APOE Haplotype Phasing Using ONT Long-Read Sequencing Reveals Two Common ε3 and ε4 intragenic haplotypes in the Spanish Population

Pablo García-González[1,2,3*], Raquel Puerta[1,3*], Amanda Cano[1,2], Claudia Olivè[1], Marta Marquié[1,2], Sergi Valero[1,2], Maitee Rosende-Roca[1,2], Montserrat Alegret[1,2], Pilar Sanz[1], Frederik Brosseron[4], Pamela Martino-Adami[5], Itziar de Rojas[1,2,6], Michael Heneka[6], Alfredo Ramírez[5,7,8,9,10], Arcadi Navarro[11,12,13,14], María Eugenia Sáez[1,15], Lluís Tárraga[1,2], José E. Cavazos[16,17], Mercè Boada[1,2], María Victoria Fernandez[1], Alfredo Cabrera-Socorro[18*], Agustín Ruiz[1,2,17,19*,¥]

1. ACE Alzheimer Center Barcelona – Universitat Internacional de Catalunya, Barcelona, Spain.
2. CIBERNED, Network Center for Biomedical Research in Neurodegenerative Diseases, National Institute of Health Carlos III, Madrid, Spain.
3. PhD program in Biotecnology, Faculty of Pharmacy and Food Sciences, University of Barcelona, 08028- Barcelona, Spain.
4. Universitätsklinikum Bonn & Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE) Bonn, Germany.
5. Division of Neurogenetics and Molecular Psychiatry, Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital Cologne, University of Cologne, 50937 Cologne, Germany
6. Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Esch-sur-Alzette/Belvaux, Luxembourg.
7. Department of Neurodegenerative Diseases and Geriatric Psychiatry, University Hospital Bonn, Medical Faculty, 53127 Bonn, Germany.
8. German Center for Neurodegenerative Diseases (DZNE), 53127 Bonn, Germany.
9. Cluster of Excellence Cellular Stress Responses in Aging-associated Diseases (CECAD), University of Cologne, 50931 Cologne, Germany.
10. Department of Psychiatry and Glenn, Biggs Institute for Alzheimer's and Neurodegenerative Diseases, San Antonio, TX, USA.

11. Institució Catalana de Recerca i Estudis Avançats (ICREA) and Universitat Pompeu Fabra. Pg. Lluís Companys 23, 08010, Barcelona, Spain
12. IBE, Institute of Evolutionary Biology (UPF-CSIC), Department of Medicine and Life Sciences, Universitat Pompeu Fabra. PRBB, C. Doctor Aiguader N88, 08003 Barcelona, Spain
13. Center for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Av. Doctor Aiguader, N88, 08003 Barcelona, Spain.
14. BarcelonaBeta Brain Research Center, Pasqual Maragall Foundation, C. Wellington 30, 08005, Barcelona, Spain
15. CAEBI, Centro Andaluz de Estudios Bioinformáticos, Sevilla, Spain.
16. South Texas Medical Science Training Program, University of Texas Health San Antonio, San Antonio, TX, USA.
17. Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, San Antonio, TX, USA.
18. Janssen Pharmaceutica NV, a Johnson & Johnson company, Beerse, Belgium.
19. Department of Microbiology, Immunology and Molecular Genetics. Long School of Medicine. University of Texas Health Science Center, San Antonio, TX, USA.

*These authors contributed equally to this work

¥Correspondence to:

Prof. Agustin Ruiz MD PhD.

Research Center & Memory Clinic. ACE Alzheimer Center Barcelona. Marques de Sentmenat 57. 08028-Barcelona. Spain. aruiz@fundacioace.org

Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases. University of Texas Health Science Center, 78229 Floyd Curl Dr. San Antonio, TX, USA. ruiza5@uthscsa.edu

Keywords: APOE, Alzheimer, epsilon 4 allele, haplotype, phasing, Oxford Nanopore, NGS, long-read sequencing, MCI, dementia, gene

Abstract

Background: The apolipoprotein E (*APOE*) gene is a key genetic determinant of Alzheimer's disease (AD) risk, with the ε4 allele significantly increasing susceptibility. While the pathogenic effects of the ε4 allele are well established, the functional impact of distinct haplotype configurations within the broader ε3 and ε4 backgrounds remains poorly understood. This study investigates the role of intragenic sub haplotypes in modulating *APOE* expression and their potential influence on AD progression.

Methods: We utilized Oxford Nanopore Technology (ONT) long-read sequencing to phase variants within a 4-kilobase comprising the *APOE* locus in a cohort of 1,265 individuals with known *APOE* genotypes. We evaluated the impact of the identified intragenic haplotypes on APOE protein levels in cerebrospinal fluid (CSF) using the Olink platform, adjusting for demographic and molecular covariates. Statistical modeling was employed to assess the independent effects of these haplotypes alongside traditional *APOE* genotypes. Additionally, their influence on dementia progression in mild cognitive impairment (MCI) subjects was analyzed using adjusted Cox proportional hazards models.

Results: Our analysis identified 48 Single Nucleotide Variants (SNVs) within a 4-kilobase region containing the *APOE* gene, including nine novel variants. Phasing of variants within the *APOE* locus revealed 59 unique haplotypes in the Spanish population, which were grouped into five major haplogroups—ε2, ε3A, ε3B, ε4A, and ε4B—including two common haplogroups for each of the ε3 and ε4 isoforms. The ε4A haplogroup was associated with a significant decrease in APOE ε4 protein levels in CSF ($p = 0.004$), suggesting a regulatory mechanism that may mitigate the toxic gain-of-function effect typically attributed to the ε4 allele. Conversely, the ε3B haplogroup was linked to increased APOE ε3 protein levels in ε3/ε4 carriers ($p = 0.025$), potentially serving a compensatory role.

These effects were independent of overall *APOE* genotype and remained significant after adjusting for covariates. Both haplogroups (ε4A and ε3B) demonstrated protective effects in the progression from MCI to dementia, underscoring their potential relevance in Alzheimer's disease.

Conclusions: This study provides new insights into the intragenic allelic variability of the *APOE* gene, demonstrating that intragenic *APOE* haplogroups within the ε3 and ε4 backgrounds can modulate *APOE* isoform expression in ways that might modulate AD. Our findings highlight the importance of considering haplotype-specific effects when interpreting the functional impact of *APOE* and in designing targeted therapeutic strategies. Further research is needed to explore the broader regulatory network of the *APOE* locus and its interaction with neighboring loci in the 19q13 region.

Keywords: Alzheimer, *APOE*, haplotype, long-read sequencing, risk, progression, protein levels

1. Introduction

Alzheimer's disease (AD) and related dementias present a growing global health challenge as populations age, with significant social and economic burdens[1,2]. AD, the most common cause of dementia, is characterized by a progressive decline in cognitive function, leading to memory impairment, loss of independence, and ultimately, death. The profound effects on patients and caregivers, coupled with the strain on healthcare systems, highlight the urgent need for effective strategies to prevent and manage this disease[3]. Genetic factors are critical in developing late-onset AD (LOAD), with the Apolipoprotein E (*APOE*) gene identified as a key contributor. Specifically, the *APOE* ε4 allele significantly increases the risk for AD, while the ε2 allele appears to have a protective effect[4,5]. However, the risk associated with the *APOE* ε4 allele varies across populations[6–11]. Notably, individuals of African ancestry exhibit a reduced risk of Alzheimer's disease (AD) associated with the ε4 allele compared to those of European descent, suggesting the presence of protective effects within specific genetic backgrounds[12] . Recent studies have highlighted a significant interaction between *APOE* ε4 and the rs10423769-A allele, which reduces AD risk by up to 75% in African-origin haplotypes[12]. Understanding the genetic

mechanisms modulating *APOE* and its association with AD—particularly the role of cis and trans regulatory elements—remains critical for developing targeted interventions aimed at mitigating the impact of this major risk factor and ultimately reducing the societal burden of the disease.

The *APOE* gene encodes a major regulator of lipid metabolism, functioning by binding and transporting lipids within the bloodstream and the brain. In neurodegenerative diseases, *APOE* lipoprotein plays pivotal roles in amyloid-β clearance, synaptic maintenance, and neuroinflammatory processes[13,14]. The *APOE* ε4 isoform has been shown to impair amyloid-β clearance, promoting its accumulation in the brain, a hallmark of AD pathology[15]. Furthermore, the ε4 isoform is associated with enhanced neuroinflammation[16] and oxidative stress[17], contributing to neuronal damage and accelerating disease progression. These distinct roles of *APOE* isoforms emphasize the importance of understanding their functional differences, particularly in the context of AD pathology and progression[18].

Oxford Nanopore Technology (ONT)[19] offers a cutting-edge solution for overcoming challenges associated with genetic analysis, particularly the issue of homoplasy, where genetic variants may independently arise but appear identical[20]. Homoplasy is a well-known issue within the *APOE* locus, complicating efforts to link genetic variants with disease phenotypes[21]. ONT's long-read sequencing technology provides the ability to phase DNA variations, enabling the precise connection of single nucleotide variants (SNVs) with haplotypes and regulatory elements that modulate *APOE* gene expression. This level of resolution facilitates a clearer distinction between benign and disease-associated variants, enhancing our understanding of how different *APOE* haplotypes influence disease risk and progression across diverse populations[22]. The ability to phase the classic *APOE* isoforms with other variants located in functional elements provides an invaluable tool for elucidating the mechanisms by which cis-genetic variants configure *APOE* sub-haplotypes. This understanding extends beyond the classical ε2/ε3/ε4 isoforms, shedding light on how specific combinations contribute to Alzheimer's disease and other neurodegenerative conditions

In this study, we explore the association between intragenic *APOE* haplotypes and Alzheimer's disease (AD) progression, with a special focus on identifying existing haplotypes within the ε3 and ε4 protein isoforms. Utilizing ONT-based long-read sequencing, we analyzed the distribution of intragenic *APOE* sub-haplotypes in a large Spanish cohort, identifying 48 SNVs, including both known and novel variants. Our analysis specifically focuses on examining common variants within regulatory motifs that influence CSF APOE levels and AD biomarkers in an isoform-specific manner. Moreover, we identified distinct ε3 and ε4 sub-haplotypes that appear to influence the rate of conversion from mild cognitive impairment (MCI) to dementia, providing potential genetic markers for individualized disease risk assessment and therapeutic intervention.

2. Materials and methods

2.a. The ACE Alzheimer's Center Barcelona CSF cohort (ACE)

Patient recruitment and evaluation were conducted at the Memory Disorders Unit of the ACE Alzheimer Center Barcelona (ACE) in Spain, spanning the years 2016 to 2021[23]. Diagnoses were established through consensus discussions involving neurologists, neuropsychologists, and social workers during case conferences. All patients classified as having mild cognitive impairment (MCI) met the criteria outlined by *Petersen et al.* for MCI diagnosis[24,25]. These criteria include subjective memory complaints, a noticeable decline from normal cognitive function, intact daily living activities, the absence of dementia, and measurable impairments in one or more cognitive domains, whether amnestic MCI (single or multiple domains). Impairment thresholds were determined according to age and education level. Specific cutoffs for the tests in the neuropsychological battery (NBACE) are detailed in prior publications[26], with individuals scoring below these thresholds classified as having MCI.

For patients with MCI who were later followed up, dementia diagnoses were made according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-V) criteria[27]. Dementia subtypes were further categorized based on the 2011 National Institute on Aging-Alzheimer's Association (NIA-AA) criteria for Alzheimer's disease[28], the NINDS-AIREN criteria

for vascular dementia[29], frontotemporal dementia[30], and the diagnostic criteria for Lewy body dementia[31].

Fasting patients provided paired CSF and plasma samples using standardized clinical methods. CSF was collected via lumbar puncture from the L3-L4 intervertebral space following standard protocols[32]. The procedure, performed by skilled neurologists under local anesthesia (1% mepivacaine), involved the patient sitting upright. Two 10-ml polypropylene tubes (Sarstedt ref 62,610,018) of CSF were collected, with one tube used for routine biochemistry tests (glucose, total protein, proteinogram, cell type, and count). The second tube was centrifuged (2000xg, 10 min at 4°C), aliquoted into polypropylene tubes (Sarstedt ref 72,694,007), and stored at -80°C within two hours of collection. On the day of biomarker analysis, an aliquot was thawed at room temperature, vortexed for 5–10 seconds, and analyzed for Aβ1-42, total tau (t-tau), and phosphorylated tau (p-tau) levels using enzyme-linked immunosorbent assays (ELISAs): Innotest Aβ1-42, Innotest hTAU Ag, and Innotest PHOSPHO-TAU (181P) (Fujirebio Europe)[32–34].

CSF protein levels of APOE were measured using conventional ELISA protocols as previously described[35]. In addition, APOE protein levels, along with the protein products of loci near the *APOE* gene (NECTIN2, BCAM, and APOC2), were also obtained from OLINK panels generated in our prior work as part of the PREADAPT JPND programs[36]. Each selected protein's measurements were adjusted for age, sex, and the principal components of the CSF proteome variance. Standardized residuals were then used as dependent variables in subsequent analyses (see below).

Informed consents for lumbar puncture, genetic analyses, and the anonymized use of clinical records for research were approved by the Ethics Committee of the Hospital Clinic i Provincial de Barcelona (Barcelona, Spain) in accordance with Spanish biomedical laws (Law 14/2007, of July 3, on biomedical research; Royal Decree 1716/2011, of November 18). The study also adhered to the recommendations of the Declaration of Helsinki. The Harpone project, encompassing all the research described in this manuscript, has also been approved by the Ethics Committee of the Hospital de Bellvitge (Barcelona, Spain) (Acta 04/21).

2.b. DNA extraction and *APOE* genotyping

DNA extraction from blood specimens was performed automatically using standard procedures with the DNA Chemagic system (Perkin Elmer) or the Maxwell RSC48 instrument (Promega). Extensive DNA quality control was conducted, and only samples with DNA concentrations greater than 10 ng/µl and high integrity were included for *APOE* genotyping. The isoforms were determined by TaqMan probe analysis in the Real-Time PCR QuantStudio3 (ThermoFisher, Waltham, Massachusetts, USA) system or extracted from Affymetrix Axiom SP biobank arrays processed as previously described [37–39].

2.c. Long PCR amplification of the *APOE* locus

Long-range PCR was performed to amplify the *APOE* locus using primer pairs designed with the National Institutes of Health (NIH) primer designing tool software (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) to cover the entire region, from the promoter to the 3' UTR. We used Roche's Expand™ Long Template PCR System kit (UNSPSC number: 41106300) for PCR amplification. To find the optimal reaction conditions, we tested in a single sample three different primer pairs, along with different DMSO concentrations (2.5, 5.0, 7.5 and 10%) and three kit-supplied reaction buffers (Supp. Table 1, Supp. Fig. 1). Based on our results, we used a 10% concentration of DMSO, kit-supplied reaction buffer 2 and the primer pair labeled "APOE_III" for optimal long-range amplification of the region. The selected primers were: forward primer 5'- ACAGGGTCAGGAAAGGAGGA -3' and reverse primer 5'- GGCTGGGGCTTAGAGGAAAT -3', spanning a 3939 bp fragment surrounding the *APOE* locus. The reaction mixture included a high-fidelity DNA polymerase, PCR dNTP mix (Roche, UNSPSC code 41106300), template DNA, primers, DMSO, and buffer (including 17.5 mM $MgCl_2$) (Supp. Table 2). The PCR was carried out on a Thermo Fisher QuantStudio 3 Real-Time Thermal Cycler with the following conditions: initial denaturation at 94°C for 2 minutes, followed by 40 cycles of denaturation at 92°C for 10 seconds, annealing at 59°C for 30 seconds, and extension at 68°C for 4 minutes. A ramp increasing the elongation time by 20 seconds per cycle starting in cycle 11, and a final 10-minute elongation step at 68ºC were included in the protocol

to ensure complete amplification of the DNA segment. Following amplification, PCR products were purified using AMpure XP beads (Beckman Coulter, A63881), following the manufacturer's protocol to remove salts, primers, nucleotides, and enzymes, ensuring high-quality DNA for downstream applications. Amplification success was verified via automated capillary electrophoresis using the LabChip GX Touch Nucleic Acid Analyzer (Revvity, CLS138162) and the DNA 12K Reagent Kit (Revvity, 760569).

2.d. ONT long-read Sequencing and Library Preparation

For the ONT sequencing of the *APOE* amplicon, we employed the ONT 96 Native Barcoding Kit (EXP-NBD196) to multiplex 96 samples in a single reaction, and the ONT Ligation Sequencing Kit (SQK-LSK109). We followed the protocol recommended by ONT based on these reagent kits: “Amplicon barcoding with Native Barcoding Expansion 96 (EXP-NBD196, and SQK-LSK109), Version: NBA_9102_v109_revA_09Jul2020”. Purified *APOE* amplicons (100 fmoles) were quantified using the Quant-iT™ PicoGreen™ dsDNA Assay Kits (Invitrogen, P7589), and 250 ng were taken forward for barcode ligation. This process involved two steps: first, the ends of the DNA fragments were repaired using the NEBNext® Ultra™ II End Repair/dA-Tailing Module (New England Biolabs, E7546), following manufacturer instructions. This reaction is optimized for the subsequent ligation reaction, yielding 5’ phosphorylated and 3’ dA-tailed ends. Then, unique barcodes were ligated to each sample using the NEB Blunt/TA Ligase Master Mix module (New England Biolabs, M0367). Subsequently, the barcoded library was pooled together and 480 uL were carried forward for purification using Ampure XP beads, removing excess barcodes and small fragments, and collected in a volume of 30 uL. Absorbance at 260 nm, A260/A280 and A260/230 ratios, measured with a BioTek Epoch Microplate Spectrophotometer, were used subsequently for rapid estimation of DNA concentration and purity. 500 ng (200 fmol) of the pulled library was then ligated to ONT sequencing adapters using the NEBNext Quick Ligation Module (New England Biolabs, E6056), following manufacturer instructions. A final AMpure XP purification step was performed to remove unwanted products. Finally, 100 ng (~40 fmoles) of the library were loaded onto primed ONT R9.4.1 flow cells. Sequencing was carried

out on a MinION platform, enabling real-time demultiplexing by Guppy, until satisfactory coverage of all barcoded samples was achieved (at least 100x coverage of the least represented barcodes).

2.e. Bioinformatics Analysis

The dry lab workflow for the identification of missense mutations in the *APOE* locus using Oxford Nanopore Technologies (ONT) sequencing data is based on the protocol described by *Leija-Salazar et al.*[40]. This bioinformatics processing involves a series of computational steps, starting from raw signal data and culminating in the generation of a multi-sample VCF file. The process begins with basecalling and demultiplexing using Guppy v4.5.2 (https://nanoporetech.com/software/other/guppy), which converts raw electrical signals from the nanopore sequencing (.fast5) into nucleotide sequences (.fastq files), while simultaneously assigning each read to its corresponding barcode. Basecalling was conducted using the dna_r9.4.1_450bps_hac.cfg --config file for high accuracy. We generated in-silico gel electrophoresis images to visually assess the read length of sequencing reads, ensuring the signal detected for each barcode was present in the expected length (3939bp). Following this, the quality of the reads is assessed; only those marked as "PASS" and within the length range of 3800 to 4000 base pairs were retained to ensure only full-length *APOE* amplicons were included in downstream analysis. Subsequently, nanopolish (https://github.com/jts/nanopolish) was utilized to index the .fastq files, linking the sequencing data back to the original raw signal data in the .fast5 files. This linkage is crucial for the later implementation of signal-based error correction during variant calling. We used two different aligners—Graphmap v0.5.2 and NGLMR v0.2.7—to generate the SAM files providing the positional information of the filtered and indexed reads within the reference genome (GRCh38.p13). Subsequently, we used samtools v1.10 (http://www.htslib.org/) to convert the files to BAM format, and then sort and index the files. nanopolish was subsequently used for variant calling, employing the raw nanopore signal to enhance the accuracy of mutation detection. The output of this step is a .vcf file containing the identified variants within the *APOE* locus for each sample. Individual .vcf files from each sample

were then merged using bcftools 1.9 (https://samtools.github.io/bcftools), and sample IDs were reassigned based on barcode correspondence, resulting in a multi-sample .vcf file that consolidates the variant information across all 96 samples. SNP identifiers and TOPMed allele frequencies were annotated based on dbSNP build 151 (GRCh38.p7). The mean ratio of the QUAL and TotalReads parameters generated by nanopolish was used as a quality measure for variant calling. Overall, this comprehensive approach allowed for the accurate detection of missense mutations in the *APOE* locus by leveraging both basecalled sequences and the inherent error-correction capabilities of Nanopolish based on the raw nanopore sequencing data.

2.f. Variant phasing and further haplotype analyses

The phasing of *APOE* haplotypes was conducted using WhatsHap (https://whatshap.readthedocs.io), a tool designed for accurate read-based phasing of long-read sequencing data[41,42]. WhatsHap performs phasing by directly utilizing DNA sequencing reads, a method known as read-based phasing or haplotype assembly, which is particularly suited for long reads but also works well with short reads. By leveraging the long-range information from ONT sequencing, WhatsHap effectively separated detected variants into their respective haplotypes, enabling the reconstruction of complete *APOE* haplotypes for each individual. To facilitate downstream analyses, we developed a custom code. This program automated the extraction of individual-level haplotypes from the phased VCFs generated by WhatsHap, calculated haplotype frequencies, and encoded individual-level subhaplotypes for subsequent use. Additionally, the code prepared nexus files for evolutionary analyses in MEGA and formatted data for phenotype-haplotype correlation studies in SPSS.

Following the phasing process, the evolutionary relationships of the *APOE* haplotypes were examined using MEGA Software v11.0.13 (https://www.megasoftware.net)[43]. MEGA was used for the alignment and comparison of haplotype sequences, enabling the construction of phylogenetic trees and the assessment of evolutionary divergence. Specifically, the evolutionary history was inferred using the Maximum Likelihood method with the Tamura-Nei model[44]. The tree was drawn to scale, with branch lengths representing the number of substitutions per site.

Sub-haplotypes derived from the different *APOE* common haplotypes were color-coded to facilitate interpretation: green for ε2, orange for ε3, red for ε4, and blue for the ancestral sequence extracted from Ensembl[45].

2.g. Statistical analysis

For statistical analysis, SPSS (version 26.0 for Windows, IBM, Armonk, New York, USA) was used. Group differences were assessed using the Chi-Square test for categorical variables and the t-test for quantitative variables. To evaluate the impact of *APOE* sub haplotypes on the risk of conversion to dementia, Kaplan–Meier curves were generated, stratifying mild cognitive impairment (MCI) subjects into sub-haplotype categories based on the presence of ε3 or ε4 variants, encoded as the presence/absence of potentially disease-modifying subhaplotypes. Cox proportional-hazard regressions were then conducted to identify potential confounders of the categories, adjusting for covariates including age at lumbar puncture, sex, baseline MMSE score, formal education (years), AD biomarker levels, and biomarker determination methods (ELISA or CLEIA). To examine the impact of subhaplotypes on CSF AD biomarkers or CSF protein levels of APOE, BCAM, APOC3, and NECTIN2, a general linear model of the standardized residuals of the target proteins was further adjusted for AD biomarkers (except when exploring the impact of *APOE* subhaplotypes on AD biomarkers), *APOE* isoforms (ε2/ε3/ε4), age, sex, and the principal components of the proteome variance as previously reported[46]. The significance level was set at $\alpha = 0.05$.

3. Results

3.a. ACE CSF cohort description

Table 1 provides a detailed descriptive overview of the demographic, clinical, neurological, and biological characteristics of the ACE CSF cohort, comprising 1,267 subjects sequenced for the *APOE* gene using ONT long-read sequencing. The ACE cohort is predominantly female (57.9%), with a mean age of 72.68 years (SD = 8.46), a mean body mass index (BMI) of 26.9 kg/m² (SD = 4.36), and an average of 8.17 years of education (SD = 4.77). Syndromic diagnoses based on

Clinical Dementia Rating (CDR) scores include 6.2% with subjective cognitive decline (SCD, CDR = 0), 62.7% with mild cognitive impairment (MCI, CDR = 0.5), and 30.1% with dementia (CDR > 0). The AT(N) classification reveals that 33% of participants fall under the A+T+N+ category, indicating the presence of both amyloid pathology and tauopathy, while 29.7% are A-T-N-, showing no evidence of AD pathology. Of note, a relatively large proportion of individuals (37.7%) displayed intermediate AT(N) profiles, with or without the presence of amyloidosis, suggesting the existence of a relatively big proportion of individuals with an ongoing neurodegenerative process that does not yet fulfill the full positive AT(N) molecular pattern (Table 1). Additional clinical and biochemical data include a mean Mini-Mental State Examination (MMSE) score of 24.4 (SD = 4.41), and CSF biomarkers with mean concentrations of 72.28 pg/mL (SD = 45.35) for phosphorylated tau (pTau181), 811.2 pg/mL (SD = 393.5) for amyloid β42, and 469.2 pg/mL (SD = 330.1) for total tau. Lastly, among the MCI group with follow-up information available (n=720), 49.1% converted to dementia, reflecting the longitudinal progression of a significant proportion of MCI subjects. Analysis of *APOE* genotype distribution shows that, as expected, the ε3/ε3 genotype is the most prevalent (57.9%), followed by ε3/ε4 (27.9%), with less frequent occurrences of ε2/ε3 (7.2%), ε4/ε4 (4.7%), ε2/ε4 (1.9%), and ε2/ε2 (0.3%).

3.b. DNA Sequencing of the *APOE* locus in the Spanish Population

We applied Oxford Nanopore Technologies (ONT) long-read DNA sequencing to a 3939 bp region surrounding the *APOE* locus (from the promoter to the 3'UTR) in our cohort of 1,267 individuals (Table 1). Our study focuses exclusively on SNVs within the *APOE* region on chromosome 19 due to the known limitations of ONT sequencing, which tends to be less reliable for detecting insertion and deletion (indel) mutations[47]. By limiting our analysis to SNVs, we leverage the strengths of ONT sequencing, ensuring higher confidence in variant detection and reducing the risk of false positives typically associated with indel identification. The results of this sequencing project are presented in Table 2 and Fig. 1. Overall, we identified 48 single nucleotide variants (SNVs) in the *APOE* gene region using the methods and criteria applied.

We categorized detected SNVs based on their frequency and novelty status: 8 were common (MAF>0.01), 7 rare (0.01>MAF>0.001), 24 very rare (MAF<0.001), and 9 unlisted in dbSNP. These classifications are derived using allele frequency data from the TOPMed[48] and GR@ACE [37–39] populations. The eight common variants are well-documented in population databases, indicating a relatively high frequency in the general population. Notably, the observed and expected frequencies of SNVs with a frequency of 1% or higher in previous studies align almost perfectly with the frequencies observed in our study, underscoring the reliability of the results for SNVs using the ONT technology implemented (Fig. 2). These variants are primarily distributed in intronic and exonic regions, suggesting potential regulatory or coding impacts.

The 7 rare variants are less prevalent, and their occurrence is limited to specific subpopulations or contexts. The 24 very rare variants exhibit an extremely low frequency (<0.1%) in global databases and are mostly singletons, underscoring their possible relevance to niche population studies or for tracking rare genetic traits. We also generated confidence ratings for each variant based on their frequency and sequencing quality, with 37 variants labeled as Very Good or Good, and all the new ones marked with caution (Fig. 1, Supp. Table 3). We labeled two additional variants as ‘caution’: rs1488164590 was detected in a single chromosome, had an exceptionally low QUAL to TotalReads ratio, and was present in a homopolymer region; rs147236548 was called in only one chromosome using the Graphmap but not the NGLMR aligner. Comparing our data with available TOPMed imputed genotypes from the GR@ACE project [37–39], we confirmed that the Graphmap call matched the genotype of this sample, so we kept this aligner over NGLMR for downstream analyses. This emphasizes the robustness of the ONT sequencing approach while highlighting new variants requiring careful interpretation. Additionally, substitution types indicate that 31 are transitions and 17 are transversions (Ts/Tv rate 1.82). We also found 9 novel variants are not found in public databases, suggesting they are unique to this dataset and could represent newly identified mutations. Of these, 6 are located within intronic regions, 1 is in the 5'UTR region, and 2 are found in exon 4 (Table 2).

We analyzed these 9 potentially novel SNVs using the Ensembl Variant effect prediction tool (VEP)[49], revealing that the majority are classified as having a modifier impact. This classification typically means that these variants do not directly alter the protein function but may influence gene regulation, splicing, or transcriptional control (Supp. Table 4). Most of these variants are in upstream, downstream, or intronic regions of the *APOE* gene, suggesting that their impact is more likely to be on gene expression rather than protein coding. Two new variants, specifically 19:44908665:C/T and 19:44908797:C/A, were identified as synonymous variants in the exon 4 of *APOE* but do not affect the resulting amino acid. These types of changes often have minimal impact on protein structure, although they could potentially affect splicing or regulatory elements. VEP confirmed that analyzed variants appear are novel, as they were not registered in the VEP databases (Supp. Table 4). We repeated the VEP analysis using the 15 coding SNVs detected in this study (Supp. Table 5). The most important VEP predictions for coding variants identified in the *APOE* locus are summarized in Fig. 3. Half of the observed variants are classified as Disease Causing (Predicted+Literature), indicating potential pathogenicity based on in silico predictions. This visualization helps in identifying the functional impact of each variant based on CADD scores, providing further insights into the severity and potential functional consequences of the observed genetic alterations. Only the *APOE* variants encoding the classic *APOE* alleles (rs429358 tagging ε4, and rs7412 tagging ε2), rs563571689 (D19N substitution) and rs121918393 (R180H substitution also known as Christchurch mutation) displayed relatively high CADD scores although with a modest impact according to standard CADD criteria even not reaching a moderately deleterious score (scoring ranged between 5 and 10)[50].

3.c. Haplotype analysis of the *APOE* locus using ONT long read sequences

In this section, we present a comprehensive analysis of the *APOE* haplotypes derived from the SNVs in the Spanish population. Specifically, we are focusing on characterizing intragenic 4KB-haplotypes derived from a detailed examination of the phase of the 48 detected variants. This analysis was run using WhatsHap[41], known for its high accuracy in phasing genomic variants through sequencing data.

We identified 59 distinct *APOE* haplotypes in the Spanish population, ranging from a maximum of 970 (38.07%) observations for the most common haplotype (Haplotype 1) to just a single observation (0.04% of the total observations) for the rarest haplotypes (Supp. Table 6). This count reflects the prevalence of each haplotype within the population, offering insights into its genetic structure in the *APOE* locus. We also categorized the haplotypes into haplogroups, considering the common *APOE* isoform backgrounds (ε2, ε3, ε4) and the presence or absence of the alternative sequence (T-allele) of the homoplasic SNV rs405509 (Supp. Fig. 2). For instance, Haplotype 1 is classified under haplogroup 3a, while Haplotype 3 falls under haplogroup 4a. This classification framework aids in understanding the evolutionary relationships and potential health implications associated with each haplotype.

A detailed examination of haplotype frequencies revealed a total of 17 distinct haplotypes that exhibited three or more occurrences among the sampled individuals (Fig. 4, Table 3). Our findings revealed that the five most prevalent haplotypes accounted for 92.74% of all haplotypes identified in this cohort. Specifically, Haplotype 1 and Haplotype 2 are associated with the ε3 allele, while Haplotype 3 and Haplotype 4 are linked to the ε4 allele. Additionally, we identified a common ε2 haplotype, designated as Haplotype 5. The predominance of these haplotypes suggests shared ancestry within this population, offering valuable insights into the genetic landscape of the *APOE* gene and its potential health implications. This haplotype frequency distribution underscores the genetic diversity present within the population, which may be influenced by historical demographic factors including migration, genetic drift, and natural selection. The high frequency of certain haplotypes suggests a possible selective advantage, which could be related to environmental adaptations or other health-related traits. This aspect of our findings provides a foundational understanding of the *APOE* intragenic variation that exists within the Spanish population, reflecting broader evolutionary processes.

Importantly, this study firmly demonstrates the existence of two common haplotypes for ε3 and ε4, just determined by the phasing of a single homoplasic variant located in the promoter region of the *APOE* gene (rs405509). We utilized this information to simplify and statistically manage a

panoply of the less common haplotypes, which were categorized into A or B haplogroups based on the correct phasing of classic *APOE* isoforms and the presence of the T-allele of the rs405509 variants (Fig. 4, Supp. Table 6).

To elucidate the evolutionary relationships among the identified haplotypes, we employed the Maximum Likelihood method in conjunction with the Tamura-Nei model implemented in the MEGA software[43] (Fig. 5). These are well-established techniques in phylogenetic analysis. The dataset for this analysis comprised a total of 60 alleles, including the 59 observed haplotypes (Supp. Table 5) and the homo sapiens ancestral *APOE* allele obtained from Ensembl[45]. This analysis yielded a phylogenetic tree with a log likelihood of -404.80, indicating robust statistical support for the inferred relationships among haplotypes. The constructed tree was scaled to represent genetic distances, with branch lengths proportional to the number of nucleotide substitutions per site. This visualization allowed for a clear interpretation of the evolutionary trajectories and relationships among the haplotypes, revealing insights into how these genetic variants may have diverged and adapted over time within the population (Fig. 5). According to the MEGA analysis the oldest haplotype observed in the Spanish population is Haplotype 4, corresponding to the 4b haplogroup of the *APOE* ε4 allele (Table 3). This information is fully concordant with previous small-scale studies in other populations[21]. Furthermore, the analysis permitted the identification of the exact haplotype context (Haplotype 26) of other missense variants of the *APOE* coding sequence, including the *APOE* Christchurch mutation[51–53], which occurs in the *APOE* ε3 isoform and haplogroup ε3b in the Spanish population (Supp. Table 6).

An intriguing aspect of our findings was the absence of a full ancestral sequence among the Spanish *APOE* haplotypes examined. Instead, our analysis indicated that all haplotypes, except for two singletons, harbored alternative sequences shared to Neanderthal-derived sequences at two specific positions within the gene (rs1212454788 and rs749102800). This observation raises the possibility of introgression, suggesting that some of the genetic material associated with the *APOE* sequences in modern humans may have originated from interbreeding events between early humans and Neanderthals. However, extended sequencing is needed to elucidate this possibility.

Furthermore, our analysis confirmed the presence of homoplasy for the variant rs405509, a phenomenon previously documented[21]. This phenomenon complicates the phasing of *APOE* haplotypes and introduces potential inconsistencies in understanding the relationships between these genetic variants located in *APOE* regulatory regions and their associated phenotypic outcomes. Homoplasy, defined as the independent occurrence of the same genetic variant in distinct lineages, can obscure the accurate tracing of inheritance patterns and complicate interpretations of their functional implications. As illustrated in the simplified Supp. Fig. 2, the homoplasy observed for rs405509 could be attributed to a single ancestral recombination (or gene conversion) event within the *APOE* gene. This event likely reshuffled SNVs in the promoters and intronic enhancers of the *APOE* gene in the ancestral ε4b and possibly ε3b backgrounds, leading to the differentiation of the ε3 and ε4 isoforms into the two (A and B) haplogroups in a single mutation step. Subsequently, further diversification of haplotypes occurred within each observed haplogroup, resulting in variants that differ by only 1-3 positions from the most common haplotypes (Haplotypes 1-5). Supp. Fig. 2 displays a simplified phylogenetic network and the corresponding genomic context of most common *APOE* intragenic haplotypes observed in the Spanish population, highlighting their evolutionary relationships and associated regulatory elements within the *APOE* locus and the classic protein isoforms.

3.d. Exploring *APOE* haplogroups effects in Alzheimer's Disease

As previously mentioned, our technique permitted us to separate ε3 and ε4 isoforms into two haplogroups. We decided to investigate the potential impact of these common haplogroups on Alzheimer's disease progression and other potentially relevant endophenotypes. We initially split our population by *APOE* haplogroups and detected the expected 15 categories resulting from the combination of the five common haplogroups (ε2, ε3a, ε3b, ε4a and ε4b). Supp. Fig. 3 illustrates the distribution and survival analysis of *APOE* haplogenotypes stratified by these *APOE* haplogroups. Panel A shows the frequency distribution of these haplogenotypes, revealing that `3a/3b`, `3a/3a`, and `3b/3b` are the most common. Panel B presents survival curves for all 15 haplogenotypes, indicating distinct survival patterns for each group. Panels C-E display specific

APOE genotypes analyses: ε3/ε3 carriers (Panel C), ε3/ε4 carriers (Panel D), and ε4/ε4 carriers (Panel E) further divided by relevant epsilon haplogroups (`3a/3a`, `3a/3b`, `3b/3b` for ε3 haplogroups and `4a/4a`, `4a/4b`, `4b/4b` for ε4 haplogroups), revealing varying disease progression probabilities among these categories. The generation of haplogroups enhances analytical power by refining the categorization of subjects. However, the excess of categories observed tends to reduce statistical power. Despite of this limitation, our initial exploration revealed a significant impact of the haplogroups in the global analysis (panel B) and for the 3b haplogroup in ε3/ε3 carriers ($p < 0.04$), along with a trend toward association for the 4A haplogroup (Supp. Fig. 3).

Overall, we observed evidence suggesting protective effects of the 4A and 3B haplogroups. Consequently, we encoded two dichotomous variables— "presence of 4A" and "presence of 3B" haplogroups—to further investigate their potential effects in detail and to assess the impact of other important covariates that might confound our observations in MCI to AD Disease conversion. Following this idea, we conducted an analysis on the impact of haplogroups on MCI to AD phenoconversion adjusted by demographics, AD biomarkers and the *APOE* genotype. Fig. 7 illustrates the results obtained for these analyses. Panel A shows the distribution of *APOE* haplogroups in the cohort, categorized into four distinct groups: presence of the 4A haplogroup, presence of the 3B haplogroup, presence of both protective haplogroups (3B/4A), and absence of both (None).

Panel B depicts the results of the Cox proportional hazard risk model to explore 4A and 3B haplogroups effect in MCI to AD phenotypic conversion, the table is showing the hazard ratios (HR) of 4A and 3B haplogroups, adjusted for the *APOE* genotype, demographics, and AD biomarkers. Both haplogroups demonstrated significant and independent protective effects in this comprehensive analysis even in a model adjusted by the other haplogroup and the *APOE* genotype. Panel C shows the survival curves specific effect of the 4A haplogroup, where carriers had a significantly reduced risk of disease progression (HR = 0.599, p = 0.002). In the same way, Panel D displays the effect of the 3B haplogroup, highlighting its protective association (HR =

506 0.79, p = 0.037). These results suggest that both haplogroups confer independent protective
507 effects and that encoding dichotomous variables based on these haplogroups might be
508 instrumental for understanding their contributions in modulating disease progression, beyond the
509 *APOE* genotype alone.

510 To gain insight into the potential mechanism of action of the haplogroups and considering that
511 sequence differences between them are located in the promoter region of the gene and other
512 intronic enhancers (Supp. Fig. 2), we hypothesized that the specific combination of these variants
513 with the ε3 and ε4 *APOE* isoforms might influence the expression levels of each isoform in the
514 central nervous system.

515 However, one critical issue for testing this hypothesis is ensuring the accuracy of protein level
516 measurements in the CSF. We reviewed results from various methods tested in our laboratory,
517 including SomaScan[54], ELISA measures previously reported by us[35], and the Olink platform[36].
518 Our exploratory study concluded that the Olink method provides the most reliable measurements,
519 as it explains a much larger proportion of variance in APOE protein levels when considering well-
520 known demographic and molecular factors, compared to other techniques (Fig. 7). Furthermore,
521 this method was also able to replicate the classic observation of decreasing APOE protein levels
522 by genotype (ε2 > ε3 > ε4), as reported by other research groups[55] (Fig. 7).

523 Thus, we decided to test our hypothesis using the Olink measurements of APOE lipoprotein in
524 CSF, which were available for 499 individuals with known *APOE* haplogroups (Fig. 7C and D).
525 To this end, we performed a general linear model using standardized residuals of Olink APOE
526 protein levels (adjusted for age, sex, principal components of the variance in the CSF proteome,
527 and AD biomarkers) as the dependent variable, with *APOE* genotype and the presence of the E4A
528 and E3B haplogroups as independent variables.

529 As expected, our results showed a strong effect of the *APOE* genotype on CSF APOE protein
530 levels (p = 2.1E-33, Fig. 7). Moreover, the model identified independent, though more modest,
531 effects of the protective haplogroups on APOE protein levels (Fig. 7). Interestingly, the direction

of these effects was opposite for the two haplogroups. Specifically, the ε4A haplogroup was associated with a decrease in APOE protein levels in ε4 carriers ($p = 0.004$, Fig. 7), while the presence of the ε3B haplogroup showed the opposite effect, increasing APOE protein levels in the CSF ($p = 0.025$), even after adjusting for the remaining covariates and *APOE* genotype. We also examined the impact of the haplogroups on AD biomarkers (Supp. Fig. 5) and other proteins located on 19q13 surrounding the *APOE* locus (Supp. Fig. 5), detecting negligible effects of the ε4A and ε3B haplogroups on amyloid beta 42 levels, phospho-Tau81, NECTIN2, BCAM, and APOC1.

Finally, by analyzing the phase of the ε3 and ε4 isoforms in combination with upstream variants that configure the detected sub-haplotypes (Supp. Fig. 2), we concluded that the upstream variants defining the ε4A haplogroup specifically reduce the levels of the APOE ε4 protein isoform in the CSF. In contrast, the 3B haplogroup uniquely increases levels of the APOE ε3 isoform, particularly in individuals with the ε3/ε4 genotype. This finding could be significant, as it suggests that elevated ε4 allele expression may be detrimental (indicative of a toxic gain of function). Thus, reducing ε4 expression in the brain could potentially benefit individuals experiencing neurodegenerative processes.

4. Discussion

The *APOE* gene remains one of the most extensively studied genetic determinants of AD due to its significant impact on AD risk. Despite decades of research since the initial identification of the *APOE* ε4 allele as a major AD susceptibility factor[4,56], the precise molecular mechanisms by which different *APOE* variants influence disease pathogenesis are still not fully understood. Additionally, the extensive pleiotropy of the *APOE* locus, which modulates the levels of numerous proteins in plasma and CSF, several cardiovascular traits, and various neurodegenerative diseases, remains a question that continues to fascinate the scientific community.

Regarding the potential mechanism of action of *APOE* in relation with Alzheimer's disease, the controversy continues. While the ADGC consortium has recently reviewed the substantial evidence suggesting that the ε4 allele may act as a toxic gain-of-function variant[57], consensus is still lacking on whether its effects result from a toxic gain or a loss of function. In fact, a large study conducted within the Copenhagen Heart Study cohort suggested that reducing APOE protein expression in plasma, regardless of the protein isoform, could have detrimental effects, thereby supporting the notion of a potential loss-of-function mechanism for the APOE protein[58,59]. This alternative mechanism has been further supported by classic studies using *APOE* knockout (KO) mouse models, which demonstrated similar effects on cardiovascular traits, particularly arteriosclerosis[60,61]. Additionally, the loss-of-function mechanisms that explain the deleterious effect of the *APOE* ε4 allele are further supported by its demonstrated partial inability to transport lipoproteins effectively. Specifically, the ε4 isoform has reduced capacity to clear chylomicrons, VLDL particles, and other triglyceride-rich particles from the bloodstream, leading to dyslipidemia and associated cardiovascular outcomes, including atherosclerosis or hypertension[62,63]. However, while these robust demonstrations of the *APOE* ε4 isoform's hypomorphism (i.e., partial loss-of-function) role are linked to cardiovascular traits, they are considered independent of its effect on increasing dementia risk in humans[64].

Connecting these controversies with our findings, it is tempting to speculate that, depending on sub-haplotype configurations and the effects of variants influencing *APOE* expression across different tissues, we might explain the paradoxical toxic gain-of-function effects of the same alleles on central nervous system physiology, alongside hypomorphic loss-of-function effects in the periphery and vascular system for the same protein isoforms. However, large-scale population studies examining central and peripheral APOE levels, as well as molecular experiments to assign tissue specificity to the array of regulatory elements controlling *APOE* gene expression, are essential. Much more research is needed to understand the precise regulation of *APOE* gene expression in cells influencing both atherosclerosis and Alzheimer's disease risk[65].

583 The existence of cis-acting modulators increasing or decreasing the inherent risk associated to
584 *APOE* ε4 isoform is largely suspected and a few studies have been conducted to uncover the
585 potential modulators surrounding the *APOE* locus[66–69]. Importantly, there are evidences in the
586 literature suggesting a differential risk of the *APOE* isoforms in different human ancestries. For
587 example, different research work identified genetic markers upstream the *APOE* locus that might
588 explain the attenuated or increase effect of the *APOE* ε4 in an ancestry-dependent manner[6,8–10].
589 Connected with this idea, there are previous descriptions of common variants surrounding the
590 *APOE* locus that might alter the transcription of the mRNA[70,71]. Some of the common variants
591 we have detected during our sequencing study such rs40550977 or rs440446[72,73] among others
592 have been proposed as candidate factors modulating *APOE* expression with different levels of
593 evidence. In addition, the 3' region of the gene including the exon 4 (which encode the *APOE*
594 isoforms itself) have been proven to contain an epigenetic modulator acting as CpG element[74–77].
595 Additionally, regional and GWAS analyses of plasma APOE protein levels also support the notion
596 that a relatively large number of independent variants surrounding the locus might alter the
597 production of the *APOE* isoform even in a tissue specific manner[69,78].

598 Previous research reinforces the importance of phasing the potential modulators of *APOE*
599 expression with APOE protein isoforms because a differential distribution of different
600 haplogroups among populations might contribute to explain differences in the effect of the *APOE*
601 risk alleles in different genetic backgrounds. The use of ONT for generating long-range
602 haplotypes has proven instrumental in studying the *APOE* locus[22] and other regions across the
603 human genome[79], although the existence of technical problems must not be ignored[80]. However,
604 due to the small scale of previously reported studies, little is known about the clinical impact of
605 the detected sub-haplotypes or common haplogroups. To our knowledge, this is the largest study
606 indicating differential effects of sub haplotypes within the ε4 and ε3 isoforms. Our strategy has
607 permitted the accurate phasing of the nearest 5' (upstream) intragenic regulatory variants in the
608 locus with the classic protein isoforms, resolving the homoplasy of key variants upstream the
609 exon 4 of the gene. This enabled us to identify distinct sub-haplotypes for the common *APOE* ε3

610 and ε4 isoforms in the Spanish population, providing for the first time insights into how these
611 haplotypic configurations might modulate disease progression in MCI subjects and influence
612 APOE protein expression in the CSF.

613 Importantly, the haplotypic architecture identified using our method is somewhat consistent with
614 other small-scale studies previously published. *Fullerton et al.* used allele-specific amplification
615 techniques twenty-four years ago to resolve homoplasy variants in the *APOE* locus and detected
616 ε3 and ε4 sub haplotypes across diverse populations[21]. Our findings confirm these early
617 observations, particularly for the common haplogroups detected. More recently, *Abondio et al.*
618 demonstrated the presence of ε3 and ε4 sub haplotypes across geographic macro areas[81]. Although
619 their study used inference methods rather than long-read sequencing, the six core variants they
620 identified (rs440446, rs565782572, rs769449, rs769450, rs429358, and rs7412) represent a shared
621 haplotype cluster that aligns well with our findings. This study reinforces the relevance of long-
622 read sequencing for characterizing *APOE* haplogroups and validating the core haplotype
623 composition in different human populations. The study of intragenic haplogroups may be
624 particularly relevant because they are not dependent on ancestry background. This implies that
625 detected haplogroups are present across different ancestries, albeit in varying proportions.
626 Furthermore, the combination of these intragenic haplogroups with other distal and more
627 ancestry-dependent elements could complicate the interpretation of studies aimed at identifying
628 ancestry-specific modulators of the ε4 allele. In any case, large scale studies are necessary to
629 elucidate the intragenic haplogroup composition in different populations and its impact in ε4
630 expressivity.

631 From an evolutionary perspective, the distinct regulatory profiles detected for ε4A and ε3B
632 haplogroups raise intriguing questions about the evolutionary trajectory of the human *APOE* gene.
633 Previous studies have suggested that the ε4 allele may have been positively selected in ancestral
634 human populations due to its advantageous effects on lipid metabolism and immune response[62].
635 However, the emergence of sub haplotypes such as ε4A, which appears to mitigate some of the

636 adverse effects of the ε4 allele, suggests a more complex evolutionary history involving balancing
637 selection and regulatory adaptations.

638 The opposing effects of the ε4A and ε3B haplogroups on CSF APOE protein levels are
639 noteworthy. The ε4A haplogroup, associated with a significant reduction in APOE ε4 isoform
640 levels, suggests a potential attenuation of the well-established toxic gain-of-function model
641 proposed for the ε4 allele[57]. Previous research has shown that the ε4 allele exacerbates amyloid
642 pathology and tau hyper-phosphorylation, thereby contributing to neurodegeneration through
643 both direct and indirect mechanisms[18]. However, our findings indicate that specific functional
644 variants within the ε4A haplogroup might counteract this effect by reducing the levels of the toxic
645 ε4 isoform, potentially mitigating its deleterious impact in ε4 carriers.

646 Conversely, the ε3B haplogroup was associated with an increase in APOE ε3 protein levels,
647 particularly in individuals carrying the ε3/ε4 genotype. This suggests that certain variants within
648 this haplogroup may act elevating ε3 isoform expression and potentially compensating for the
649 presence of the ε4 allele. These isoform-specific regulatory effects highlight the complex
650 interplay between promoter variants and overall *APOE* function, underscoring the need to
651 consider haplotype configurations in future genetic studies.

652 Our study suggests that the combination of intragenic regulatory variants may fine-tune *APOE*
653 expression. This is supported by our observation that the ε4A and ε3B haplogroups exert opposing
654 effects despite sharing a common genetic background, including the T-allele of the homoplasic
655 rs405509 with the ε4 and ε3 alleles, respectively. However, each is uniquely combined with
656 different SNPs in intronic enhancers of the gene (the C-allele of rs440646 in haplotype 2 and its
657 derivatives [ε3B haplogroup] and the A-allele of rs769449 in the ε4A haplogroup). It is possible
658 that the specific combination of these variants within enhancer-active regions may induce
659 paradoxical effects on the gene expression of homoplasic variants. However, an exhaustive
660 functional characterization of the entire intragenic haplotypes is necessary to understand the
661 mechanisms associated with our findings.

Present results may have significant implications for the design of *APOE*-targeted gene therapies. Gene-editing strategies aimed at controlling ε4 expression, such as CRISPR-Cas9 or epigenome-based approaches[82], should carefully consider the potential for unintended effects on neighboring regulatory elements within the less harmful ε4A haplogroup. Likewise, therapeutic strategies that seek to upregulate ε3 expression must account for the presence of enhancers within the ε3B haplogroup, which could amplify ε3 isoform levels. Again, a comprehensive understanding of *APOE* haplotype structure is thus essential for the rational design of future isoform-specific interventions.

Despite the novel insights provided by our study, several limitations must be acknowledged. First, our sample size, while robust for a single-cohort study, may not capture the full spectrum of haplotypes present across diverse populations. Ancestry-specific effects, highlighted in several studies [11,83,84], also influence the impact of these *APOE* intragenic haplogroups on AD risk. Future studies should aim to replicate our findings in larger, multi-ethnic cohorts to assess the generalizability of these results. Second, measuring APOE protein levels in CSF remains challenging, as different platforms yield varying results depending on the assay method used (e.g., SomaScan, ELISA, Olink). Therefore, while our findings are robust, they should be interpreted with caution until validated using complementary methodologies. Third, our study focused exclusively on APOE protein levels in the CSF. The effects of these intragenic haplogroups in other tissues, particularly circulating APOE measured in plasma and produced in the liver, remain to be determined. A more comprehensive analysis of the impact of these intragenic haplogroups across multiple cell types and tissues is warranted. Fourth, this study centers on the *APOE* gene itself without accounting for the broader *APOE* region, which includes other potentially AD-related genes such as *TOMM40*, *BCAM*, or *APOC1*. These neighboring loci are known to interact with *APOE* at the regulatory level. Future research should aim to map the full haplotype structure of the entire 19q13 region to provide a more complete understanding of its functional and genetic landscape. Fifth, due to the inherent limitations of ONT, we focused our sequencing effort solely on SNV sites. Indel variants might also contribute to intragenic haplotype diversity of the *APOE*

689 locus. Further sequencing efforts combining short-read and long-read sequencing are necessary
690 to fully characterize the exact sequence composition of *APOE* intragenic haplogroups. This
691 information will be essential and relevant for elucidating the functional mechanisms behind the
692 observed effects on APOE protein isoform expressivity.

693 In summary, our high-resolution intragenic haplotype analysis of the *APOE* gene revealed
694 different alleles within the ε3 and ε4 backgrounds that have distinct effects on APOE protein
695 levels in the CSF and on MCI to AD progression. These findings provide new insights into the
696 functional regulation of *APOE* in AD, suggesting that precise modulation of isoform-specific
697 expression may be a promising therapeutic strategy. Further research is needed to explore the
698 broader implications of these haplogroups in AD risk and pathogenesis and to refine our
699 understanding of the regulatory mechanisms that govern *APOE* function in the brain. The
700 prevalence of specific haplotypes, alongside insights into their evolutionary trajectories and
701 possible Neanderthal genetic contributions, opens new research avenues aimed at elucidating the
702 genetic elements that influence the expression of *APOE* locus. Our findings lay the groundwork
703 for further exploration of the implications of intragenic *APOE* subhaplotypes in human
704 populations, emphasizing the importance of continued research in *APOE* regulation and its impact
705 in AD and other human conditions.

706 **5. References**

707 1. 2023 Alzheimer's disease facts and figures. *Alzheimer's & Dementia* **19**, 1598–1695
708 (2023).
709 2. 2024 Alzheimer's disease facts and figures. *Alzheimer's & Dementia* **20**, 3708–3821
710 (2024).
711 3. Rodríguez-Gómez, O., Palacio-Lacambra, M. E., Palasí, A., Ruiz-Laza, A. & Boada-
712 Rovira, M. Prevention of Alzheimer's Disease: A Global Challenge for Next Generation
713 Neuroscientists. *JAD* **42**, S515–S523 (2014).
714 4. Corder, E. H. *et al.* Gene Dose of Apolipoprotein E Type 4 Allele and the Risk of
715 Alzheimer's Disease in Late Onset Families. *Science* **261**, 921–923 (1993).

5. Moreno–Grau, S. *et al.* Exploring *APOE* genotype effects on Alzheimer's disease risk and amyloid β burden in individuals with subjective cognitive decline: The FundacioACE Healthy Brain Initiative (FACEHBI) study baseline results. *Alzheimer's & Dementia* **14**, 634–643 (2018).
6. Celis, K. *et al.* Ancestry-related differences in chromatin accessibility and gene expression of *APOE* ε4 are associated with Alzheimer's disease risk. *Alzheimer's & Dementia* **19**, 3902–3915 (2023).
7. Reitz, C., Pericak-Vance, M. A., Foroud, T. & Mayeux, R. A global view of the genetic basis of Alzheimer disease. *Nat Rev Neurol* **19**, 261–277 (2023).
8. Nuytemans, K. *et al.* Identifying differential regulatory control of *APOE* ε4 on African versus European haplotypes as potential therapeutic targets. *Alzheimer's & Dementia* **18**, 1930–1942 (2022).
9. Rajabli, F. *et al.* A locus at 19q13.31 significantly reduces the ApoE ε4 risk for Alzheimer's Disease in African Ancestry. *PLoS Genet* **18**, e1009977 (2022).
10. Griswold, A. J. *et al.* Increased *APOE* ε4 expression is associated with the difference in Alzheimer's disease risk from diverse ancestral backgrounds. *Alzheimer's & Dementia* **17**, 1179–1188 (2021).
11. Farrer, L. A. *et al.* Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA* **278**, 1349–1356 (1997).
12. Bertholim-Nasciben, L. *et al.* African origin haplotype protective for Alzheimer's disease in *APOE* ε4 carriers: exploring potential mechanisms. Preprint at https://doi.org/10.1101/2024.10.24.619909 (2024).
13. Alzheimer's Disease Neuroimaging Initiative (ADNI)* *et al.* Multiomics integrative analysis identifies APOE allele-specific blood biomarkers associated to Alzheimer's disease etiopathogenesis. *Aging* **13**, 9277–9329 (2021).
14. Belloy, M. E., Napolioni, V. & Greicius, M. D. A Quarter Century of APOE and Alzheimer's Disease: Progress to Date and the Path Forward. *Neuron* **101**, 820–838 (2019).

15. Holtzman, D. M., Herz, J. & Bu, G. Apolipoprotein E and Apolipoprotein E Receptors: Normal Biology and Roles in Alzheimer Disease. *Cold Spring Harbor Perspectives in Medicine* **2**, a006312–a006312 (2012).
16. Lynch, J. R., Morgan, D., Mance, J., Matthew, W. D. & Laskowitz, D. T. Apolipoprotein E modulates glial activation and the endogenous central nervous system inflammatory response. *Journal of Neuroimmunology* **114**, 107–113 (2001).
17. Butterfield, D. A. & Mattson, M. P. Apolipoprotein E and oxidative stress in brain with relevance to Alzheimer's disease. *Neurobiology of Disease* **138**, 104795 (2020).
18. Liu, C.-C., Liu, C.-C., Kanekiyo, T., Xu, H. & Bu, G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat Rev Neurol* **9**, 106–118 (2013).
19. Clarke, J. *et al.* Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotech* **4**, 265–270 (2009).
20. Van Dijk, E. L., Jaszczyszyn, Y., Naquin, D. & Thermes, C. The Third Revolution in Sequencing Technology. *Trends in Genetics* **34**, 666–681 (2018).
21. Fullerton, S. M. *et al.* Apolipoprotein E Variation at the Sequence Haplotype Level: Implications for the Origin and Maintenance of a Major Human Polymorphism. *The American Journal of Human Genetics* **67**, 881–900 (2000).
22. Maestri, S. *et al.* A Long-Read Sequencing Approach for Direct Haplotype Phasing in Clinical Settings. *IJMS* **21**, 9177 (2020).
23. Boada, M. *et al.* Design of a comprehensive Alzheimer's disease clinic and research center in Spain to meet critical patient and family needs. *Alzheimer's and Dementia* (2014) doi:10.1016/j.jalz.2013.03.006.
24. Petersen, R. C. Mild cognitive impairment as a diagnostic entity. in *Journal of Internal Medicine* (2004). doi:10.1111/j.1365-2796.2004.01388.x.
25. Petersen, R. C. *et al.* Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology* (1999) doi:10.1001/archneur.56.3.303.
26. Alegret, M. *et al.* Cut-off Scores of a Brief Neuropsychological Battery (NBACE) for Spanish Individual Adults Older than 44 Years Old. *PLoS ONE* **8**, e76436 (2013).

27. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. (American Psychiatric Association, 2013). doi:10.1176/appi.books.9780890425596.
28. McKhann, G. M. *et al.* 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 and Dementia* (2011) doi:10.1016/j.jalz.2011.03.005.
29. Román, G. C. *et al.* Vascular dementia: Diagnostic criteria for research studies: Report of the ninds-airen international workshop*. *Neurology* (1993) doi:10.1212/wnl.43.2.250.
30. Mesulam, M.-M., Grossman, M., Hillis, A., Kertesz, A. & Weintraub, S. The core and halo of primary progressive aphasia and semantic dementia. *Ann Neurol.* **54**, S11–S14 (2003).
31. McKeith, I. G. *et al.* Diagnosis and management of dementia with Lewy bodies: Fourth consensus report of the DLB Consortium. *Neurology* **89**, 88–100 (2017).
32. Vanderstichele, H. *et al.* Standardization of preanalytical aspects of cerebrospinal fluid biomarker testing for Alzheimer's disease diagnosis: A consensus paper from the Alzheimer's Biomarkers Standardization Initiative. *Alzheimer's & Dementia* **8**, 65–73 (2012).
33. Blennow, K. & Zetterberg, H. The Application of Cerebrospinal Fluid Biomarkers in Early Diagnosis of Alzheimer Disease. *Medical Clinics of North America* **97**, 369–376 (2013).
34. Orellana, A. *et al.* Establishing In-House Cutoffs of CSF Alzheimer's Disease Biomarkers for the AT(N) Stratification of the Alzheimer Center Barcelona Cohort. *IJMS* **23**, 6891 (2022).
35. Brosseron, F. *et al.* Soluble TAM receptors sAXL and sTyro3 predict structural and functional protection in Alzheimer's disease. *Neuron* **110**, 1009-1022.e4 (2022).
36. Martino Adami, P. V. *et al.* Matrix metalloproteinase 10 is linked to the risk of progression to dementia of the Alzheimer's type. *Brain* **145**, 2507–2517 (2022).
37. Moreno-Grau, S. *et al.* Genome-wide association analysis of dementia and its clinical endophenotypes reveal novel loci associated with Alzheimer's disease and three causality networks: The GR@ACE project. *Alzheimer's & Dementia* **15**, 1333–1347 (2019).

38. Rojas, I. de *et al.* Common variants in Alzheimer's disease and risk stratification by polygenic risk scores. *Nature Communications* **12**, 3417 (2021).
39. García-González, P. *et al.* A Novel Susceptibility Locus in *NFASC* Highlights Oligodendrocytes and Myelination in Progressive Supranuclear Palsy Pathology. Preprint at https://doi.org/10.1101/2024.06.21.24309279 (2024).
40. Leija-Salazar, M. *et al.* Evaluation of the detection of *GBA* missense mutations and other variants using the Oxford Nanopore MinION. *Molec Gen & Gen Med* **7**, e564 (2019).
41. Martin, M. *et al.* WhatsHap: fast and accurate read-based phasing. Preprint at https://doi.org/10.1101/085050 (2016).
42. Patterson, M. *et al.* W HATS H AP : Weighted Haplotype Assembly for Future-Generation Sequencing Reads. *Journal of Computational Biology* **22**, 498–509 (2015).
43. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution* **38**, 3022–3027 (2021).
44. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* (1993) doi:10.1093/oxfordjournals.molbev.a040023.
45. Harrison, P. W. *et al.* Ensembl 2024. *Nucleic Acids Research* **52**, D891–D899 (2024).
46. Morató, X. *et al.* Associations of plasma SMOC1 and soluble IL6RA levels with the progression from mild cognitive impairment to dementia. *Brain, Behavior, & Immunity - Health* **42**, 100899 (2024).
47. Darian, J. C., Kundu, R., Rajaby, R. & Sung, W.-K. Constructing telomere-to-telomere diploid genome by polishing haploid nanopore-based assembly. *Nat Methods* **21**, 574–583 (2024).
48. Taliun, D. *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* **590**, 290–299 (2021).
49. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).
50. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310–315 (2014).

828 51. Arboleda-Velasquez, J. F. *et al.* Resistance to autosomal dominant Alzheimer's disease in
829 an APOE3 Christchurch homozygote: a case report. *Nat Med* **25**, 1680–1683 (2019).
830 52. Hernandez, I. *et al.* Heterozygous *APOE* Christchurch in familial Alzheimer's disease
831 without mutations in other Mendelian genes. *Neuropathology Appl Neurobio* **47**, 579–582
832 (2021).
833 53. Haidar, M., Schmid, B., Ruiz, A., Ebneth, A. & Cabrera-Socorro, A. Generation of three
834 isogenic, gene-edited iPSC lines carrying the APOE-Christchurch mutation into the three
835 common APOE variants: APOE2Ch, APOE3Ch and APOE4Ch. *Stem Cell Research* **77**,
836 103414 (2024).
837 54. Puerta, R. *et al.* Head-to-Head Comparison of Aptamer- and Antibody-Based Proteomic
838 Platforms in Human Cerebrospinal Fluid Samples from a Real-World Memory Clinic
839 Cohort. *IJMS* **26**, 286 (2024).
840 55. Cruchaga, C. *et al.* Cerebrospinal fluid APOE levels: an endophenotype for genetic studies
841 for Alzheimer's disease. *Human Molecular Genetics* **21**, 4558–4571 (2012).
842 56. Corder, E. H. *et al.* Protective effect of apolipoprotein E type 2 allele for late onset
843 Alzheimer disease. *Nat Genet* **7**, 180–184 (1994).
844 57. Vance, J. M. *et al.* Report of the APOE4 National Institute on Aging/Alzheimer Disease
845 Sequencing Project Consortium Working Group: Reducing APOE4 in Carriers is a
846 Therapeutic Goal for Alzheimer's Disease. *Annals of Neurology* **95**, 625–634 (2024).
847 58. Rasmussen, K. L., Tybjærg-Hansen, A., Nordestgaard, B. G. & Frikke-Schmidt, R. *APOE*
848 and dementia – resequencing and genotyping in 105,597 individuals. *Alzheimer's &*
849 *Dementia* **16**, 1624–1637 (2020).
850 59. Rasmussen, K. L., Luo, J., Nordestgaard, B. G., Tybjærg-Hansen, A. & Frikke-Schmidt, R.
851 APOE and vascular disease: Sequencing and genotyping in general population cohorts.
852 *Atherosclerosis* **385**, 117218 (2023).
853 60. Plump, A. S. *et al.* Severe hypercholesterolemia and atherosclerosis in apolipoprotein E-
854 deficient mice created by homologous recombination in ES cells. *Cell* **71**, 343–353 (1992).

855 61. Zhang, S. H., Reddick, R. L., Piedrahita, J. A. & Maeda, N. Spontaneous
856 Hypercholesterolemia and Arterial Lesions in Mice Lacking Apolipoprotein E. *Science* **258**,
857 468–471 (1992).
858 62. Mahley, R. W. & Rall, S. C. Apolipoprotein E: Far More Than a Lipid Transport Protein.
859 *Annu. Rev. Genom. Hum. Genet.* **1**, 507–537 (2000).
860 63. Gregg, R. E. *et al.* Abnormal in vivo metabolism of apolipoprotein E4 in humans. *J. Clin.*
861 *Invest.* **78**, 815–821 (1986).
862 64. Prince, M. *et al.* The association between *APOE and dementia does not seem to be*
863 *mediated by vascular factors*. *Neurology* **54**, 397–397 (2000).
864 65. Van Den Kommer, T. N. *et al.* The role of extracerebral cholesterol homeostasis and ApoE
865 e4 in cognitive decline. *Neurobiology of Aging* **33**, 622.e17-622.e28 (2012).
866 66. Ramos, M. C. *et al.* Neuronal specific regulatory elements in apolipoprotein E gene
867 proximal promoter: *NeuroReport* **16**, 1027–1030 (2005).
868 67. Artiga, M. J. *et al.* Risk for Alzheimer's disease correlates with transcriptional activity of
869 the APOE gene. *Human Molecular Genetics* **7**, 1887–1892 (1998).
870 68. Town, T. *et al.* The −491A/T apolipoprotein E promoter polymorphism association with
871 Alzheimer's disease: independent risk and linkage disequilibrium with the known APOE
872 polymorphism. *Neuroscience Letters* **252**, 95–98 (1998).
873 69. Chen, Q. *et al.* Identification of a specific APOE transcript and functional elements
874 associated with Alzheimer's disease. *Mol Neurodegeneration* **19**, 63 (2024).
875 70. Lambert, J. C. *et al.* Contribution of *APOE* promoter polymorphisms to Alzheimer's
876 disease risk. *Neurology* **59**, 59–66 (2002).
877 71. Bullido, M. J. *et al.* A polymorphism in the regulatory region of APOE associated with risk
878 for Alzheimer's dementia. *Nat Genet* **18**, 69–71 (1998).
879 72. Rantalainen, V. *et al.* APOE ε4, rs405509, and rs440446 promoter and intron-1
880 polymorphisms and dementia risk in a cohort of elderly Finns—Helsinki Birth Cohort
881 Study. *Neurobiology of Aging* **73**, 230.e5-230.e8 (2019).

882 73. Prada, D. *et al.* Influence of multiple APOE genetic variants on cognitive function in a
883 cohort of older men – results from the Normative Aging Study. *BMC Psychiatry* **14**, 223
884 (2014).
885 74. Wang, S.-C., Oelze, B. & Schumacher, A. Age-Specific Epigenetic Drift in Late-Onset
886 Alzheimer's Disease. *PLoS ONE* **3**, e2698 (2008).
887 75. Foraker, J. *et al.* The APOE Gene is Differentially Methylated in Alzheimer's Disease. *JAD*
888 **48**, 745–755 (2015).
889 76. Lee, E.-G. *et al.* Redefining transcriptional regulation of the APOE gene and its association
890 with Alzheimer's disease. *PLoS ONE* **15**, e0227667 (2020).
891 77. Mancera-Páez, O. *et al.* Differential Methylation in APOE (Chr19; Exon Four; from
892 44,909,188 to 44,909,373/hg38) and Increased Apolipoprotein E Plasma Levels in Subjects
893 with Mild Cognitive Impairment. *IJMS* **20**, 1394 (2019).
894 78. Aslam, M. M. *et al.* Genome-wide analysis identifies novel loci influencing plasma
895 apolipoprotein E concentration and Alzheimer's disease risk. *Mol Psychiatry* **28**, 4451–
896 4462 (2023).
897 79. Bowden, R. *et al.* Sequencing of human genomes with nanopore technology. *Nat Commun*
898 **10**, 1869 (2019).
899 80. Laver, T. W. *et al.* Pitfalls of haplotype phasing from amplicon-based long-read
900 sequencing. *Sci Rep* **6**, 21746 (2016).
901 81. Abondio, P., Bruno, F. & Luiselli, D. Apolipoprotein E (APOE) Haplotypes in Healthy
902 Subjects from Worldwide Macroareas: A Population Genetics Perspective for
903 Cardiovascular Disease, Neurodegeneration, and Dementia. *CIMB* **45**, 2817–2831 (2023).
904 82. Kantor, B., Rittiner, J., Nicholls, P. J. & Chiba-Falek, O. *APOE* -targeted epigenome
905 therapy for late onset Alzheimer's disease. *Alzheimer's & Dementia* **18**, e060974 (2022).
906 83. Mahley, R. W. & Huang, Y. Apolipoprotein E Sets the Stage: Response to Injury Triggers
907 Neuropathology. *Neuron* **76**, 871–885 (2012).
908 84. Belloy, M. E. *et al.* *APOE* Genotype and Alzheimer Disease Risk Across Age, Sex, and
909 Population Ancestry. *JAMA Neurol* **80**, 1284 (2023).

85. Schwarz, J. M., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* **7**, 575–576 (2010).

6. Funding & Acknowledgments

The present work has been performed as part of the Doctoral thesis of Pablo García-González at the University of Barcelona (Barcelona, Spain). We would like to thank patients and controls who participated in this project. They were processed following standard operating procedures with the appropriate approval of the Ethical and Scientific Committee. Authors acknowledge the support of the Agency for Innovation and Entrepreneurship (VLAIO) grant Nº PR067/21 and Janssen for the HARPONE project and the ADAPTED project the EU/EFPIA Innovative Medicines Initiative Joint Undertaking Grant Nº 115975. Also, the Spanish Ministry of Science and Innovation, Proyectos de Generación de Conocimiento grants PID2021-122473OA-I00, PID2021-123462OB-I00 and PID2019-106625RB-I00. ISCIII, Acción Estratégica en Salud integrated in the Spanish National R+D+I Plan and financed by ISCIII Subdirección General de Evaluación and the Fondo Europeo de Desarrollo Regional (FEDER “Una manera de hacer Europa") grants PI17/01474, PI19/00335, PI22/01403 and PI22/00258. The support of CIBERNED (ISCIII) under the grants CB06/05/2004 and CB18/05/00010. The support from PREADAPT project, Joint Program for Neurodegenerative Diseases (JPND) grant Nº AC19/00097, and from DESCARTES project, German Research Foundation (DFG). The support of Fundación bancaria “La Caixa”, Fundación ADEY, Fundación Echevarne and Grifols SA (GR@ACE project). ACF received support from the Instituto de Salud Carlos III (ISCIII) under the grant Sara Borrell (CD22/00125). PGG was supported by CIBERNED employment plan (CNV-304-PRF-866). IdR is supported by the ISCIII 706 under the grant FI20/00215. AR is also supported by STAR Award. University of Texas System. Tx, United States, The South Texas ADRC. National Institute of Aging. National Institutes of Heath. USA. (P30AG066546), the Keith M. Orme and Pat Vigeon Orme Endowed Chair in Alzheimer’s and Neurodegenerative Diseases (2024-2025) and Patricia Ruth Frederick Distinguished Chair for Precision Therapeutics in Alzheimer’s and Neurodegenerative Diseases (2025-2028). Part of this study was also funded

937 by the German Federal Ministry of Education and Research (BMBF) within the EU program
938 JPND (Grant number: PreADAPT project 01ED2007A) and by BMBF (Grant numbers:
939 DESCARTES project 01EK2102B and 01EK2102A).

7. Figures

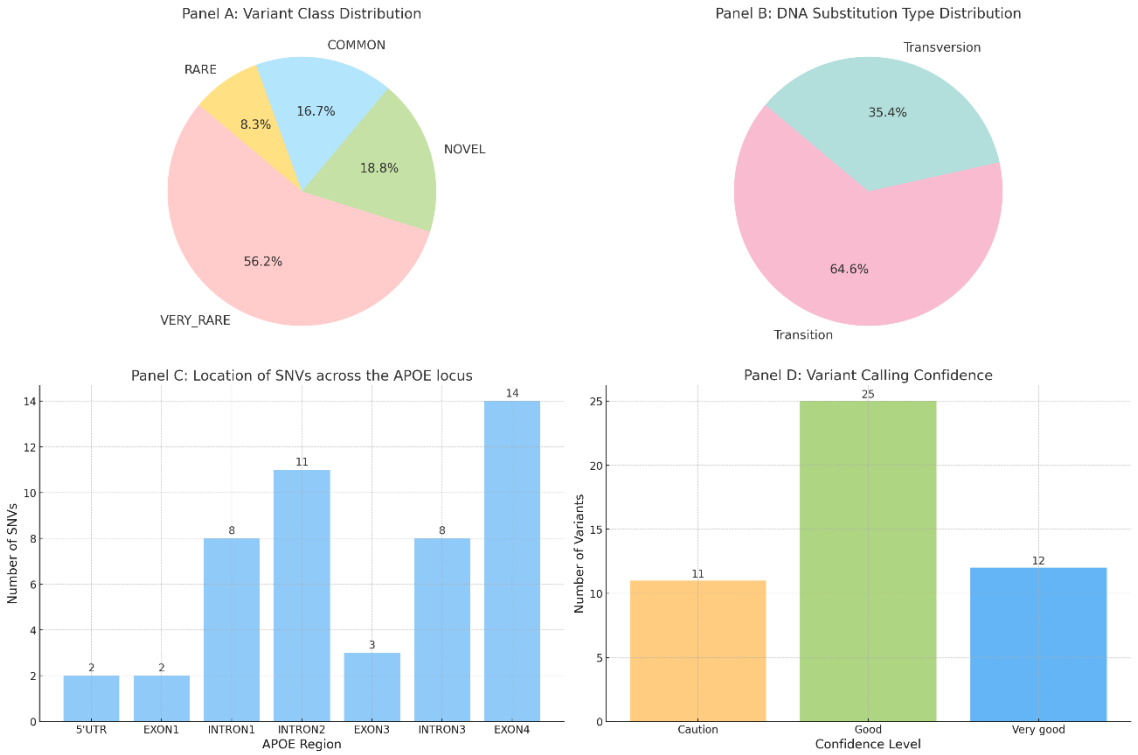


Figure 1: Summary of Variant Class, Substitution Types, Genomic Location, and Annotation Confidence for *APOE* Gene Variants.

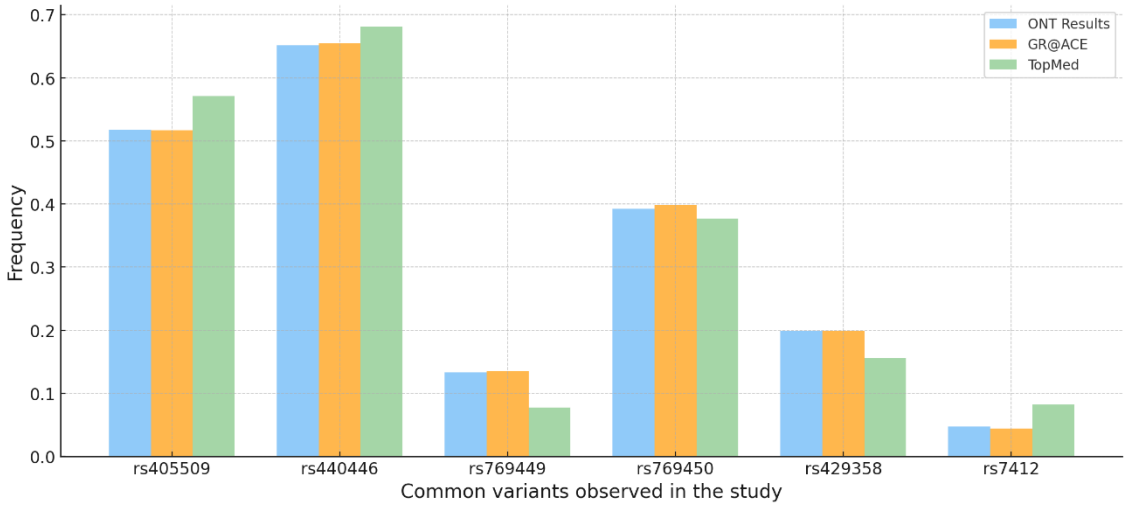


Figure 2: Comparison of common variant frequencies Across ONT Results, GR@ACE samples and the TOPMed database. The GR@ACE data reflects the frequencies in 1026 overlapping individuals genotyped using the Affymetrix Axiom Spain Biobank array and imputed with the TOPMed reference panel, as described elsewhere [37–39]. TOPMed frequencies reflect those reported by dbSNP build 151.

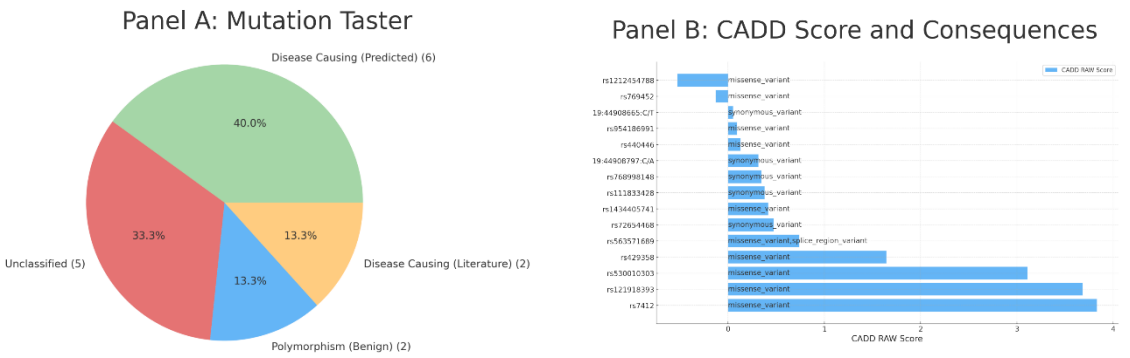
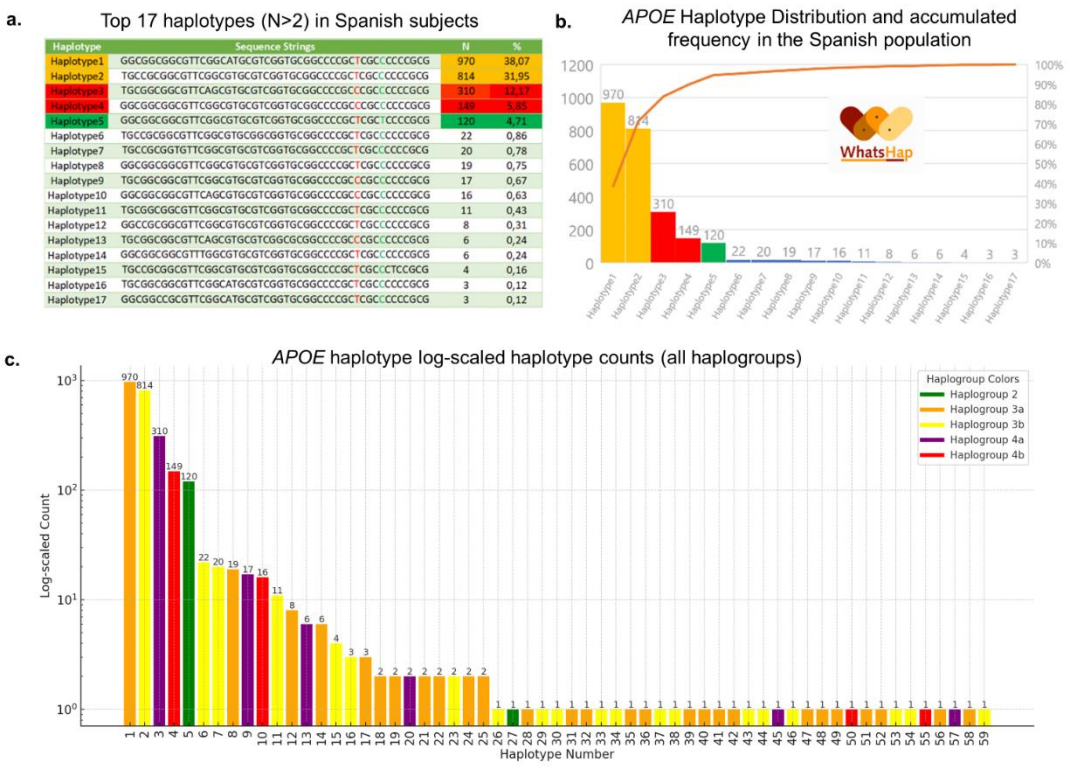


Figure 3: Comparison of Mutation Taster Predictions and CADD Scores for Observed *APOE* Coding Variants. Panel A illustrates the distribution of Mutation Taster[85] predictions for the coding variants observed in the gene. Variants are categorized into four groups: Disease Causing (Predicted), Disease Causing (based on Literature records), Polymorphism (Benign), and Unclassified. panel B presents the CADD scores for each observed variant in the *APOE* gene (most of them are very low). The scores are plotted alongside their corresponding consequence annotations predicted by VEP, such as missense variant, synonymous variant, or missense with potential splice region consequence.



Haplotype	Sequence Strings	N	%
Haplotype1	GGCGGCGGCGTTCGGCATGCGTCGGTGCGGCCCCGCTCGCCCCCGCG	970	38,07
Haplotype2	TGCCGCGGCGTTCGGCGTGCGTCGGTGCGGCCCCGCTCGCCCCCGCG	814	31,95
Haplotype3	TGCGGCGGCGTTCAGCGTGCGTCGGTGCGGCCCCGCCCGCCCCCGCG	310	12,17
Haplotype4	GGCGGCGGCGTTCGGCGTGCGTCGGTGCGGCCCCGCCCGCCCCCGCG	149	5,85
Haplotype5	GGCGGCGGCGTTCGGCGTGCGTCGGTGCGGCCCCGCTCGCCCCCGCG	120	4,71
Haplotype6	TGCCGCGGCGTTCGGCGTGCGGCGGTGCGGCCCCGCTCGCCCCCGCG	22	0,86
Haplotype7	TGCCGCGGTGTTCGGCGTGCGTCGGTGCGGCCCCGCTCGCCCCCGCG	20	0,78
Haplotype8	GGCGGCGGCGTTCGGCGTGCGTCGGTGCGGCCCCGCTCGCCCCCGCG	19	0,75
Haplotype9	TGCGGCGGCGTTCGGCGTGCGTCGGTGCGGCCCCGCCCGCCCCCGCG	17	0,67
Haplotype10	GGCGGCGGCGTTCAGCGTGCGTCGGTGCGGCCCCGCCCGCCCCCGCG	16	0,63
Haplotype11	TGCGGCGGCGTTCGGCGTGCGTCGGTGCGGCCCCGCTCGCCCCCGCG	11	0,43
Haplotype12	GGCCGCGGCGTTCGGCGTGCGTCGGTGCGGCCCCGCTCGCCCCCGCG	8	0,31
Haplotype13	TGCGGCGGCGTTCAGCGTGCGTCGGCGCGGCCCCGCCCGCCCCCGCG	6	0,24
Haplotype14	GGCGGCGGCGTTTGGCGTGCGTCGGTGCGGCCCCGCTCGCCCCCGCG	6	0,24
Haplotype15	TGCCGCGGCGTTCGGCGTGCGTCGGTGCGGCCCCGCTCGCCTCCGCG	4	0,16
Haplotype16	TGCGGCGGCGTTCGGCATGCGTCGGTGCGGCCCCGCTCGCCCCCGCG	3	0,12
Haplotype17	GGCGGCCGCGTTCGGCATGCGTCGGTGCGGCCCCGCTCGCCCCCGCG	3	0,12

Figure 4: *APOE* Haplotype Analysis in a Spanish Population. The figure illustrates the distribution and characteristics of the *APOE* haplotypes identified in a Spanish population. Panel A shows the top 17 haplotypes with a count greater than two, detailing their specific sequences, frequencies, and classifications based on ε2/ε3/ε4 isoform information. Panel B provides a graphical representation of the haplotype distribution, illustrating both the count of individuals carrying each haplotype (bars) and the accumulated frequency of these haplotypes in the population (line). Haplotype backgrounds are color-coded: red indicates ε4, orange indicates ε3, and green indicates ε2. Panel C provides the distribution of haplotype counts across different haplogroups using log-scaled visualization. Numbers at the tip of each bar indicate the count of individuals for each haplotype.

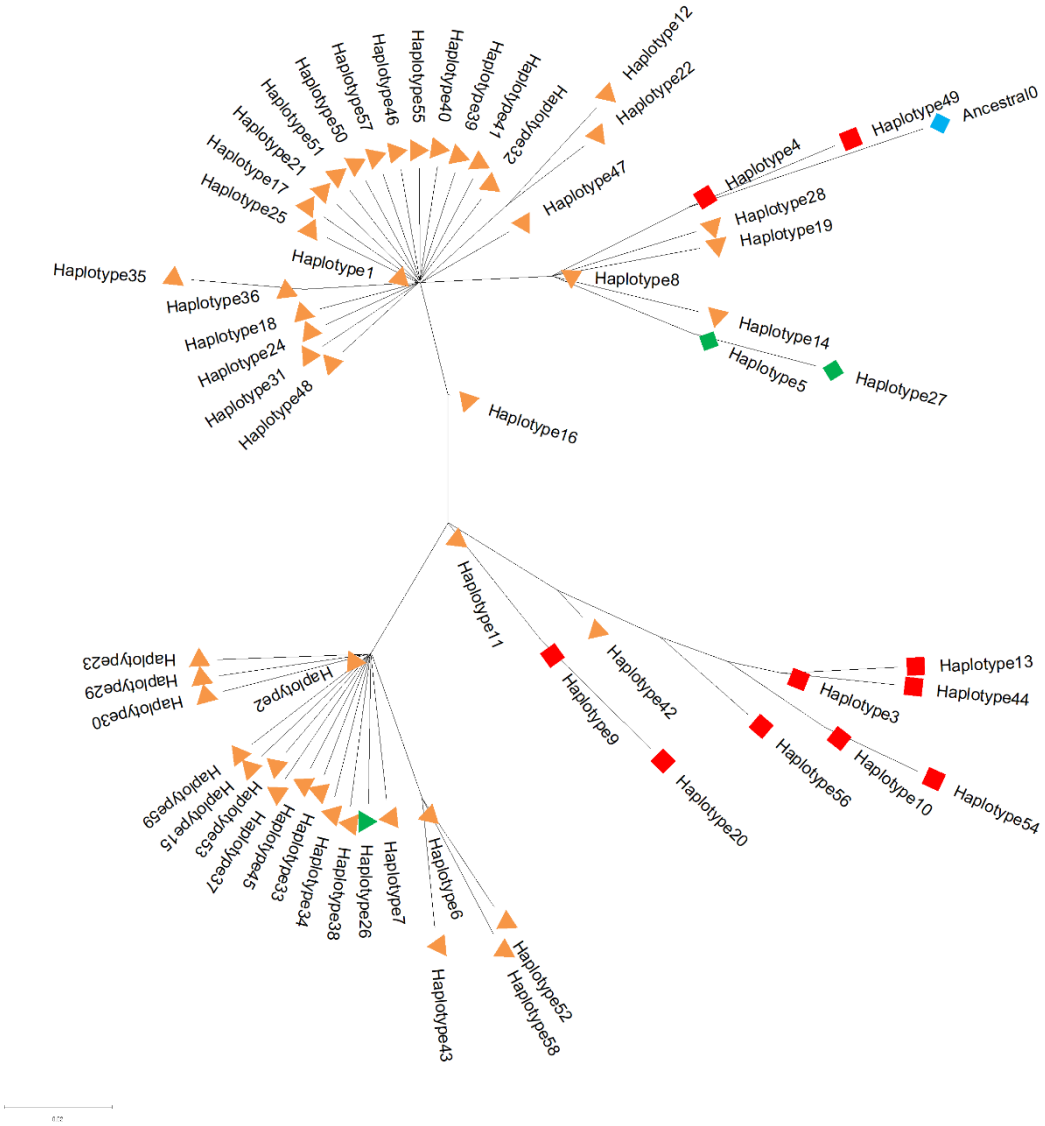
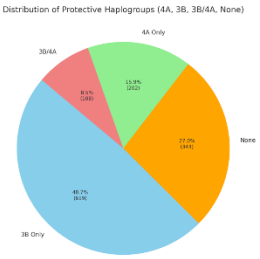


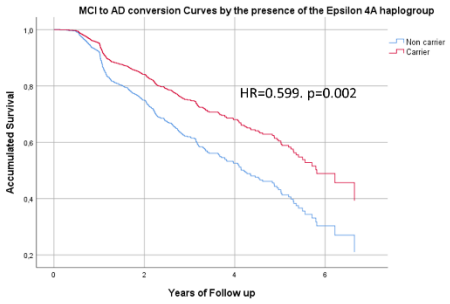
Figure 5: Phylogenetic Network of *APOE* intragenic Haplotypes Based on SNV Variations in the Spanish Population. This phylogenetic network illustrates the evolutionary relationships among the 59 *APOE* haplotypes identified in the Spanish population. Each node represents a distinct haplotype, and branches reflect the genetic distances based on single nucleotide variant (SNV) differences. The shapes and colors of the nodes correspond to specific isoform backgrounds: Orange triangles: ε3 background. Green triangle: Christchurch mutation under the isoform ε3 background. Green diamond: Haplogroup 2 (ε2 background). Red squares: ε4 background). Blue diamond is the ancestral allele sequence according to EMBL database. The

clustering of nodes indicates shared evolutionary paths, highlighting how specific haplotypes are grouped based on shared genetic variations.

A



C



B

Cox propotional Hazard risk Models adjusted by Age, sex, AD biomarkers in CSF and APOE genotypes

	B	SE	Wald	df	Sig.	H.R.	95,0% CI for H.R. lower	upper
Age	,057	,009	43,746	1	,000	1,058	1,041	1,076
Sex	,004	,109	,002	1	,968	1,004	,812	1,242
Abeta_42_CSF	-,002	,000	68,733	1	,000	,998	,998	,999
P_tau181_CSF	,008	,001	52,964	1	,000	1,008	1,006	1,011
Presence_4a	-,512	,169	9,203	1	,002	,599	,431	,834
Presence_3b	-,236	,113	4,336	1	,037	,790	,633	,986
APOE gentoype	,228	,114	3,996	1	,046	1,256	1,004	1,571

D

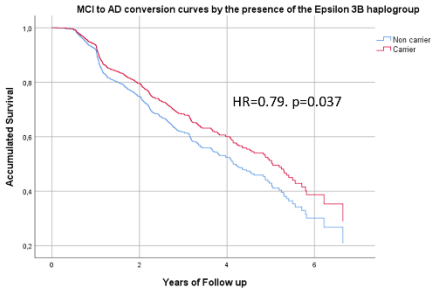


Figure 6: Distribution and Protective Effects of *APOE* Haplogroups (4A and 3B) on Disease Progression Analyzed Using Cox Proportional Hazard Models. Panel A: Pie chart showing the distribution of subjects based on the presence of 4A, 3B, both haplogroups (3B/4A), or neither (None). Panel B: Cox proportional hazard model displaying adjusted hazard ratios for the presence of 4A and 3B haplogroups along with other covariates (*APOE* genotype, demographics, and AD biomarkers). Panel C: Survival analysis highlighting the protective effect of the 4A haplogroup (HR = 0.599, p = 0.002). Panel D: Survival analysis showing the protective impact of the 3B haplogroup (HR = 0.79, p = 0.037).

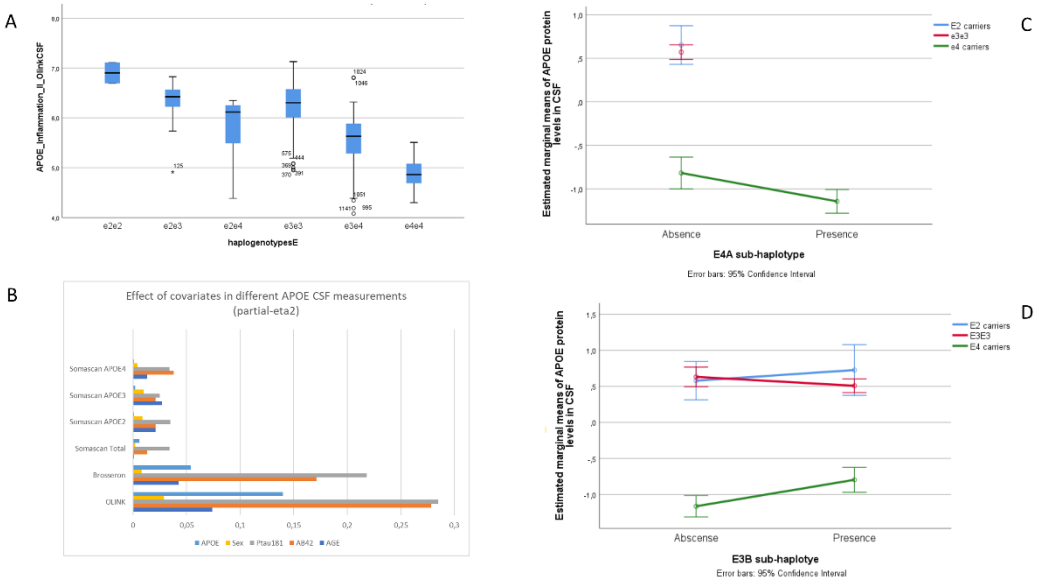


Figure 7: Effect of *APOE* Genotype and Haplogroups on CSF APOE Protein Levels Using Olink Measurements. This figure illustrates the impact of *APOE* genotype and the presence of ε4A and ε3B haplogroups on APOE protein levels in CSF measured using the Olink platform. Panel A: Boxplots showing CSF APOE protein levels across *APOE* genotypes (ε2/ε2, ε2/ε3, ε2/ε4, ε3/ε3, ε3/ε4, and ε4/ε4). The observed decreasing trend in APOE protein levels by genotype (ε2 > ε3 > ε4) replicates previous findings by other groups, confirming the reliability of the Olink measurements. Panel B: Partial eta-squared values indicating the effect of demographic and molecular covariates on CSF APOE levels measured using different platforms (Somascan APOE4, APOE3, and APOE2[54]; ELISA method[35]; and Olink[36]). The Olink platform explains the highest variance, supporting its use for accurate quantification of APOE protein levels. Panel C: Estimated marginal means of APOE protein levels by presence or absence of the ε4A haplogroup across different genotypes (ε2 carriers, ε3/ε3, and ε4 carriers). The ε4A haplogroup is associated with a significant decrease in APOE levels in ε4 carriers ($p = 0.004$), indicating a specific regulatory effect on the ε4 isoform. Panel D: Estimated marginal means of APOE protein levels by presence or absence of the ε3B haplogroup across different genotypes. The presence of the ε3B haplogroup is associated with increased APOE protein levels in CSF ($p = 0.025$), particularly in ε3/ε4 carriers, suggesting a specific upregulation of the ε3 isoform.

8. Tables

Table 1: Demographic, clinical, neurological, and biomarker Profiles of the ACE CSF Cohort

Demographics. N=1267	
Female (n, %)	733 (57.9%)
Age (years, sd)	72.68 (8.46)
BMI (kg/m3, sd)	26.9 (4.36)
Education (years, sd)	8.17 (4.77)
Syndromic diagnosis N(%)	
SCD (CDR=0)	79 (7.2%)
MCI (CDR=0.5)	795 (62.7%)
Dementia (CDR≥1)	381 (30.1%)
NA	12 (0.9%)
AT(N) classification N(%)	
A-T-N-	376 (29.7%)
A-T-N+	13 (1.0%)
A-T+N-	39 (3.1%)
A-T+N+	162 (12.8%)
A+T-N-	183 (14.4%)
A+T-N+	10 (0.8%)
A+T+N-	65 (5.1%)
A+T+N+	418 (33.0%)
NA	1 (0.1%)
APOE status (N,%)	
e2e2	4 (0.3%)
e2e3	91 (7.2%)
e2e4	24 (1.9%)
e3e3	734 (57.9%)
e3e4	354 (27.9%)
e4e4	60 (4.7%)
Other Clinical and Biochemical Data	
MMSE (mean, sd)	24.4 (4.41)
MCI converted/non conv (%)	354/366 (49.1%)
APOE CSF levels (n=499)(mean, sd)	6.02 (0.58)
pTau181 CSF levels	72.28 (45.35)
Amyloid b42 CSF levels	811.2 (393.5)
Total tau CSF Levels	469.2 (330.1)

Table 2: Summary of Single Nucleotide Variants Observed in the *APOE* Gene Region Using ONT Sequencing. rsIDs, genomic coordinates, reference alleles and alternative alleles are based on dbSNP build 151. CHR, chromosome; POS, position; REF, reference allele; ALT, alternative allele; AF, alternative allele frequency; AF TOPmed, TOPMed alternative allele frequency (as reported by dbSNP build 151); AF GR@ACE, alternative allele frequency in 1026 overlapping individuals from the GR@ACE[37–39] cohort; Qual/TotalReads, Mean QUAL to TotalReads ratio.

rsid	CHR	POS	REF	ALT	CLASS	AF	AF TOPMed	AF GR@ACE	APOE region	Substitution type	Confidence	Qual/TotalReads
rs405509	19	44905579	T	G	COMMON	0,517954723	0,57121241	0,51712963	5'UTR	Transversion	Very good	8,80
19:44905709:G:A	19	44905709	G	A	NOVEL	0,00039	0	0	5'UTR	Transition	Caution	1,91
rs9282609	19	44905856	C	T	RARE	0,00078064	0,00739042	0,000459559	INTRON1	Transition	Very good	4,82
rs440446	19	44905910	C	G	COMMON	0,651834504	0,68091393	0,65437788	EXON1	Transversion	Very good	7,29
rs563571689	19	44905923	G	A	VERY_RARE	0,00078064	0,00064507	0,000377929	EXON1	Transition	Good	5,91
rs877973	19	44906026	C	A	COMMON	0,00039032	0,01068744	0,000459559	INTRON1	Transversion	Very good	4,33
rs528229851	19	44906085	G	C	VERY_RARE	0,00117096	*NA*	0,000755858	INTRON1	Transversion	Good	3,85
19:44906188:G:A	19	44906188	G	A	NOVEL	0,00039	0	0	INTRON1	Transition	Caution	3,56
rs769448	19	44906322	C	T	COMMON	0,007806401	0,01403224	0,007359706	INTRON1	Transition	Very good	2,70
rs147236548	19	44906338	G	T	VERY_RARE	0,00039032	0,004881817	0	INTRON1	Transversion	Caution	5,23
rs568489254	19	44906449	T	C	VERY_RARE	0,00078064	0,00057339	0	INTRON1	Transition	Good	2,46
19:44906536:T:C	19	44906536	T	C	NOVEL	0,00039	0	0	INTRON1	Transition	Caution	2,60
rs143063029	19	44906731	C	T	RARE	0,00234192	0,00148127	0,002757353	INTRON2	Transition	Very good	3,05
rs769449	19	44906745	G	A	COMMON	0,133099141	0,0772012	0,135272392	INTRON2	Transition	Very good	4,03
rs1370426995	19	44906802	G	C	VERY_RARE	0,00039032	3,98E-05	0	INTRON2	Transversion	Good	5,04
rs1488788938	19	44906941	C	T	VERY_RARE	0,00039032	*NA*	0	INTRON2	Transition	Good	2,38
rs769450	19	44907187	G	A	COMMON	0,392661983	0,3771343	0,398344066	INTRON2	Transition	Very good	5,87
19:44907281:T:A	19	44907281	T	A	NOVEL	0,00039	0	0	INTRON2	Transversion	Caution	4,98
rs1383989611	19	44907326	G	T	VERY_RARE	0,00078064	3,19E-05	0,000377644	INTRON2	Transversion	Good	4,36
19:44907563:C:G	19	44907563	C	G	NOVEL	0,00039	0	0	INTRON2	Transversion	Caution	9,30
rs1307012197	19	44907629	G	A	VERY_RARE	0,00039032	1,59E-05	0	INTRON2	Transition	Good	4,17
rs769451	19	44907654	T	G	RARE	0,009758002	0,00535965	0,009199632	INTRON2	Transversion	Very good	6,37

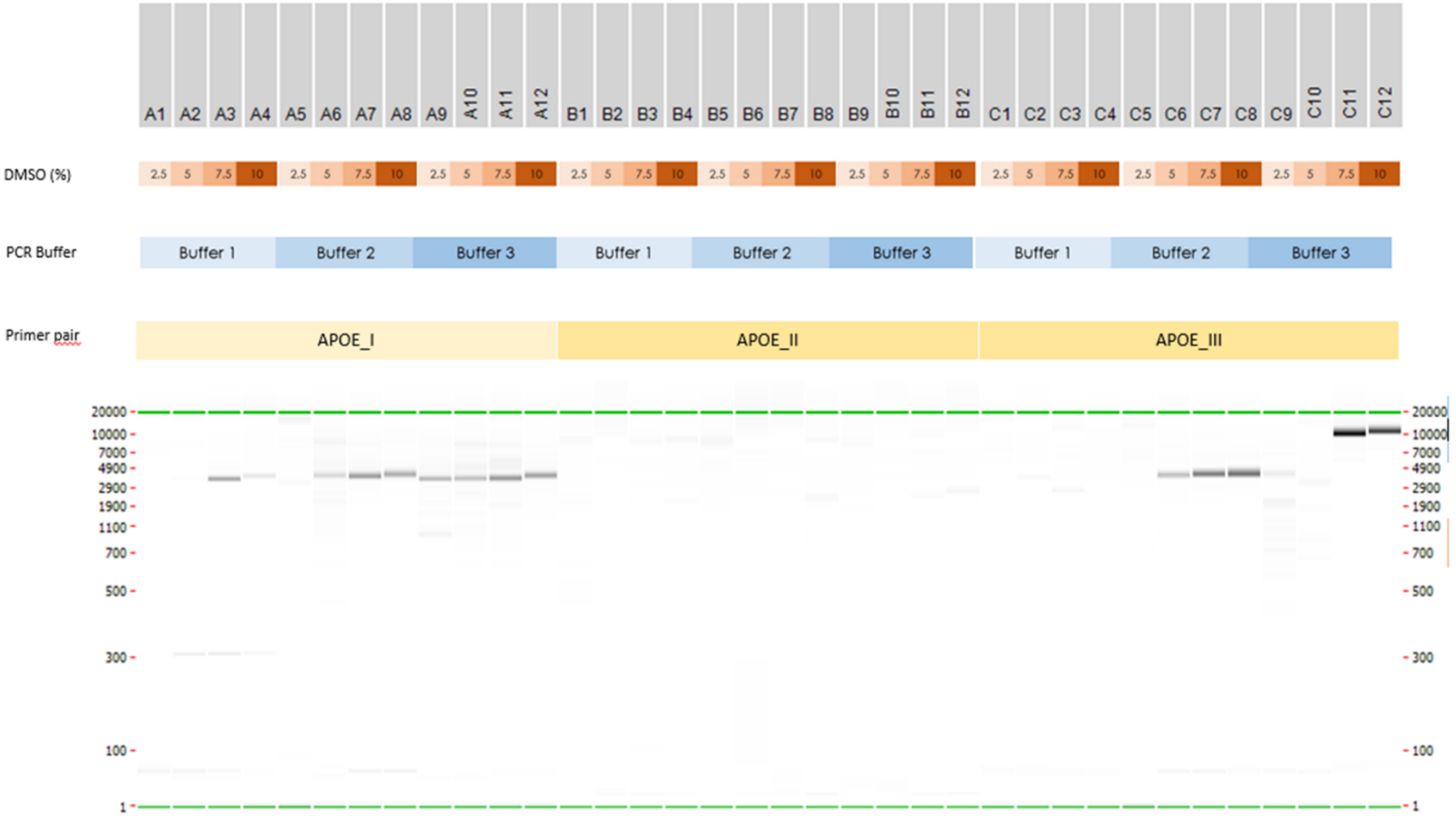
rs780533924	19	44907715	C	G	VERY_RARE	0,00039032	7,96E-06	0	INTRON2	Transversion	Good	4,63
rs111833428	19	44907785	G	A	VERY_RARE	0,00078064	0,00025484	0	EXON3	Transition	Good	1,75
rs1212454788	19	44907814	G	A	VERY_RARE	0,00039032	1,59E-05	0	EXON3	Transition	Good	2,37
rs769452	19	44907853	T	C	RARE	0,00273224	0,00134588	0,001382488	EXON3	Transition	Very good	2,79
19:44907977:G:C	19	44907977	G	C	NOVEL	0,00039	0	0	INTRON3	Transversion	Caution	5,58
rs560770492	19	44908033	C	T	VERY_RARE	0,00039032	9,56E-05	0,000377644	INTRON3	Transition	Good	4,76
rs746300299	19	44908048	G	A	VERY_RARE	0,00078064	*NA*	0	INTRON3	Transition	Good	3,32
rs952830931	19	44908176	G	A	VERY_RARE	0,00039032	3,98E-05	0	INTRON3	Transition	Good	2,07
rs865935864	19	44908380	C	T	VERY_RARE	0,00039032	1,59E-05	0	INTRON3	Transition	Good	3,35
rs1488164590	19	44908468	C	T	VERY_RARE	0,00039032	7,96E-06	0	INTRON3	Transition	Caution	0,20
rs1212541952	19	44908472	C	A	VERY_RARE	0,00039032	1,59E-05	0	INTRON3	Transversion	Good	3,72
19:44908477:C:T	19	44908477	C	T	NOVEL	0,00039	0	0	INTRON3	Transition	Caution	3,01
rs768998148	19	44908650	G	T	VERY_RARE	0,00039032	7,96E-06	0	EXON4	Transversion	Good	5,09
19:44908665:C:T	19	44908665	C	T	NOVEL	0,00039	0	0	EXON4	Transition	Caution	2,95
rs429358	19	44908684	T	C	COMMON	0,198672912	0,15559697	0,199355433	EXON4	Transition	Very good	2,50
rs121918393	19	44908756	C	A	VERY_RARE	0,00039032	7,96E-06	0	EXON4	Transversion	Good	4,27
rs954186991	19	44908762	G	A	VERY_RARE	0,00039032	7,96E-06	0	EXON4	Transition	Good	2,27
19:44908797:C:A	19	44908797	C	A	NOVEL	0,00039	0	0	EXON4	Transversion	Caution	5,22
rs7412	19	44908822	C	T	COMMON	0,047228728	0,08213876	0,044577206	EXON4	Transition	Very good	4,90
rs1434405741	19	44908837	C	G	VERY_RARE	0,00039032	7,96E-06	0	EXON4	Transversion	Good	3,64
rs72654468	19	44908947	C	T	VERY_RARE	0,00156128	0,00081231	0,000378788	EXON4	Transition	Good	1,72
rs530010303	19	44908999	C	T	VERY_RARE	0,00039032	5,57E-05	0	EXON4	Transition	Good	2,59
rs750138933	19	44909152	C	T	VERY_RARE	0,00039032	7,96E-06	0	EXON4	Transition	Good	3,15
rs749102800	19	44909209	G	A	VERY_RARE	0,00039032	3,19E-05	0,000377644	EXON4	Transition	Good	2,37
rs374329439	19	44909275	C	T	VERY_RARE	0,00078064	0,00045394	0,000755287	EXON4	Transition	Good	3,67
rs951089672	19	44909353	G	A	VERY_RARE	0,00039032	1,59E-05	0,000377644	EXON4	Transition	Good	3,47

Table 3. Complete list of the 17 Spanish *APOE* haplotypes with more than two occurrences in our population. Each row of the table corresponds to a distinct haplotype, specifying its sequence across the 48 positions within the APOE locus containing detected SNVs, along with its count, frequency and other classifications. The count column indicates the number of chromosomes carrying each specific haplotype. We also categorized the haplotypes into haplogroups, considering the common APOE isoform backgrounds (ε2, ε3, ε4) and the presence or absence of the alternative sequence (T-allele) of the homoplasic SNV rs405509 (Supp. Fig. 2), specifying the isoform background and the corresponding haplogroup for each haplotype. This classification framework aids in understanding the evolutionary relationships and potential health implications associated with each haplotype.

Haplotype	Sequence	Count	Frequency (%)	Background	Haplogroup
Haplotype1	GGCGGCGGCGTTCGGCATGCGTCGGTGCGGCCCCGCTCGCCCCCCGCG	970	38,069	ε3	3a
Haplotype2	TGCCGCGGCGTTCGGCGTGCGTCGGTGCGGCCCCGCTCGCCCCCCGCG	814	31,947	ε3	3b
Haplotype3	TGCGGCGGCGTTCAGCGTGCGTCGGTGCGGCCCCGCCCGCCCCCCGCG	310	12,166	ε4	4a
Haplotype4	GGCGGCGGCGTTCGGCGTGCGTCGGTGCGGCCCCGCCCGCCCCCCGCG	149	5,848	ε4	4b
Haplotype5	GGCGGCGGCGTTCGGCGTGCGTCGGTGCGGCCCCGCTCGCTCCCCGCG	120	4,710	ε2	2
Haplotype6	TGCCGCGGCGTTCGGCGTGCGGCGGTGCGGCCCCGCTCGCCCCCCGCG	22	0,863	ε3	3b
Haplotype7	TGCCGCGGTGTTCGGCGTGCGTCGGTGCGGCCCCGCTCGCCCCCCGCG	20	0,785	ε3	3b
Haplotype8	GGCGGCGGCGTTCGGCGTGCGTCGGTGCGGCCCCGCTCGCCCCCCGCG	19	0,746	ε3	3a
Haplotype9	TGCGGCGGCGTTCGGCGTGCGTCGGTGCGGCCCCGCCCGCCCCCCGCG	17	0,667	ε4	4a
Haplotype10	GGCGGCGGCGTTCAGCGTGCGTCGGTGCGGCCCCGCCCGCCCCCCGCG	16	0,628	ε4	4b
Haplotype11	TGCGGCGGCGTTCGGCGTGCGTCGGTGCGGCCCCGCTCGCCCCCCGCG	11	0,432	ε3	3b
Haplotype12	GGCCGCGGCGTTCGGCGTGCGTCGGTGCGGCCCCGCTCGCCCCCCGCG	8	0,314	ε3	3a
Haplotype13	TGCGGCGGCGTTCAGCGTGCGTCGGCGCGGCCCCGCCCGCCCCCCGCG	6	0,235	ε4	4a
Haplotype14	GGCGGCGGCGTTTGGCGTGCGTCGGTGCGGCCCCGCTCGCCCCCCGCG	6	0,235	ε3	3a
Haplotype15	TGCCGCGGCGTTCGGCGTGCGTCGGTGCGGCCCCGCTCGCCCTCCGCG	4	0,157	ε3	3b
Haplotype16	TGCGGCGGCGTTCGGCATGCGTCGGTGCGGCCCCGCTCGCCCCCCGCG	3	0,118	ε3	3b

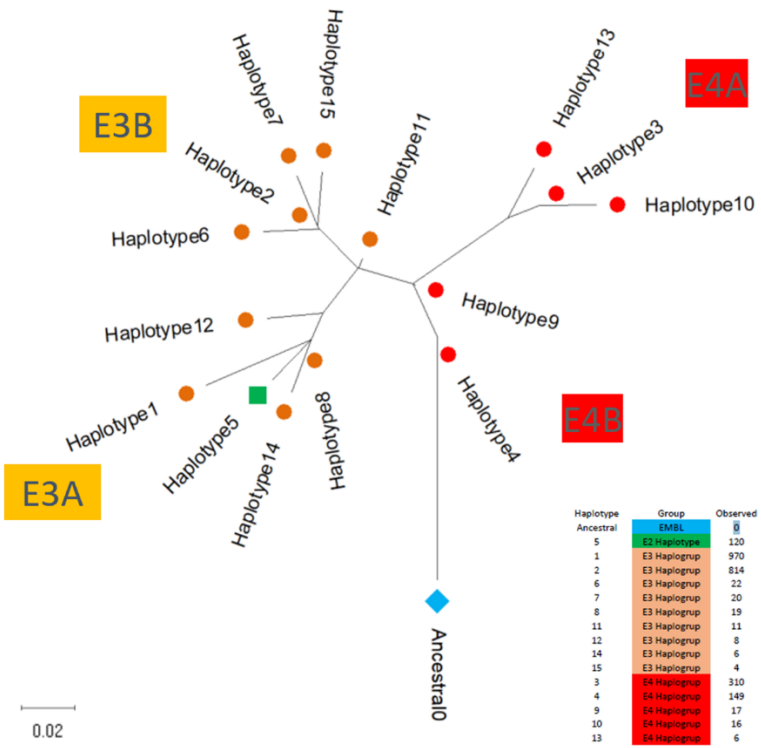
Haplotype17	GGCGGCCGCGTTCGGCATGCGTCGGTGCGGCCCCGCTCGCCCCCCGCG	3	0,118	ε3	3a

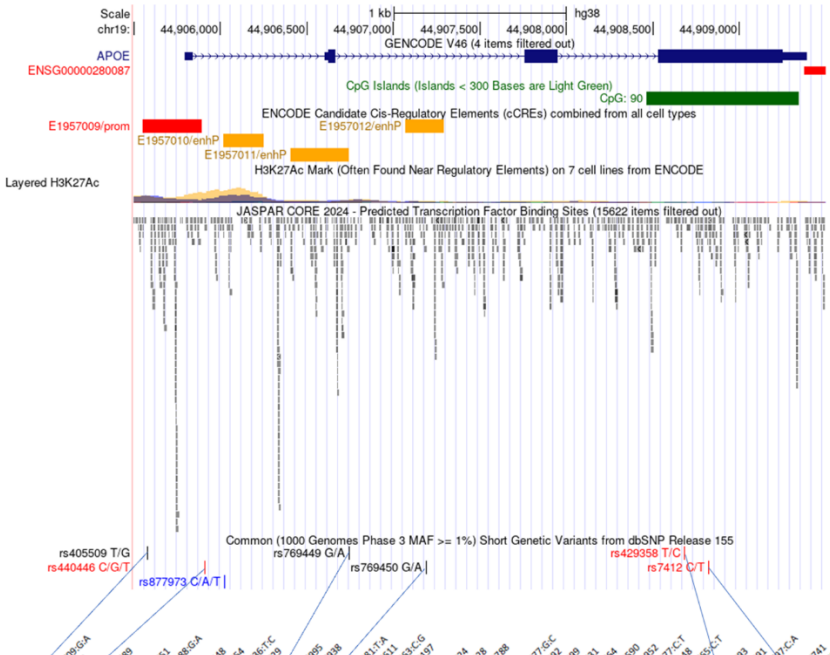
9. Supplementary Figures



Supplementary figure 1. PCR optimization. Different conditions tested in a single sample to search for optimal reaction. The optimal conditions were primer pair APOE_III with 10% DMSO and buffer 2. The PCR products were examined by automated capillary electrophoresis using the LabChip GX Touch Nucleic Acid Analyzer (Revvity, CLS138162) and the DNA 12K Reagent Kit (Revvity, 760569).

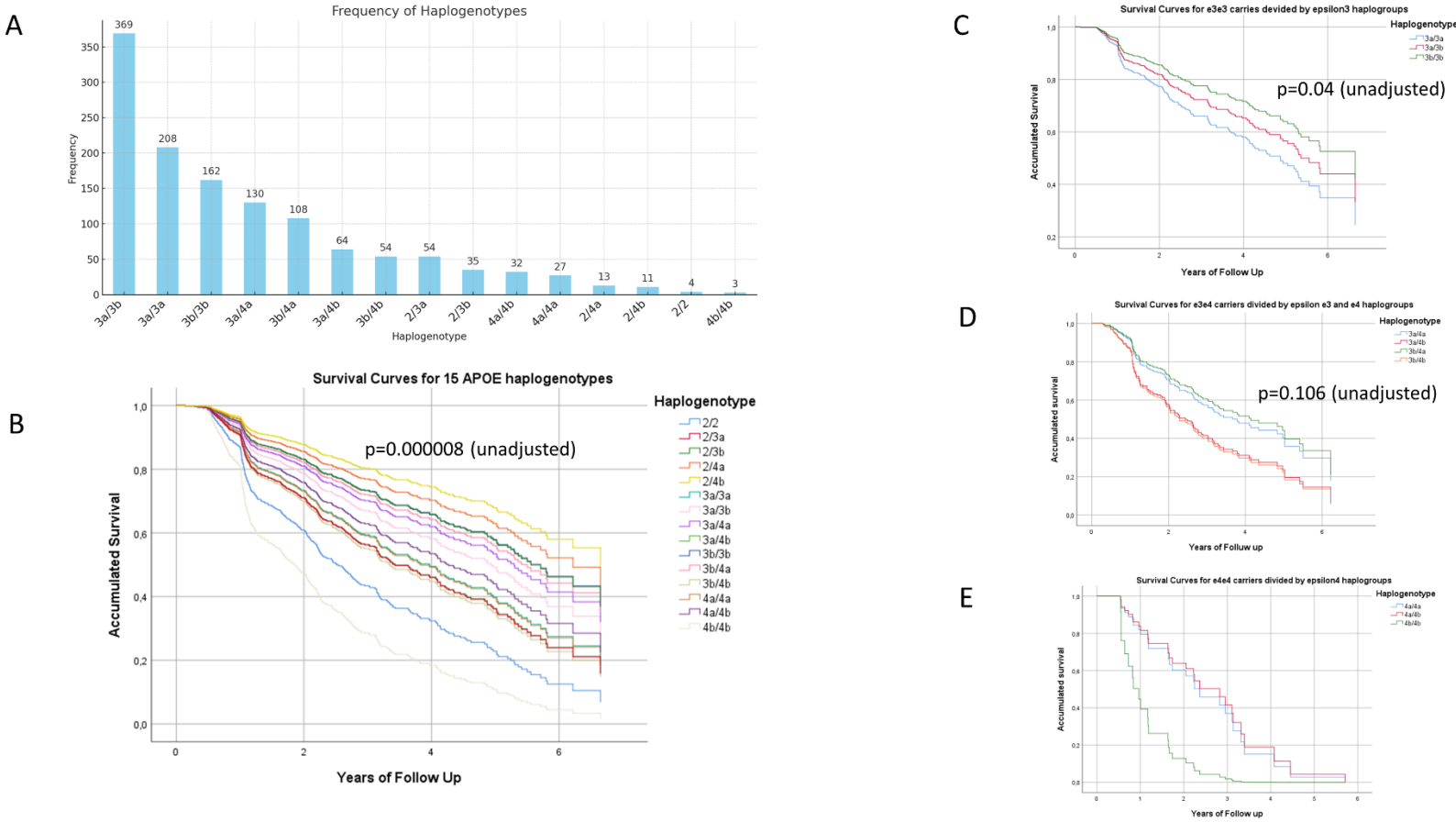
Evolutionary analysis of APOE haplotypes observed in the Spanish population restricted to top 15 haplotypes (Maximum Likelihood method)





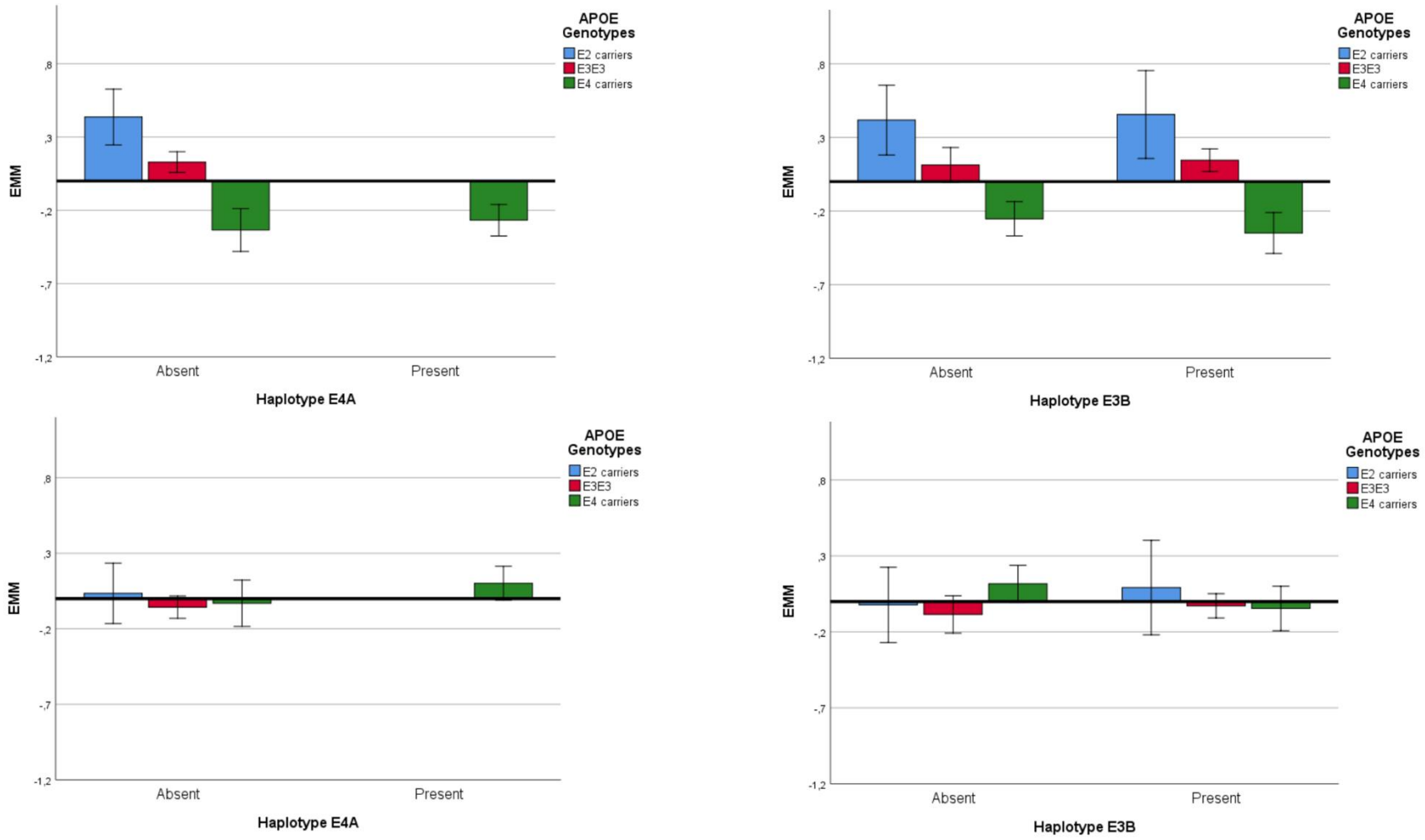
Haplotype	Group	Observed	%	rs405509	19:44905709:G:A	rs9282609	rs440446	rs535371689	rs877973	rs532239851	19:44906186:G:A	rs769448	rs147236548	rs568489254	19:44906536:T:C	rs143060029	rs769449	rs113704426995	rs1488788938	rs769450	19:44907281:T:A	rs1383989611	19:44907569:C:G	rs13307011197	rs769451	rs780533924	rs111833428	rs121245788	rs769452	19:44907977:G:C	rs540770492	rs746380299	rs952830931	rs865938864	rs1488764590	rs121241952	19:44908477:C:T	rs768998148	19:44908665:C:T	rs429358	rs121918393	rs954186991	19:44909079:T:C/A	rs7412	rs143440541	rs765454468	rs530010303	rs750138933	rs749103800	rs374359439	rs955089672
Ancestral	EMBL	0	0	G	G	C	G	G	C	G	G	C	G	T	T	C	G	G	C	G	T	G	C	G	T	C	G	A	T	G	C	G	G	C	C	C	C	G	C	C	G	C	C	C	C	C	C	A	C	G
5	E2 Haplotype	120	4,71																									G									T			T			G		
1	E3 Haplogrup	970	38,07																A									G									T						G		
2	E3 Haplogrup	814	31,95	T			C																					G									T						G		
6	E3 Haplogrup	22	0,86	T			C																	G				G									T						G		
7	E3 Haplogrup	20	0,78	T			C				T																	G									T						G		
8	E3 Haplogrup	19	0,75																									G									T						G		
11	E3 Haplogrup	11	0,43	T																								G									T						G		
12	E3 Haplogrup	8	0,31				C																					G									T						G		
14	E3 Haplogrup	6	0,24												T													G									T						G		
15	E3 Haplogrup	4	0,16	T			C																					G									T					T	G		
3	E4 Haplogrup	310	12,17	T												A												G															G		
4	E4 Haplogrup	149	5,85																									G															G		
9	E4 Haplogrup	17	0,67	T																								G															G		
10	E4 Haplogrup	16	0,63													A												G															G		
13	E4 Haplogrup	6	0,24	T												A												G	C															G		

Supplementary Figure 2: **Phylogenetic and Genomic Analysis of the most common APOE Haplotypes in the Spanish Population.** This figure illustrates the evolutionary relationships and genomic context of the top 15 most frequent *APOE* haplotypes identified in the Spanish population. The left panel shows a phylogenetic network constructed using the Maximum Likelihood (ML) method, where haplotypes are grouped into distinct haplogroups based on their isoform backgrounds (E3A, E3B, E4A, and E4B). Each haplotype is represented as a node, color-coded according to its classification, with orange nodes for E3 haplotypes, red for E4 haplotypes, and green for the epsilon 2 haplogroup. The ancestral haplotype, labelled as "Ancestral0" (blue diamond), serves as the root from which other haplotypes diverged. The right panel provides a detailed map of the 4kb genomic region (chr19: 44,905,500-44,909,500) covering the *APOE* gene, with annotations highlighting promoters, enhancers, CpG islands, and transcription factor binding sites that may impact gene regulation. A summary table below the phylogenetic network details the SNV positions for each haplotype, observed count, and relative frequency in the population. This figure effectively integrates phylogenetic analysis with functional genomic annotations, providing insights into the evolutionary diversification and regulatory landscape of *APOE* haplotypes within the Spanish population.

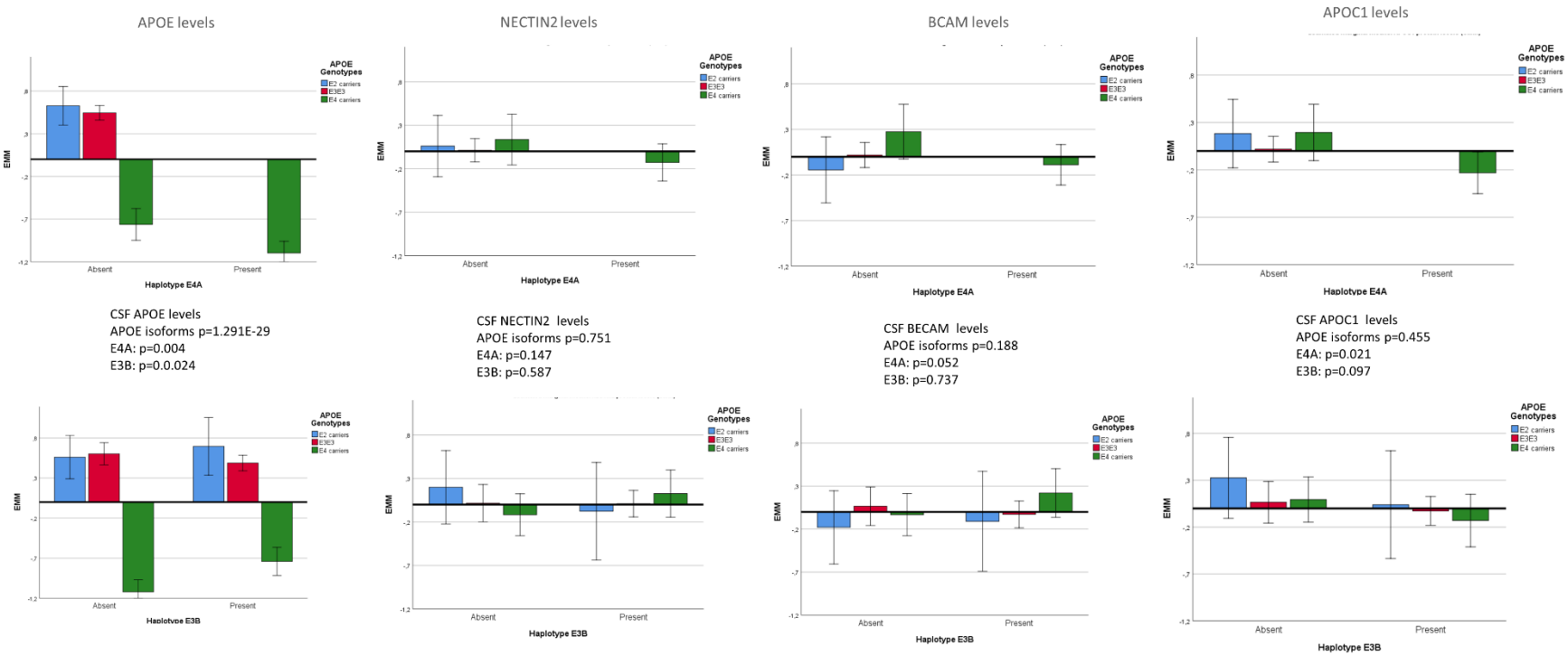


Supplementary Figure 3: Distribution and Survival Analysis of APOE Haplogenotypes. Panel A shows a bar chart depicting the frequency distribution of 15 different *APOE* haplogenotypes, with the x-axis representing distinct haplogenotypes and the y-axis indicating their respective counts within the sample population. The most common haplogenotypes are 3a/3b, 3a/3a, and 3b/3b. Panel B presents survival curves illustrating accumulated survival probabilities over

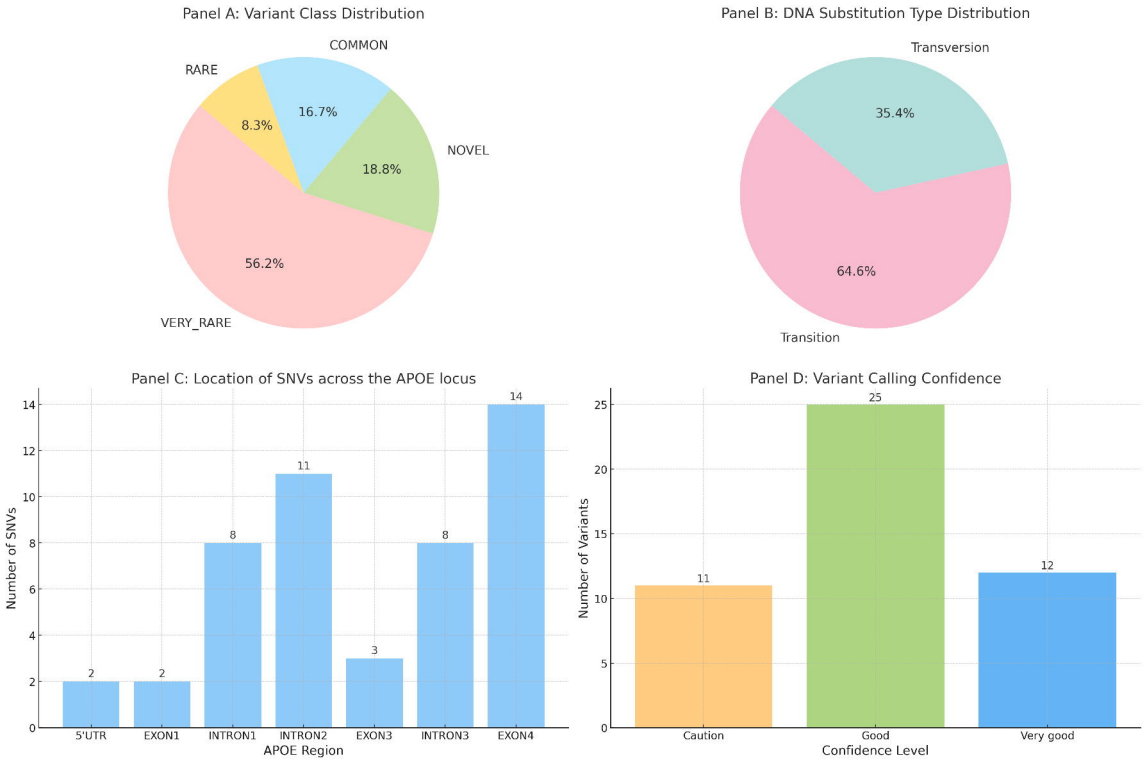
time for the 15 different *APOE* haplogenotypes, where each coloured line corresponds to a unique haplogenotype, demonstrating variation in survival across different genetic profiles. Panels C-E display survival curves for specific subgroups of carriers: Panel C shows ε3/ε3 carriers divided by epsilon3 haplogroups (3a/3a, 3a/3b, 3b/3b), Panel D shows ε3/ε4 carriers divided by epsilon3 and epsilon4 haplogroups (3a/3a, 3a/3b, 3b/3b), and Panel E shows e4e4 carriers divided by epsilon4 haplogroups (4a/4a, 4a/4b, 4b/4b), revealing varying survival trends based on distinct haplogenotype groupings.

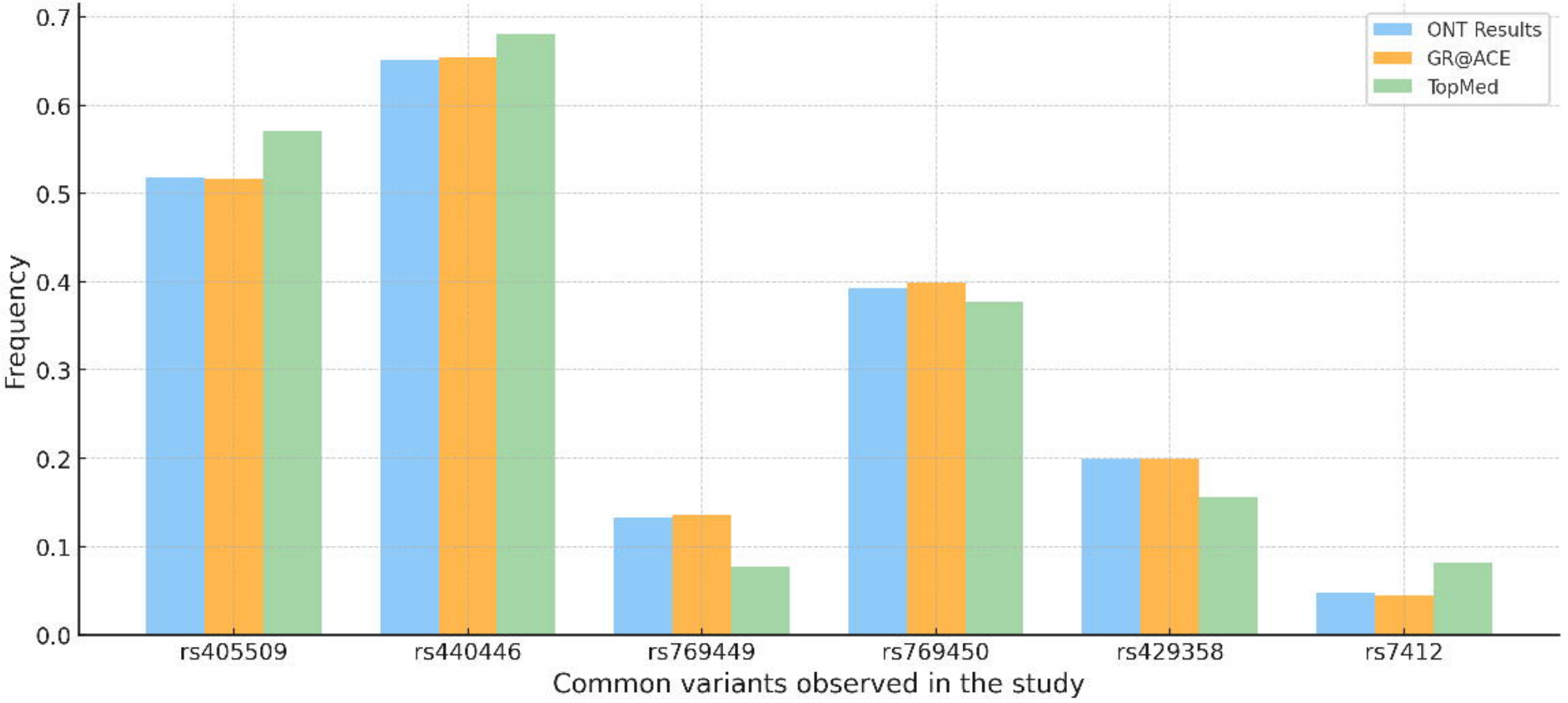


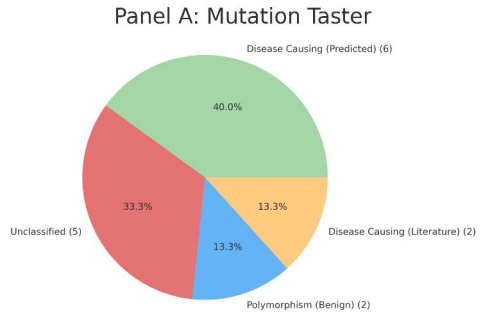
Supplementary Figure 4. Influence of haplogroups in AD biomarkers.

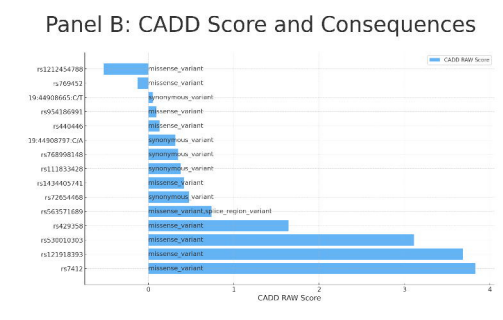


Supplementary Figure 5. Influence of haplogroups in other 19q13 genes (Olink platform).


Panel A: Variant Class Distribution
COMMON
RARE
NOVEL
VERY_RARE
16.7%
8.3%
18.8%
56.2%
Panel B: DNA Substitution Type Distribution
Transversion
Transition
35.4%
64.6%
Panel C: Location of SNVs across the APOE locus
Number of SNVs
APOE Region
5'UTR
EXON1
INTRON1
INTRON2
EXON3
INTRON3
EXON4
2
2
8
11
3
8
14
Panel D: Variant Calling Confidence
Number of Variants
Confidence Level
Caution
Good
Very good
11
25
12


ONT Results
GR@ACE
TopMed
Frequency
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
rs405509
rs440446
rs769449
rs769450
rs429358
rs7412
Common variants observed in the study


Panel A: Mutation Taster
Disease Causing (Predicted) (6)
40.0%
33.3%
13.3%
Unclassified (5)
Disease Causing (Literature) (2)
13.3%
Polymorphism (Benign) (2)

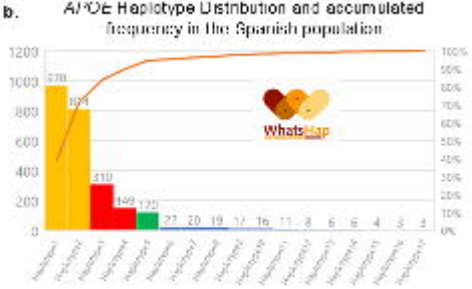

Panel B: CADD Score and Consequences
CADD RAW Score
rs1212454788
rs769452
19:44908665:C/T
rs954186991
rs440446
19:44908797:C/A
rs768998148
rs111833428
rs1434405741
rs72654468
rs563571689
rs429358
rs200100303
rs121918393
rs7412
missense_variant
missense_variant
synonymous_variant
missense_variant
missense_variant
synonymous_variant
synonymous_variant
synonymous_variant
missense_variant
synonymous_variant
missense_variant,splice_region_variant
missense_variant
missense_variant
missense_variant
missense_variant
0
1
2
3
4
CADD RAW Score

a.

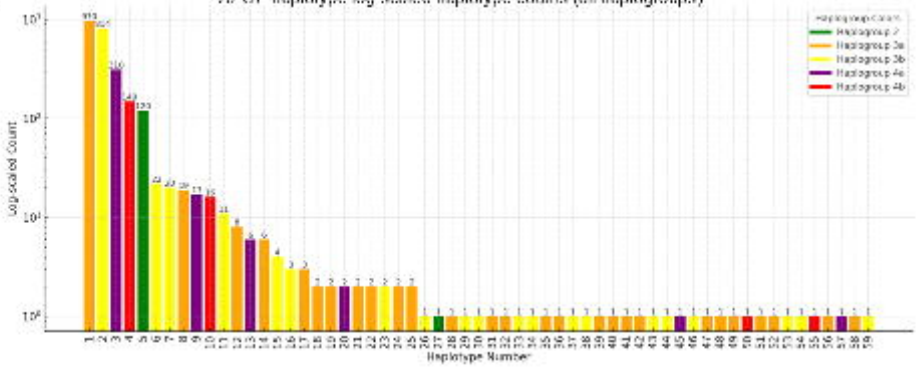
Top 17 haplotypes (N=2) in Spanish subjects

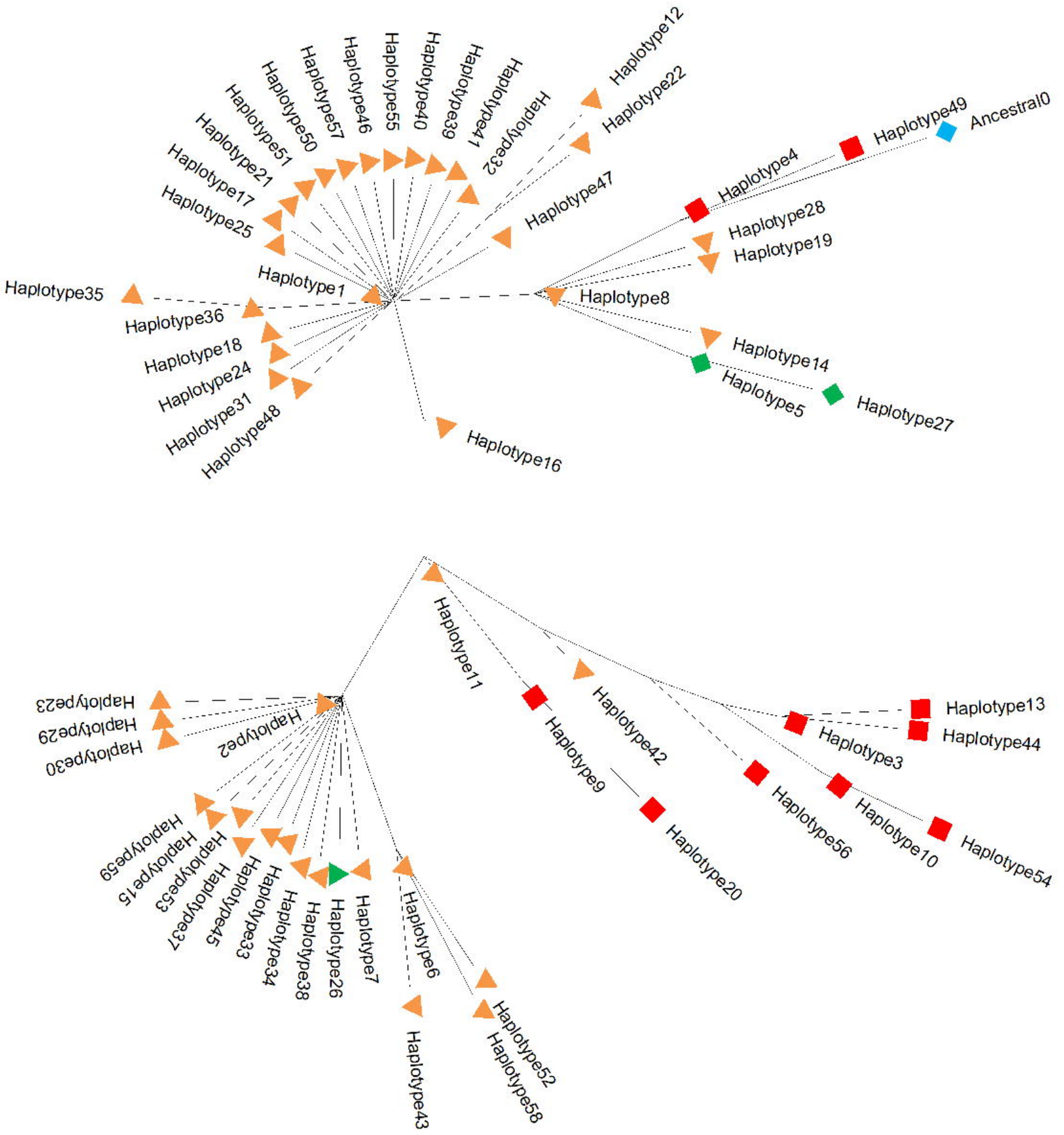
Haplotype	Sequence String	N	%
Haplotype1	GGCGCGCGGCGTTCGGCATGCGTCGGTGCGCGCGCGTCGGCCGCCGCG	972	36,07
Haplotype2	TGCCGCGCGTTCGGCGTGCGTCGGTGCGCGCGCGTCGGCCGCCGCG	834	31,55
Haplotype3	TGCGCGCGGCGTTCAGCGTGCGTCGGTGCGCGCGCGTCGGCCGCCGCG	433	16,07
Haplotype4	GGCGCGCGCGTTCGCGGTGCGTCGGTGCGCGCGCGCGGCCGCCGCG	188	6,98
Haplotype5	GGCGCGCGGCGTTCGCGGTGCGTCGGTGCGCGCGCGTCGGCCGCCGCG	128	4,75
Haplotype6	TGCGCGCGGCGTTCGGCGTGCGTCGGTGCGCGCGCGTCGGCCGCCGCG	22	0,82
Haplotype7	TGCGCGCGGCGTTCAGCGTGCGTCGGTGCGCGCGCGTCGGCCGCCGCG	20	0,74
Haplotype8	GGCGCGCGCGTTCGCGGTGCGTCGGTGCGCGCGCGCGGCCGCCGCG	19	0,71
Haplotype9	TGCGCGCGCGTTCGCGGTGCGTCGGTGCGCGCGCGCGGCCGCCGCG	17	0,67
Haplotype10	GGCGCGCGGCGTTCAGCGTGCGTCGGTGCGCGCGCGTCGGCCGCCGCG	16	0,63
Haplotype11	TGCGCGCGGCGTTCGGCGTGCGTCGGTGCGCGCGCGTCGGCCGCCGCG	11	0,45
Haplotype12	GGCGCGCGGCGTTCGGCGTGCGTCGGTGCGCGCGCGTCGGCCGCCGCG	9	0,41
Haplotype13	TGCGCGCGGCGTTCAGCGTGCGTCGGCGCGCGCGTCGGCCGCCGCG	6	0,34
Haplotype14	GGCGCGCGGCGTTCGCGGTGCGTCGGTGCGCGCGCGTCGGCCGCCGCG	6	0,34
Haplotype15	TGCGCGCGCGTTCGCGGTGCGTCGGTGCGCGCGCGCGGCCGCCGCG	4	0,26
Haplotype16	TGCGCGCGGCGTTCGGCATGCGTCGGTGCGCGCGCGTCGGCCGCCGCG	3	0,12
Haplotype17	CGCGCGCGGCGTTCGGCATGCGTCGGTGCGCGCGCGTCGGCCGCCGCG	3	0,12

b.

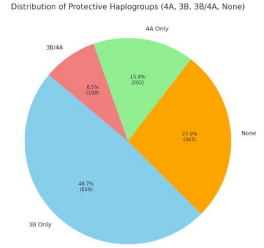


c.

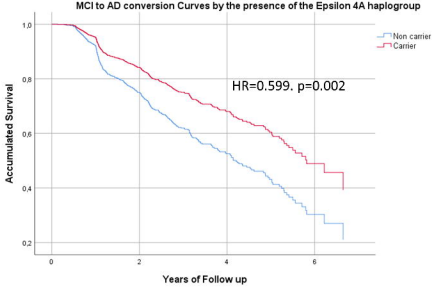



Haplotype12
Haplotype22
Haplotype41
Haplotype32
Haplotype39
Haplotype40
Haplotype55
Haplotype46
Haplotype57
Haplotype50
Haplotype51
Haplotype21
Haplotype17
Haplotype25
Haplotype47
Haplotype4
Haplotype49
Ancestral0
Haplotype28
Haplotype19
Haplotype35
Haplotype1
Haplotype8
Haplotype36
Haplotype18
Haplotype24
Haplotype31
Haplotype48
Haplotype14
Haplotype5
Haplotype27
Haplotype16
Haplotype11
Haplotype42
Haplotype9
Haplotype20
Haplotype13
Haplotype44
Haplotype3
Haplotype56
Haplotype10
Haplotype54
Haplotype23
Haplotype29
Haplotype30
Haplotype2
Haplotype59
Haplotype15
Haplotype53
Haplotype37
Haplotype45
Haplotype33
Haplotype34
Haplotype38
Haplotype26
Haplotype7
Haplotype6
Haplotype43
Haplotype52
Haplotype58

A



C

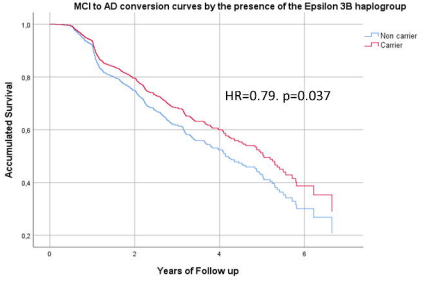


B

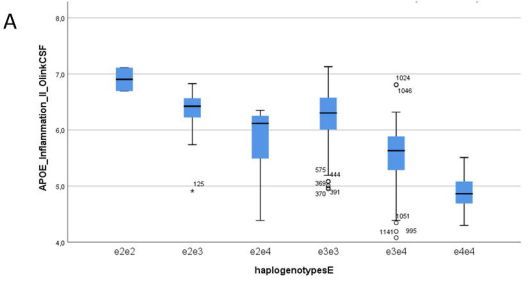
Cox propotional Hazard risk Models adjusted by Age, sex, AD biomarkers in CSF and APOE genotypes

	B	SE	Wald	df	Sig.	H.R.	95,0% CI for H.R. lower	upper
Age	,057	,009	43,746	1	,000	1,058	1,041	1,076
Sex	,004	,109	,002	1	,968	1,004	,812	1,242
Abeta_42_CSF	-,002	,000	68,733	1	,000	,998	,998	,999
P_tau181_CSF	,008	,001	52,964	1	,000	1,008	1,006	1,011
Presence_4a	-,512	,169	9,203	1	,002	,599	,431	,834
Presence_3b	-,236	,113	4,336	1	,037	,790	,633	,986
APOE gentoype	,228	,114	3,996	1	,046	1,256	1,004	1,571

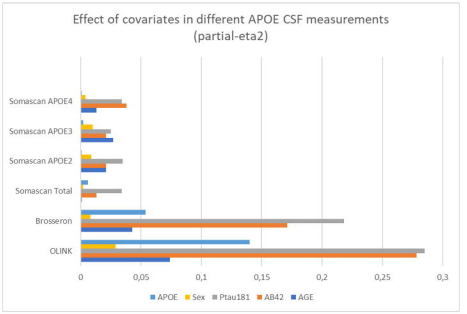
D



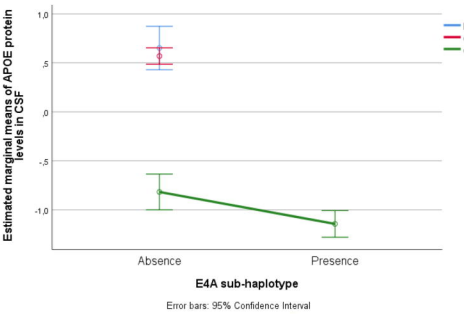
A


APOE_Inflammation_II_OlinkCSF
8,0
7,0
6,0
5,0
4,0
e2e2
e2e3
e2e4
e3e3
e3e4
e4e4
haplogenotypesE
1024
1046
125
575
444
369
370
391
1051
1141
995

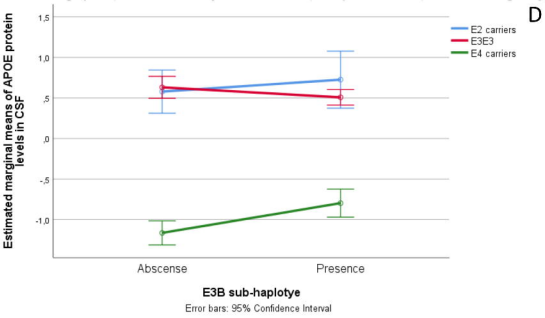
B


Effect of covariates in different APOE CSF measurements
(partial-eta2)
Somascan APOE4
Somascan APOE3
Somascan APOE2
Somascan Total
Brosseron
OLINK
0
0,05
0,1
0,15
0,2
0,25
0,3
APOE
Sex
Ptau181
AB42
AGE

C


Estimated marginal means of APOE protein levels in CSF
1,0
,5
,0
-,5
-1,0
E2 carriers
e3e3
e4 carriers
Absence
Presence
E4A sub-haplotype
Error bars: 95% Confidence Interval

D


Estimated marginal means of APOE protein levels in CSF
1,5
1,0
,5
,0
-,5
-1,0
E2 carriers
E3E3
E4 carriers
Abscense
Presence
E3B sub-haplotye
Error bars: 95% Confidence Interval