

Article

Apolipoprotein E (APOE) Allele Frequencies and Genotypic Distribution in Huambo, Angola

Job Pakisi ^{1,2} , Vicente Martín-Sánchez ^{2,3,*} , Cruz S. Sebastião ^{4,5,6} , Sergio Rodríguez Núñez ² ,
Celina Capingala-Pakisi ¹, Victor Moreno ^{3,7,8,9}  and Enrique Bayón-Darkistade ² 

- ¹ Faculdade de Medicina (FM), Universidade José Eduardo dos Santos (UJES), Huambo, Angola; jobpakisi19@gmail.com (J.P.); celinapakisi40@gmail.com (C.C.-P.)
- ² Instituto de Biomedicina, Universidad León, 24071 León, Spain; srodrn@unileon.es (S.R.N.); jebayd@unileon.es (E.B.-D.)
- ³ Consortium for Biomedical Research in Epidemiology & Public Health (CIBER Epidemiología y Salud Pública-CIBERESP), 28029 Madrid, Spain; v.moreno@iconcologia.net
- ⁴ Centro de Investigação em Saúde de Angola (CISA), Instituto Nacional de Investigação em Saúde (INIS), Luanda, Angola; cruzdossantos10@gmail.com
- ⁵ Centro Nacional de Investigação Científica (CNIC), Luanda, Angola
- ⁶ Global Health and Tropical Medicine, GHTM, Associate Laboratory in Translation and Innovation Towards Global Health, LA-REAL, Instituto de Higiene e Medicina Tropical, IHMT, Universidade NOVA de Lisboa, UNL, Rua da Junqueira 100, 1349-008 Lisboa, Portugal
- ⁷ Catalan Institute of Oncology, Institut d'Investigació Biomèdica de Bellvitge, 08908 Barcelona, Spain
- ⁸ Colorectal Cancer Group, ONCOBELL Program, Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), L'Hospitalet de Llobregat, 08908 Barcelona, Spain
- ⁹ Department of Clinical Sciences, Faculty of Medicine and Health Sciences and Universitat de Barcelona Institute of Complex Systems (UBICS), University of Barcelona (UB), L'Hospitalet de Llobregat, 08908 Barcelona, Spain
- * Correspondence: vmars@unileon.es; Tel.: +34-987293110

Abstract

Background/Objectives: Non-communicable diseases (NCDs) are a leading cause of death globally, particularly in low- and middle-income countries undergoing rapid epidemiological transition. Apolipoprotein E (APOE) polymorphisms have been implicated in modulating risk for cardiovascular and neurodegenerative diseases. This study characterizes the distribution of APOE alleles and genotypes in an Angolan population sample and explores associations with demographic variables. **Methods:** A cross-sectional study was conducted in Huambo Province, Angola, with 200 unrelated adults aged 40–70 years. Genotyping for $\epsilon 2$, $\epsilon 3$, and $\epsilon 4$ was performed via real-time PCR. Allelic and genotypic frequencies were calculated and tested for Hardy–Weinberg equilibrium. Associations with sex, age (≤ 54 vs. > 54 years), and urban vs. rural residence were analyzed. **Results:** $\epsilon 3$ was the most frequent allele (63%), followed by $\epsilon 4$ (24.8%) and $\epsilon 2$ (12.2%). The $\epsilon 3$ homozygous genotype was prevalent (44%), with $\epsilon 3/\epsilon 4$ (19%) and $\epsilon 2/\epsilon 3$ (18%) also common. Significant Hardy–Weinberg disequilibrium was observed due to excess $\epsilon 4$ homozygotes and deficit of certain heterozygotes; possible contributing factors, including population structuring, convenience sampling, and technical aspects of genotyping, are discussed. $\epsilon 4$ prevalence was higher in urban residents and older individuals, while $\epsilon 3$ and $\epsilon 2$ were more common in rural areas and younger participants. **Conclusions:** Our results reveal notable genetic heterogeneity and highlight the epidemiological and evolutionary importance of $\epsilon 4$ in Angola. The findings underscore the need for integrating population genomics into public health strategies targeting NCD prevention, especially in rapidly urbanizing African contexts.

Keywords: APOE; allele frequency; Angola; non-communicable diseases



Received: 23 June 2026

Revised: 28 July 2026

Accepted: 29 July 2026

Published: 3 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Non-communicable diseases (NCDs) account for three out of every four deaths worldwide, and three out of four of these deaths occur in low- and middle-income countries [1]. Demographic changes and lifestyle transitions underlie the high prevalence of these diseases and their concerning rapid increase in emerging countries [2]. Additionally, interactions between genetic and environmental factors play a crucial role in the incidence and course of these diseases and must be considered when designing prevention and control strategies [3,4].

Apolipoprotein E (*APOE*) contributes to the modulation of immune and inflammatory responses, and its different polymorphisms can have differential effects on some of the most relevant NCDs such as cardiovascular, autoimmune, and neurodegenerative diseases [5,6]. The clinical relevance of *APOE* alleles lies in their pleiotropic impact: the $\epsilon 3$ allele is usually associated with a balanced lipid profile, $\epsilon 2$ with some cardiovascular protection in heterozygotes (though with a risk of dyslipidemia in homozygotes), and $\epsilon 4$ with a pro-atherogenic phenotype characterized by increased LDL cholesterol, decreased HDL cholesterol, and a greater predisposition to both atherosclerosis and cognitive decline including Alzheimer's disease [7].

The global distribution of *APOE* alleles shows marked geographic and ethnic gradients [8]. Worldwide, the $\epsilon 3$ allele, with prevalences exceeding 70%, is the most frequent and considered the wild-type allele. The $\epsilon 4$ allele shows variable frequency and can reach up to 30% in some African populations. The $\epsilon 2$ allele is the least common, generally not exceeding 10%. These differences reflect patterns of positive selection, environmental pressures, and human migrations [8].

Angola, a Southwestern African country characterized by remarkable ethnic heterogeneity and a history marked by migrations and admixture [9], constitutes a unique setting to analyze the genetic variability of *APOE* and its relationship with emerging chronic diseases, which cause one in four deaths in the country [10]. This study aims to characterize the frequencies of *APOE* polymorphisms in Angola and their relationship with demographic variables.

2. Materials and Methods

2.1. Study Design and Setting

A cross-sectional study was conducted in Huambo Province, Angola, including a convenience sample of 200 unrelated adults (100 men and 100 women), half residing in urban areas and half in rural areas. Participants were aged between 40 and 70 years (mean 55 ± 9 years; median 54). The population of Huambo Province is predominantly composed of the Bantu Ovimbundo ethnic group; the sample is therefore relatively ethnically homogeneous and is not intended to capture the full ethnic diversity of Angola. Urban participants were recruited in the municipality of Huambo through public institutions after an informational session and written informed consent; rural participants were recruited from among residents of several rural municipalities who were staying at a religious institution at the time of data collection. Exclusion criteria included major neurological diseases, known high risk of severe cardiovascular disease, pregnancy, and HIV infection. The study was approved by the Scientific Council of the Faculty of Medicine of Huambo (Deliberation No. 003/CC/23) and the Ethics Committee of the Angolan Ministry of Health (031/C.E.M.S./2023).

2.2. Sample Collection and Processing

Peripheral blood was collected in EDTA tubes. Genomic DNA was extracted using the QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany). *APOE* genotyping for alleles $\epsilon 2$, $\epsilon 3$, and $\epsilon 4$ was performed by real-time PCR (StepOne 7500, Applied Biosystems, Waltham,

MA, USA) with SYBR Green and specific primers designed from the human APOE reference sequence (NG_007084.2, NCBI). Allele-specific discrimination was achieved with three primer combinations, each producing a 178-bp amplicon: the $\epsilon 2$ reaction (primers ApoE-112C and ApoE-158C), the $\epsilon 3$ reaction (primers ApoE-112C and ApoE-158R), and the $\epsilon 4$ reaction (primers ApoE-112R and ApoE-158R), corresponding to the two polymorphic sites that jointly define the $\epsilon 2/\epsilon 3/\epsilon 4$ alleles [10]. Each 20- μ L PCR reaction contained 1 $\times$ Power SYBR Green PCR Master Mix (Applied Biosystems, Waltham, MA, USA), 0.3 μ M of each primer, and 2 μ L of genomic DNA; cycling conditions were 95 $^{\circ}$ C for 10 min followed by 40 cycles of 95 $^{\circ}$ C for 15 s and 62 $^{\circ}$ C for 1 min (StepOne Real-Time PCR System, 96-well format), using the comparative Ct method. All reactions were run in duplicate together with no-template negative controls and a previously genotyped positive control; a Ct value < 25 indicated allele positivity, following the validated allele-specific PCR protocol of Calero et al. [10]. Demographic data collected included sex, age (categorized as ≤ 54 or > 54 years), and place of residence (urban or rural). Allelic and genotypic frequencies were calculated from the PCR results [10].

2.3. Statistical Analysis

Descriptive statistics were used to calculate allelic and genotypic frequencies, expressed with 95% confidence intervals (Clopper–Pearson exact method) [11]. Hardy–Weinberg equilibrium was assessed with chi-square tests ($\alpha = 0.05$). The inbreeding coefficient (F-IS) was computed as $F-IS = (H_e - H_o)/H_e$, where H_e is expected heterozygosity and H_o is observed heterozygosity [12]. Allelic frequencies were compared using contingency chi-square tests; genotypic frequencies were compared using Fisher’s exact test. Differences in sex and age-group distribution between urban and rural participants were assessed with the chi-square test, and mean age was compared with the Student’s *t*-test. To evaluate the joint associations of sex, age group (≤ 54 vs. > 54 years), and residence (urban vs. rural) with APOE status, multivariable logistic regression models were fitted with $\epsilon 4$ -carrier status ($\epsilon 2/\epsilon 4$, $\epsilon 3/\epsilon 4$, or $\epsilon 4/\epsilon 4$ vs. other genotypes), and separately, $\epsilon 2$ -carrier status ($\epsilon 2/\epsilon 2$, $\epsilon 2/\epsilon 3$, or $\epsilon 2/\epsilon 4$ vs. other genotypes) as binary outcomes, with sex, age group, and residence entered simultaneously as covariates; results are reported as odds ratios (ORs) with 95% confidence intervals. A full multinomial model across all six genotypes was not pursued because the rarest genotype ($\epsilon 2/\epsilon 4$) comprised a single participant, which would produce unstable estimates; the carrier-status approach is standard in the APOE epidemiological literature and directly addresses the sex/age/residence associations of interest. All analyses were performed using Stata version 16 (StataCorp, College Station, TX, USA).

3. Results

The $\epsilon 3$ allele had the highest frequency, present in approximately two-thirds of participants, followed by $\epsilon 4$ (24.8%) and $\epsilon 2$ (12.2%) (Table 1, Figure 1). Genotypic analysis revealed 62.5% homozygotes and 37.5% heterozygotes. The most frequent genotype was $\epsilon 3$ homozygous (44.0%; 95% CI: 37.6–50.4), followed by $\epsilon 3/\epsilon 4$ (19.0%; 95% CI: 13.2–24.8), $\epsilon 2/\epsilon 3$ (18.0%; 95% CI: 12.3–23.7), and $\epsilon 4/\epsilon 4$ (15.0%; 95% CI: 10.1–19.9). $\epsilon 2/\epsilon 2$ and $\epsilon 2/\epsilon 4$ genotypes were the least frequent (3.5% and 0.5%, respectively). Hardy–Weinberg equilibrium testing showed significant deviation ($\chi^2 = 51.73$; $df = 3$; $p < 0.001$), driven by an excess of $\epsilon 4$ homozygotes and a deficit of $\epsilon 2/\epsilon 4$ and $\epsilon 3/\epsilon 4$ heterozygotes. The inbreeding coefficient ($F-IS = 0.295$) indicated a substantial deficiency of heterozygotes. This pattern persisted across sexes, urban/rural residence, and age groups. Significant differences were observed in allelic and genotypic distributions by residence and age but not sex. $\epsilon 3$ was more prevalent in rural areas (72.0% vs. 53.0% urban), while $\epsilon 4$ was more prevalent in urban residents (34.0% vs. 15.5%

rural). Similarly, $\epsilon 3/\epsilon 3$ homozygotes predominated in rural (59.0%) versus urban (29.0%) areas, with $\epsilon 3/\epsilon 4$ heterozygotes more common in urban settings (28.0% vs. 10.0%). Age comparisons showed higher $\epsilon 4$ frequency in older participants (>54 years: 34.3% vs. ≤ 54 years: 15.4%) and higher $\epsilon 2$ in younger individuals (16.8% vs. 8.6%), with corresponding genotypic differences ($\epsilon 2/\epsilon 3$ and $\epsilon 4/\epsilon 4$). In Figure 2, the distribution of genotypic variables according to the different categories analyzed can be found.

Table 1. Distribution of genotypic and allelic variants according to various variables.

Genotypes and Alleles	1. Urban \$	2. Rural \$	0. Male *	1. Female *	0. Age &	1. Age &	N	Total, % (95% CI)
$\epsilon 2/\epsilon 2$	3.0 (0.6–8.4)	4.0 (1.1–9.8)	3.0 (0.6–8.4)	4.0 (1.1–9.8)	4.0 (1.1–9.9)	3.0 (0.6–8.5)	7	3.5 (0.9–6.1)
$\epsilon 2/\epsilon 3$	20.0 (12.6–29.3)	16.0 (9.4–24.8)	19.0 (11.6–28.2)	17.0 (10.0–26.2)	25.0 (17.1–34.1)	11.0 (5.5–18.7)	36	18.0 (12.3–23.7)
$\epsilon 2/\epsilon 4$	0.0 (0.0–3.6)	1.0 (0.0–5.6)	0.0 (0.0–3.6)	1.0 (0.0–5.6)	1.0 (0.0–5.4)	0.0 (0.0–3.6)	1	0.5 (0.0–1.5)
$\epsilon 3/\epsilon 3$	29.0 (20.4–39.1)	59.0 (48.6–67.0)	45.0 (35.1–55.1)	43.0 (33.0–53.2)	49.0 (39.1–57.9)	39.0 (29.2–48.6)	88	44.0 (37.6–50.4)
$\epsilon 3/\epsilon 4$	28.0 (19.6–37.7)	10.0 (4.9–18.0)	14.0 (8.0–22.3)	24.0 (15.5–33.8)	14.0 (8.0–22.1)	24.0 (15.9–33.3)	38	19.0 (13.2–24.8)
$\epsilon 4/\epsilon 4$	20.0 (12.6–29.3)	10.0 (4.9–18.0)	19.0 (11.6–28.2)	11.0 (5.6–18.7)	8.0 (3.4–15.1)	22.0 (13.8–31.6)	30	15.0 (10.1–19.9)
$\epsilon 2$	13.0 (8.3–17.7)	12.5 (7.9–17.1)	12.5 (7.9–17.1)	13.0 (8.3–17.7)	16.8 (11.7–21.9)	8.6 (4.7–12.4)	51	12.8 (9.6–16.0)
$\epsilon 3$	53.0 (46.1–59.9)	72.0 (65.8–78.2)	61.5 (54.8–68.2)	63.5 (56.9–70.1)	67.8 (61.4–74.2)	57.1 (50.2–64.0)	250	62.5 (57.8–67.3)
$\epsilon 4$	34.0 (27.5–40.5)	15.5 (10.5–20.5)	26.0 (19.9–32.0)	23.5 (17.7–29.3)	15.4 (10.3–20.3)	34.3 (27.7–40.9)	99	24.7 (20.6–28.9)
Genotypes	\$ $p = 0.0001$		* $p = 0.2813$		& $p = 0.0025$			
Alleles	\$ $p < 0.0001$		* $p = 0.85$		& $p < 0.0001$			

Note: “\$” stands for the comparison between rural vs urban areas; “*” stands for Sex comparison; “&” stands for age comparison.

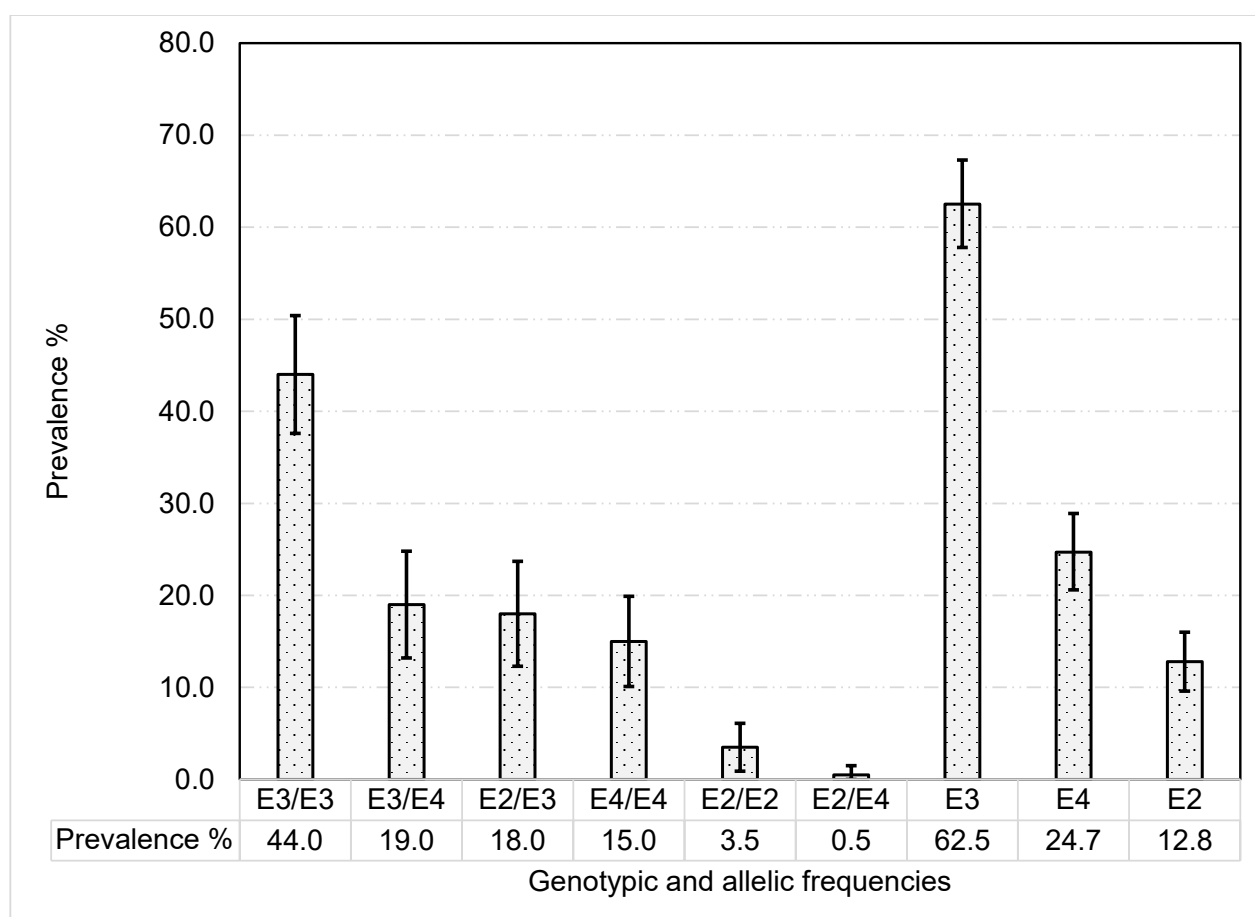


Figure 1. Distribution of genotypes and allelic frequencies and their 95% confidence intervals.

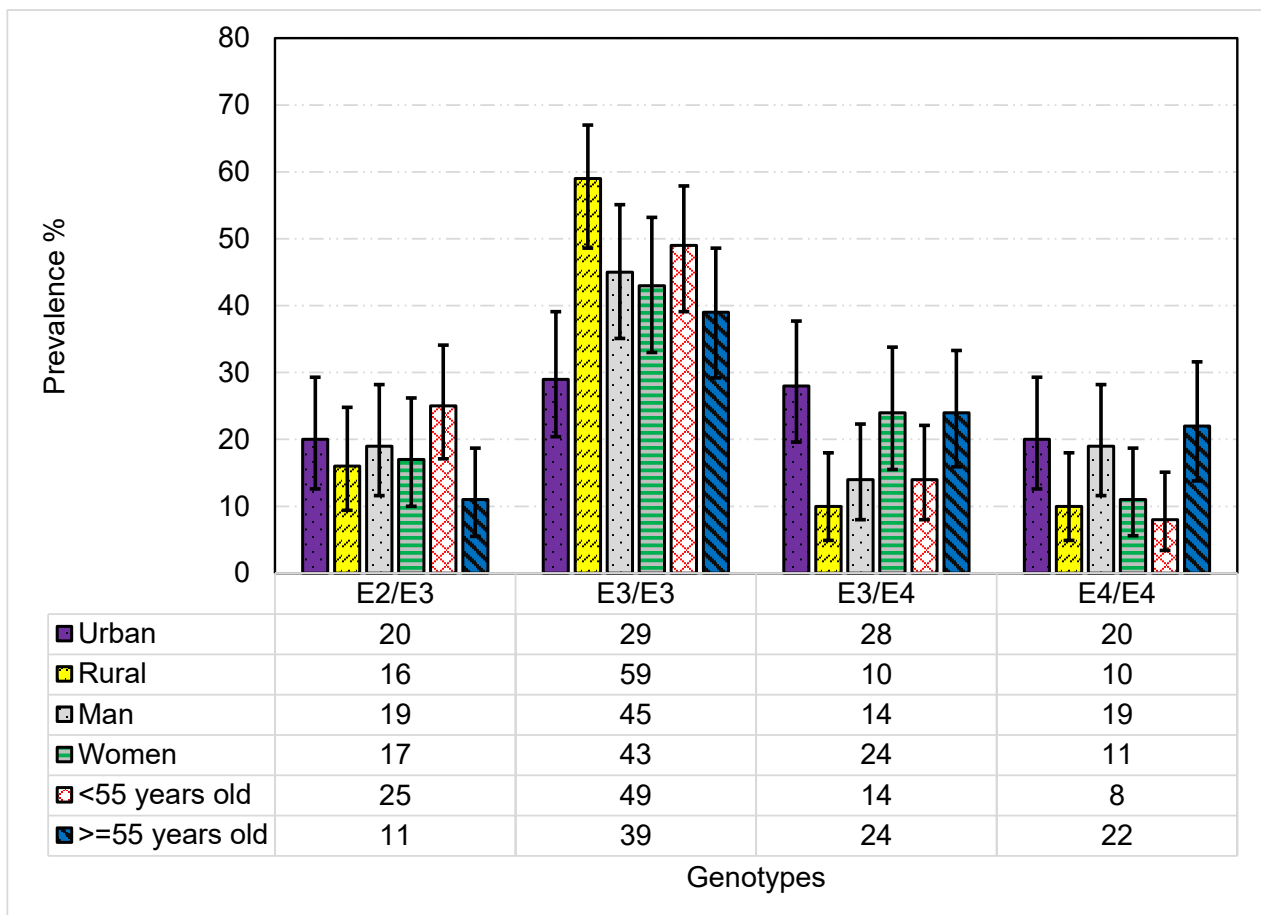


Figure 2. Distribution of genotypic variables according to the different categories analyzed.

Because the urban and rural subsamples were recruited through different channels (see Section 2.1), their comparability with respect to sex and age was formally tested (Table 2). Sex was perfectly balanced by design between residence groups (50.0% male in both urban and rural participants; $\chi^2 = 0.00$, $p = 1.000$). Age-group distribution, however, differed significantly by residence: participants >54 years were more frequent among urban residents (57.0%) than rural residents (42.0%) ($\chi^2 = 3.92$, $p = 0.048$), although the mean continuous age did not differ significantly between groups (55.3 ± 8.7 years urban vs. 53.1 ± 8.8 years rural; t -test $p = 0.085$). This age imbalance is potentially relevant, since age was independently associated with $\epsilon 4$ carriage in the multivariable model below, and is addressed further in the Discussion section.

Table 2. Sex and age distribution of the urban and rural participants.

Characteristic	Urban (n = 100)	Rural (n = 100)	p-Value
Sex, n (%)			
Male	50 (50.0)	50 (50.0)	1.000 χ^2
Female	50 (50.0)	50 (50.0)	
Age group, n (%)			
≤ 54 years	43 (43.0)	58 (58.0)	0.048 χ^2
> 54 years	57 (57.0)	42 (42.0)	
Age, years (mean $\pm$ SD)	55.3 $\pm$ 8.7	53.1 $\pm$ 8.8	0.085 t -test

To evaluate whether sex, age, and residence were independently associated with APOE status when considered simultaneously, multivariable logistic regression models were fitted for $\epsilon 4$ -carrier status and $\epsilon 2$ -carrier status (Table 3). $\epsilon 4$ carriage was independently associated

with older age (>54 years vs. ≤ 54 years: OR = 2.77, 95% CI: 1.46–5.25, $p = 0.002$) and with urban residence (OR = 3.19, 95% CI: 1.69–6.04, $p < 0.001$), but not with sex (OR = 1.33, 95% CI: 0.70–2.50, $p = 0.381$). Notably, the association between urban residence and $\epsilon 4$ carriage remained statistically significant after adjusting for the age imbalance described above, indicating that residence contributes independently of age to $\epsilon 4$ carriage in this sample. $\epsilon 2$ carriage was independently associated with younger age (>54 years vs. ≤ 54 years: OR = 0.37, 95% CI: 0.18–0.76, $p = 0.007$) but not with sex ($p = 0.743$) or residence ($p = 0.440$).

Table 3. Multivariable logistic regression for $\epsilon 4$ -carrier and $\epsilon 2$ -carrier status ($n = 200$).

Outcome	Covariate	OR (95% CI)	<i>p</i> -Value
$\epsilon 4$ carrier vs. non-carrier	Female vs. Male	1.33 (0.70–2.50)	0.381
	Age > 54 vs. ≤ 54 y	2.77 (1.46–5.25)	0.002
	Urban vs. Rural	3.19 (1.69–6.04)	<0.001
$\epsilon 2$ carrier vs. non-carrier	Female vs. Male	0.89 (0.45–1.77)	0.743
	Age > 54 vs. ≤ 54 y	0.37 (0.18–0.76)	0.007
	Urban vs. Rural	1.31 (0.66–2.63)	0.440

OR: odds ratio; CI: confidence interval. Each outcome was modeled separately by multivariable logistic regression with sex, age group, and residence entered simultaneously. Reference categories: Male, ≤ 54 years, Rural.

4. Discussion

This is the first study to describe the allelic and genotypic prevalences of apolipoprotein E in the Angolan population. Our findings highlight the remarkable genetic and phenotypic heterogeneity present in the African continent, as well as the coexistence of historical adaptive advantages with emerging risks stemming from changes in lifestyle [13–15]. The $\epsilon 4$ allele, which confers an advantage in environments characterized by high infectious burden and nutritional limitations, is significantly associated with adverse impacts in urbanized settings with Western lifestyles, marked sedentarism, and high-calorie diets [7,15]. This represents a critical challenge for public health strategies aimed at preventing and managing emerging diseases, particularly cardiovascular and neurodegenerative disorders.

In our sample, the $\epsilon 3$ allele was the most prevalent, with an approximate frequency of 63%, a figure that falls within the commonly reported range in global and regional studies, where $\epsilon 3$ typically varies from 55% to 90% across different human populations, showing closer similarity to prevalences reported for other sub-Saharan groups [8,16]. It should be noted that from an evolutionary standpoint, $\epsilon 4$ rather than $\epsilon 3$ is generally considered the ancestral allele at this locus, with $\epsilon 3$ and $\epsilon 2$ arising later through sequential mutation and subsequently rising to high frequency in most modern human populations [17]. The high frequency of $\epsilon 3$ observed here therefore reflects its status as the currently dominant, but not ancestral, allele, and is broadly consistent with frequencies reported for other African and Afro-descendant groups [8,16], supporting the internal consistency of our results.

The frequency of the $\epsilon 4$ allele shows considerable worldwide variation, ranging from 0% in some Indian populations to nearly 50% in certain Brazilian or African tribes [18]. In our case, the observed value of 24.8% represents an intermediate level within the African spectrum. While North African regions, such as Morocco and Tunisia, report low frequencies (<10%), other West African countries (such as Senegal: 3%) also exhibit considerably lower levels. In contrast, countries like Uganda (25%), Rwanda (24%), South Africa (25.4%), Nigeria (30%), Sudan (29%), and particularly Khoisan groups (37%) and Central African pygmies (41%) present high prevalences [13,19,20]. This finding underscores the genetic heterogeneity within the African continent and emphasizes the uniqueness of the Angolan population structure.

The high prevalence of the $\epsilon 4$ allele, especially in its homozygous form, is one plausible contributor to the significant genotype disequilibrium observed. However, several alternative and non-mutually exclusive explanations must be considered before attributing this deviation to a single biological cause. First, the convenience sampling strategy used here may have introduced selection bias, as participants were not randomly drawn from the source population. Second, undetected genotyping error is a known risk in *APOE* typing, since the $\epsilon 2$, $\epsilon 3$, and $\epsilon 4$ alleles are distinguished by only two single-nucleotide polymorphisms (rs429358 and rs7412), and misclassification can spuriously inflate apparent homozygote excess. Third, population stratification arising from Angola's marked ethnic heterogeneity could generate a Wahlund-type effect that mimics inbreeding at the population level even in the absence of true consanguinity. Only after these technical and sampling explanations are reasonably excluded can population structure and endogamy, consistent with localized marital practices reported in other African settings [13,14,21], be considered a likely contributor to the elevated inbreeding coefficient observed. A formal comparison of our allele frequencies with publicly available reference datasets supports this cautious interpretation: our results are broadly consistent with the largest available reference for continental African ancestry: in the gnomAD database (v4.1.0, June 2025), individuals of African ancestry show allele frequencies of approximately 22% for $\epsilon 4$ and 11% for $\epsilon 2$, with the remainder corresponding to $\epsilon 3$ ($\approx 67\%$) [22]. Our observed frequencies ($\epsilon 4 = 24.8\%$; $\epsilon 2 = 12.2\%$; $\epsilon 3 = 62.5\%$) fall close to these continental African estimates, suggesting that the Huambo sample is broadly representative of the wider pattern described for African populations, even though Sub-Saharan Africa itself shows considerable inter-ethnic variation, as illustrated by the range reported across Sub-Saharan cohorts ($\epsilon 4$ from 3% in Senegal to over 40% in Central African Pygmy groups) [13,18–20]. A dedicated comparison against Angola-specific or Southern African 1000 Genomes Project subpopulations was not possible, as no such reference panel currently exists for Angola specifically; this is itself a gap in the literature that our study helps to address.

This dynamic balance of risks and benefits underscores the importance of a personalized approach in medicine and public health for countries undergoing epidemiological transition.

It is important to emphasize that the present study collected only genotypic and demographic data; no phenotypic, biochemical, or clinical measurements (e.g., lipid profile, cardiovascular events, cognitive assessment) were obtained from the participants. Consequently, any discussion of disease risk below draws on the established literature rather than on direct evidence from this cohort, and should be interpreted as hypothesis-generating rather than as a demonstrated finding in this population. The epidemiological relevance of the $\epsilon 4$ allele lies in its well-established association, reported elsewhere, with increased risk of cardiovascular and neurodegenerative diseases in Western lifestyle contexts. However, in Africa, phenotypic plasticity attributable to environmental, infectious, and behavioral factors has been reported [23–25]. As documented by Masemola et al. in South Africa, rural–urban transitions and lifestyle changes may precipitate future cardiovascular epidemics in $\epsilon 4$ carriers, emphasizing the urgency of genetic–environmental surveillance in Angola to anticipate disease burden evolution [15].

Analyzing the biomedical impact of the $\epsilon 4$ allele, several studies have confirmed its relationship with elevated total cholesterol, LDL-c concentrations, and reduced HDL-c, findings consistent with extensive international evidence [26–29]. This positions $\epsilon 4$ as a potential driver of the epidemiological transition in developing countries, accelerating the prevalence of non-communicable diseases (NCDs). On the other hand, the $\epsilon 2$ allele, while usually associated with a favorable lipid profile in heterozygotes, can predispose to severe dyslipidemias (type III) in the presence of factors such as diabetes and excessive alcohol intake, reinforcing the multifactorial nature of clinical risk [30–32].

Regarding urban–rural differences, the predominance of $\epsilon 3$ in rural areas (72%) contrasts with the higher frequency of $\epsilon 4$ in urban environments (34%), probably due to internal migrations, inequalities in healthcare access, and lifestyle variations. This pattern suggests that urban exposure may amplify cardiovascular and neurodegenerative risk for $\epsilon 4$ carriers, in line with the global findings [33]. It is worth clarifying that a higher frequency of $\epsilon 4$ among urban residents does not imply that the urban environment causes or selects for $\epsilon 4$; rather, the allele's adverse metabolic effects are thought to be more likely to manifest phenotypically under an urban, Western-type lifestyle (sedentarism, energy-dense diets) while remaining comparatively silent under the more physically active, traditional lifestyle prevalent in rural areas. The higher observed frequency of $\epsilon 4$ in the urban subsample most plausibly reflects the demographic history of rural-to-urban migration in Huambo Province, in which lineages or ethnic groups with a naturally higher baseline $\epsilon 4$ frequency may be over-represented among urban residents, rather than any selective process favoring $\epsilon 4$ carriers who move to cities. Because our design is cross-sectional and convenience-based, we cannot rule out that this urban–rural contrast partly reflects recruitment bias (e.g., differential willingness to participate, or differences in the ethnic composition of the urban versus rural sampling sites) rather than a true population-level difference; distinguishing between these possibilities would require a randomly sampled, ethnically-stratified cohort with residential-history data, which was beyond the scope of this study. Notably, because urban participants in our sample were, on average, somewhat older than rural participants (Table 2), age is a plausible confounder of the urban– $\epsilon 4$ association; however, multivariable logistic regression showed that urban residence remained independently associated with $\epsilon 4$ carriage after adjusting for age and sex (OR = 3.19, 95% CI: 1.69–6.04; Table 3), indicating that age alone does not explain this pattern.

Age distribution is also relevant: the higher prevalence of $\epsilon 4$ among older individuals (>54 years) and predominance of $\epsilon 2$ in younger cohorts is a cross-sectional observation that may reflect cohort effects (e.g., generational differences in ancestry or admixture) rather than differential survival, and any inference regarding selection or mortality would require longitudinal data that this study did not collect. While $\epsilon 4$ has traditionally been linked to reduced longevity [34], Lane et al. demonstrated that its impact on mortality varies significantly according to ethnic and geographic context, potentially exhibiting neutral effects in African and Afro-descendant groups [21]. A more recent study by Vivian et al. found no association between $\epsilon 4$ genotypes and mortality in elderly individuals ≥ 80 years; however, classical cardiovascular factors such as smoking and diabetes increased mortality risk, whereas physical activity and elevated systolic blood pressure reduced it [35]. Therefore, it is essential to deepen longitudinal studies incorporating environmental and clinical variables to clarify these effects [36].

Limitations

Several limitations should be considered when interpreting these findings. First, the sample size ($n = 200$) was relatively small for population genetics purposes and limits the precision of allele- and genotype-frequency estimates, as reflected in the width of several confidence intervals; results should therefore be interpreted as preliminary. Second, participants were recruited as a convenience sample rather than through random or stratified population sampling. The sample was drawn predominantly from the Bantu Ovimbundo ethnic group, the majority population of Huambo Province, and is therefore relatively ethnically homogeneous rather than representative of Angola's full ethnic diversity; findings should not be generalized to other Angolan ethnic groups without further study. Urban participants were recruited through public institutions in the municipality of Huambo following informational sessions and written informed consent, whereas rural participants

were recruited among residents of different rural municipalities who were staying at a single religious institution at the time of data collection; this difference in recruitment channel could itself introduce clustering or selection effects that are difficult to fully disentangle from genuine urban–rural biological differences. Third, no phenotypic, biochemical, or clinical data were collected, so associations with cardiovascular or neurodegenerative disease risk discussed above are drawn from the external literature and were not directly tested in this cohort. Fourth, the cross-sectional design could not distinguish cohort effects from selection or survival effects in the age-related comparisons. Fifth, the real-time PCR genotyping method, while previously validated elsewhere [10], was not accompanied in this study by an independent confirmatory method for a subset of samples; genotyping error therefore cannot be entirely excluded as a contributor to the observed Hardy–Weinberg disequilibrium. Finally, although multivariable logistic regression was used to assess the independent associations of sex, age, and residence with $\epsilon 4$ - and $\epsilon 2$ -carrier status (Table 3), the modest sample size limits statistical power to detect smaller effects or to test interactions between these covariates (e.g., whether the age-residence association with $\epsilon 4$ differs by sex); larger, multicenter studies are needed to confirm and extend these associations.

5. Conclusions

In conclusion, these findings emphasize the urgency of public health policies integrating population genomics as a tool for risk assessment and healthcare planning, especially in rapidly changing urban African contexts. Targeted preventive interventions should be prioritized, as recommended by Masemola et al., to mitigate the rise of NCDs in genetically susceptible groups and to test new personalized medicine strategies. Future research should expand sample sizes, incorporate additional genetic markers and clinical-metabolic variables, and conduct multicenter and trans-ethnic meta-analyses following models such as Marini et al., which enable a comprehensive understanding of genetic, environmental, and population health interactions. This study contributes to describing the *APOE* genomic profile in Angola, demonstrating the predominance of the $\epsilon 3$ allele and a comparatively high frequency of $\epsilon 4$, and highlighting population structuring as a plausible, though not the only possible, contributor to the Hardy–Weinberg disequilibrium observed. These findings, together with their potential public health implications for African populations, should be confirmed in larger, representative samples incorporating phenotypic data before firm conclusions are drawn.

Author Contributions: Conceptualization: J.P., C.C.-P. and E.B.-D. Investigation: J.P., C.C.-P., V.M.-S., C.S.S., V.M.-S., and E.B.-D. Methodology: J.P., C.C.-P., V.M.-S., and E.B.-D. Validation: J.P. and E.B.-D. Data curation: J.P., V.M.-S., C.S.S., and E.B.-D. Formal analysis: J.P., V.M.-S., and E.B.-D. Data collection: J.P. Supervision: E.B.-D. Writing—original draft: J.P., V.M.-S., C.S.S., V.M. and E.B.-D. Writing—review & editing: J.P., C.C.-P., S.R.N., V.M.-S., C.S.S., V.M., and E.B.-D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was approved by the Scientific Council of the Faculty of Medicine of Huambo (Deliberation No. 003/CC/23) and the Ethics Committee of the Angolan Ministry of Health (031/C.E.M.S./2023).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data supporting this study’s findings are available on request from the corresponding author.

Acknowledgments: The authors thank the patients for their participation. We also thank UJES, Huambo Health Delegation, The Unileon infectious disease laboratory team, CISA I INIS, and CNIC for the logistical, administrative, and scientific support.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. WHO. *Global Challenges in Cardiovascular Prevention in Populations with Low Socioeconomic Status*; Springer: Cham, Switzerland, 2025.
2. National Institute of Environmental Health Sciences Environmental Factor-February 2022: Ethical Issues Emerge in Gene-Environment Interactions Research. Available online: <https://factor.niehs.nih.gov/2022/2/feature/2-feature-research-ethics> (accessed on 23 July 2025).
3. Peng, J.; Xu, H.; Tang, X. Global Inequalities in the Burden of Digestive Diseases From 1990 to 2019: Findings from the Global Burden of Disease Study 2019. *Gastroenterology* **2024**, *166*, 223–224.e1. [CrossRef] [PubMed]
4. Matsunaga, A.; Saito, T. Impact of Apolipoprotein E Variants: A Review of Naturally Occurring Variants and Clinical Features. *J. Atheroscler. Thromb.* **2025**, *32*, 281–303. [CrossRef] [PubMed]
5. Civeira-Marín, M.; Cenarro, A.; Marco-Benedí, V.; Bea, A.M.; Mateo-Gallego, R.; Moreno-Franco, B.; Ordovás, J.M.; Laclaustra, M.; Civeira, F.; Lamiquiz-Moneo, I. APOE Genotypes Modulate Inflammation Independently of Their Effect on Lipid Metabolism. *Int. J. Mol. Sci.* **2022**, *23*, 12947. [CrossRef] [PubMed]
6. Islam, S.; Noorani, A.; Sun, Y.; Michikawa, M.; Zou, K. Multi-Functional Role of Apolipoprotein E in Neurodegenerative Diseases. *Front. Aging Neurosci.* **2025**, *17*, 1535280. [CrossRef] [PubMed]
7. Mahley, R.W. Apolipoprotein E: From Cardiovascular Disease to Neurodegenerative Disorders. *J. Mol. Med.* **2016**, *94*, 739–746. [CrossRef] [PubMed]
8. Corbo, R.M.; Scacchi, R. Apolipoprotein E (APOE) Allele Distribution in the World. Is APOE*4 a “thrifty” Allele? *Ann. Hum. Genet.* **1999**, *63*, 301–310. [CrossRef] [PubMed]
9. Campbell, M.C.; Tishkoff, S.A. African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. *Annu. Rev. Genom. Hum. Genet.* **2008**, *9*, 403–433. [CrossRef] [PubMed]
10. Calero, O.; Hortigüela, R.; Bullido, M.J.; Calero, M. Apolipoprotein E Genotyping Method by Real Time PCR, a Fast and Cost-Effective Alternative to the TaqMan and FRET Assays. *J. Neurosci. Methods* **2009**, *183*, 238–240. [CrossRef] [PubMed]
11. Clopper, C.J.; Pearson, E.S. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika* **1934**, *26*, 404–413. [CrossRef]
12. Wright, S. Coefficients of Inbreeding and Relationship. *Am. Nat.* **1922**, *56*, 330–338. [CrossRef] [PubMed]
13. Van Exel, E.; Koopman, J.J.E.; Van Bodegom, D.; Meij, J.J.; De Knijff, P.; Ziem, J.B.; Finch, C.E.; Westendorp, R.G.J. Effect of APOE E4 Allele on Survival and Fertility in an Adverse Environment. *PLoS ONE* **2017**, *12*, e0179497. [CrossRef] [PubMed]
14. Octavio-Aguilar, P.; Ramos-Frías, J. Aplicación de La Genética de Poblaciones En El Ámbito de La Medicina. *Biomédica* **2014**, *34*, 171–180. [CrossRef]
15. Masemola, M.L.; Alberts, M.; Urdal, P. Apolipoprotein E Genotypes and Their Relation to Lipid Levels in a Rural South African Population. *Scand. J. Public Health* **2007**, *35*, 60–65. [CrossRef] [PubMed]
16. Martínez-Martínez, A.B.; Torres-Perez, E.; Devanney, N.; Del Moral, R.; Johnson, L.A.; Arbones-Mainar, J.M. Beyond the CNS: The Many Peripheral Roles of APOE. *Neurobiol. Dis.* **2020**, *138*, 104809. [CrossRef] [PubMed]
17. Fullerton, S.M.; Clark, A.G.; Weiss, K.M.; Nickerson, D.A.; Taylor, S.L.; Stengard, J.H.; Salomaa, V.; Vartiainen, E.; Perola, M.; Boerwinkle, E.; et al. Apolipoprotein E Variation at the Sequence Haplotype Level: Implications for the Origin and Maintenance of a Major Human Polymorphism. *Am. J. Hum. Genet.* **2000**, *67*, 881–900. [CrossRef] [PubMed]
18. Ruiz, M.; Arias, I.; Rolón, G.; Hernández, E.; Garavito, P.; Silvera-Redondo, C. Análisis Del Polimorfismo Del Gen APOE En La Población de Barranquilla, Colombia. *Biomédica* **2016**, *36*, 52–58. [CrossRef] [PubMed]
19. Atadzhanov, M.; Mwaba, M.H.; Mukomena, P.N.; Lakhi, S.; Mwaba, P.; Rayaprolu, S.; Meschia, J.F.; Ross, O.A. Frequency of APOE, MTHFR and ACE Polymorphisms in the Zambian Population. *BMC Res. Notes* **2014**, *7*, 194. [CrossRef] [PubMed]
20. Soo, C.C.; Farrell, M.T.; Tollman, S.; Berkman, L.; Nebel, A.; Ramsay, M. Apolipoprotein E Genetic Variation and Its Association with Cognitive Function in Rural-Dwelling Older South Africans. *Front. Genet.* **2021**, *12*, 689756. [CrossRef] [PubMed]
21. Lane, K.A.; Gao, S.; Hui, S.L.; Murrell, J.R.; Hall, K.S.; Hendrie, H.C. Apolipoprotein E and Mortality in African-Americans and Yoruba. *J. Alzheimer's Dis.* **2003**, *5*, 383. [CrossRef] [PubMed]
22. Broad Institute of MIT and Harvard. *Genome Aggregation Database (gnomAD)*; v4.1.0; Broad Institute of MIT and Harvard: Cambridge, MA, USA, 2025; Available online: <https://gnomad.broadinstitute.org> (accessed on 26 July 2026).
23. Mooijaart, S.P.; Berbée, J.F.P.; Van Heemst, D.; Havekes, L.M.; De Craen, A.J.M.; Slagboom, P.E.; Rensen, P.C.N.; Westendorp, R.G.J. ApoE Plasma Levels and Risk of Cardiovascular Mortality in Old Age. *PLoS Med.* **2006**, *3*, e176. [CrossRef] [PubMed]

24. Lindner, K.; Gavin, A.C. Isoform- and Cell-State-Specific APOE Homeostasis and Function. *Neural Regen. Res.* **2024**, *19*, 2456. [[CrossRef](#)] [[PubMed](#)]
25. Wang, Y.; Lou, H.; Wang, M.; Mei, J.; Xing, T.; Wang, F.; Dong, Z.; Wang, L.; Cao, R.; Yao, L.; et al. Correlation between Genetic Polymorphisms in Apolipoprotein E and Atrial Fibrillation. *Rev. Port. Cardiol.* **2022**, *41*, 417–423. [[CrossRef](#)] [[PubMed](#)]
26. Leskinen, H.; Tringham, M.; Karjalainen, H.; Iso-Touru, T.; Hietaranta-Luoma, H.L.; Marnila, P.; Pihlava, J.M.; Hurme, T.; Puolijoki, H.; Åkerman, K.; et al. APOE Genotypes, Lipid Profiles, and Associated Clinical Markers in a Finnish Population with Cardiovascular Disease Risk Factors. *Lifestyle Genom.* **2022**, *15*, 45–54. [[CrossRef](#)] [[PubMed](#)]
27. Guo, J.L.; Braun, D.; Fitzgerald, G.A.; Hsieh, Y.T.; Rougé, L.; Litvinchuk, A.; Steffek, M.; Propson, N.E.; Heffner, C.M.; Discenza, C.; et al. Decreased Lipidated ApoE-Receptor Interactions Confer Protection against Pathogenicity of ApoE and Its Lipid Cargoes in Lysosomes. *Cell* **2025**, *188*, 187–206. [[CrossRef](#)] [[PubMed](#)]
28. Matsuo, H.; Kaneko, T.; Fujikawa, K.; Ohara, H.; Mashimo, T.; Nabika, T. Development of a New Model for Combined Atherosclerosis and Hypertension: ApoE-Knockout Spontaneously Hypertensive Rat. *Atherosclerosis* **2017**, *263*, e86. [[CrossRef](#)]
29. Cahua-Pablo, G.; Cruz, M.; Moral-Hernández, O.D.; Leyva-Vázquez, M.A.; Antúnez-Ortiz, D.L.; Cahua-Pablo, J.A.; Alarcón-Romero, L.D.C.; Ortuño-Pineda, C.; Moreno-Godínez, M.E.; Hernández-Sotelo, D.; et al. Elevated Levels of LDL-C Are Associated with ApoE4 but Not with the Rs688 Polymorphism in the LDLR Gene. *Clin. Appl. Thromb. Hemost.* **2016**, *22*, 465–470. [[CrossRef](#)] [[PubMed](#)]
30. Hu, J.; Liu, C.C.; Chen, X.F.; Zhang, Y.W.; Xu, H.; Bu, G. Opposing Effects of Viral Mediated Brain Expression of Apolipoprotein E2 (ApoE2) and ApoE4 on ApoE Lipidation and A β Metabolism in ApoE4-Targeted Replacement Mice. *Mol. Neurodegener.* **2015**, *10*, 6. [[CrossRef](#)] [[PubMed](#)]
31. Littlejohns, T.J.; Collister, J.A.; Liu, X.; Clifton, L.; Tapela, N.M.; Hunter, D.J. Hypertension, a Dementia Polygenic Risk Score, APOE Genotype, and Incident Dementia. *Alzheimer's Dement.* **2023**, *19*, 467–476. [[CrossRef](#)] [[PubMed](#)]
32. Lozano, S.; Padilla, V.; Avila, M.L.; Gil, M.; Maestre, G.; Wang, K.; Xu, C. APOE Gene Associated with Cholesterol-Related Traits in the Hispanic Population. *Genes* **2021**, *12*, 1768. [[CrossRef](#)] [[PubMed](#)]
33. Freeman, E.E.; Bastasic, J.; Grant, A.; Leung, G.; Li, G.; Buhrmann, R.; Roy-Gagnon, M.H. Inverse Association of APOE E4 and Glaucoma Modified by Systemic Hypertension: The Canadian Longitudinal Study on Aging. *Investig. Ophthalmol. Vis. Sci.* **2022**, *63*, 9. [[CrossRef](#)] [[PubMed](#)]
34. Wang, Y.Z.; Zhao, W.; Moorjani, P.; Gross, A.L.; Zhou, X.; Dey, A.B.; Lee, J.; Smith, J.A.; Kardina, S.L.R. Effect of Apolipoprotein E E4 and Its Modification by Sociodemographic Characteristics on Cognitive Measures in South Asians from LASI-DAD. *Alzheimers Dement.* **2024**, *20*, 4854–4867. [[CrossRef](#)] [[PubMed](#)]
35. Vivian, L.; Bruscato, N.M.; Werle, B.M.; de Carli, W.; Soares, R.A.G.; de Lima Santos, P.C., Jr.; Moriguchi, E.H. Association of Cardiovascular Risk Factors and Apoe Polymorphism with Mortality in the Oldest Old: A 21-Year Cohort Study. *Arq. Bras. Cardiol.* **2020**, *115*, 873–881. [[CrossRef](#)] [[PubMed](#)]
36. Marini, S.; Crawford, K.; Morotti, A.; Lee, M.J.; Pezzini, A.; Moomaw, C.J.; Flaherty, M.L.; Montaner, J.; Roquer, J.; Jimenez-Conde, J.; et al. Association of Apolipoprotein E With Intracerebral Hemorrhage Risk by Race/Ethnicity: A Meta-Analysis. *JAMA Neurol.* **2019**, *76*, 480–491. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.