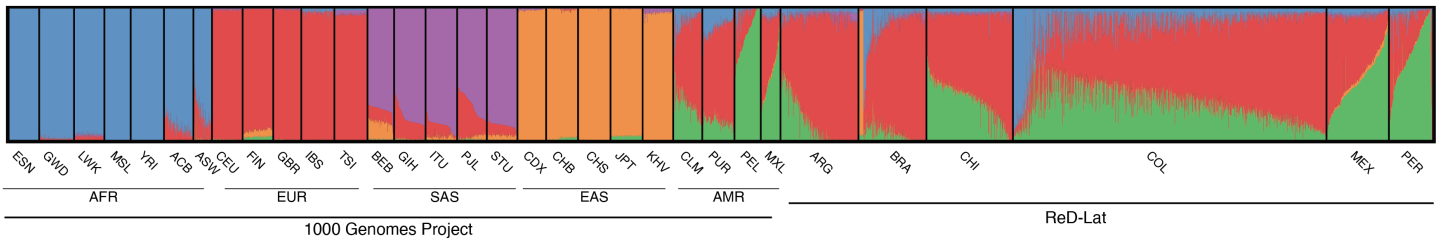


Figure 3

Global ancestry proportions of the ReDLat cohort represented by ADMIXTURE Q. values assuming 5 ancestral populations (K)

The 1000 Genomes project includes African (AFR): GWD: Gambian in Western Divisions in the Gambia, LWK: Luhya in Webuye, MSL: Mende in Sierra Leone, YRI: Yoruba in Ibadan, Nigeria, ACB: African Caribbean in Barbados, ASW: African Ancestry in Southwest US. European (EUR): CEU: Utah Residents (CEPH) with Northern and Western European Ancestry, FIN: Finnish in Finland, GBR: British in England and Scotland, IBS: Iberian Population in Spain, TSI: Tuscany in Italia. South Asian (SAS): BEB: Bengali in Bangladesh, GHI: Gujarati Indians in Houston, ITU: Indian Telugu in the U.K., PJL: Punjabi in Lahore, STU: Sri Lankan Tamil in the UK. East Asian (EAS): CDX: Chinese Dai in Xishuangbanna, China, CHB: Han Chinese in Beijing, CHS: Han Chinese South, JPT: Japanese, Kyushu, KHV: Kinh Vietnamese. Admixed American (AMR): CLM: Colombian from Medellin, PUR: Puerto Rican from Puerto Rico. ReDLat subcohorts: ARG: Argentina, BRA: Brazil, CHI: Chile, COL: Colombia, MEX: Mexico, PER: Peru

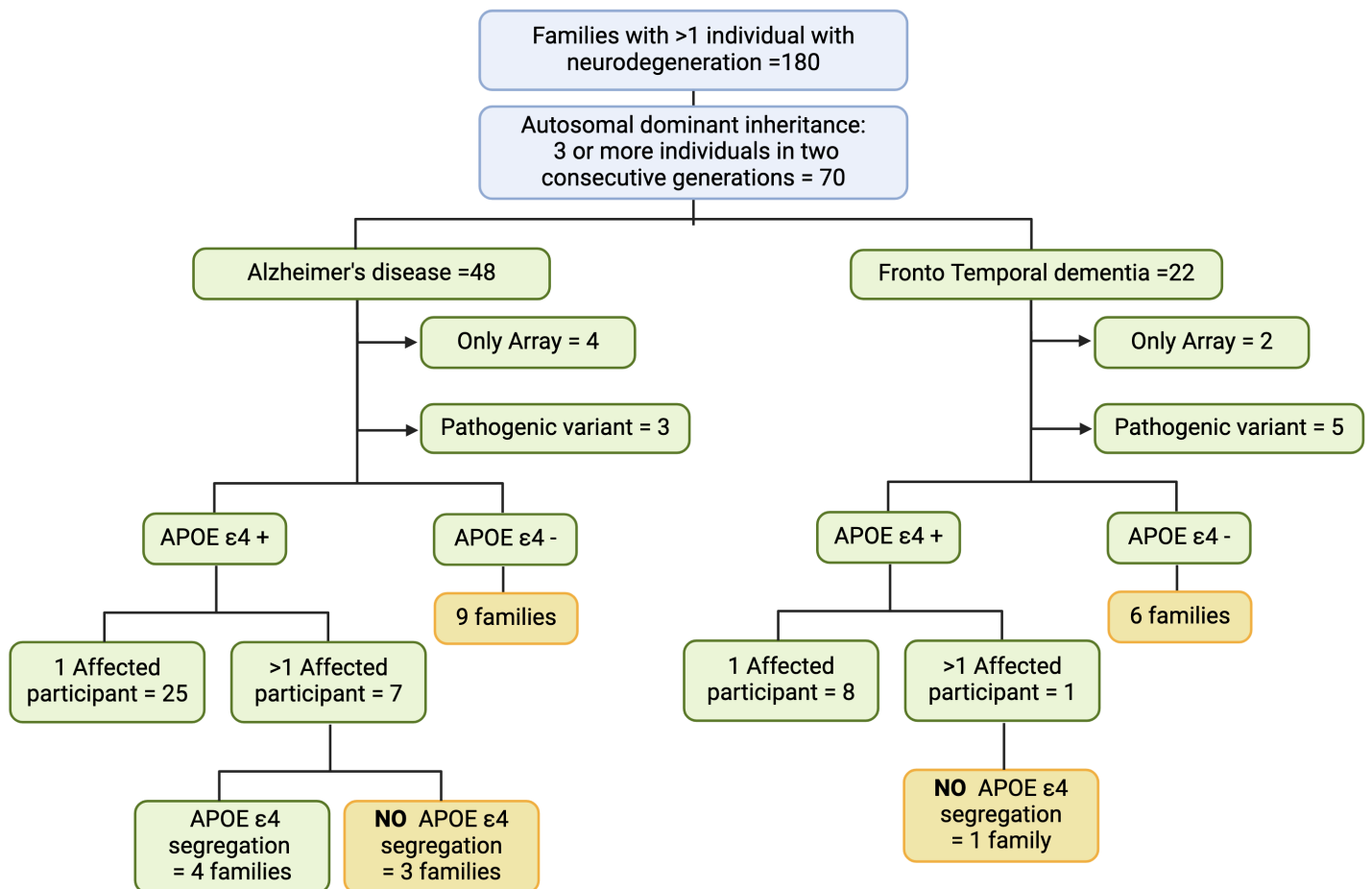


Figure 4

Cosegregation analysis of *APOE* ε4 and neurodegeneration in families with multiple affected individuals

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryData.pdf](#)