EBioMedicine 59 (2020) 102954

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.elsevier.com/locate/ebiom

Research paper

Apolipoprotein E (*APOE*) genotype-associated disease risks: a phenome-wide, registry-based, case-control study utilising the UK Biobank

Amanda L. Lumsden[a,b], Anwar Mulugeta[a,b,c], Ang Zhou[a,b], Elina Hyppönen[a,b,*]

[a] Australian Centre for Precision Health, Cancer Research Institute, University of South Australia, Adelaide, SA 5001, Australia
[b] South Australian Health and Medical Research Institute, Adelaide, SA 5000, Australia
[c] Department of Pharmacology and Clinical Pharmacy, College of Health Sciences, Addis Ababa University, Addis Ababa, Ethiopia

ARTICLE INFO

Article History:
Received 24 June 2020
Revised 27 July 2020
Accepted 4 August 2020
Available online xxx

Keywords:
Phenome-wide association
PheWAS
APOE
Apolipoprotein E
Disease risk
Biomarkers

ABSTRACT

Background: The three main alleles of the *APOE* gene (ε4, ε3 and ε2) carry differential risks for conditions including Alzheimer's disease (AD) and cardiovascular disease. Due to their clinical significance, we explored disease associations of the *APOE* genotypes using a hypothesis-free, data-driven, phenome-wide association study (PheWAS) approach.

Methods: We used data from the UK Biobank to screen for associations between *APOE* genotypes and over 950 disease outcomes using genotype ε3ε3 as a reference. Data was restricted to 337,484 white British participants (aged 37–73 years).

Findings: After correction for multiple testing, PheWAS analyses identified associations with 37 outcomes, representing 18 distinct diseases. As expected, ε3ε4 and ε4ε4 genotypes associated with increased odds of AD ($p \leq 7.6 \times 10^{-46}$), hypercholesterolaemia ($p \leq 7.1 \times 10^{-17}$) and ischaemic heart disease ($p \leq 2.3 \times 10^{-4}$), while ε2ε3 provided protection for the latter two conditions ($p \leq 3.7 \times 10^{-10}$) compared to ε3ε3. In contrast, ε4-associated disease protection was seen against obesity, chronic airway obstruction, type 2 diabetes, gallbladder disease, and liver disease (all $p \leq 5.2 \times 10^{-4}$) while ε2ε2 homozygosity increased risks of peripheral vascular disease, thromboembolism, arterial aneurysm, peptic ulcer, cervical disorders, and *hallux valgus* (all $p \leq 6.1 \times 10^{-4}$). Sensitivity analyses using brain neuroimaging, blood biochemistry, anthropometric, and spirometric biomarkers supported the PheWAS findings on *APOE* associations with respective disease outcomes.

Interpretation: PheWAS confirms strong associations between *APOE* and AD, hypercholesterolaemia, and ischaemic heart disease, and suggests potential ε4-associated disease protection and harmful effects of the ε2ε2 genotype, for several conditions.

Funding: National Health and Medical Research Council of Australia.

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license. (http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

Apolipoprotein E (APOE) is a glycoprotein involved in cholesterol homeostasis and lipid metabolism, that is produced mainly by hepatocytes and astrocytes, and is found in plasma and cerebrospinal fluid. The three main *APOE* alleles (ε4, ε3, and ε2) are defined by the combination of variants at 2 single nucleotide polymorphisms (*rs429358* and *rs7412*). The ε3 allele is the most common (~78% globally), followed by ε4 (~14%), and ε2 (~8%). Chronologically, ε4 is believed to be the ancestral allele from which the ε3 and ε2 variants sequentially evolved over 200,000 years ago [1]. The APOE protein has lipid-binding and receptor-binding domains that enable its role in directing the uptake of chylomicrons and very low density lipoprotein (VLDL) remnant particles from the circulation via specific receptors; the key one being the low density lipoprotein receptor (LDLR). The polymorphisms affect the binding affinities of APOE for lipid and the LDLR (appendix p 34 figure S1), resulting in various effects on biomarkers of lipid and cholesterol health, and differential risks of a variety of health outcomes [2]. Serum APOE protein levels increase progressively across the genotypes from ε4ε4 to ε2ε2 [3,4] which is also likely to contribute to differing effects of the *APOE* variants. The *APOE* ε4 allele is one of the most notorious common genetic risk factors, with the potential to increase AD risk up to 15-fold when homozygous, and further adverse effects on lipid profiles and

* Corresponding author.
E-mail address: elina.hypponen@unisa.edu.au (E. Hyppönen).

https://doi.org/10.1016/j.ebiom.2020.102954
2352-3964/© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license. (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Research in context

Evidence before this study

We searched the PubMed database using the search terms ("APOE" OR "apolipoprotein e") AND ("meta-analysis" OR "pooled analysis"). The review was restricted to meta-analyses of *APOE* allele health outcome associations published up to August 31st, 2019, totalling 501 papers whose abstracts were reviewed for further information. We identified 184 meta-analyses that investigated disease outcome associations between *APOE* genotypes or alleles in humans. A summary of the most extensive and up-to-date published meta-analysis associations for all major conditions is presented in table 1 (and see extended, referenced table S1 appendix p 5).

The ε4 allele was associated with increased risk for several mental, neurological, and cerebrovascular disorders, with the strongest association seen with Alzheimer's disease (AD). In addition, there was some evidence suggesting the ε2 allele was associated increased risk for Parkinson's disease, and multiple sclerosis, and that both ε2 and ε4 were risky for intracerebral haemorrhage.

There was consistent evidence of ε4-associated elevated risk of various cardiovascular diseases. In contrast, the ε2 allele showed protective effects in coronary heart disease and myocardial infarction, however it appeared to have divergent effects on risk for premature coronary heart disease in Caucasian and Asian populations for which odds decreased, and increased, respectively.

In addition to cerebral and cardiovascular diseases, *APOE* variants associated with several other disorders. ε4 increased the risk of nephrotic syndrome, while ε2 elevated the odds of nephropathy in type 2 diabetes (T2D), and psoriasis. ε2 and ε4 alleles increased risk of T2D, particularly among Asian cohorts. The ε4 allele also increased risk of gallstone disease and breast cancer in Asian populations. Not all ε4 associations were adverse, with this allele offering protection against age-related macular degeneration, end stage renal disease, and proximal colorectal cancer.

Evidence from meta-analyses also supported *APOE* associations with health-related outcomes that were not diseases, but were of interest for the current study. For example, ε4 carriers tended to have a less favourable lipid/cholesterol profile while the reverse was generally true for ε2 carriers. ε4 was associated with lower body mass index (BMI), and reduced longevity, while possession of an ε2 allele promoted longevity. Both ε2 and ε4 were also associated with recurrent pregnancy loss.

Added value of this study

Our large-scale data-driven analyses allowed us to investigate phenome-wide risks associated with each *APOE* genotype compared to ε3ε3. *APOE* genotypes were associated with 37 outcomes representing 18 distinct diseases, supporting well-established increased odds of AD, hypercholesterolaemia and ischaemic heart disease (IHD) for ε3ε4 and ε4ε4 genotypes, and beneficial effects of the ε2ε3 genotype against hypercholesterolaemia and IHD. Additionally, we uncovered ε4 associations with protection against obesity, chronic airway obstruction, liver disease, T2D, and gallstone disease. ε2 homozygosity increased risk of aortic aneurysm, peripheral arterial thromboembolism, peripheral vascular disease, peptic ulcers, cervical disorders and *hallux valgus*.

Implications of all the available evidence

While ε2 is generally regarded as favourable, our study exposes potential health risks of ε2 homozygosity that may help explain the limited prevalence of the ε2 variant in the population, and which should be considered in therapeutic attempts to promote the beneficial effects of the ε2 variant. Intriguingly, our findings suggest that the AD risk-associated ε4 allele provides protection against conditions such as obesity, T2D, chronic airway obstruction, gallstone disease and liver disease. Since ε4 is associated with increased T2D and gallstone disease in Asian cohorts, our findings support possible ethnic differences in physiological consequences of *APOE* status. The outcomes of this agnostic survey extend our knowledge of *APOE* associations, and warrant replication and validation in future studies.

cardiovascular diseases. In contrast, the rare ε2 allele is often found to be protective, while ε3 is considered neutral with respect to AD, lipid profiles and cardiovascular health.

Despite some evidence for influences of *APOE* genotype on multiple other conditions, results are often inconsistent, particularly due to inadequately powered studies looking into the effects of least frequent genotypes (~0.6% for ε2ε2). In this hypothesis-free phenome-wide association study, we will explore the effects of *APOE* genotype across the spectrum of human disease. For background, the *Research In Context* section summarises all disease associations for *APOE* alleles which have arisen from the most recently published meta-analyses with respect to each disease. Subsequently, in our study we used information from 337,484 UK Biobank participants to screen for *APOE* genotype associations with over 950 disease outcomes covering all conditions within hospital inpatient records and mortality registrations, using the most common genotype, ε3ε3, as a reference.

2. Methods

2.1. UK Biobank Cohort

The UK Biobank consists of 503,000 participants who were aged 37–73 years (99.5% between 40–69 years) when recruited between 2006 and 2010 [5]. The study includes extensive self-reported information collected using touchscreen questionnaires and verbal interviews, and information on genetics and biochemical marker levels through sampling of blood, urine and saliva. Our analyses were restricted to 337,484 unrelated white British individuals (established by self-report and genetic data) [6] who had consistent information between self-reported and genetic sex (appendix p 35 figure S2). Ethical approval for the UK Biobank was granted by the National Information Governance Board for Health and Social Care and North West Multicentre Research Ethics Committee (11/NW/0382). This research was conducted under application 10171. Participants provided electronic consent to use their anonymised data and samples for health-related research, to be re-contacted for further sub-studies, and for the UK Biobank to access their health-related records [5].

2.2. APOE *genotyping*

Genetic data was available for 488,377 UK Biobank participants, of whom 49,950 were genotyped using a UK BiLEVE array while the remaining 438,427 were genotyped using the UK biobank Axiom array, with the two arrays having 95% marker content similarity [6]. The combination of variants at two single nucleotide polymorphisms (SNPs *rs429358* and *rs7412*) within the *APOE* gene define the three main *APOE* alleles (ε4, ε3, and ε2). We extracted *rs429358* and *rs7412* variants which were directly genotyped and did not deviate from the Hardy-Weinberg equilibrium (both $p > 0.05$). Depending on the

Table 1
APOE allelic associations with major conditions, summarised from the most current and comprehensive meta-analyses. (Refer to table S1 appendix p 5 for extended information.)

Disease outcomes	# Studies in meta-analysis	Population size (×1000)	Ethnicity	ε4	Reported reference group	ε2
Mental, neurological, and cerebrovascular disorders						
Sporadic late onset Alzheimer's disease	21	5 to 10	Mixed	↑	Non carriers	••
Alzheimer's disease	20	1 to 5	Chinese	↑	Non carriers	ns
Severe cerebral amyloid angiopathy (vs mild/moderate)	5	<1	Mixed	↑	Non carriers	ns
Mild cognitive impairment	18	5 to 10	Mixed	↑	ε3	ns
Ischaemic stroke	81	10 to 70	Mixed	↑	Non carriers	ns
Vascular dementia	29	5 to 10	Mixed	↑	ε3	ns
Frontotemporal lobar degeneration	34	10 to 70	Mixed	↑	ε3	ns
Lobar intracerebral haemorrhage	4	5 to 10	Mixed	↑	ε3	↑
Creutzfeldt-Jakob disease	11	1 to 5	Mixed	↑	Non carriers	ns
Depression	20	5 to 10	Mixed	ns	ε3	ns
Depression (age ≥ 50 years)	13	NA	Mixed	↑	ε3	ns
Depression	9	1 to 5	Chinese Han	↑	Non carriers	••
Epilepsy	9	1 to 5	Mixed	↑	ε3ε3	ns
Aneurysmal subarachnoid haemorrhage	8	1 to 5	Mixed	↑	Non carriers	ns
Subjective cognitive decline	13	5 to 10	Mixed	↑	Non carriers	••
Parkinson's disease	47	10 to 70	Mixed	ns	Non carriers	↑
Multiple sclerosis	20	5 to 10	Mixed	ns	ε3	↑
Cardiovascular diseases						
Ischaemic heart disease	18	1 to 5	Chinese	↑	ε3	ns
Hypertension	28	10 to 70	Mixed	↑	ε3	ns
Premature coronary heart disease	18	5 to 10	Mixed	↑	ε3	ns
Premature coronary heart disease	12	1 to 5	Caucasian	↑	ε3	↓
Premature coronary heart disease	5	<1	Asian	↑	ε3	↑
Coronary heart disease	30	10 to 70	Mixed	↑	ε3	ns
Coronary heart disease	22	10 to 70	Caucasoid	↑	ε3	↓
Coronary heart disease	8	5 to 10	Mongoloid	↑	ε3	ns
Myocardial infarction	20[a]	10 to 70	Mixed	↑	ε3	↓
Other						
Nephrotic syndrome	12	1 to 5	••	↑	Non carriers	ns
T2D	26	5 to 10	Chinese Han	↑	ε3	↑
T2D	30	10 to 70	Mixed	••	Non carriers	↑
T2D	11	5 to 10	Caucasian	••	Non carriers	ns
T2D	15	5 to 10	Asian	••	Non carriers	↑
T2D nephropathy (vs T2D)	17	1 to 5	Mixed	ns	ε3	↑
Gallstone disease	14	5 to 10	Mixed	ns	Non carriers	ns
Gallstone disease	17	1 to 5	Mixed	ns	ε3	ns
Gallstone disease	7	1 to 5	Chinese	↑	ε3	ns
Psoriasis	7	1 to 5	Mixed	ns	Other alleles[b]	↑
Breast cancer	11	1 to 5	Mixed	ns	ε3	ns
Breast cancer	4	1 to 5	Caucasian	ns	ε3	ns
Breast cancer	3	1 to 5	Asian	↑	ε3	ns
Proximal colorectal neoplasm	3	<1	Mixed	↓	ε3	ns
End stage renal disease	16	10 to 70	Mixed	↓	Non carriers	↑
Age-related macular degeneration	12	10 to 70	Mixed	↓	ε3	↑

[a] : reported as 22 studies in body text, but information is provided for only 20.
[b] : reported as (ε4 vs ε2+ε3) and (ε2 vs ε3+ε4)
ns: not significant. T2D: type 2 diabetes. "••": data not presented. Arrows indicate whether an *APOE* allele increases (↑) or decreases (↓) the odds of the corresponding condition in reference to the reported reference group.

combination of alleles at *rs429358* and *rs7412* variants, an individual could possess one of six common *APOE* genotypes (ε4ε4, ε4ε3, ε4ε2, ε3ε3, ε2ε3 and ε2ε2, appendix p 6 table S2). The ε1ε4 and ε1ε2 genotypes were detected in 15 and two individuals, respectively. These individuals were excluded from this analysis due to small sample size. Ambiguous ε2ε4/ε1ε3 genotypes were coded as ε2ε4 since the ε1 allele is so rare. We used the ε3ε3 genotype as a reference since it is the most frequent genotype. We generated a binary variable for each of the six common genotypes, creating five indicator variables, and the ε3ε3 reference group.

2.3. Phenome generation

Information on disease outcomes and underlying causes of death were obtained through linkage to hospital episode statistics (HES) and mortality registrations [6]. We included all entries until March 31st, 2017 resulting in 15,119 disease outcomes recorded according to the International Classification of Diseases, ninth/tenth revision (ICD-9/10) codes. Before analyses, ICD codes were converted into 1,859 phecodes which provide classifications more closely aligned with diseases commonly cited in clinical practice and genomic studies [7]. For each phecode, we coded individuals with the phecode-of-interest as cases, whilst participants without a phecode within the same category were considered the control group [8]. For each PheWAS analysis, we excluded phecodes with less than 200 cases, leaving 958, 1070, 960, 1013 and 950 unique phecodes for analyses involving ε4ε4, ε3ε4, ε2ε4, ε2ε3 and ε2ε2 (versus ε3ε3), respectively.

2.4. Biomarkers for sensitivity analyses

To further explore some of the PheWAS outcomes, we utilised blood biomarkers and brain neuroimaging data from the UK Biobank Assessment Centre to assess their association with *APOE* genotypes. Outcomes included brain health biomarkers (total brain, white matter, grey matter, and whole hippocampus volumes, and volume of white matter hyperintensities), cardiovascular biomarker levels (total

cholesterol, low density lipoprotein (LDL), high density lipoprotein-cholesterol (HDL-cholesterol), triglycerides, apolipoprotein A (APOA), apolipoprotein B (APOB), lipoprotein A (Lp(A)), and C-reactive protein (CRP)), diabetes markers (glucose, and glycated haemoglobin (HbA1c)), obesity measures (BMI, waist circumference (WC), waist-hip ratio (WHR)), and spirometry measures (forced expiratory volume in 1-second (FEV_1), forced vital capacity (FVC), and FEV_1/FVC ratio). Brain volume measures were normalised to head size, and white matter hyperintensity data were inverse normal transformed due to left skewness, to approximate normal distribution. For FEV_1 and FVC, we used 'best measure' values and the FEV_1/FVC ratios were calculated as the ratio of these values. Further details on the biomarker measures and detection methods are provided in the supplementary methods (appendix p 4).

2.5. Statistical analysis

We used the R package phewas [9] to run logistic regression of each phecode against each *APOE* genotype (in comparison to reference genotype $\varepsilon3\varepsilon3$), adjusting for demographics (age and sex), genotyping array, and population structure (dummy indicators for each assessment centre, and top 40 genetic principal components). Prior to undertaking the phenome-wide analyses, each *APOE* genotype was run (as described above) for pre-selected control phecodes. These included three positive control outcomes for which possession of *APOE* $\varepsilon4$ is known to increase the odds (namely dementia, hyperlipidaemia, and ischaemic heart disease (IHD)), and one negative control with no known or likely association with *APOE* (diaphragmatic hernia). We used a false discovery rate (FDR $q = 0.05$) corrected p-value threshold of 6.1×10^{-4} to control for multiple testing [10], accounting for all comparisons across the five PheWASs. In sensitivity analyses using the biomarker data, linear regressions of each biomarker were fitted against each *APOE* genotype in models adjusted for age, sex, assessment centres (as dummy variable), genotyping array and 40 principal components. For analysis of glucose levels, we further adjusted for fasting time.

3. Results

Population characteristics are shown in table 2. The most prevalent *APOE* genotype was $\varepsilon3\varepsilon3$ (58.2%), followed by $\varepsilon3\varepsilon4$ (23.9%), $\varepsilon2\varepsilon3$ (12.3%), $\varepsilon2\varepsilon4$ (2.6%), $\varepsilon4\varepsilon4$ (2.4%), and $\varepsilon2\varepsilon2$ (0.6%). We observed no differences in genotypic frequencies based on sex ($p = 0.68$, likelihood ratio test; P_{LRT}), history of comorbidity ($P_{LRT} = 0.08$) or general health ($P_{LRT} = 0.07$), although there was a slight underrepresentation of participants with the $\varepsilon4$ allele in the older age groups ($P_{LRT} = 0.0008$).

3.1. Positive and negative controls

APOE genotypes showed the expected associations with all three positive control disease outcomes; dementia, hyperlipidaemia, and ischaemic heart disease (IHD) (appendix p 7 table S3). For diaphragmatic hernia, the negative control, there were no *APOE* genotype associations that passed the PheWAS FDR threshold.

3.2. Genotype PheWAS analyses

Manhattan plots for each genotype are shown in figures S3–S7 (appendix pp 36–40), and PheWAS results for all outcomes can be found in table S4 (appendix pp 8–32). Compared to $\varepsilon3\varepsilon3$, genotypes $\varepsilon3\varepsilon4$ and $\varepsilon4\varepsilon4$ were associated with increased odds of 21 outcomes representing six distinct diseases, and lower odds of seven outcomes representing five diseases. Compared to $\varepsilon3\varepsilon3$, the odds of eight outcomes representing six diseases were elevated in the presence of $\varepsilon2$ homozygosity. Overall, PheWAS identified *APOE* genotypic associations with 37 outcomes, representing 18 distinct diseases. Odds ratios (OR) and 95% confidence intervals (95% CI) for each genotype, for a representative outcome from each distinct disease are presented as forest plots in Fig. 1.

For brain conditions including dementias, neurological disorders, and cerebral degenerations, the odds increased with the number of $\varepsilon4$ alleles. The strongest associations were with AD, with OR 3.69 (95% CI 3.08–4.40) for $\varepsilon3\varepsilon4$ and OR 13.52, (10.64–17.18) for $\varepsilon4\varepsilon4$

Table 2
General characteristics of the white British UK Biobank population across *APOE* genotypes

	n (%)	*APOE* genotype, *n* (%)						*p*
		$\varepsilon4\varepsilon4$	$\varepsilon3\varepsilon4$	$\varepsilon2\varepsilon4$	$\varepsilon3\varepsilon3$	$\varepsilon2\varepsilon3$	$\varepsilon2\varepsilon2$	
All	337,484	8,179 (2.4)	80,499 (23.9)	8,616 (2.6)	196,306 (58.2)	41,695 (12.3)	2,172 (0.6)	
Sex								0.68
Women	181,236 (53.7)	4,360 (2.4)	43,202 (23.8)	4,574 (2.5)	105,535 (58.2)	22,368 (12.3)	1188 (0.7)	
Men	156,248 (46.3)	3,819 (2.4)	37,297 (23.9)	4042 (2.6)	90,771 (58.1)	19,327 (12.4)	984 (0.6)	
Age (in years)								0.0008
39-44.9	31,719 (9.4)	771 (2.4)	7,590 (23.9)	830 (2.6)	18,457 (58.2)	3,861 (12.2)	208 (0.7)	
45-49.9	42,130 (12.5)	1,048 (2.5)	10,281 (24.4)	1,082 (2.6)	24,213 (57.5)	5,199 (12.3)	305 (0.7)	
50-54.9	50,240 (14.9)	1,224 (2.4)	12,118 (24.1)	1,297 (2.6)	29,133 (58.0)	6,156 (12.3)	308 (0.6)	
55-59.9	61,032 (18.1)	1,489 (2.4)	14,497 (23.8)	1,580 (2.6)	35,488 (58.1)	7,561 (12.4)	414 (0.7)	
60-64.5	85,434 (25.3)	2,087 (2.4)	20,206 (23.7)	2,135 (2.5)	49,871 (58.4)	10,607 (12.4)	525 (0.6)	
65-73	66,929 (19.8)	1,560 (2.3)	15,807 (23.6)	1,692 (2.5)	39,144 (58.5)	8,311 (12.4)	412 (0.6)	
History of Comorbidity								0.08
No	77,043 (22.8)	1,816 (2.4)	18,405 (23.9)	2,033 (2.6)	44,701 (58.0)	9,566 (12.4)	515 (0.7)	
One	44,974 (13.3)	1,097 (2.4)	10,735 (23.9)	1,136 (2.5)	26,181 (58.2)	5,570 (12.4)	251 (0.6)	
Two to three	70,419 (20.9)	1,733 (2.5)	16,788 (23.8)	1,800 (2.6)	40,714 (57.8)	8,906 (12.6)	476 (0.7)	
Four to five	47,305 (14.0)	1,150 (2.4)	11,329 (23.9)	1,215 (2.6)	27,487 (58.1)	5,851 (12.4)	272 (0.6)	
Six or more	97,743 (29.0)	2,383 (2.4)	23,242 (23.8)	2,432 (2.4)	57,223 (58.5)	11,802 (12.0)	658 (0.7)	
General health								0.07
Excellent	56,531 (16.8)	1,352 (2.4)	13,389 (23.7)	1,495 (2.6)	33,041 (58.4)	6,905 (12.2)	347 (0.6)	
Good	197,169 (58.4)	4,832 (2.5)	47,120 (23.9)	5,037 (2.6)	114,414 (58.0)	24,523 (12.4)	1,232 (0.6)	
Fair	68,621 (20.3)	1,664 (2.4)	16,335 (23.8)	1,709 (2.5)	40,022 (58.3)	8,412 (12.3)	476 (0.7)	
Poor	13,983 (4.1)	308 (2.2)	3,362 (24.0)	342 (2.4)	8,138 (58.2)	1,723 (12.3)	109 (0.8)	
Missing	1180 (0.3)	23 (1.9)	293 (24.8)	333 (2.8)	691 (58.6)	132 (11.2)	8 (0.7)	

APOE genotypes were coded as 0, 1, 2, 3, 4, and 5 for *APOE* $\varepsilon3\varepsilon3$ (reference), $\varepsilon2\varepsilon2$, $\varepsilon2\varepsilon3$, $\varepsilon2\varepsilon4$, $\varepsilon3\varepsilon4$, and $\varepsilon4\varepsilon4$.
p-values were generated by a likelihood ratio test from logistic regression. For all analyses, adjustments were made for genotyping array, 40 principal components, and birth location. For comorbidity and general health, further adjustment was made for sex and age. Total *n* includes 15 individuals that had the $\varepsilon1\varepsilon4$ genotype (0.004%), and two individuals had the $\varepsilon1\varepsilon2$ genotype (0.0006%), that were excluded from further analyses.

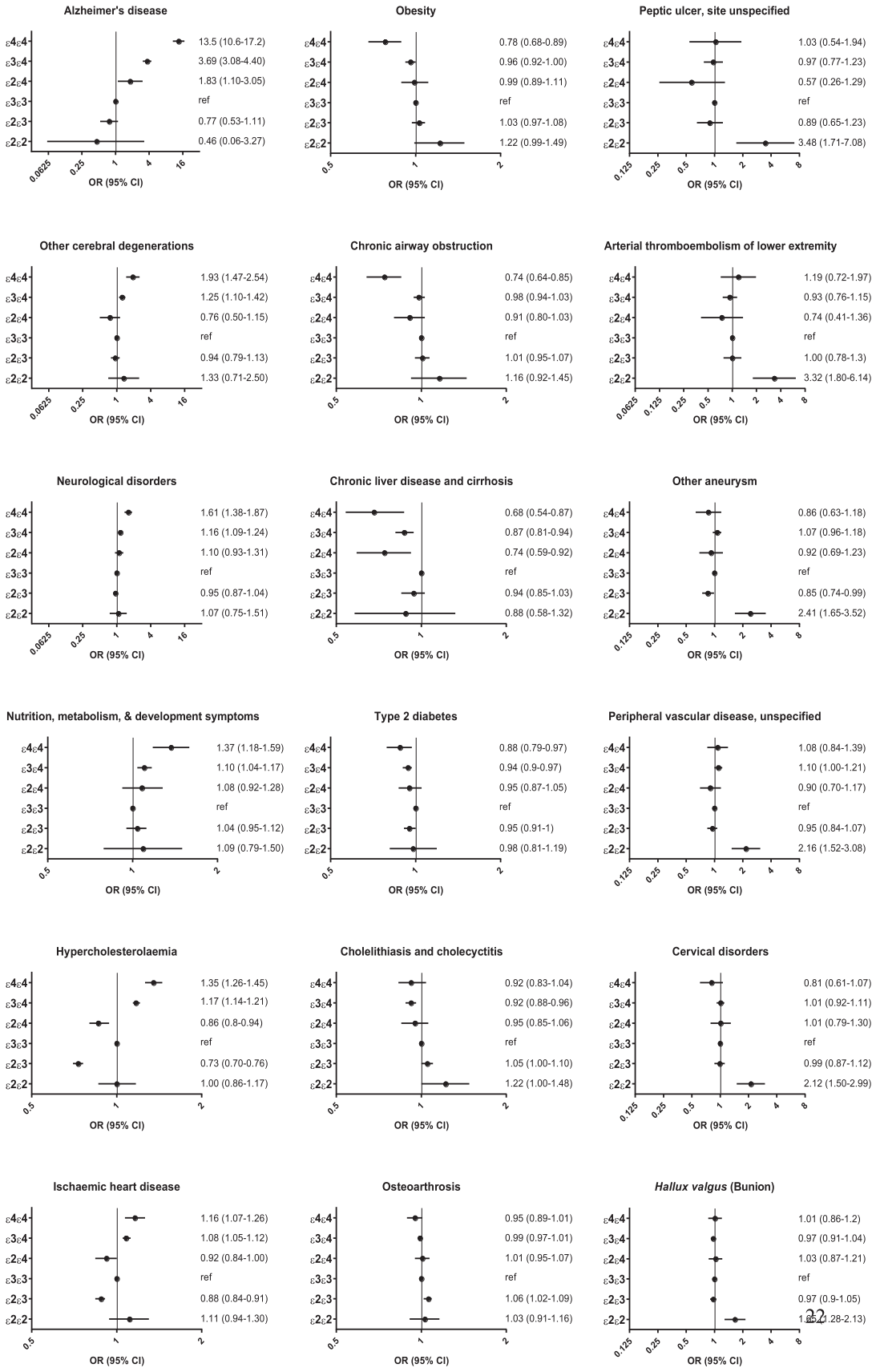

Fig. 1. *APOE* genotypes and risk of disease.

Forest plots depicting the OR (black box symbols) and 95% CI (horizontal lines) for each genotype compared to reference genotype, ε3ε3. Data is presented for representative disease outcomes where at least one genotype showed a signal in the PheWAS. Actual values are shown to the right of each graph. Case and control numbers and *p*-values (logistic regression) for each comparison group can be found in table S4 (appendix pp 8–32), with further breakdown by genotype shown within figures S3–S7 (appendix pp 36–40). "Other aneurysm" encompasses aortic, and other arterial aneurysms, but not cerebral, or heart aneurysms. "Other cerebral degenerations" includes gangliosidosis, sphingolipidosis,

compared to ε3ε3 (Fig. 1 and appendix pp 8–32, table S4). The odds also increased with number of ε4 alleles for 'symptoms of nutrition, metabolism, and development' which includes descriptors such as unspecified severe protein-energy malnutrition, feeding difficulties and mismanagement, and abnormal weight loss. For hypercholesterolaemia and IHD, the odds of disease increased with number of ε4 alleles, while the odds were lower for carriers of one ε2 allele, but not for ε2ε2 homozygotes. ε4ε4 homozygotes had a lower odds of obesity (0.78, 0.68–0.89), with little evidence of an association for ε4 heterozygotes. ε4ε4 homozygotes also had lower odds of chronic airway obstruction (0.74, 0.64–0.85) compared to ε3ε3. The ε3ε4 genotype had protective effects against chronic liver disease and cirrhosis, T2D, and cholelithiasis and cholecystitis (gallstone diseases) while the ε2ε3 genotype increased the risk of osteoarthrosis (Fig. 1).

We observed evidence for elevated odds of six diseases for ε2ε2 homozygotes compared to ε3ε3, including peptic ulcer, arterial thromboembolism of lower extremity, "other aneurysm" (encompassing aortic, and other arterial aneurysms, but not cerebral, or heart aneurysms), peripheral vascular disease (unspecified), noninflammatory disorders of the cervix, and *hallux valgus* (bunion; Fig. 1). While the large population allowed detection of associations with this low frequency genotype (0.6%), it should be noted that the number of ε2ε2 homozygote cases with these diseases was still limited (8, 11, 30, 34, 35 and 63, respectively).

3.3. Sensitivity analyses of disease biomarkers

We next assessed associations of *APOE* genotypes with biological and phenotypic markers of diseases identified by PheWAS (Fig. 2 and appendix p 33 table S5). Firstly, we looked at neuroimaging markers of brain health (Fig. 2A–E) since *APOE* genotypes were associated with risk of dementia and some neurological disorders. While no significant differences in volume of total brain, grey matter, or white matter were observed across the *APOE* genotypes compared to ε3ε3 (Fig. 2A–C), *APOE* ε3ε4 and ε4ε4 were associated with decreases in hippocampal volume, with the greatest effect size in ε4ε4 (Fig. 2D) which also showed an increase in white matter hyperintensity (Fig. 2E). Next we assessed biomarkers related to cardiovascular health (Fig. 2F–M), since several *APOE* genotypes were associated with hypercholesterolaemia and vascular diseases. Both ε3ε4 and ε4ε4 were associated with an unfavourable lipid profile (high LDL and triglycerides, and low HDL-cholesterol), while ε2ε3 had a favourable profile (low LDL and high HDL-cholesterol) and ε2ε2 was associated with very low LDL and very high triglycerides compared to ε3ε3 (Fig. 2F–L). Levels of CRP decreased by ε4 dosage and were modestly increased for ε2ε3 compared to ε3ε3 (Fig. 2M).

We also looked at markers for diabetes, obesity and lung function (Fig. 2N–U). In line with PheWAS findings on T2D and obesity, there was an ε4 dose-dependent decrease in HbA1c (Fig. 2N), and all markers of adiposity (Fig. 2P–R). Despite some evidence for an association between ε4ε4 and chronic airway obstruction, there was little evidence for differences in the measures of lung function by *APOE* genotype (Fig. 2S–U); compared to ε3ε3, the ε4ε4 genotype was associated with an increase in forced vital capacity (FVC; Fig. 2T; β = 19.97 mL, 2.89 to 37.06 mL).

4. Discussion

APOE ε4 is one of the most notorious common variants affecting the risks of chronic diseases, while ε2 is often perceived as the protective rare variant. In this study, we have assessed *APOE*-associated risks across the disease spectrum, with our large sample and hypothesis-free approach revealing pleiotropic effects of *APOE* variants. We confirmed many of the known adverse effects of the ε4 variant, including risk of AD, heart disease, and adverse lipid profile. However, we found that not all associations with the ε4 allele were adverse, with protection seen against obesity, chronic airway obstruction, T2D, gallstones, and liver disease. On the other hand, we found evidence to suggest that the ε2 allele, which is typically considered to be beneficial, increases the risks of several conditions when homozygous, including peripheral thromboembolism, aneurysms, peptic ulcers, cervical disorders, and *hallux valgus*.

In this study, increased risk of dementia and AD in ε4-associated genotypes was supported by neuroimaging markers showing decreased volume of the hippocampus (the memory-associated region of the brain most affected in AD) and increased white matter hyperintensities (which signify brain lesions). However, not all associations with the ε4 genotypes were found to be detrimental. Indeed, possession of the ε4 allele has previously been associated with benefits such as improved fitness during foetal development, infancy and youth [11], and endows carriers with the ability to thrive when faced with severe infections, such as has been observed in children with enteric infections and heavy diarrhoea [12], and adults with high parasitic burden [13]. Our study has revealed a number of additional beneficial effects of the ε4 allele. The reduced risk of obesity in ε4ε4 homozygotes relative to ε3ε3 is consistent with previous reports that BMI and other measures of obesity increase across the *APOE* allelic spectrum in the direction ε4 < ε3 < ε2 [14,15]. It is conceivable that the observed ε4-associated reduced risk of T2D may be related to the decrease in obesity risk, since T2D can often be alleviated with weight loss. It could also be that other mechanisms are at play; for example, APOE protein within the extracellular matrix of pancreatic islets isolated from rats, directly stimulates expression of important genes for β-cell function [16], although the impact of different APOE isoforms on this stimulation *in vivo* (and in humans) is not yet known. Reduced risk of gallstone disease may be a consequence of diminished T2D risk, consistent with a previous study that identified APOE ε4, and lower levels of insulin resistance markers, as being associated with protection against gallstone disease in a Caucasian (Danish) population [17]. Our observation of reduced odds of chronic liver disease and cirrhosis provides experimental support for the findings of a recent literature review suggesting the ε4 allele has protective effects in the progression of liver cirrhosis [18]. Although spirometric measures of lung function provided limited support for the protective effect we observed between the ε4ε4 genotype and chronic airway obstruction (also known as chronic obstructive pulmonary disease, COPD), our findings add to previous observations of marginally increased lung function in ε4 carriers [19], strengthening the suggestion of an underlying difference in lung physiology.

Levels of plasma biomarkers of cardiovascular health generally supported the signals observed in the PheWAS, with ε4 carriers having particularly high levels of LDL, a marker for LDL-cholesterol which is known to activate pro-inflammatory macrophage responses involved in the development of atherosclerosis that leads to IHD [20]. In line with previous findings of low plasma CRP levels in ε4 carriers [21], we found that CRP levels decreased with each additional ε4 allele. Conversely, we saw marginal increases in CRP for ε2 carriers. The informativeness of CRP as an indicator of cardiovascular risk is largely based upon the association of increased CRP levels with obesity, and our findings are indeed consistent with this (genotypes with lower BMI have lower CRP). Further studies are warranted to elucidate the mechanisms by which the ε4 variant leads to lower plasma

neuronal ceroid lipofuscinosis, Rett syndrome, Reye syndrome, systemic atrophy primarily affecting the central nervous system, and degeneration of the nervous system due to alcohol. "Nutrition, metabolism and development symptoms" pertains to symptoms including severe protein-energy malnutrition, feeding difficulties and mismanagement, and abnormal weight loss.

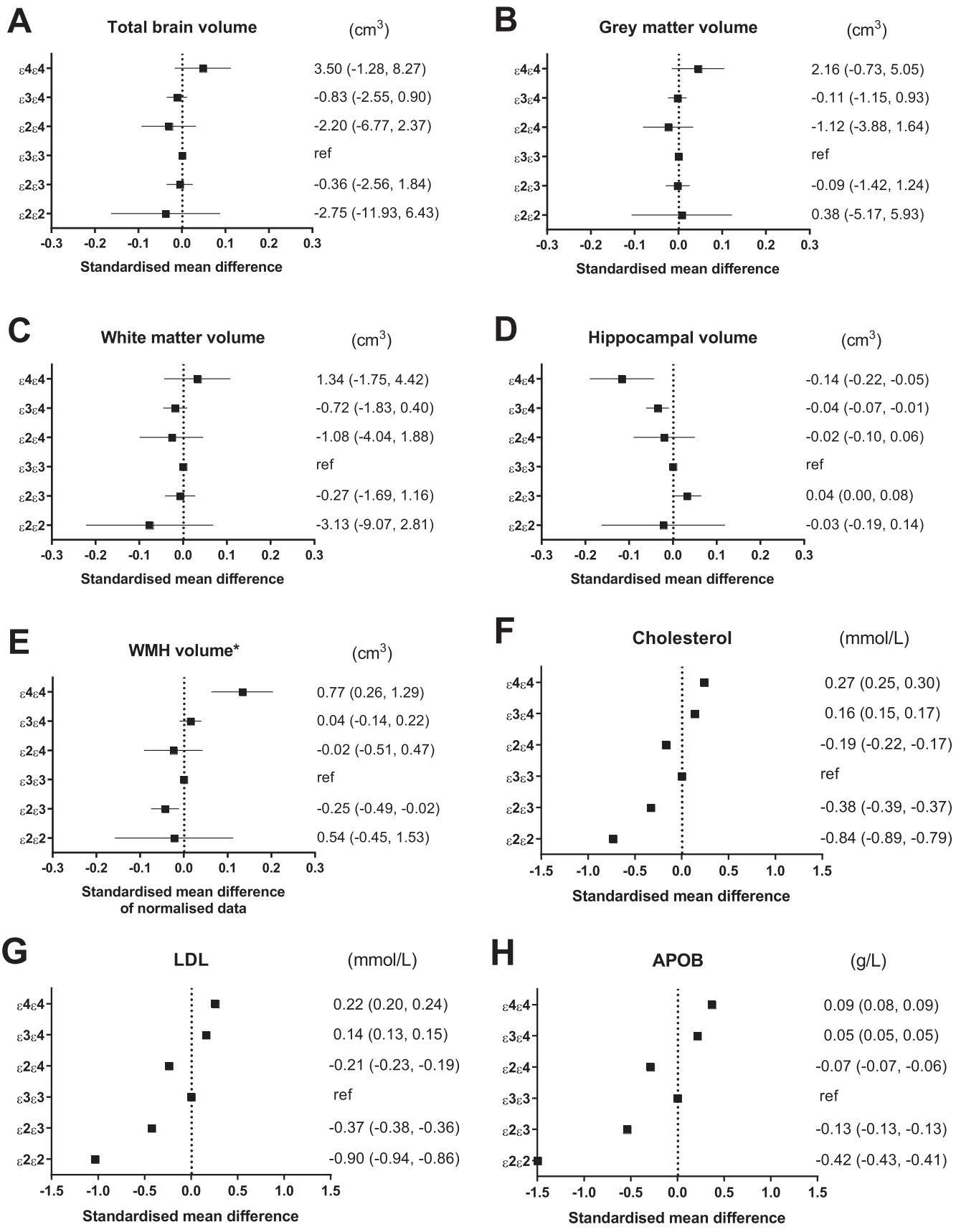

Fig. 2. *APOE* genotypes and differences in biomarkers

Graphs are displayed as standardised mean differences in biomarker levels for each *APOE* genotype with reference to ε3ε3. The values on the right correspond to absolute mean difference values (β) and 95% CI. *Note that for white matter hyperintensity (WMH) volume (E), the plotted data were inverse normal transformed to approximate normal distribution, while the volume values presented to the right are natural values. The panel includes markers related to brain health (A–E), cardiovascular risk (F–M), diabetes (N–O), obesity (P–R), and lung health (S–U). Population numbers and *p*-value (linear regression) for each biomarker can be found in table S5 (appendix p 33). APOA: apolipoprotein A. APOB:

CRP levels and how/whether these relate to ε4-associated risks of diseases such as AD and IHD.

As is typically reported, ε2ε3 and ε2ε4 were protective for hypercholesterolaemia compared to ε3ε3 (with lower levels of "bad" cholesterol indicators; total cholesterol, LDL and APOB). ε2ε3 also had increased levels of "good" cholesterol indicators (HDL-cholesterol and APOA) and was associated with reduced risk of IHD, however it slightly increased (by 6%) the odds of osteoarthrosis; a joint disease characterised by degeneration of joint cartilage and underlying bone.

A key advantage of our study was the large sample size, which enabled investigation of the effects of the ε2ε2 genotype that represents only ~0.6% of the population. The homozygous ε2ε2 genotype increased the odds for a distinct set of disease outcomes not associated with other genotypes, suggesting that these conditions stem from a homozygous loss of *APOE* function specific to the ε2 variant, such as its inability to bind the LDL receptor [22]. The ε2ε2 genotype was associated with more than a two-fold increase in risk for diseases relating to blockage or rupture of peripheral vasculature. Our findings are consistent with a 1.5-fold increase in prevalence of peripheral artery disease that has been reported to associate with the ε2ε2 genotype amongst patients with high risk of cardiovascular disease [23], and extend this risk to ε2ε2 carriers in the wider population. While the low cholesterol and LDL indicators in the ε2ε2 group can generally be considered a favourable profile, the ε2ε2 genotype is known to be associated with the highest circulating APOE protein levels of all the genotypes [3,4], and APOE levels positively associate with coagulation markers [24] which may contribute to blood viscosity. This genotype also had the highest level of triglycerides; an independent risk factor for stroke [25]. With regard to the increase in risk of peptic ulcer associated with the ε2ε2 genotype, we suspect (due to increased odds of thromboembolism-related disease in this group) that this may be a consequence of blood thinning medication leading to peptic ulcer bleeding and hospitalisation, rather than susceptibility to infection with the bacteria *Helicobacter pylori* which underlies most cases of peptic ulcer, since we found no evidence to suggest increased susceptibility to bacterial infection in this group. Homozygous ε2ε2 females had a greater than two-fold increased risk of non-inflammatory cervical disorders compared to ε3ε3 females. The phecode encompassed cervical erosion and ectropion, stricture and stenosis, cervical incompetence, and requirement of pregnancy-related care for cervical abnormalities, suggesting this finding may be related to risk of recurrent pregnancy loss, that has previously been associated with the ε2 and ε4 alleles [26]. The ε2ε2 genotype was also associated with *hallux valgus*, suggesting the *APOE* ε2 allele may represent another risk allele for this highly heritable foot disorder that has been linked to susceptibility loci near genes encoding Axin 2 (*AXIN2*), Esterase D (*ESD*) [27], vitamin D receptor (*VDR*) [28], and tumour necrosis factor (*TNF*) [29]. It should be noted that while significant associations were found with ε2ε2, the case numbers for some outcomes were limited, and further targeted studies with greater population sizes are warranted.

While utilisation of the white British contingent of the UK Biobank aims to provide an ethnically homogenous population in order to increase sensitivity of detection and avoid confounding due to population stratification, the corollary is that not all findings from the study may be applicable to other populations. For example, our study suggests ε4-associated protection against T2D and gallstones, while meta-analyses of Asian and predominantly Asian populations have reported the reverse; ε4-associated increased risk of T2D [30], and gallstone disease [31]. Another limitation is that, despite the large population size, we may have been underpowered for disease outcomes with low case numbers, especially if the true effect sizes for *APOE*-disease associations are relatively small (although clinical significance of such small effects may be questionable). Another limitation is that no selection by severity has been done and not all cases can be captured in this analysis. It is also important to bear in mind that the majority of phenotypes that comprise the phenome are derived from hospital records, and are recorded for a patient only if the outcome has been noted during a hospital visit. This method could potentially lead to some degree of differential reporting of outcomes across *APOE* genotypes. Although it is reassuring that *APOE* genotypes are not associated with overall hospitalisation in the UK Biobank (data not shown), reporting bias may still occur for some conditions; in particular secondary diagnoses, which are reported only if the person is hospitalised for other primary reasons. For example, hypercholesterolemia, which is relatively common in the general population, is more likely to be recorded for hospital visits relating to vascular health, than those unrelated to vascular health. In the comparison of ε2ε2 versus ε3ε3, hypercholesterolaemia may be more likely to be reported in the ε2ε2 group which has increased risk of peripheral vascular diseases, than in the ε3ε3 control group. Indeed, we believe this reporting bias may underlie the lack of association with hospital-diagnosed hypercholesterolaemia for ε2ε2, which conflicted with our analyses of cholesterol biomarkers (recorded at baseline), and previous studies by others, that have shown low cholesterol levels in this genotype group compared to ε3ε3 [4]. Finally, while previous knowledge of *APOE* function and clinical significance aids in the interpretation of the PheWAS analyses, we cannot rule out the possibility that linked polymorphisms unrelated to *APOE* function may contribute to the clinical associations observed, such that SNPs used to define the *APOE* alleles are in partial linkage disequilibrium with other causative/functional polymorphisms. That said, we did not observe any significant *APOE* associations with our negative control. Further functional studies are warranted to validate the associations detected in this study.

In conclusion, while the ε4 allele is generally thought of unfavourably, particularly for being the greatest genetic risk factor for late-onset AD, our current findings suggest the ε4 allele is protective against several metabolic and respiratory conditions in Caucasians. The ε2 allele, on the other hand, is typically considered beneficial, especially in individuals possessing only a single ε2 allele. Yet, homozygosity was found to be associated with increased risk of peripheral vascular disorders and other undesirable disease outcomes such as cervical disorders that could reduce the chance of successful pregnancy in ε2ε2 females, and may contribute to the low prevalence of the ε2 variant in the population despite the apparent general health linked to ε2 heterozygosity. The adverse effects associated with ε2 homozygosity also suggest that attempts to therapeutically mimic the beneficial effects of ε2 to counter ε4-associated diseases should be approached with caution.

Contributors

AL wrote the first draft, reviewed literature, prepared figures and tables, and conceptualised the study with EH. EH led the study, advised on statistical analyses and presentation. AM managed data, conducted statistical analyses, prepared tables and figures. AZ managed data and advised on statistical analyses. All authors interpreted results, revised paper and approved the manuscript for submission.

apolipoprotein B. BMI: Body mass index. CRP: C-reactive protein. FEV1: Forced expiratory volume during one second. FVC: Forced vital capacity. HbA1c: Haemoglobin A1c (glycated haemoglobin). HDL: high density lipoprotein. LDL: low density lipoprotein. Lp(A): lipoprotein A. WC: waist circumference. WHR: waist-hip ratio. WMH: white matter hyperintensity.

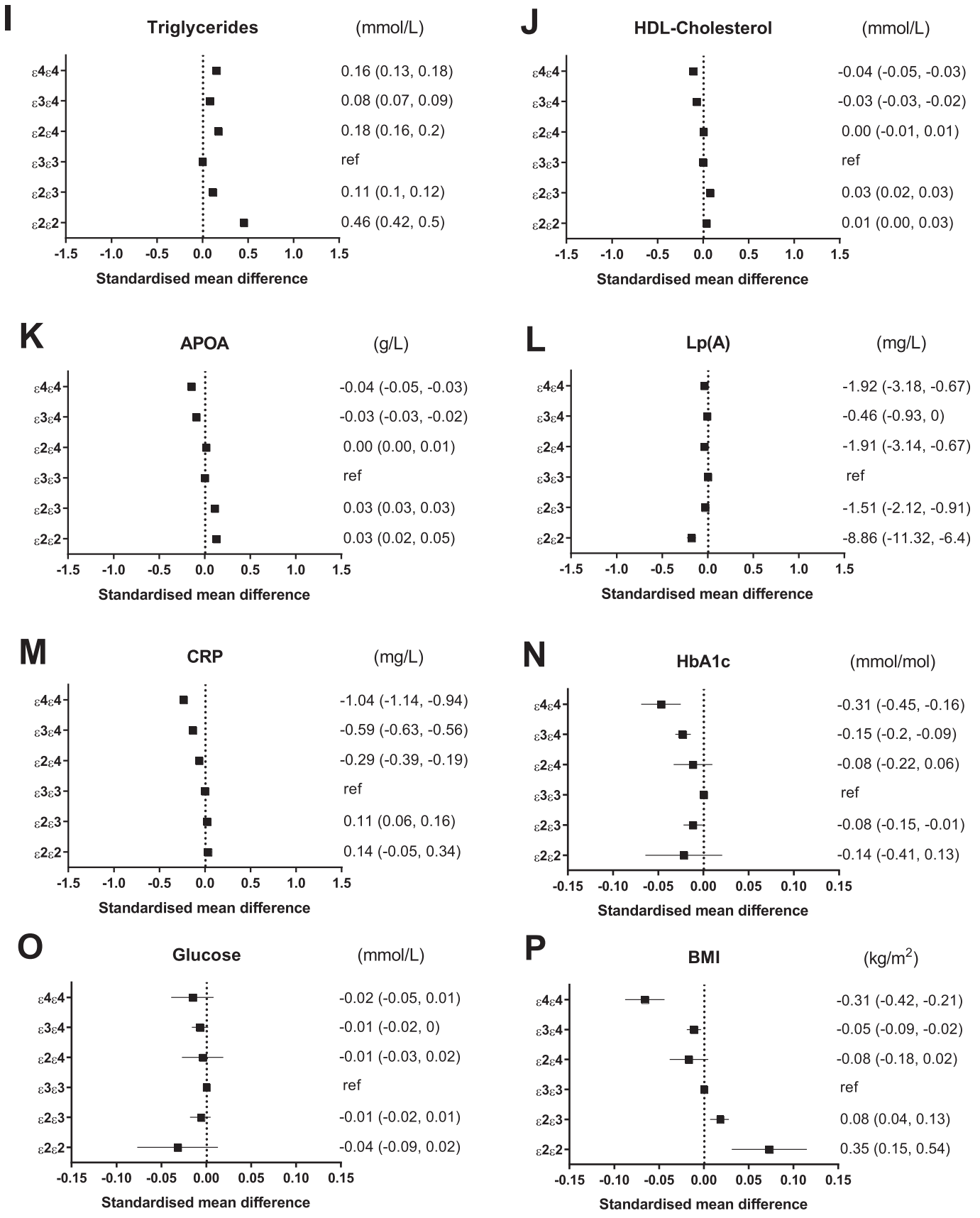

Fig. 2 Continued.

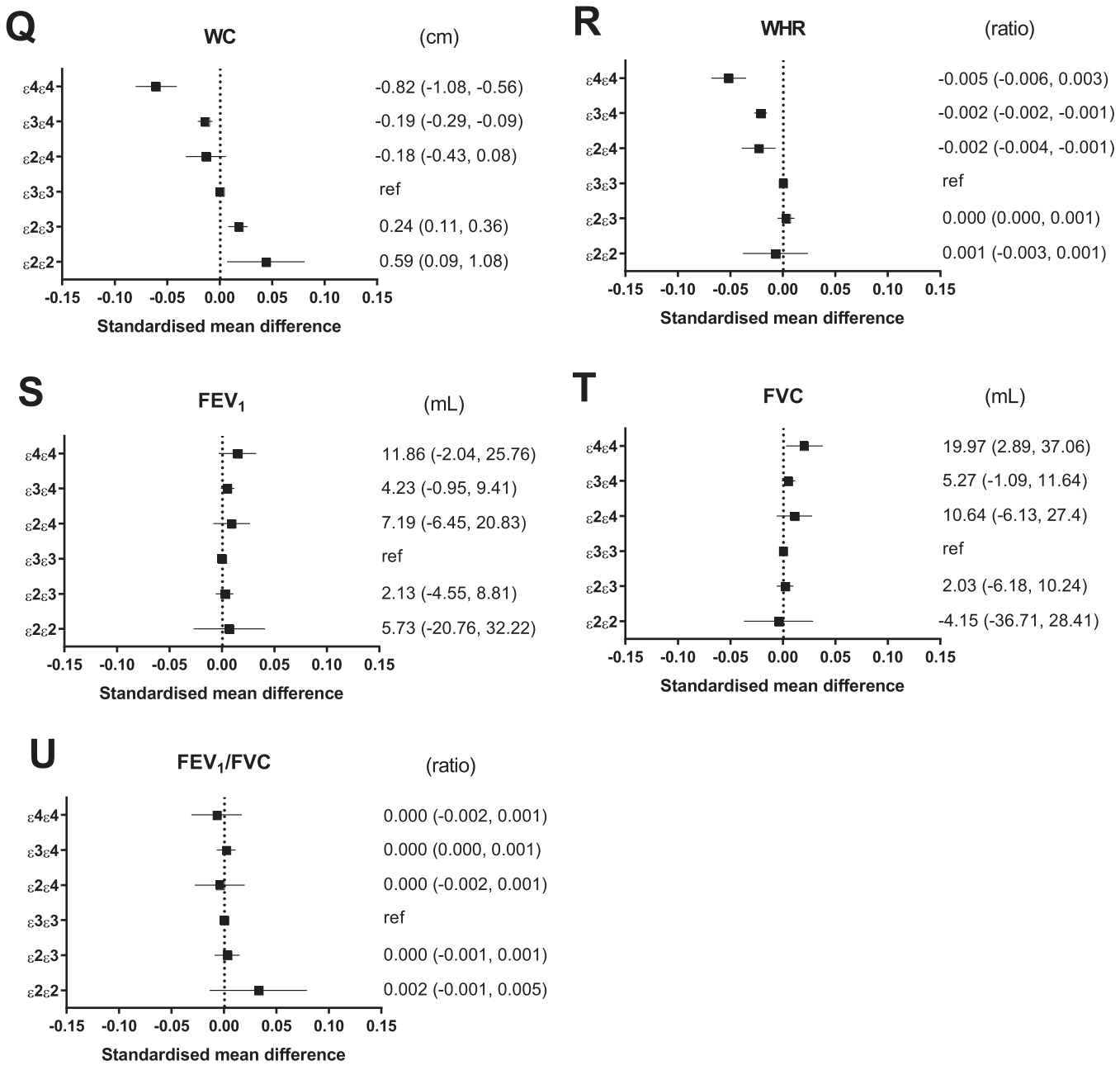

Fig. 2 Continued.

Declaration of Interests

The authors declare no competing interests.

Acknowledgments

The authors extend sincere thanks to the participants of UK Biobank, who made this work possible, and are grateful to the National Health and Medical Research Council Australia for the grant funding that supported this work (GNT1123603 and GNT1157281).

Role of Funding Source

The funding body played no role in the study design, the collection, analysis or interpretation of data, the writing of the report, or the decision to submit this paper for publication.

Data Sharing Statement

All data will be available to approved users of the UK Biobank upon application.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ebiom.2020.102954.

References

[1] Huebbe P, Rimbach G. Evolution of human apolipoprotein E (APOE) isoforms: Gene structure, protein function and interaction with dietary factors. Ageing Res Rev 2017;37:146–61.
[2] Tudorache IF, Trusca VG, Gafencu AV. Apolipoprotein E - a multifunctional protein with implications in various pathologies as a result of its structural features. Comput Struct Biotechnol J 2017;15:359–65.

[3] Rasmussen KL, Tybjaerg-Hansen A, Nordestgaard BG, Frikke-Schmidt R. Plasma levels of apolipoprotein E and risk of dementia in the general population. Ann Neurol 2015;77(2):301–11.
[4] Khan TA, Shah T, Prieto D, Zhang W, Price J, Fowkes GR, et al. Apolipoprotein E genotype, cardiovascular biomarkers and risk of stroke: systematic review and meta-analysis of 14,015 stroke cases and pooled analysis of primary biomarker data from up to 60,883 individuals. Int J Epidemiol 2013;42(2):475–92.
[5] Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015;12(3):e1001779.
[6] Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018;562(7726):203–9.
[7] Wei WQ, Bastarache LA, Carroll RJ, Marlo JE, Osterman TJ, Gamazon ER, et al. Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. PLoS One 2017;12(7):e0175508.
[8] Denny JC, Bastarache L, Ritchie MD, Carroll RJ, Zink R, Mosley JD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol 2013;31(12):1102–10.
[9] Carroll RJ, Bastarache L, Denny JC. R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment. Bioinformatics 2014;30(16):2375–6.
[10] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B (Methodological) 1995;57(No. 1):12.
[11] Smith CJ, Ashford JW, Perfetti TA. Putative Survival Advantages in Young Apolipoprotein varepsilon4 Carriers are Associated with Increased Neural Stress. J Alzheimers Dis 2019;68(3):885–923.
[12] Oria RB, Patrick PD, Zhang H, Lorntz B, de Castro Costa CM, Brito GA, et al. APOE4 protects the cognitive development in children with heavy diarrhea burdens in Northeast Brazil. Pediatr Res 2005;57(2):310–6.
[13] Trumble BC, Stieglitz J, Blackwell AD, Allayee H, Beheim B, Finch CE, et al. Apolipoprotein E4 is associated with improved cognitive function in Amazonian forager-horticulturalists with a high parasite burden. FASEB J 2017;31(4):1508–15.
[14] Tejedor MT, Garcia-Sobreviela MP, Ledesma M, Arbones-Mainar JM. The apolipoprotein E polymorphism rs7412 associates with body fatness independently of plasma lipids in middle aged men. PLoS One 2014;9(9):e108605.
[15] Volcik KA, Barkley RA, Hutchinson RG, Mosley TH, Heiss G, Sharrett AR, et al. Apolipoprotein E polymorphisms predict low density lipoprotein cholesterol levels and carotid artery wall thickness but not incident coronary heart disease in 12,491 ARIC study participants. Am J Epidemiol 2006;164(4):342–8.
[16] Mahmoud AI, Galdos FX, Dinan KA, Jedrychowski MP, Davis JC, Vujic A, et al. Apolipoprotein E is a pancreatic extracellular factor that maintains mature beta-cell gene expression. PLoS One 2018;13(10):e0204595.
[17] Shabanzadeh DM, Skaaby T, Sorensen LT, Eugen-Olsen J, Jorgensen T. Metabolic biomarkers and gallstone disease - a population-based study. Scand J Gastroenterol 2017;52(11):1270–7.
[18] Nascimento JCR, Matos GA, Pereira LC, Mourao A, Sampaio AM, Oria RB, et al. Impact of apolipoprotein E genetic polymorphisms on liver disease: an essential review. Ann Hepatol 2019.
[19] Kulminski AM, Barochia AV, Loika Y, Raghavachari N, Arbeev KG, Wojczynski MK, et al. The APOE epsilon4 allele is associated with a reduction in FEV1/FVC in women: A cross-sectional analysis of the Long Life Family Study. PLoS One 2018;13(11):e0206873.
[20] Lara-Guzman OJ, Gil-Izquierdo A, Medina S, Osorio E, Alvarez-Quintero R, Zuluaga N, et al. Oxidized LDL triggers changes in oxidative stress and inflammatory biomarkers in human macrophages. Redox Biol 2018;15:1–11.
[21] Martiskainen H, Takalo M, Solomon A, Stancakova A, Marttinen M, Natunen T, et al. Decreased plasma C-reactive protein levels in APOE epsilon4 allele carriers. Ann Clin Transl Neurol 2018;5(10):1229–40.
[22] Weisgraber KH, Innerarity TL, Mahley RW. Abnormal lipoprotein receptor-binding activity of the human E apoprotein due to cysteine-arginine interchange at a single site. J Biol Chem 1982;257(5):2518–21.
[23] Koopal C, Geerlings MI, Muller M, de Borst GJ, Algra A, van der Graaf Y, et al. The relation between apolipoprotein E (APOE) genotype and peripheral artery disease in patients at high risk for cardiovascular disease. Atherosclerosis 2016;246:187–92.
[24] Orsi FA, Lijfering WM, Van der Laarse A, Ruhaak LR, Rosendaal FR, Cannegieter SC, et al. Association of apolipoproteins C-I, C-II, C-III and E with coagulation markers and venous thromboembolism risk. Clin Epidemiol 2019;11:625–33.
[25] Tanne D, Koren-Morag N, Graff E, Goldbourt U. Blood lipids and first-ever ischemic stroke/transient ischemic attack in the Bezafibrate Infarction Prevention (BIP) Registry: high triglycerides constitute an independent risk factor. Circulation 2001;104(24):2892–7.
[26] Li J, Chen Y, Wu H, Li L. Apolipoprotein E (Apo E) gene polymorphisms and recurrent pregnancy loss: a meta-analysis. J Assist Reprod Genet 2014;31(2):139–48.
[27] Hsu YH, Liu Y, Hannan MT, Maixner W, Smith SB, Diatchenko L, et al. Genome-wide association meta-analyses to identify common genetic variants associated with hallux valgus in Caucasian and African Americans. J Med Genet 2015;52(11):762–9.
[28] Tao T, Jiang Y, Li W, Li Y, Du J, Gui J. Association of Vitamin D Receptor Gene TaqI, BsmI, FokI, and ApaI Polymorphisms and Susceptibility to Hallux Valgus in the Chinese Population. J Foot Ankle Surg 2018;57(4):753–8.
[29] Yang J, Wang J, Liang X, Zhao H, Lu J, Ma Q, et al. Relationship Between Genetic Polymorphisms of the TNF Gene and Hallux Valgus Susceptibility. Genet Test Mol Biomark 2019;23(6):380–6.
[30] Yin YW, Qiao L, Sun QQ, Hu AM, Liu HL, Wang Q, et al. Influence of apolipoprotein E gene polymorphism on development of type 2 diabetes mellitus in Chinese Han population: a meta-analysis of 29 studies. Metabolism 2014;63(4):532–41.
[31] Xue P, Niu WQ, Jiang ZY, Zheng MH, Fei J. A meta-analysis of apolipoprotein E gene epsilon2/epsilon3/epsilon4 polymorphism for gallbladder stone disease. PLoS One 2012;7(9):e45849.