Analysis of whole genome sequenced cases and controls shows that the association of variants in *TOMM40, BCAM, NECTIN2* and *APOC1* with late onset Alzheimer's disease is driven by linkage disequilibrium with *APOE* ε2/ε3/ε4 alleles.

David Curtis and the Alzheimer's Disease Neuroimaging Initiative

UCL Genetics Institute, UCL, Darwin Building, Gower Street, London WC1E 6BT.
Centre for Psychiatry, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ.

Correspondence:
David Curtis *d.curtis@ucl.ac.uk*
00 44 7973 906 143

Running head: Role of *APOE* in Alzheimer's disease

Abstract

Variants in *APOE* are associated with risk of late onset Alzheimer's disease (LOAD) but the magnitude of the effect has been reported to vary across ancestries. Also, other variants in the region have been reported to show association though it has been unclear whether this was secondary to their linkage disequilibrium with the *APOE* variants rs429358 and rs7412. Previous analyses of exome-sequenced samples have identified other genes in which rare variants impact risk of disease. In this study 2000 whole genome sequenced cases and controls with different ancestries were subjected to gene-based weighted burden analysis to identify risk genes. Additionally, individual variants in the *APOE* region were tested for association with LOAD. When using the *APOE* variants as covariates no individual genes showed statistically significant evidence for association after Bonferroni correction for multiple testing, which may well be a consequence of the modest sample size. Likewise, for those variants initially showing evidence of association with LOAD incorporating the *APOE* variants as covariates dramatically reduced the strength of association. These results demonstrate that the differential association of *APOE* across ancestries does not appear to be driven by another variant in the region. It seems likely that no other genes in the region have a direct effect on LOAD risk.

Keywords

LOAD; *BCAM*; *NECTIN2*; *TOMM40*; *APOC1*; *APOE*.

Introduction

The first report of association between *APOE* alleles and late onset Alzheimer's disease (LOAD) noted that there were three isoforms of the ApoE protein, referred to as ApoE-E2, ApoE-E3 and ApoE-E4, and that restriction-based isotyping showed that the frequency of the *APOE* ε4 allele in 30 unrelated late onset familial cases was significantly different from that in 91 controls (0.50 v. 0.16, p=0.014) (Strittmatter et al., 1993). A table in the same paper showed the allele frequency of the *APOE* ε2 allele to be 0.04 in cases versus 0.08 in controls although this was not commented on at the time but a protective effect for the *APOE* ε2 allele was reported the following year (Corder et al., 1994). Apo-E4 contains an arginine at residue 112 whereas E3 has a cysteine and the variant encoding this substitution is now named rs429358:T>C. Apo-E2 contains a cysteine at residue 158 while E3 has an arginine and this is coded for by rs7412:C>T. The two variants are in strong LD with each other and it is thought that the ε3 allele derived from the ancestral ε4 allele and that the ε2 allele subsequently derived from the ε3 allele, with the ε4 allele possibly being under negative selection owing to its association with a less favourable lipid profile and increased risk of cardiovascular disease (Fullerton et al., 2000).

The association of these *APOE* variants with LOAD was confirmed and functional studies in mice have shown that the human APOE isoforms are associated with differences in amyloid plaque deposition (E4 > E3 > E2) and differences in clearance of amyloid-β (Aβ) peptide (E2 > E3 > E4) (Bales et al., 2009; Castellano et al., 2011). A very rare allele of *APOE* is referred to as the *APOE3* Christchurch (R136S) variant, GRCh38:g.19-44908756C>A (Wardell et al., 1987). There is a recent case report of an elderly subject with the pathogenic *PSEN1* E280A variant, an autosomal dominant cause of presenile Alzheimer's disease, who was also homozygous for the *APOE3* Christchurch variant and who had high brain amyloid levels with only mild cognitive impairment (Arboleda-Velasquez et al., 2019). Investigations revealed she had high amyloid-β plaque burden but limited neurodegeneration or tau accumulation. In vitro experiments showed that the APOE3ch isoform triggered less Aβ42 aggregation than APOE3, suggesting a possible mechanism underlying a protective effect. They also showed that APOE3ch has the lowest heparin binding ability of all APOE isoforms, whereas APOE4 had been shown to have a higher heparin affinity than APOE3 (Yamauchi et al., 2008).

Although the association of APOE variants with LOAD was repeatedly replicated, differences related to ancestry were soon reported. For purposes of clarity, the original nomenclature is retained here even though it might differ from the language we would use to describe ancestry effects in today's literature. An early meta-analysis reported that the *APOE* ε4 AD association was weaker among African Americans and Hispanics than for Caucasians (Farrer et al., 1997) and a recent study confirmed that rs429358 was more strongly associated with LOAD in whites than in Hispanics, with the effect in African-Americans being intermediate (Kulminski et al., 2019). This study analysed variants from the region including *BCAM*, *NECTIN2*, *TOMM40*, *APOE*, and *APOC1* and observed that rs429358 and rs7412 were in LD with nearby SNPs and that the patterns of LD varied between ethnicities. If nearby SNPs independently influenced LOAD risk then such differences could theoretically account for

different strengths of association between rs429358, the *APOE* ε4 variant, and LOAD in different ancestries. For example, there have been inconsistent claims that rs449647 and rs405509 are independently associated with LOAD, with one study reporting that the haplotype of rs405509-T with rs429358-C increased risk whereas the rs405509-G with rs429358-C did not, although this effect was restricted to subjects aged over 75 (Lescai et al., 2011). A study in 525 whole genome sequenced Chinese subjects and large samples of Chinese and non-Asian subjects identified haplotypes in the region exerting independent effects on risk, apparently exerted via modulating gene expression, but this study did not explicitly address the question of whether these haplotypes might explain the differential effects of APOE ε4 in different ancestries (Zhou et al., 2019). The haplotypes reported were made up of multiple non-coding variants, none of which individually had a substantial effect. Two additional recent studies have demonstrated that local ancestry around *APOE* moderates the risk associated with the *APOE* alleles, again suggesting that other nearby variants might have an effect (Blue et al., 2019; Rajabli et al., 2018).

One way to explain the observation that rs429358-C is more strongly associated with LOAD risk in Caucasians than Hispanics would be to postulate another locus influencing risk in linkage disequilibrium (LD) with it. LD describes the situation where there are two polymorphic loci which are close together on the same chromosome for which particular pairs of alleles occur together in the same haplotype more often than would be expected by random segregation. Envisioning this scenario, the high risk allele at the second locus might tend to occur in the same haplotype as rs429358-C in Caucasians and in the same haplotype as rs429358-T in Hispanics. Thus it would amplify the effect of rs429358-C in Caucasians and counteract it in Hispanics. If the LD between the loci was strong within each group then in a homogeneous sample of either Caucasians or Hispanics it might not be possible to detect an independent effect of the second variant. However if one analysed a sample consisting of a mixture of Caucasian and Hispanic cases and controls then in a multivariate analysis the separate contributions of both loci should become apparent. The present report describes such analyses in a mixed sample. The purpose of these analyses was firstly to determine whether it was possible to identify genes apart from APOE affecting LOAD risk and secondly to determine whether variants close to APOE exerted independent effects on LOAD risk, especially in a way that could explain differences in APOE associations between different ancestries.

Methods

Data used in the preparation of this article was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early

Alzheimer's disease. The investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report (Hurko et al., 2012). A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

Phenotype information and variant calls based on whole genome sequencing using the GRCh38 assembly for case-control subjects of the extension phase of the Alzheimer's Disease Sequencing Project (ADSP) were downloaded from the NIAGADS website (https://www.niagads.org/adsp/content/study-design). Ethical approval and informed consent had been obtained by the researchers who generated this dataset. As described previously, participants were at least 60 years old and although different diagnostic procedures were used by different contributing studies all cases met NINCDS-ADRDA criteria for possible, probable or definite AD based on clinical assessment, or had presence of AD (moderate or high likelihood) upon neuropathology examination (Beecham et al., 2017). The subjects are described as being of Non-Hispanic White (NHW), Caribbean Hispanic (CH), and African American (AA) descent. Whole genome sequencing was carried out using standard methods as described previously (Naj et al., 2019; Vardarajan et al., 2018) and on the ADSP website: https://www.niagads.org/adsp/content/sequencing-pipelines. The PrevAD (prevalent Alzheimer's disease) field was used to define phenotype and subjects who had been included in the earlier exome-sequenced datasets were excluded, yielding 985 cases and 2,101 controls.

To obtain population principal components, version 1.90beta of *plink* (https://www.cog-genomics.org/plink2) was run on the chromosome 22 genotypes with the options *--maf 0.1 --pca header tabs --make-rel* (Chang et al., 2015; Purcell et al., 2007, 2009). This is a standard approach whereby the genotypes of a large number of loci are entered into a principal components analysis. The main principal components obtained are taken to reflect any geographical and ancestry-related sub-stratification within the samples and incorporating them into association analyses as covariates can mitigate problems caused by inadequate matching of cases and controls.

The analyses were carried out in two stages. Initially a weighted burden analysis of each gene was carried out using GENEVARASSOC and SCOREASSOC in the same way as had previously been applied to the ADSP exome-sequenced case-control samples (Curtis et al., 2019). For each gene, variants in the exomes and splice regions were extracted and annotated using VEP, PolyPhen and SIFT (Adzhubei et al., 2013; Kumar et al., 2009; McLaren et al., 2016). These programs assess the likely impact of DNA changes on the protein coded for by the gene and in particular whether the changes produced are likely to disrupt the normal functioning of the protein. SCOREASSOC was then used to carry out a weighted burden analysis to test whether, in a particular gene, variants which were rarer and/or predicted to have more severe functional effects occurred more commonly in cases than

controls. Variants were weighted according to their functional annotation using the default weights provided with the GENEVARASSOC program, which was used to generate input files for weighted burden analysis by SCOREASSOC (Curtis, 2016, 2012). For example, a weight of 5 was assigned for a synonymous variant, 10 for a non-synonymous variant and 20 for a stop gained variant. Additionally, 10 was added to the weight if the PolyPhen annotation was possibly or probably damaging and also if the SIFT annotation was deleterious, meaning that a non-synonymous variant annotated as both damaging and deleterious would be assigned an overall weight of 30. The full set of weights is shown in Supplementary Table 1, copied from the previous report (Curtis et al., 2018). Variants were excluded if they did not have a PASS in the information field or if there were more than 10% of genotypes missing in either cases or controls or if the heterozygote count was smaller than both homozygote counts in both cohorts. For each subject a gene-wise weighted burden score was derived as the sum of the variant-wise weights, each multiplied by the number of alleles of the variant which the given subject possessed. If a subject was not genotyped for a variant then they were assigned the subject-wise average score for that variant.

As described previously, ridge regression analysis with lamda=1 was used to test whether the gene-wise score was associated with caseness (Curtis et al., 2018). To do this, SCOREASSOC first calculates the likelihood for the phenotypes as predicted by the first 20 principal components and then calculates the likelihood using a model which additionally incorporates the gene-wise scores. It then carries out a likelihood ratio test assuming that twice the natural log of the likelihood ratio follows a chi-squared distribution with one degree of freedom to produce a p value. Results are expressed as signed log p (SLP) which is the logarithm base 10 of this p value and is positive if the scores are higher in cases and negative if they are higher in controls. This analysis was carried out for every autosomal gene listed in the RefSeq GRCh38. All the analyses were then repeated using rs429358 and rs7412 genotypes as covariates as well as the principal components.

Data handling and generic statistical analyses were carried out using R (R Core Team, 2014).

In the initial weighted burden analyses without including rs429358 and rs7412 as covariates only two genes showed evidence for association which was statistically significant after Bonferroni correction for the number of genes tested, *APOE* and the neighbouring gene *TOMM40*. Therefore a second set of analyses was carried out in order to explore whether the result obtained for *TOMM40* was likely to be due to direct functional effects of variants in this gene or whether it was more likely a consequence of LD relationships with *APOE* variants. Because previous reports have also claimed that variants in the neighbouring genes *APOC1, BCAN* and *NECTIN2* might influence LOAD risk, all variants in the region including these five genes plus a margin of 10 kb at either end were extracted, between genomic coordinates 44799059 and 44929344. For each of the 718 variants with a minor allele frequency of at least 0.01 in cases and/or controls in the combined sample an individual ridge regression analysis was carried out including population principal components and

testing the effect of the variant with and without also including as covariates the genotypes for rs429358 and rs7412. In order to assess ancestry effects, analyses were carried out without including population principal components and raw odds ratio (OR) for association with LOAD was obtained for each variant. All variants with OR>2 or OR<0.6 were selected for further analysis. For these variants showing association with LOAD, the count of the number of alternate alleles was obtained and this allele count was then treated as a quantitative variable in joint ridge regression analyses with PrevAD as the outcome. In an attempt to identify those variants which might be exerting an independent effect on risk, a subset of 14 variants was selected using an arbitrary threshold of |beta/SE(beta)| > 1.3, with beta being the coefficient for the variant in the fitted multivariate model.

In order to explore whether any of these variants was exerting a direct effect on LOAD risk, likelihood ratio tests were carried out for each associated variant by comparing the likelihoods of models which did or did not include the variant in question, and which included the following as covariates:

1. None.

2. Each of the 14 variants identified above as possibly having an independent effect.

3. Rs429358 and rs7412 together.

These analyses were then repeated for each of the three ancestry groups separately. The results of these analyses are expressed as minus log p (MLP), which is the minus logarithm base 10 of the p value for the likelihood ratio test.

In order to explore the possible independent contributions of individual variants further, the subjects were subdivided according to the rs429358 genotype. Based on allele counts in cases and controls an odds ratio and Pearson chi-squared statistic were calculated for each variant in each group. Again, this process was then repeated with the subjects additionally divided by ancestry groups.

Results

Table 1 presents the age and sex distribution of the samples used. The weighted burden tests evaluated 1,662,028 variants in 28,862 autosomal genes. *TOMM40* produced an SLP of 20.43 and *APOE* produced an SLP of 13.1. No other gene produced results which would be considered statistically significant after correction for the number of genes tested but it may be worth noting that *APOC1* had the fourth highest SLP, of 4.2, and that *TREM2* had the eighth highest SLP, of 3.5. When these analyses were repeated using rs429358 and rs7412 as covariates no gene produced significant results. The SLP for *APOE* was 0.93 and for *TOMM40* was 0.86, suggesting no evidence for an independent effect of any other variants. The SLP of *APOC1* reduced to 1.94 while the SLP for *TREM2* fell only slightly to 3.3, ninth highest of all genes. No subject carried an *APOE3* Christchurch variant.

There were 718 genotyped common variants in the region between 10kb proximal of *BCAM* and 10kb distal of *APOC1*. In the analysis including principal components, 91 produced an MLP greater than 4.15, the critical value to be considered significant at p<0.05 after correction for multiple testing, when the *APOE* alleles were not included as covariates. For all of these the MLP reduced, sometimes dramatically, when APOE alleles were included in the model, indicating that initial result had been driven via LD. The highest MLP obtained after inclusion of the *APOE* alleles was 3.72 for rs59007384 and Table 2 shows the results for all the variants which produce an MLP of greater than 2.5 in the analyses including *APOE*. Important, none of the 718 variants produced a markedly increased MLP when the *APOE* alleles were included. The largest increase was for rs145449661, from 0.59 to 2.24.

In the analyses not including principal components, the OR was over 2 for 58 variants and under 0.6 for 10. When all 68 of these variants were entered into a ridge regression analysis there were 14 for which |beta/SE(beta)| exceeded 1.3, consisting of rs140824606, rs149626525, rs7258166, rs147711004, rs116094317, rs41289512, rs142778802, rs79701229, rs6857, rs34404554, rs11556505, rs429358, rs7412 and rs1081105. However in preliminary analyses rs66626994 provided some evidence of being independently associated with LOAD so it was added to the list of variants to be used as covariates in the likelihood ratio tests for each of the 68 individually associated variants. The results of these likelihood ratio tests are in Supplementary Table S3. The main finding of note is that for every variant initially showing strong evidence of association with LOAD the MLP dramatically decreased when rs429358 was included as a covariate. In the adjusted analyses the strongest evidence for independent association was for rs66626994, which produced MLP=3.93 when rs429358 was included and MLP=3.58 when both rs429358 and rs7412 were included. By contrast, the MLP for rs429358 itself remained over 20 no matter which other variants were included as covariates. The MLP for rs7412 was 6.70 on its own and fell to 2.56 when rs429358 was included. The MLPs obtained when both rs429358 and rs7412 were included as covariates were not markedly different from those obtained when including only rs429358.

In terms of effect size, the estimate for the OR for rs429358 on its own was 3.05 with 95% confidence interval 2.65-3.51. Including other variants as covariates had very little impact on this estimate, the lowest value of OR = 2.55 (2.15-3.02) being obtained when including rs6857. Thus there is no suggestion that any other variants in the region have a modifying effect on rs429358. By contrast, there were striking differences in the effect sizes for rs429358 between different ancestry groups, with OR = 5.36 (4.18-6.87) in NHW subjects, OR = 1.72 (1.35-2.20) in CH subjects and OR = 3.15 (2.41-4.11) in AA subjects. Within ancestry groups, these estimates were not impacted by including other variants as covariates.

For rs7412, the unadjusted estimate was OR = 0.53 (0.42-0.68). With rs429358 also included this changed to 0 .68 (0.53-0.88) but including other variants had little effect. Including

rs429358 produced OR = 0.62 (0.37-1.05) in NHW subjects, OR = 0.83 (0.55-1.26) in CH subjects and 0.65 (0.42-0.98) for AA subjects. Again, within ancestry groups variants apart from rs429358 had little effect on the estimates.

For rs66626994, the estimate on its own was OR = 2.13 (1.84-2.48). With rs429358 and rs7412 included this fell to OR = 1.37 (1.15-1.62) in the whole cohort and was OR = 1.34 (.95-1.89) in NHW subjects, OR = 1.00 (0.73-1.38) in CH subjects and OR = 1.51 (1.09-2.08) in AA subjects.

Analyses of variants subdivided by rs429358 genotype did not reveal any which showed independent evidence for association with LOAD risk, either in the sample as a whole or in individual ancestry groups.

The results from the various analyses detailed above show that, after correction for multiple testing, none of the common variants tested had a statistically significant independent association with LOAD risk apart from rs429358. Further examination of results broken down by ancestry and by rs429358 genotype did not reveal evidence that other variants were independently exerting an effect within ancestry cohorts.

Discussion

Given the sample size, the results from the weighted burden analyses are consistent with expectations, in that no gene produces statistically significant results once the effect of rs429358 is incorporated. This suggests that the initially significant result obtained for *TOMM40* simply results from LD between variants in it and rs429358. The fact that other genes with prior convincing evidence for involvement in LOAD do not produce significant results is unsurprising. When similar methods were applied to a sample approximately five times larger significant results were obtained for *TREM2*, *ABCA7* and *SORL1* (Curtis et al., 2019). Of these, the most significant was *TREM2*, which produced a likelihood ratio statistic of 57.9. Since this is distributed as a chi-squared statistic with one degree of freedom, the expectation for the likelihood ratio statistic to be produced from the current sample is approximately 1+(57.9-1)/5=12.4. In fact, the actual likelihood ratio statistic produced was only slightly lower than this, 12.0, indicating that the current results are entirely consistent with those obtained previously but that the sample size is too small to expect to obtain significant results.

More detailed analysis of variants in the region did not produce any convincing evidence for any other variant to exert a causative effect on LOAD risk. (Here, we use the term causative to mean that a variant itself makes a direct contribution to risk rather than showing association for some other reason.) Of course it is not possible to completely exclude the possibility that one or more of them might exert some small effect which might be detected in a larger sample, either on their own or through epistatic effects on the APOE variants. However for all variants showing initially strong evidence of association with LOAD, the

evidence dramatically diminishes when rs429358 is included as a covariate. This implies that their association is at least primarily driven by LD relationships with rs429358. Conversely, the evidence supporting rs429358 remains very strong no matter which other variants are included as covariates, consistent with the notion that it exerts a direct causative effect. We might note that the current study does not provide support for a protective role for rs7412-T, since the initial MLP was only 6.70 and this fell to 2.56 when rs429358 was included. Using the standard allelic notation, th is is equivalent to initially comparing the effects of *APOE ε2* against *ε3* and *ε4* combined and then subsequently comparing *ε2* against *ε3* alone. The lower risk associated with *ε2* against *ε3* is a consistent finding of other studies and the relatively weak signal in the current study is expected given the low allele frequency of rs7412-T and should not be taken to undermine support for it having a real protective effect. The very low risk of LOAD in rs7412-T homozygotes has recently been confirmed in a large study (Reiman et al., 2020).

In order to asses whether the current sample would have the power to detect a variant modifying the effect of rs429358 in such a way at to explain the differential effect across ancestries, we should begin by noting that in this heterogenous sample the effect of such a variant should emerge more strongly when the *APOE* variants are included as covariates. In fact, there is no variant for which the MLP increases substantially. Instead, we tend to observe the opposite effect, explained by variants simply being in LD but having no direct causative effect themselves. To quantify this further, we can consider the situation in which there is a second variant affecting risk which has different LD relationships with rs429358 in different ancestry groups. The effect on differential risk will be maximal if it is in complete LD with rs429358 with reciprocal relationships such that in the NHW subjects the higher risk allele always occurs in phase with the C allele of rs429358 (which confers increased risk) while in CH subjects it is the lower risk allele of the second variant which always occurs in phase with the C allele of rs429358. In this scenario, we observe that the two variants together, tagged by rs429358, have an OR of 5.36 in NHW subjects and 1.72 in CH subjects, equivalent to logistic regression coefficients of the natural logs of these ORs, 1.68 and 0.53. We could then break this down as a contribution from the two loci separately, with rs429358 having a coefficient of 1.105 and the putative second variant having a coefficient of 0.575 to fit the observed values (since 1.105+0.575=1.68 and 1.105-0.575=0.53). This is the minimum effect size which a single variant explaining the differential risk between ancestries could have. If the LD relationships were weaker the effect size of the variant would need to be larger. In an analysis which included rs429358 genotype as a covariate the effect of this variant would emerge and the OR associated with this coefficient of 0.575 would be exp(0.575)=1.8. Using a standard sample size calculator (https://www.stat.ubc.ca/~rollin/stats/ssize/caco.html), we find that in these samples a common variant with OR of 1.8 has over 90% power to produce an MLP of 6 whereas the highest MLP we observe when rs429358 genotype is included as a covariate is only 3.93 (as shown in Supplementary Table 3). From these considerations, we conclude that there is

good power to detect a variant in the region which would explain the differential effect of rs429358 across ancestries and that it is unlikely that such a variant is in fact present.

In this study we do observe that the effect of rs429358 differs between the ancestry groups. However the low magnitude of the effect noted in the CH sample could be a simple consequence of the lower age of controls relative to cases in this sample, since it would be expected that in this situation any signal from any genetic effect on risk would be reduced. There is no obvious explanation for the intermediate effect size observed in the AA subjects and this might represent some real biological phenomenon or might be due to other, unknown mechanisms such as ascertainment procedures or differential cardiovascular mortality. From other studies, the difference in rs429358 effects between ancestries seems to be a fairly consistent finding which, from the current study, cannot be explained by the presence of another causal variant in LD with it. It is difficult to exclude the possibility that a number of variants might cumulatively have some effect, although the present study does not detect any evidence for this. This would be consistent with the previous findings that local ancestry is associated with rs429358 effect size. If the previously reported haplotypes associated with LOAD risk via effects on gene expression of *APOE* were in different LD relationships with rs429358 in different ancestry groups then this might potentially account for the differential association (Zhou et al., 2019). This could be investigated using large multi-ethnic samples genotyped for the relevant markers.

Theoretically it remains a possibility that there could be an unknown variant which was driving this effect. An example would be the poly-T polymorphism in intron 6 of *TOMM40*, rs10524523, which is in LD with rs429358 and which has been inconsistently reported to be associated with risk of LOAD and also a large number of other related phenotypes as reviewed recently (Chiba-Falek et al., 2018). This variant was not provided in the data release, presumably because of problems calling a single nucleotide repeat accurately from next generation sequencing pipelines.

When looked at in their entirety, reports of involvement of individual genes and variants in this region resemble the findings which were reported in the candidate gene era. Because of LD relationships with rs429358 they do show a primary association with LOAD. But when studies try to examine whether they exert independent effects one runs into problems with small sample size, subgroup analyses, application to different phenotypes and, presumably, publication bias. We should start from the position that it is inherently unlikely that by chance there will be a gene close to rs429358 which happens to be causative. It could also be argued that it is *a priori* quite unlikely that variants such as intronic single nucleotide repeats and intergenic variants would have important functional effects. Taking this consideration into account and incorporating the results from the current study it seems reasonable to conclude that, aside from rs429358 and rs7412, other variants in this region do not individually have a substantial effect on risk of LOAD.

References

Adzhubei, I., Jordan, D.M., Sunyaev, S.R. (2013) Predicting functional effect of human missense mutations using PolyPhen-2. Curr. Protoc. Hum. Genet. 7 Unit7.20.

Arboleda-Velasquez, J.F., Lopera, F., O'Hare, M., Delgado-Tirado, S., Marino, C., Chmielewska, N., Saez-Torres, K.L., Amarnani, D., Schultz, A.P., Sperling, R.A., Leyton-Cifuentes, D., Chen, K., Baena, A., Aguillon, D., Rios-Romenets, S., Giraldo, M., Guzmán-Vélez, E., Norton, D.J., Pardilla-Delgado, E., Artola, A., Sanchez, J.S., Acosta-Uribe, J., Lalli, M., Kosik, K.S., Huentelman, M.J., Zetterberg, H., Blennow, K., Reiman, R.A., Luo, J., Chen, Y., Thiyyagura, P., Su, Y., Jun, G.R., Naymik, M., Gai, X., Bootwalla, M., Ji, J., Shen, L., Miller, J.B., Kim, L.A., Tariot, P.N., Johnson, K.A., Reiman, E.M., Quiroz, Y.T. (2019) Resistance to autosomal dominant Alzheimer's disease in an APOE3 Christchurch homozygote: a case report. Nat. Med. 25, 1680–1683.

Bales, K.R., Liu, F., Wu, S., Lin, S., Koger, D., DeLong, C., Hansen, J.C., Sullivan, P.M., Paul, S.M. (2009) Human APOE isoform-dependent effects on brain β-amyloid levels in PDAPP transgenic mice. J. Neurosci. 29, 6771–6779.

Beecham, G.W., Bis, J.C., Martin, E.R., Choi, S.-H., DeStefano, A.L., van Duijn, C.M., Fornage, M., Gabriel, S.B., Koboldt, D.C., Larson, D.E., Naj, A.C., Psaty, B.M., Salerno, W., Bush, W.S., Foroud, T.M., Wijsman, E., Farrer, L.A., Goate, A., Haines, J.L., Pericak-Vance, M.A., Boerwinkle, E., Mayeux, R., Seshadri, S., Schellenberg, G. (2017) The Alzheimer's Disease Sequencing Project: Study design and sample selection. Neurol. Genet. 3, e194.

Blue, E.E., Horimoto, A.R.V.R., Mukherjee, S., Wijsman, E.M., Thornton, T.A. (2019) Local ancestry at APOE modifies Alzheimer's disease risk in Caribbean Hispanics. Alzheimers. Dement. 15, 1524–1532.

Castellano, J.M., Kim, J., Stewart, F.R., Jiang, H., DeMattos, R.B., Patterson, B.W., Fagan, A.M., Morris, J.C., Mawuenyega, K.G., Cruchaga, C., Goate, A.M., Bales, K.R., Paul, S.M., Bateman, R.J., Holtzman, D.M. (2011) Human apoE isoforms differentially regulate brain amyloid-β peptide clearance. Sci. Transl. Med. 3, 89ra57.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7.

Chiba-Falek, O., Gottschalk, W.K., Lutz, M.W. (2018) The effects of the TOMM40 poly-T alleles on Alzheimer's disease phenotypes. Alzheimer's Dement.

Corder, E.H., Saunders, A.M., Risch, N.J., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Rimmler, J.B., Locke, P.A., Conneally, P.M., Schmader, K.E., Small, G.W., Roses, A.D., Haines, J.L., Pericak-Vance, M.A. (1994) Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. Nat. Genet. 7, 180–184.

Curtis, D. (2012) A rapid method for combined analysis of common and rare variants at the level of a region, gene, or pathway. Adv Appl Bioinform Chem 5, 1–9.

Curtis, D. (2016) Pathway analysis of whole exome sequence data provides further support for the involvement of histone modification in the aetiology of schizophrenia. Psychiatr.

Genet. 26, 223–7.

Curtis, D., Bakaya, K., Sharma, L., Bandyopadhay, S. (2019) Weighted burden analysis of exome-sequenced late onset Alzheimer's cases and controls provides further evidence for involvement of PSEN1 and demonstrates protective role for variants in tyrosine phosphatase genes. Ann Hum Genet 84, 291–302.

Curtis, D., Coelewij, L., Liu, S.-H., Humphrey, J., Mott, R. (2018) Weighted Burden Analysis of Exome-Sequenced Case-Control Sample Implicates Synaptic Genes in Schizophrenia Aetiology. Behav. Genet. 43, 198–208.

Farrer, L.A., Cupples, L.A., Haines, J.L., Hyman, B., Kukull, W.A., Mayeux, R., Myers, R.H., Pericak-Vance, M.A., Risch, N., Van Duijn, C.M. (1997) Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: A meta-analysis. J. Am. Med. Assoc.

Fullerton, S.M., Clark, A.G., Weiss, K.M., Nickerson, D.A., Taylor, S.L., Stengård, J.H., Salomaa, V., Vartiainen, E., Perola, M., Boerwinkle, E., Sing, C.F. (2000) Apolipoprotein E variation at the sequence haplotype level: Implications for the origin and maintenance of a major human polymorphism. Am. J. Hum. Genet. 67, 881–900.

Hurko, O., Black, S.E., Doody, R., Doraiswamy, P.M., Gamst, A., Kaye, J., Obisesan, T.O., Rusinek, H., Scharre, D., Sperling, R., Weiner, M.W., Green, R.C. (2012) The ADNI Publication Policy: Commensurate recognition of critical contributors who are not authors. Neuroimage.

Kulminski, A.M., Shu, L., Loika, Y., Nazarian, A., Arbeev, K., Ukraintseva, S., Yashin, A., Culminskaya, I. (2019) APOE region molecular signatures of Alzheimer's disease across races/ethnicities. Neurobiol. Aging.

Kumar, P., Henikoff, S., Ng, P.C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 4, 1073–1081.

Lescai, F., Chiamenti, A.M., Codemo, A., Pirazzini, C., D'agostino, G., Ruaro, C., Ghidoni, R., Benussi, L., Galimberti, D., Esposito, F., Marchegiani, F., Cardelli, M., Olivieri, F., Nacmias, B., Sorbi, S., Tagliavini, F., Albani, D., Boneschi, F.M., Binetti, G., Santoro, A., Forloni, G., Scarpini, E., Crepaldi, G., Gabelli, C., Franceschi, C. (2011) An APOE Haplotype Associated with Decreased 4 Expression Increases the Risk of Late Onset Alzheimer's Disease. J. Alzheimer's Dis. 24, 235–245.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., Cunningham, F. (2016) The Ensembl Variant Effect Predictor. Genome Biol. 17, 122.

Naj, A.C., Lin, H., Vardarajan, B.N., White, S., Lancour, D., Ma, Y., Schmidt, M., Sun, F., Butkiewicz, M., Bush, W.S., Kunkle, B.W., Malamon, J., Amin, N., Choi, S.H., Hamilton-Nelson, K.L., van der Lee, S.J., Gupta, N., Koboldt, D.C., Saad, M., Wang, B., Nato, A.Q., Sohi, H.K., Kuzma, A., Wang, L.S., Cupples, L.A., van Duijn, C., Seshadri, S., Schellenberg, G.D., Boerwinkle, E., Bis, J.C., Dupuis, J., Salerno, W.J., Wijsman, E.M., Martin, E.R., DeStefano, A.L. (2019) Quality control and integration of genotypes from two calling pipelines for whole genome sequence data in the Alzheimer's disease sequencing project. Genomics 111, 808–818.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–75.

Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., Sklar, P., Purcell Leader, S.M., Ruderfer, D.M., McQuillin, A., Morris, D.W., O'Dushlaine, C.T., Corvin, A., Holmans, P. a, Macgregor, S., Gurling, H., Blackwood, D.H.R., Craddock, N.J., Gill, M., Hultman, C.M., Kirov, G.K., Lichtenstein, P., Muir, W.J., Owen, M.J., Pato, C.N., Scolnick, E.M., St Clair, D., Sklar Leader, P., Williams, N.M., Georgieva, L., Nikolov, I., Norton, N., Williams, H., Toncheva, D., Milanova, V., Thelander, E.F., Sullivan, P.F., Kenny, E., Quinn, E.M., Choudhury, K., Datta, S., Pimm, J., Thirumalai, S., Puri, V., Krasucki, R., Lawrence, J., Quested, D., Bass, N., Crombie, C., Fraser, G., Leh Kuan, S., Walker, N., McGhee, K. a, Pickard, B., Malloy, P., Maclean, A.W., Van Beck, M., Pato, M.T., Medeiros, H., Middleton, F., Carvalho, C., Morley, C., Fanous, A., Conti, D., Knowles, J. a, Paz Ferreira, C., Macedo, A., Helena Azevedo, M., Kirby, A.N., Ferreira, M. a R., Daly, M.J., Chambert, K., Kuruvilla, F., Gabriel, S.B., Ardlie, K., Moran, J.L. (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460, 748–752.

R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria., Austria.

Rajabli, F., Feliciano, B.E., Celis, K., Hamilton-Nelson, K.L., Whitehead, P.L., Adams, L.D., Bussies, P.L., Manrique, C.P., Rodriguez, A., Rodriguez, V., Starks, T., Byfield, G.E., Sierra Lopez, C.B., McCauley, J.L., Acosta, H., Chinea, A., Kunkle, B.W., Reitz, C., Farrer, L.A., Schellenberg, G.D., Vardarajan, B.N., Vance, J.M., Cuccaro, M.L., Martin, E.R., Haines, J.L., Byrd, G.S., Beecham, G.W., Pericak-Vance, M.A. (2018) Ancestral origin of ApoE ε4 Alzheimer disease risk in Puerto Rican and African American populations. PLOS Genet. 14, e1007791.

Reiman, E.M., Arboleda-Velasquez, J.F., Quiroz, Y.T., Huentelman, M.J., Beach, T.G., Caselli, R.J., Chen, Y., Su, Y., Myers, A.J., Hardy, J., Paul Vonsattel, J., Younkin, S.G., Bennett, D.A., De Jager, P.L., Larson, E.B., Crane, P.K., Keene, C.D., Kamboh, M.I., Kofler, J.K., Duque, L., Gilbert, J.R., Gwirtsman, H.E., Buxbaum, J.D., Dickson, D.W., Frosch, M.P., Ghetti, B.F., Lunetta, K.L., Wang, L.-S., Hyman, B.T., Kukull, W.A., Foroud, T., Haines, J.L., Mayeux, R.P., Pericak-Vance, M.A., Schneider, J.A., Trojanowski, J.Q., Farrer, L.A., Schellenberg, G.D., Beecham, G.W., Montine, T.J., Jun, G.R. (2020) Exceptionally low likelihood of Alzheimer's dementia in APOE2 homozygotes from a 5,000-person neuropathological study. Nat. Commun. 11, 667.

Strittmatter, W.J., Saunders, A.M., Schmechel, D., Pericak-Vance, M., Enghild, J., Salvesen, G.S., Roses, A.D. (1993) Apolipoprotein E: High-avidity binding to β-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc. Natl. Acad. Sci. U. S. A. 90, 1977–1981.

Vardarajan, B.N., Barral, S., Jaworski, J., Beecham, G.W., Blue, E., Tosto, G., Reyes-Dumeyer, D., Medrano, M., Lantigua, R., Naj, A., Thornton, T., DeStefano, A., Martin, E., Wang, L.S., Brown, L., Bush, W., van Duijn, C., Goate, A., Farrer, L., Haines, J.L., Boerwinkle, E., Schellenberg, G., Wijsman, E., Pericak-Vance, M.A., Mayeux, R., Mosley, Cantwell, L., Childress, M., Chou, Y.F., Cweibel, R., Gangadharan, P., Kuzma, A., Lin, H.J., Malamon, J.,

Mlynarski, E., Naj, A., Qu, L., Schellenberg, G., Valladares, O., Wang, L.S., Wang, W., Zhang, N., Below, Boerwinkle, E., Bressler, J., Fornage, M., Jian, X., Liu, X., Bis, Blue, E., Brown, L., Day, T., Dorschner, M., Nafikov, Navas, P., Nguyen, H., Psaty, B., Rice, K., Saad, M., Sohi, H., Thornton, T., Tsuang, D., Wang, B., Wijsman, E., Appelbaum, E., Cruchaga, C., Koboldt, D.C., Waligorski, J. (2018) Whole genome sequencing of Caribbean Hispanic families with late-onset Alzheimer's disease. Ann. Clin. Transl. Neurol. 5, 406–417.

Wardell, M.R., Brennan, S.O., Janus, E.D., Fraser, R., Carrell, R.W. (1987) Apolipoprotein E2-Christchurch (136 Arg----Ser). New variant of human apolipoprotein E in a patient with type III hyperlipoproteinemia. J. Clin. Invest. 80, 483–490.

Yamauchi, Y., Deguchi, N., Takagi, C., Tanaka, M., Dhanasekaran, P., Nakano, M., Handa, T., Phillips, M.C., Lund-Katz, S., Saito, H. (2008) Role of the N- and C-Terminal Domains in Binding of Apolipoprotein E Isoforms to Heparan Sulfate and Dermatan Sulfate: A Surface Plasmon Resonance Study [†]. Biochemistry 47, 6702–6710.

Zhou, X., Chen, Yu, Mok, K.Y., Kwok, T.C.Y., Mok, V.C.T., Guo, Q., Ip, F.C., Chen, Yuewen, Mullapudi, N., Weiner, M.W., Aisen, P., Petersen, R., Jack, C.R., Jagust, W., Trojanowski, J.Q., Toga, A.W., Beckett, L., Green, R.C., Saykin, A.J., Morris, J., Shaw, L.M., Khachaturian, Z., Sorensen, G., Kuller, L., Raichle, M., Paul, S., Davies, P., Fillit, H., Hefti, F., Holtzman, D., Mesulam, M.M., Potter, W., Snyder, P., Schwartz, A., Montine, T., Thomas, R.G., Donohue, M., Walter, S., Gessert, D., Sather, T., Jiminez, G., Harvey, D., Bernstein, M., Thompson, P., Schuff, N., Borowski, B., Gunter, J., Senjem, M., Vemuri, P., Jones, D., Kantarci, K., Ward, C., Koeppe, R.A., Foster, N., Reiman, E.M., Chen, K., Mathis, C., Landau, S., Cairns, N.J., Householder, E., Taylor-Reinwald, L., Lee, V., Korecka, M., Figurski, M., Crawford, K., Neu, S., Foroud, T.M., Potkin, S.G., Shen, L., Faber, K., Kim, S., Nho, K., Thal, L., Buckholtz, N., Albert, M., Frank, R., Hsiao, J., Kaye, J., Quinn, J., Lind, B., Carter, R., Dolen, S., Schneider, L.S., Pawluczyk, S., Beccera, M., Teodoro, L., Spann, B.M., Brewer, J., Vanderswag, H., Fleisher, A., Heidebrink, J.L., Lord, J.L., Mason, S.S., Albers, C.S., Knopman, D., Johnson, Kris, Doody, R.S., Villanueva-Meyer, J., Chowdhury, M., Rountree, S., Dang, M., Stern, Y., Honig, L.S., Bell, K.L., Ances, B., Carroll, M., Leon, S., Mintun, M.A., Schneider, S., Oliver, A., Marson, D., Griffith, R., Clark, D., Geldmacher, D., Brockington, J., Roberson, E., Grossman, H., Mitsis, E., de Toledo-Morrell, L., Shah, R.C., Duara, R., Varon, D., Greig, M.T., Roberts, P., Onyike, C., D'Agostino, D., Kielb, S., Galvin, J.E., Cerbone, B., Michel, C.A., Rusinek, H., de Leon, M.J., Glodzik, L., De Santi, S., Doraiswamy, P.M., Petrella, J.R., Wong, T.Z., Arnold, S.E., Karlawish, J.H., Wolk, D., Smith, C.D., Jicha, G., Hardy, P., Sinha, P., Oates, E., Conrad, G., Lopez, O.L., Oakley, M.A., Simpson, D.M., Porsteinsson, A.P., Goldstein, B.S., Martin, K., Makino, K.M., Ismail, M.S., Brand, C., Mulnard, R.A., Thai, G., McAdams-Ortiz, C., Womack, K., Mathews, D., Quiceno, M., Diaz-Arrastia, R., King, R., Weiner, M., Martin-Cook, K., DeVous, M., Levey, A.I., Lah, J.J., Cellar, J.S., Burns, J.M., Anderson, H.S., Swerdlow, R.H., Apostolova, L., Tingus, K., Woo, E., Silverman, D.H.S., Lu, P.H., Bartzokis, G., Graff-Radford, N.R., Parfitt, F., Kendall, T., Johnson, H., Farlow, M.R., Hake, A.M., Matthews, B.R., Herring, S., Hunt, C., van Dyck, C.H., Carson, R.E., MacAvoy, M.G., Chertkow, H., Bergman, H., Hosein, C., Hsiung, G.Y.R., Feldman, H., Mudge, B., Assaly, M., Bernick, C., Munic, D., Kertesz, A., Rogers, J., Trost, D., Kerwin, D., Lipowski, K., Wu, C.K., Johnson, N., Sadowsky, C., Martinez, W., Villena, T., Turner, R.S., Johnson, Kathleen, Reynolds, B., Sperling, R.A., Johnson, K.A., Marshall, G., Frey, M., Lane, B., Rosen, A., Tinklenberg, J., Sabbagh, M.N., Belden, C.M., Jacobson, S.A., Sirrel, S.A., Kowall, N., Killiany, R., Budson, A.E., Norbash, A., Johnson, P.L., Allard, J., Lerner,

A., Ogrocki, P., Hudson, L., Fletcher, E., Carmichael, O., Olichney, J., DeCarli, C., Kittur, S., Borrie, M., Lee, T.Y., Bartha, R., Johnson, S., Asthana, S., Carlsson, C.M., Preda, A., Nguyen, D., Tariot, P., Reeder, S., Bates, V., Capote, H., Rainka, M., Scharre, D.W., Kataki, M., Adeli, A., Zimmerman, E.A., Celmins, D., Brown, A.D., Pearlson, G.D., Blank, K., Anderson, K., Santulli, R.B., Kitzmiller, T.J., Schwartz, E.S., Sink, K.M., Williamson, J.D., Garg, P., Watkins, F., Ott, B.R., Querfurth, H., Tremont, G., Salloway, S., Malloy, P., Correia, S., Rosen, H.J., Miller, B.L., Mintzer, J., Spicer, K., Bachman, D., Pasternak, S., Rachinsky, I., Drost, D., Pomara, N., Hernando, R., Sarrael, A., Schultz, S.K., Boles Ponto, L.L., Shim, H., Smith, K.E., Relkin, N., Chaing, G., Raudin, L., Smith, A., Fargher, K., Raj, B.A., Neylan, T., Grafman, J., Davis, M., Morrison, R., Hayes, J., Finley, S., Friedl, K., Fleischman, D., Arfanakis, K., James, O., Massoglia, D., Fruehling, J.J., Harding, S., Peskind, E.R., Petrie, E.C., Li, G., Yesavage, J.A., Taylor, J.L., Furst, A.J., Giusti-Rodríguez, P., Sullivan, P.F., Hardy, J., Fu, A.K.Y., Li, Y., Ip, N.Y. (2019) Non-coding variability at the APOE locus contributes to the Alzheimer's risk. Nat. Commun. 10.

Acknowledgments

Conflict of interest statement

The author declares no conflict of interest.

Data availability statement

Data sharing is not applicable to this article as no new data was created in this study. However scripts, supporting files, interim files and full results will be deposited at the NIAGADS site: https://www.niagads.org/adsp/content/home.

Table 1

Breakdown of samples by ethnicity, age and sex. AA = African American; NHW = Non-Hispanic White; CH = Caribbean Hispanic.

Cohort	Status	Male:Female	Age mean (SD)
AA	Controls	41:160	76.4 (6.5)
	Cases	83:165	72.7 (6.9)
NHW	Controls	250:394	80.1 (6.4)
	Cases	185:154	73.4 (7.6)
CH	Controls	133:268	69.3 (6.4)
	Cases	93:177	74.7 (6.8)
All	Controls	424:822	76.0 (8.0)
	Cases	361:496	73.6 (7.2)

Table 2

The table shows the MLP for the likelihood ratio test for each variant as a predictor of LOAD including 20 principal components as covariates, with and without also including rs429358 and rs7412. All variants for which the second MLP was greater than 2.5 are shown.

, with different covariates included in the model, for all variants which initially produced and MLP > 20.

SNP	rs429358 and rs7412 not included	rs429358 and rs7412 included
rs59007384	18.97	3.72
rs66626994	21.78	3.66
rs11556505	23.57	3.2
rs2075650	21.88	3.11
rs6857	36.45	2.97
rs12721046	30.47	2.96
rs186113697	1.93	2.95
rs205909	5.71	2.9
rs34954997	20	2.87
rs111789331	30.45	2.86
rs149626525	3.59	2.82
rs434132	9.17	2.6
rs157587	3.43	2.6
rs5117	20.05	2.58