# scientific reports



## **OPEN** Search for the elusive haplotype of the APOE polymorphism associated with Alzheimer's disease

Asma Naseer Cheema, Kang-Hsien Fan, Elizabeth Lawrence, Narges Zafari, Ruyu Shi, Muhammad Muaaz Aslam, Vibha Acharya, Alayna Jean Holderman, Annie Bedison, Eleanor Feingold & M. Ilyas Kamboh<sup>⊠</sup>

The common APOE2/E3/E4 polymorphism is determined by two-site haplotypes: C112R and R158C. Due to strong linkage disequilibrium (LD) between the two sites, three of the four expected haplotypes/alleles (E2, E3, E4) have been observed. Compared to the most common E3 haplotype (C112–R158), E4 (R112–R158) results from a mutation at codon 112, while E2 (C112–C158) results from a mutation at codon 158. The fourth haplotype (E5) having mutations at both sites (R112-C158) has been reported only as an incidental finding in three kindreds. To our knowledge, no systematic search has been done to determine its distribution in the general population. The objective of this study was to search for the elusive haplotype by subcloning a DNA fragment of 177 bp from 355 subjects with the APOE 2/4 genotype followed by sequencing as well as in 11,647 subjects genotyped by TaqMan assays. No example of the E5 haplotype was observed, suggesting it might have a minimum effect, if any, on Alzheimer's disease risk. Under the assumption of strong LD between the two sites, the estimated probability for the occurrence of the E5 haplotype by recombination event is 3.31E-08, which is similar to 5.58E-08 probability obtained by recurrent point mutation.

**Keywords** APOE haplotype, Cloning, Sequencing, Genotyping, Recombination, point mutation

The common APOE2/E3/E4 polymorphism is determined by two-site haplotypes at codon 112 [cysteine (C) > arginine (R)] and codon 158 (R > C), resulting in six genotypes. Due to strong linkage disequilibrium (LD) between the two sites, three of the four expected haplotypes or alleles (E2, E3, and E4) have been observed and extensively studied in relation to the risk of Alzheimer's disease (AD) risk<sup>1,2</sup>. Compared to the most common haplotype of E3 (C112 – R158), the mutant E4 (R112 – R158) and E2 (C112 – C158) haplotypes are determined by single-point mutations: T > C transition at codon 112 (TGC to CGC for E4) and C >  $\overline{T}$  transition at codon158 (CGC to TGC for E2)<sup>2</sup>. The fourth haplotype, which we denote here as APOE5 and explained later, having amino acid substitutions at both positions (R112 - C158) has been reported only as an incidental finding in three kindreds<sup>3-5</sup>. Genetically determined structural variation in APOE was initially identified using two-dimensional gel electrophoresis or isoelectric focusing (IEF) from ultracentrifuged and delipidated plasma<sup>6-8</sup>. This method was later simplified for APOE genotyping directly from plasma, eliminating the need for ultracentrifugation and delipidation in large-scale population screening<sup>9</sup>.Gel electrophoresis that separates plasma APOE isoforms based on their size, charge, or isoelectric point differences, can also enable the identification of new APOE allelic isoforms is not possible to determine using the current TaqMan assays as they are specific for detecting genomic variations at codons 112 and 158 only<sup>10</sup>.

Individuals with the APOE 2/4 genotypes are double heterozygotes at codons 112 and 158 and due to strong LD between the two sites, it is always assumed that the alleles for E4 (R112) and E2 (C158) are present on opposite chromosomes<sup>2</sup>. Since TaqMan assays do not reveal the two-site haplotype phases, it is possible that some of the APOE 2/4 individuals may carry the APOE5 haplotype where both alleles (R112 and C158) are present on the same haplotype and inherited from a single parent. This could only be accomplished unequivocally by either subcloning followed by single-strand sequencing or next-generation whole-genomic sequencing (WGS). To our knowledge, no such systematic effort has been made to determine the haplotype phases and the occurrence of the APOE5 haplotype in the general population. The objective of this study was to search for the elusive APOE5 haplotype by using combination of subcloning, WGS or restriction fragment length polymorphism (RFLP) assays in a large number of APOE 2/4 subjects and determine its potential association with AD.

Department of Human Genetics, School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA. <sup>™</sup>email: kamboh@pitt.edu

	Total Sample ( <i>N</i> = 14819)	AD Cases (N= 3280)	Controls (N=11535)	<i>P</i> -value between cases and controls
Mean age ± SD	$72.7 \pm 14.2$	$72.1 \pm 9.8$	72.9 ±15.2	4.94E-04
Female, N (%)	8142 (56.6)	2000 (61.0)	6141 (55.4)	1.20E-08
White/Black/ Others (%)	86.8/12.7/0.5	93.4/6.3/0.3	84.8/14.6/0.6	1.36E-36
APOE Genotype N (%)				1.95E-238
2/2	88 (0.6)	8(0.2)	80 (0.7)	
2/3	1557 (10.5)	163(5.0)	1394(12.1)	
2/4	355 (2.4)	88(2.7)	267(2.3)	
3/3	8254 (55.7)	1280(39.0)	6973(60.5)	
3/4	3948 (26.6)	1415(43.1)	2531(21.9)	
4/4	617 (4.2)	326(10.0)	290(2.5)	
APOE Allele Frequency (%)				7.08E-254
E2	7.1	4.1	7.9	
E3	74.2	63.1	77.5	
E4	18.7	32.8	14.6	

 Table 1. Demographic and APOE genotype/allele data in study participants.

APOE Genotype	ORs (95% confidence interval)	P-Value
2/2	0.73 (0.32–1.45)	0.41
2/3	0.65(0.54-0.77)	1.64E-06
2/4	2.01(1.55-2.58)	8.92E-08
3/4	3.26(2.97-3.58)	2.72E-137
4/4	6.87(5.70-8.30)	4.50E-90

Table 2. Odds ratios (ORs) for APOE genotypes using APOE 3/3 as reference.

### Methods

This study was approved by the University of Pittsburg institutional Review Board and informed consent was obtained from all the subjects prior to their participation in the study. All the experimental protocols were reviewed and approved by Department of Human Genetics, School of Public Health University of Pittsburgh and were performed in accordance with the approved guidelines and regulations of the department. APOE genotype data on 14,819 subjects (mean age = 72.7 years; female = 56.6%; Table 1,2) derived from multiple studies  $10^{-13}$  was used in this study. 11,647 subjects were genotyped by TaqMan assays for the two APOE SNPs: rs429358 (E4) and rs7412 (E2) to confirm the APOE 2/4 genotype. For subcloning, A 177 bp product with single deoxyadenosine (A) overhang on 3' end was PCR amplified from genomic DNA (0.5-1.0 µg) using primers [5'GCGGACAT GGAGGACGTG-3'; 5'GGCCTGGTACACTGCCAG-3']<sup>14</sup>. and confirmed by running on agarose gel with 1X TBE buffer. The PCR product was purified using QIAGEN purification kit and ligated into 4 kb PCR™ II-TOPO linearized vector with single 3' deoxythymidine(T) residue and Ampicillin resistance ORF (bases 2173-3033), then transformed into chemically competent bacterial cells (E. coli DH5aTM -T1R) to construct the phased haplotype clones of 177 bp derived from the maternal or paternal chromosome. The clones were cultured on Luria-Bertani (LB) agar containing ampicillin to get multiple copies of the constructed clone. To differentiate the bacterial colonies from the successful clone constructs, 40 mg/ml X-gal was spread onto the plates before culturing the clone. To create a stock of each colony, the colonies with successful constructs were incubated separately overnight at 37 °C into 3 mL LB broth containing ampicillin. DNA was isolated from the cultured colony cells using Biolab miniprep plasmid DNA kit followed by sequencing with M13 forward and reverse primers. The Sequencher (v5.4.6) software was used to analyze the inserted phased haplotype.

We also employed an RFLP assay, utilizing two different enzymes to discern the *APOE5* haplotype<sup>4</sup>. We amplified a 221 bp PCR product using forward (5"CTGTCCAAGGAGCTGCAG3") and reverse (5'GCCCCGGCCTGGTACACTGCCAG3") primers followed by overnight digestion at 37 °C with a 3:2 ratio of smartcut enzymes *AfIIII* (for codon 112 digestion) and *HaeII* (for codon 158 digestion). The samples were loaded onto a 3.5% metaphor GEL using 1XTBE buffer alongside a 50 bp ladder. The *AfIIII* enzyme cleaves at codon112 only in the presence of nucleotide *T*GC(C112), while the *HaeII* enzyme cleaves only if there is nucleotide *C*GC(R158) at codon158; however, the cutting sites at both positions would be abolished in the *E5* haplotype (*C*GC/R112 and *T*GC/C158).

#### Results

#### Demographic characteristics of the subjects

Demographics information along with the *APOE* genotype/allele frequencies in the total sample and sample stratified by case-control are provided in Table 1. Of the14,819 individuals, four subjects lacked case control status. As expected, the frequency of *APOE4* carriers (2/4, 3/4, and 4/4 genotypes) was higher in AD cases than in controls, the difference of which is also reflected in the *APOE4* allele frequency (32.8% vs. 14.6%). Likewise, the frequency of *APOE2* carriers (2/2 and 2/3 genotypes) was lower in AD cases than in controls, as also reflected in the *APOE4* allele frequency (32.8% vs. 14.6%). Likewise, the frequency of *APOE2* carriers (2/2 and 2/3 genotypes) was lower in AD cases than in controls, as also reflected in the *APOE2* allele frequency difference (4.1% vs. 7.9%). Although the 2/4 genotype carries opposite protective and risk alleles, it was a risk factor for AD (odds ratio (OR) = 2.01, *p* = 8.92E-08), which was similar to the OR of 2.6 reported earlier in more than 17,000 cases and controls<sup>15</sup>. Since the *E2* and *E4* alleles in the 2/4 genotype are assumed to be inherited from both parents on different haplotypes, it would be intriguing to investigate the effect of the elusive *E5* haplotype on AD risk where both alleles are inherited from a single parent. Of the 355 subjects with the 2/4 genotype (see Table 1,2), DNA was not available from 15 controls and thus we proceeded with the 340 samples (17% Black) for the search for the elusive haplotype via subcloning, WGS or RFLP assays.

#### Subcloning

Results from subcloning followed by sequencing are shown in Fig. 1. The plasmid DNA sequence from each clone revealed the presence of the following combination of nucleotide bases corresponding to the 1 st base of codon 112 and 1 st base of codon 158: C - C for the E4 haplotype or T - T for the *E2* haplotype. Bold C represents the E4 mutation and bold T the E2 mutation in the aforementioned haplotypes (Fig. 1). In no instance, the C - T combination corresponding to the *E5* haplotype was observed. Identical results were obtained with WGS (results not shown).

#### **RFLP** analysis

The full-length gel image of the RFLP assay is shown in Fig. 2. In the RFLP analysis of the 221 bp PCR amplified fragment followed by double enzyme digestion, the tested samples with the *APOE* 2/4 genotype gave the expected diagnostic bands of 168 bp and 198 bp (Fig. 2). No sample revealed the expected uncut band of 221 bp corresponding to the *E5* haplotype due to loss of restriction sites at codon 112 and codon 158. It is noteworthy that in some samples with the 2/4 genotype, we visualized a faint band at position 221 bp, which was due to incomplete digestion (Fig. 2) as we confirmed it upon sequencing. One study also reported such an undigested band in their one subject with the 2/4 genotype and suggested it to be due to the formation of heterodimers during amplification<sup>4</sup>. However, we think this vague band is most likely due to incomplete digestion rather than being heterodimers because this band was observed only in few 2/4 subjects. This cautionary note may be helpful for those seeking to detect the *E5* haplotype using the RFLP assay alone.

#### Recombinant and recurrent point mutation hypotheses

Next, we estimated the probability of recombination versus recurrent point mutation hypotheses for the generation of the *E5* haplotype.

*Recombinant hypothesis.* We estimated recombination between the two sites based on the assumptions of linkage equilibrium and linkage disequilibrium (LD).

i) Linkage equilibrium: Based on the allele frequencies of E2 (0.0664) and E4 (0.1532) in Europeans in the 1000G-30× project and considering that the two polymorphic sites are in linkage equilibrium, below are the expected four haplotype frequencies.

E2 0.0664 \* (1-0.1532) = 0.0562.

*E3* (1-0.0664) \* (1-0.1532) = 0.7905.

- E4 (1-0.0664) \* 0.1532 = 0.14303.
- $E5 \ 0.0664 * 0.1532 = 0.01017248.$

Thus, the expected recombinant frequency of the *E5* haplotype in Europeans is 1% (2% carriers). However, we did not observe any example of *E5* in TaqMan genotyped 11,647 subjects where the expected number of *E5* carriers was about 236. Similarly, the probability of *E5* haplotype in 340 subjects with the *APOE* 2/4 genotype was 50%. The probability of observing no *E5* haplotype by chance in 11,647 subjects is 4.32E-103and in 340 subjects with *APOE* 2/4 genotype is 4.46E-103.

ii) Linkage disequilibrium: The *E2* allele (rs7412) and the *E4* allele (rs429358) are only 138 bp apart, making them highly likely to be in LD, as we concluded in this study. Therefore, the *E5* haplotype could arise from a rare crossover event between these two alleles during meiosis. Considering that one centimorgan corresponds to approximately a 1% recombination rate and that, on average, one centimorgan spans roughly 1 million base pairs in the human genome, the expected recombination frequency between two loci 138 bp apart is approximately 1.38E-06. Based on the data in Tables 1, 2, the frequency of individuals heterozygous for *E2/E4* is 0.024. Thus, the probability that the *E5* haplotype is generated by a recombination event in an *E2/E4* individual is 0.024 × 1.38E-06 = 3.31E-08. Then, the probability of observing no *E5* haplotype by chance in 14,891 subjects is 0.99, which matches our conclusion.

*Recurrent point mutation.* Assuming that the human DNA point mutation rate is  $2.8 \times 10E-07$  within the same generation<sup>16</sup> we estimated the probability of the *E5* haplotype occurring by point mutation on the *E2* 





and *E4* haplotypes that require one mutational event each, and on the *E3* haplotype that would require two mutational events, and the outcomes are as follows:

E2 > E5: 0.0562 × 2.8E-07 = 1.57E-08.

 $E4 > E5: 0.14303 \times 2.8E-07 = 4.00E-08.$ 

*E3*>*E5*: 0.7905 × 2.8EE-07 × 2.8E-07 = 6.19E-14.

Since these three are independent events, the total *E5* haplotype by recurrent point mutations will be (1) + (2) + (3) = 5.579E-08.

In summary, the probability of the occurrence of E5 by recombination is 1.017% when E2 and E4 are not in LD and 3.31E-08 when they are in LD. The latter probability is similar to the recurrent point mutation of 5.58E-08. Since we did not observe any example of E5 haplotype, it indicates an almost complete LD between the two polymorphic sites where the occurrence of E2 and E4 mutations on the same chromosome (E5 haplotype) is an ultra-rare event.



codon 112 is highlighted.

The first nucleotide base of codon 158 is highlighted.

**Fig. 2.** Full length gel running after restriction enzyme digestion (*Hae*II and *Afl*III) of 221 bp long PCR product for six *APOE* genotypes (2/2, 2/3, 2/4, 3/3, 3/4, 4/4). The respective band sizes for each genotype are given below.*APOE*2/2 168,53; *APOE*2/3 168, 145, 53, 23;*APOE*2/4 198,168, 53, 23; *APOE*3/3 145, 53, 23;*APOE*3/4 198, 145, 53, 23; *APOE*4/4 198, 23 Note: The band size of "23 base pairs" is very small in size so, is invisible for most of the samples and can only be seen for two of the samples (sample no. 6 and 8). We can also see the undigested product of band size "221 base pairs" for the sample no. 1 and 5.

#### TaqMan genotype assays to detect APOE5 haplotype

Although we subcloned and sequenced only *APOE* 2/4 subjects, we employed TaqMan assays to genotype two *APOE* SNPs at codon 112 (rs429358 for T > C mutation, *E*4) and codon 158 (rs7412 for C > T mutation, E2)<sup>10</sup> in 11,647 of the 14,819 subjects included in Table 1,2; the remaining were genotyped using RFLP. The TaqMan genotypes at both sites can enable the determination of which individual might have the *E5* haplotype. By definition, the *E5* haplotype carries the mutant alleles at both sites; *E4* at codon 112 and *E2* at codon 158. If we observe heterozygotes at both sites (double heterozygotes) like in *APOE* 2/4, then an individual could be a potential carrier of the *E5* haplotype. If we observe heterozygote at one site and homozygote at the other site, then we certainly have the *E5* carrier. For example, under the current observation of 3 of the 4 expected *APOE* haplotypes (*E2, E3, E4*) due to LD between the two sites, giving rise to 6 possible genotypes (*E2/E2, E2/ E3, E2/E4, E3/E3, E3/E4, E4/E4*), then only the *E2/E4* genotype would have the *E5* haplotype by having the *E4* allele at codon 112 and *E2* at codon 158. If we consider all the expected haplotypes (*E2, E3, E4, E5*) due to linkage equilibrium between the two sites, then there will be additional genotypes in combination with *E5* (*E2/ E5, E3/E5, E4/E5, E5/E5*). Below are the expected TaqMan genotypes at codons 112 and 158 when *E5* occurs in combination with the other 3 haplotypes (3 = wild type at both sites; 4 = mutant type at codon 112, and 2 = mutant type at codon 158).

Haplotype	Codon 112 genotype (T>C)	Codon 158 genotype (C>T)
E2/E5	T/C (3/4)	T/T (2/2)
E3/E5	T/C (3/4)	C/T (2/3)
E4/E5	C/C (4/4)	C/T (2/3)
E5/E5	C/C (4/4)	T/T (2/2)

Of the above expected TaqMan genotypes among 11, 647 subjects, double heterozygotes at two sites would be only in the E3/E5 haplotype combination; the rest are homozygotes at one site and heterozygotes at the other site. The expected E3/E5 genotype carrying the E5 haplotype accounts for 50% of APOE 2/4 subjects as mentioned above, which we would have observed had there been no LD between the two sites. If we have had detected the E5 haplotype in any of the sequenced APOE 2/4 samples, it would have the E3/E5 haplotype combination, which we did not observe in our investigation.



**Fig. 3.** Schematic representation of potential crossover between two *APOE* polymorphic sites represented by amino acid changes from cysteine (C) to arginine (R) at codon 112 (C112 > R112) and from arginine (R) to cysteine (C) at codon 158 (R158 > C158) that lead to the formation of four expected two-site haplotypes.

#### Discussion

Previously, the *E5* haplotype has been reported in only 5 subjects from three unrelated kindreds. The first case was found in an autistic Italian child and his unaffected mother while investigating the potential association of *APOE* alleles with primary autism in trios<sup>3</sup>. The authors named this haplotype as *E3r* because it possesses reverse arrangement of the cysteine and arginine residues at codons 112 and 158 (R112 - C158) compared to the common *E3* haplotype (C112 - R158). The second case was reported in a 70-year-old healthy Yoruban female with normal lipid profile and in her 34-year-old son from Ibadan, Nigeria and this was named as *E1Y*<sup>4</sup> to differentiate it from the previously reported rare E1 isoform (Asp127 – Cys158)<sup>17</sup>. The third unrelated case of *E5* was observed in a 77-year-old Caucasian patient with motor neuron disease but with normal cognition and lipid profile<sup>5</sup>. This haplotype was not inherited by his two children.

We refer to this elusive haplotype as APOE5 because its earlier designation as E3r or E1Y is misleading, suggesting it may not belong to the APOE2/E3/E4 polymorphism. The original nomenclature of three APOE isoforms was based on their structure and isoelectric focusing (IEF) point differences on gel electrophoresis where IEF point of E2 isoform was more acidic and the IEF point of E4 was more basic compared to the common E3 isoform<sup>6</sup>. The E1 isoform (G127D – R158 C) differs from E2 at amino acid position 127 where glycine is replaced with aspartic acid, causing one negative charge difference from  $E2^{17}$ . Thus, the genetic determinant of E1 is a point mutation at codon 127 and its designation represents its relative IEF position to E2 on gel electrophoresis. This situation is like a rare APOE4Pittsburgh variant (L28P – C112R) which differs from E4 at amino acid position 28 where leucine is replaced with proline<sup>18</sup>. Most importantly, E1 and APOE4Pittsburgh do not correspond to the elusive E5 haplotype, which is part of the common APOE2/E3/E4 polymorphism determined by variation at codons 112 and 158; E5 has point mutations at both codons.

The most likely explanation for the observation of 4 two-site *APOE* haplotypes, *E2*, *E3*, *E4*, and *E5*, is due to intragenic crossover between the nucleotide sequence of codons 112 and 158 (Fig. 3). An alternative explanation for the generation of the *E5* haplotype may be the recurrent point mutations on the background of *E2* or *E4* haplotypes after the split of human races, like the example of the *APOE4Pittsburgh* mutation that occurred on the *E4* background<sup>18</sup>. However, the observation of *E5* haplotype in one African kindred along with two kindreds of European descent belies this hypothesis. We have estimated the probabilities of recombination versus recurrent point mutation hypotheses for the occurrence of the *E5* haplotype, which are similar. While the probability of the *E5* occurrence due to recombination event is 3.31E-08 under the assumption of LD between the two polymorphic sites, it is 5.58E-08 due to recurrent point mutations. These probabilities along with the lack of

detection of any examples of *E5* haplotype in this study confirm the strong LD between the two polymorphic sites and highlight that the occurrence of the *E5* haplotype is an ultra-rare event.

Although *E3* is considered as the parent haplotype or allele in humans because of its common occurrence followed by *E4* and *E2*, *E4* has been postulated as the ancestral allele because all the great apes code for arginine with the identical codon sequence (CGC) at positions 112 and 158 corresponding to the human *E4* haplotype<sup>19</sup>. Accordingly, it has been hypothesized that human *E3* evolved from primate *E4* by a C to T point mutation coding for cysteine at codon 112 (*T*GC) and then *E2* evolved from *E3* by a C to T transition coding for cysteine at codon 158 (*T*GC). As above, we hypothesize that the *E5* haplotype (R112 – C158) was formed most likely due to crossover between the *E4* and *E2* haplotypes (Fig. 3).

The prevalence of *APOE4* varies across ethnic groups and it may contribute to the diverse patterns of AD susceptibility. The average allele frequency of *APOE4* ranges from 7% in Asians 14% in Europeans and 26–37% in Africans, Australian Aborigines, and Pacific populations<sup>2</sup>. Despite Asians have lower prevalence of *APOE4* than Europeans, their genetic risk for developing AD is similar. On the other hand, the *APOE4*-associated AD risk in Africans is lower than Europeans despite having higher prevalence of *APOE4*<sup>2</sup>. These differences suggest that the genetic architecture of AD may be shaped by population-specific genetic architecture, environmental factors, gene-gene and gene environmental interactions. Understanding such variations is crucial for improving the accuracy of genetic risk prediction models and developing tailored therapeutic and preventive strategies for AD across diverse ethnic groups.

In conclusion, we have performed a systematic and focused search to identify the elusive E5 haplotype in the general population by cloning and sequencing a large number of subjects heterozygous for the APOE 2/4 genotype but found no such example. For this reason, we could not examine its role in AD risk. Our data suggests that the occurrence of E5 is extremely rare, and it might have a minimum effect, if any, on disease risk.

#### Data availability

Sequencing data has been submitted to the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) and is available at https://dss.niagads.org/studies/sa000012/. The principal investigator (corresponding author) of the study may be contacted if someone needs to acquire the data.

Received: 12 August 2024; Accepted: 5 May 2025 Published online: 15 May 2025

#### References

- 1. Bellenguez, C. et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* 54, 412–436 (2022).
- 2. Kamboh, M. I. Genomics and functional genomics of Alzheimer's disease. *Neurotherapeutics* **19**, 152–172 (2022).
- 3. Persico, A. M. et al. Enhanced APO\*E2 transmission rates in families with autistic probands. *Psychiat Genet.* 14, 73–82 (2004).
- 4. Murrell, J. R. et al. The fourth Apolipoprotein E haplotype found in the Yoruba of Ibadan. Am. J. Med. Genet. B141, 426-427 (2006).
- 5. Seripa, D. et al. The missing APOE allele. Ann. Hum. Genet. 71, 496-500 (2007).
- 6. Zannis, V. I., Breslow, J. L. & Apolipoprotein, E. Mol. Cell. Bioch 42, 3-20 (1982).
- 7. Menzel, H., Kladetzky, R. & Assmann, G. One-step screening method for the polymorphism of apolipoproteins AI, A-II, and A-IV. *J. Lipid Res.* 23, 915–922 (1982).
- Ehnholm, C., Lukka, M., Kuusi, T., Nikkilä, E. & Utermann, G. Apolipoprotein E polymorphism in the Finnish population: gene frequencies and relation to lipoprotein concentrations. J. Lipid Res. 27, 227–235 (1986).
- 9. Kamboh, M. I., Ferrell, R. E. & Kottke, B. Genetic studies of human Apolipoproteins. V. A novel rapid procedure to screen Apolipoprotein E polymorphism. J. Lipid Res. 29, 1535–1543 (1988).
- Fan, K. H. et al. Investigation of the independent role of a rare APOE variant (L28P; APOE\* 4Pittsburgh) in late-onset alzheimer disease. Neurobiol. Aging. 122, 107–111 (2023).
- Kamboh, M. I. et al. Genome-wide association study of Alzheimer's disease. *Transl Psychiat.* 2, 117. https://doi.org/10.1038/tp.201 2.45 (2012).
- 12. Pirim, D. et al. Apolipoprotein E-C1-C4-C2 gene cluster region and inter-individual variation in plasma lipoprotein levels: a comprehensive genetic association study in two ethnic groups. *PloS One.* **14**, 0214060. https://doi.org/10.1371/journal.pone.0214 060 (2019).
- 13. Harper, J. D. et al. Genome-wide association study of incident dementia in a community-based sample of older subjects. J. Alzheimers Dis. 88, 787-798 (2022).
- Kamboh, M. I., Aston, C. E. & Hamman, R. F. The relationship of APOE polymorphism and cholesterol levels in normoglycemic and diabetic subjects in a biethnic population from the San Luis Valley, Colorado. Atherosclerosis 112, 145–159 (1995).
- 15. Genin, E. et al. APOE and alzheimer disease: a major gene with semi-dominant inheritance. Mol. Psychiat. 16, 903-907 (2011).
- 16. Milholland, B. et al. Differences between germline and somatic mutation rates in humans and mice. *Nat. Commun.* **8**, 15183 (2017).
- 17. Weisgraber, K. H. et al. A novel electrophoretic variant of human Apolipoprotein E. Identification and characterization of Apolipoprotein E1. J. Clin. Invest. 73, 1024–1033 (1984).
- 18. Kamboh, M. I. et al. A novel mutation in the Apolipoprotein E gene (*APOE\*4Pittsburgh*) is associated with the risk of late-onset Alzheimer's disease. *Neurosci. Lett.* **263**, 129–132 (1999).
- Finch, C. E. & Sapolsky, R. M. The evolution of alzheimer disease, the reproductive schedule, and ApoE isoforms☆. Neurobiol. Aging. 20, 407–428 (1999).

#### Acknowledgements

The study was supported in part by NIH grants R01 AG064877 and P30 AG066468.

#### Author contributions

Conceptualization, Funding acquisition, Supervision: M. Ilyas KambohData curation: Asma Naseer Cheema, Elizabeth Lawrence, Narges Zafari, Kang-Hsien Fan, Ruyu Shi, Muaaz Aslam, Vibha Acharya, Alayna Jean Hol-

derman, Annie Bedison, Eleanor Feingold Formal analysis: M. Ilyas Kamboh, Asma Naseer Cheema, Ruyu Shi, Kang-Hsien Fan Methodology and Investigation: Asma Naseer Cheema, Elizabeth Lawrence, Narges Zafari, Kang-Hsien FanWriting – original draft: Asma Naseer Cheema Writing – review and editing: M. Ilyas Kamboh, Asma Naseer Cheema; Ruyu Shi, Muhammad Muaaz Aslam, Vibha Acharya.

#### Funding sources

NIH grants R01 AG064877 and P30 AG066468.

#### Declarations

#### Competing interests

The authors declare no competing interests.

#### Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/1 0.1038/s41598-025-01263-0.

Correspondence and requests for materials should be addressed to M.I.K.

Reprints and permissions information is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025