Identification of epigenome-wide DNA methylation differences between carriers of *APOE* ε4 and *APOE* ε2

Authors

Rosie M. Walker[a,1*], Kadi Vaher[a,2], Mairead L. Bermingham[a], Stewart W. Morris[a], Andrew D. Bretherick[b], Yanni Zeng[b,3], Konrad Rawlik[c], Carmen Amador[b], Archie Campbell[d], Chris S. Haley[b], Caroline Hayward[b], David J. Porteous[a,d], Andrew M. McIntosh[e], Riccardo E. Marioni[a,4], Kathryn L. Evans[a,4*]

[1]Present address: Centre for Clinical Brain Sciences, Chancellor's Building, 49 Little France Crescent, Edinburgh BioQuarter, Edinburgh, EH16 4SB

[2]Present address: MRC Centre for Reproductive Health, The Queen's Medical Research Institute, Edinburgh BioQuarter, 47 Little France Crescent, Edinburgh, EH16 4TJ

[3]Present address: Faculty of Forensic Medicine, Zhongshan School of Medicine, Sun Yat-Sen University, 74 Zhongshan 2nd Road, Guangzhou 510080, China.

[4]Joint last authors

* Corresponding authors

Affiliations

[a]Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK

[b]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK

[c]Division of Genetics and Genomics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Roslin, UK

[d]Generation Scotland, Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK

[e]Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, EH10 5HF, UK

Email addresses

Rosie M. Walker: rwalke13@staffmail.ed.ac.uk; Kadi Vaher: kadi.vaher@ed.ac.uk; Mairead L. Bermingham: mairead.bermingham@igmm.ed.ac.uk, Stewart W. Morris: Stewart.Morris@igmm.ed.ac.uk, Andrew Bretherick: a.bretherick@ed.ac.uk, Yanni Zeng: zengyn5@mail.sysu.edu.cn, Konrad Rawlik: Konrad.Rawlik@roslin.ed.ac.uk, Carmen Amador: carmen.amador@igmm.ed.ac.uk, Archie Campbell: archie.campbell@ed.ac.uk, Chris S. Haley: chris.haley@igmm.ed.ac.uk, Caroline Hayward: Caroline.Hayward@igmm.ed.ac.uk, David J. Porteous: david.porteous@igmm.ed.ac.uk, Andrew M. McIntosh: andrew.mcintosh@ed.ac.uk, Riccardo E. Marioni: Riccardo.Marioni@ed.ac.uk, Kathryn L. Evans: Kathy.Evans@igmm.ed.ac.uk

Abstract

BACKGROUND: The *Apolipoprotein E* (*APOE*) ε4 allele is the strongest genetic risk factor for late onset Alzheimer's disease, while the ε2 allele confers protection. Previous studies report differential DNA methylation of *APOE* between ε4 and ε2 carriers, but associations with epigenome-wide methylation have not previously been characterised.

METHODS: Using the EPIC array, we investigated epigenome-wide differences in whole blood DNA methylation patterns between Alzheimer's disease-free *APOE* ε4 (n=2469) and ε2 (n=1118) carriers from the two largest single-cohort DNA methylation samples profiled to date. Using a discovery, replication and meta-analysis study design, methylation differences were identified using epigenome-wide association analysis and differentially methylated region (DMR) approaches. Results were explored using pathway and methylation quantitative trait loci (meQTL) analyses.

RESULTS: We obtained replicated evidence for DNA methylation differences in a ~169kb region, which encompasses part of *APOE* and several upstream genes. Meta-analytic approaches identified DNA methylation differences outside of *APOE*: differentially methylated positions were identified in *DHCR24*, *LDLR* and *ABCG1* ($2.59 \times 10^{-100} \leq P \leq 2.44 \times 10^{-8}$) and DMRs were identified in *SREBF2* and *LDLR* ($1.63 \times 10^{-4} \leq P \leq 3.01 \times 10^{-2}$). Pathway and meQTL analyses implicated lipid-related processes and high density lipoprotein cholesterol was identified as a partial mediator of the methylation differences in *ABCG1* and *DHCR24*.

CONCLUSIONS: *APOE* ε4 vs. ε2 carrier status is associated with epigenome-wide methylation differences in the blood. The loci identified are located in *trans* as well as *cis* to *APOE* and implicate genes involved in lipid homeostasis.

KEYWORDS: Alzheimer's disease, APOE, Apolipoprotein E, DNA methylation, cholesterol, lipids

1. Background

The ε4 allele of the *apolipoprotein E* gene (*APOE*) is the strongest genetic risk factor for late-onset (>65 years) Alzheimer's disease (AD) (1-3). Inheritance of one copy of this allele increases late-onset AD risk by two to four-fold, with two copies conferring an eight to twelve–fold increase in risk compared to the ε3/ε3 genotype (4, 5). The ε4 allele is also associated with a younger age-of-onset, with ε4 homozygotes having an average age-of-onset of 68 years compared to 84 years for ε3 homozygotes (4). In contrast, the ε2 allele has been associated with a ~50% reduction in AD risk compared to the ε3/ε3 genotype (5).

The three *APOE* alleles (ε2/ε3/ε4) are defined by two *APOE* exon 4 single nucleotide polymorphisms (SNPs) and encode functionally distinct ApoE isoforms. Isoform-dependent behaviours have been observed for many ApoE functions, including lipid metabolism, amyloid beta (Aβ) metabolism, tau phosphorylation, inflammation, and synaptic plasticity, with ApoE4 and ApoE2 conferring effects consistent with increased and reduced AD risk, respectively (6, 7).

Despite the wealth of evidence linking ApoE to processes implicated in AD pathogenesis, understanding of the specific mechanism(s) by which genetic variation at this locus alters risk remains incomplete. *APOE* genotype acts in conjunction with other genetic and/or environmental factors to confer AD risk: the lifetime risk of dementia or mild cognitive impairment is 31%-40% for ε4/ε4 homozygotes (8) but the effects of *APOE* ε4 have been shown to be modified by ethnic background and sex (5, 9). DNA methylation is associated with both genetic and environmental factors, and previous studies have identified associations with AD and neuropathological hallmarks of AD (10-12), AD risk factors (e.g. ageing (13), obesity (14) and lipid levels (15)), as well as modifiers of *APOE* genotype effects (e.g. sex (16) and ethnicity (17, 18)).

The two *APOE* haplotype-defining SNPs are located in a CpG island and have a direct effect on methylation by creating/destroying CpG sites (19). The *APOE* ε2/ε3/ε4 haplotype is associated with

methylation at other CpG sites within *APOE* (20, 21) but, to date, associations with methylation across the epigenome have not been assessed. We hypothesised that characterising these associations would yield insights into the biological context in which *APOE* acts, thus facilitating the search for mechanisms conferring risk/resilience for AD. Importantly, by studying individuals who are free from AD, we have the potential to identify pathogenic processes that precede the onset of irreversible neurodegeneration.

2. Methods

2.1. Participants

The participants were selected from the Generation Scotland: Scottish Family Health Study (GS:SFHS) cohort (~24,000 participants aged ≥18 years at recruitment), which has been described previously (22, 23). The participants included in this study were of European (predominantly British) ancestry, following the exclusion of participants with likely recent Italian or African/Asian ancestry by principal components (PC) analysis (24). Participants attended a baseline clinical appointment at which they were phenotyped for social, demographic, health and lifestyle factors, completed cognitive assessments, and provided physical measurements and samples for DNA extraction. GS:SFHS obtained ethical approval from the NHS Tayside Committee on Medical Research Ethics, on behalf of the National Health Service (reference: 05/S1401/89) and has Research Tissue Bank Status (reference: 15/ES/0040).

2.2. Blood sample collection and DNA extraction

DNA was extracted from blood (9ml) collected in EDTA tubes using the Nucleon BACC3 Genomic DNA Extraction Kit (Fisher Scientific), following the manufacturer's instructions (25).

2.3. Genotyping of *APOE*

The *APOE* ε2/ε3/ε4 haplotypes are defined by two SNPs, rs429358 and rs7412, which were genotyped using TaqMan probes at the Clinical Research Facility, Edinburgh.

2.4. Measurement of cholesterol levels

Total and high density lipoprotein (HDL) cholesterol were measured at the GS:SFHS baseline appointment and non-HDL cholesterol levels were calculated by subtracting HDL cholesterol from total cholesterol. The non-HDL cholesterol level reflects a combination of low density lipoprotein (LDL) cholesterol and very low-density lipoprotein.

2.5. Genome-wide DNA methylation profiling for EWAS analyses

DNA methylation was profiled using the Infinium MethylationEPIC BeadChip (Illumina Inc.) in a discovery (n=5190) and replication (n=4583) sample, as described previously (26-28) (Supplementary Methods). The discovery and replication samples were normalised separately and converted to M-values. The discovery data was corrected for relatedness (Supplementary Methods). Participants in the replication sample were unrelated (SNP-based relatedness<0.05) to each other and/or discovery sample participants.

Poor performing probes, X/Y chromosome probes and participants with unreliable self-report data or potential XXY genotype were excluded (Supplementary Methods). The final discovery dataset comprised M-values at 760,943 loci for 5087 participants; the replication dataset comprised M-values at 758,332 loci for 4450 participants. All subsequent analyses of the DNA methylation data were carried out using R versions 3.6.0., 3.6.1., or 3.6.2. (29, 30).

2.6. Statistical analyses

A flow chart indicating all analyses is presented in Figure 1.

2.7. Epigenome-wide association studies

EWASs were implemented using limma (31). CpG M-values were the dependent variable and *APOE* ε4 vs. ε2 carrier status (a binary variable indicating *APOE* ε4 carriers with a "1" and *APOE* ε2 with a "0"; ε4/ε2 and ε3/ε3 participants were excluded) was the predictor-of-interest. Participants self-reporting AD (n=five) were excluded. Additional covariates were included as below:

Discovery sample

CpG site (pre-corrected for relatedness, estimated cell counts and processing batch) ~ *APOE* ε4 vs. ε2 + age + sex + smoking status + pack years + 20 methylation PCs

Replication sample

CpG site (M-values) ~ *APOE* ε4 vs. ε2 + age + sex + smoking status + pack years + estimated cell counts (granulocytes, natural killer cells, B-lymphocytes, CD4+T-lymphocytes and CD8+T-lymphocytes) + processing batch + 20 methylation PCs

The variables "smoking status", "pack years" and the methylation PCs are explained in the Supplementary Methods.

An additional sensitivity analysis of the replication sample was performed in which the first 10 genetic PCs, calculated using GCTA (32), were included. The decision to include 10 PCs was based on inspection of a scree plot (Supplementary Figure 1).

Limma was used to calculate empirical Bayes moderated t-statistics from which *P*-values were obtained. The significance threshold in the discovery sample was $P \leq 3.6 \times 10^{-8}$ (33). Sites attaining significance in the discovery sample were assessed in the replication sample using a Bonferroni-corrected threshold of 0.05/no. sites assessed.

2.8. EWAS meta-analysis

Inverse variance-weighted fixed effects meta-analyses of 756,971 sites common to the discovery and replication EWAS results were performed using METAL (34). Sites attaining a meta-analysis $P \leq 3.6 \times 10^{-8}$ were considered significant.

2.9. Comparison of DNA methylation levels between *APOE* haplotypes

For the differentially methylated positions (DMPs) identified through the EWAS meta-analysis, pairwise differences in methylation levels between carriers of the *APOE* ε2/ε2, ε2/ε3, ε3/ε3, ε3/ε4, and ε4/ε4 haplotypes in the discovery sample were investigated, using the R package lsmeans (35). *P*-values were adjusted using a Bonferroni correction to account for the 10 within-CpG comparisons performed for each of the 20 CpGs assessed (i.e. an adjustment was performed for 200 tests). Corrected $P \leq 0.05$ was considered statistically significant.

2.10. Identification of differentially methylated regions

DMRs associated with *APOE* ε4 vs. ε2 carrier status were identified using the dmrff.meta function from the dmrff R package (36). Putative DMRs were defined as regions containing two to thirty sites separated by ≤500 bp with EWAS meta-analysis $P \leq .05$ and methylation changes in a consistent direction. Following dmrff's subregion selection step, DMRs with Bonferroni-adjusted $P \leq .05$ were declared significant.

2.11. Gene ontology/KEGG pathway analyses

Gene ontology (GO) and KEGG pathway analyses were implemented using missMethyl's gometh function (37). The target list comprised probes that were suggestively associated with the phenotype-of-interest ($P \leq 1 \times 10^{-5}$) in the meta-EWAS or that contributed to a significant DMR (adjusted $P \leq 0.05$) and the gene universe included all analysed probes. Enrichment was assessed using a hypergeometric test, accounting for the bias arising from the variation in the number of

168 probes-per-gene. Bonferroni-corrected significance thresholds of $P \leq 2.21 \times 10^{-6}$ and $P \leq 1.48 \times 10^{-4}$
169 were applied to account for the 22,578 GO terms and 337 KEGG pathways assessed.

170 **2.12. Bootstrap mediation analysis**

171 The roles of cholesterol levels (total cholesterol, HDL cholesterol and non-HDL cholesterol) in
172 mediating any observed associations between *APOE* ε4 vs. ε2 carrier status and DNA methylation
173 were assessed by bootstrap mediation analysis using the R package "mediation" (38). The analyses
174 were performed using 10000 bootstrap samples in the discovery and replication samples separately
175 and these results were then meta-analysed using inverse variance-weighted fixed effects meta-
176 analyses to obtain meta-analyses *P*-values and effect estimates. Significant mediation was declared
177 when the meta-analysis *P*-value met a Bonferroni-adjusted (to account for the assessment of 20
178 DMPs) significance threshold of $P \leq .05$.

179 **2.13. Genotyping and imputation**

180 The genotyping and imputation of GS:SFHS to the Haplotype Reference Consortium reference panel
181 release 1.1 (39) has been described previously (25, 40) (Supplementary Methods).

182 **2.14. Identification of methylation quantitative trait loci**

183 Methylation quantitative trait loci (meQTLs) were identified using the discovery sample. Following
184 quality control, the data was normalised and corrected as described previously (41) (Supplementary
185 Methods). Normalised and corrected data was available for 26 of the 31 CpGs-of-interest in this
186 study. The resulting residuals were inverse rank normal transformed and entered as the dependent
187 variable in simple linear model GWASs to identify meQTLs. GWASs were implemented using
188 REGSCAN v0.5 (42). SNPs that were associated with a CpG with $P \leq 1.92 \times 10^{-9}$ ($5 \times 10^{-8}/26$) were
189 declared to be meQTLs. SNPs located within one megabase up- or downstream of their associated
190 CpG were defined as *cis* meQTLs; all other associated SNPs were defined as *trans* meQTLs. A look-up

191 analysis of the GWAS catalog (43) (GWAS catalog v1.0.2., downloaded 07/09/20) was performed in
192 which SNPs identified as meQTLs for the CpGs of interest were queried for their significant ($P \leq 5 \times 10^{-8}$)
193 disease or trait associations in the GWAS catalog.

194 ### 2.15. Association analyses of *APOE* ε4 vs. ε2 carrier status

195 Association analyses were performed to assess whether meQTLs for the meta-analysis DMPs are
196 associated with *APOE* ε4 vs. ε2 carrier status and, therefore, might contribute to the differences in
197 methylation observed between *APOE* ε4 and ε2 carriers. Association tests used BOLT-LMM (44) to
198 perform linear mixed models in participants with available *APOE* genotypes (ε2 n=2613; ε4 n=5401).
199 BOLT-LMM adjusts for population structure and relatedness between individuals whilst assessing
200 association. Sex was included as a covariate. Associations were considered significant when $P \leq 5 \times 10^{-8}$
201 .

202 ## 3. Results

203 ### 3.1. Sample demographics

204 The EWAS discovery sample comprised 1253 *APOE* ε4 and 596 *APOE* ε2 allele carriers and the
205 replication sample comprised 1216 *APOE* ε4 and 522 *APOE* ε2 allele carriers. Twenty-seven ε2/ε2,
206 569 ε2/ε3, 2926 ε3/ε3, 1128 ε3/ε4 and 125 ε4/ε4 participants from the discovery sample were
207 available for the pairwise analysis of genotypes. Key sample demographic information is presented
208 in Supplementary Table 1.

209 ### 3.2. Identification of differentially methylated positions and regions in *APOE* ε4 vs. ε2 carriers

210 An EWAS of *APOE* ε4 vs. ε2 carriers in the discovery sample identified eight significant DMPs, of
211 which half were hypermethylated in *APOE* ε4 carriers. These DMPs had a mean absolute effect size
212 of 0.070 (range: 0.033 – 0.103) and *P*-values ranging from 6.40×10^{-56} to 8.81×10^{-9}. All eight sites
213 were also significant ($8.60 \times 10^{-49} \leq P \leq 7.25 \times 10^{-6}$) in the replication sample with a consistent direction

of effect (mean absolute effect size = 0.102; range: 0.049 – 0.170; Supplementary Table 2). The eight sites are located in a ~169kb region on chromosome 19 (chr. 19: 45,242,346-45,411,802; GRCh37/hg19), which spans a region of the genome upstream of and including part of the *APOE* gene (chr19: 45,409,039-45,412,650; GRCh37/hg19). A sensitivity analysis of the discovery sample in which a methylation-based smoking score (45) was included as a covariate instead of the smoking covariates included in the original analysis ("smoking status" and "pack years") produced highly similar results across all measured CpGs (correlation between effect sizes = 0.99, 95% confidence interval (CI): 0.99-0.99; $P<2.2 \times 10^{-16}$; Supplementary Table 2). An additional sensitivity analysis in which the first 10 genetic PCs were included as additional covariates in the analysis of the replication sample also produced results that were highly correlated with those from the original replication sample analysis ($r = 1.00$, 95% CI: 1.00-1.00, $P<2.2 \times 10^{-16}$; Supplementary Table 2).

Inverse variance-weighted fixed effects meta-analysis of the discovery and replication samples identified 20 DMPs, with *APOE* ε4 carrier status associated with hypomethylation at 13 (65%) of these sites. Across all 20 DMPs, the mean absolute effect size was 0.052 (range: 0.022 – 0.11) with *P*-values ranging from 2.80×10^{-100} to 2.4×10^{-8} (Table 1; Figure 2). Sixteen of these sites are located on chromosome 19q in a ~233kb region (chr19: 45,221,584 – 45,454,752; GRCh37/hg19) encompassing *APOE* and several surrounding genes (Supplementary Figure 2). Henceforth, the region containing *APOE* and neighbouring genes will be referred to as the "*APOE* locus". The most significant DMP, cg13375295, is located ~4.5kb upstream of *Poliovirus Receptor-related 2* (*PVRL2*), a gene situated ~16.5kb upstream of *APOE*. Four other DMPs (cg10762466, cg10178308, cg11643040 and cg06198803) are located either upstream or in the gene body of *PVRL2*. Two DMPs (cg06750524 and cg16471933) are located in *APOE*: cg06750524, the DMP with the largest effect size, in the intron between exons 2 and 3; and cg16471933 in exon 4, 139bp 5' of rs429358, one of the *APOE* ε4/ε2-defining SNPs. Although both the *APOE* DMPs are more highly methylated in *APOE* ε4 carriers; the DMPs in the surrounding region do not show a consistent direction of effect.

Four DMPs are located outside of chromosome 19q: cg17901584, 785bp upstream of the *24-dehydrocholesterol reductase* (*DHCR24*) gene on chromosome 1; cg19751789, 94bp upstream of the *low density lipoprotein receptor* (*LDLR*) gene on chromosome 19p; and two, cg16740586 and cg06500161, are located 668bp apart in the same intron of multiple *ATP Binding Cassette Subfamily G Member 1* (*ABCG1*) isoforms.

To further investigate the pattern of methylation observed at these 20 DMPs, pairwise comparisons were performed between carriers of the following *APOE* haplotypes: ε2/ε2, ε2/ε3, ε3/ε3, ε3/ε4, and ε4/ε4. These analyses revealed a range of allele-associated methylation patterns, which are depicted in Supplementary Figure 3 and described in Supplementary Table 3. Carriers of the *APOE* ε2 allele (ε2/ε2 or ε2/ε3) differed from ε3/ε3 homozygotes at 14 of the DMPs, whilst carriers of the *APOE* ε4 allele (ε4/ε4 or ε3/ε4) differed from ε3/ε3 homozygotes at four DMPs. Dosage effects were observed at two DMPs for ε2 carriers (Supplementary Figure 3 A and S) and one DMP for ε4 carriers (Supplementary Figure 3B), although the small numbers of participants who are homozygous for *APOE* ε2 (n = 27) and ε4 (n = 128) likely rendered our study underpowered to detect all dosage effects. For the two DMPs located within the *APOE* gene (cg06750524 and cg16471933), an increase in mean methylation levels was observed from ε2/ε2 homozygotes to ε3/ε3 homozygotes, with a further increase to the ε4/ε4 group (Supplementary Figure 3B and E). At the four DMPs outside of the *APOE* locus, the methylation differences appear to be predominantly driven by the ε2 allele (Supplementary Figure 3K, M, O and S).

Differentially methylated regions (DMRs) were identified using a meta-analysis approach, which identified six significant regions (Supplementary Figure 4). Across all the DMRs, the mean absolute effect size was 0.182 (range: 0.135 – 0.231) and Bonferroni-adjusted *P*-values ranged from 1.63×10^{-4} to 3.01×10^{-2} (Table 2).Three of the DMRs are located at the *APOE* locus, two are in the first intron of *Sterol Regulatory Element Binding Transcription Factor 2* (*SREBF2*) on chromosome 22, and the other is in the putative promoter of *LDLR* on chromosome 19p. All but one of the DMRs, which is

264 located 190 bp upstream of the *apolipoprotein C1 pseudogene 1* (*APOC1P1*) at the *APOE* locus, are
265 hypomethylated in *APOE* ε4 carriers. Only one of the DMRs, located in an exon of a read-through
266 transcript involving *apolipoprotein C2* (*APOC2*) and *apolipoprotein C4* (*APOC4*), contains CpGs that
267 were identified as DMPs (cg13119609 and cg09555818).

268 GO analysis was carried out using the 19 Entrez IDs mapping to the 46 CpG sites with a meta-EWAS
269 $P \leq 1 \times 10^{-5}$ or that contributed to a significant DMR. This identified 78 significant GO terms (Table 3;
270 Supplementary Table 4), the most significant of which was "cholesterol metabolic process" ($P=2.00 \times$
271 10^{-11}). Significant enrichment for the KEGG pathways "cholesterol metabolism" ($P=5.93 \times 10^{-10}$) and
272 "steroid biosynthesis" ($P=1.22 \times 10^{-4}$) was also observed.

273 **3.3. Assessment of the role of cholesterol in mediating methylation differences between *APOE***
274 **ε4 and ε2 carriers**

275 Given the well-established role of ApoE in cholesterol metabolism (6), bootstrap mediation analyses
276 were performed to assess the role of cholesterol levels (total, HDL or non-HDL cholesterol) in
277 mediating the association between *APOE* ε4 vs. ε2 carrier status and methylation at the 20 meta-
278 analysis DMPs. Inverse variance-weighted fixed effects meta-analysis of the bootstrap mediation
279 analyses in the discovery and replication samples identified HDL cholesterol as a significant mediator
280 of the associations with the two *ABCG1* DMPs cg06500161 (effect size = 0.006; effect size standard
281 error = 0.001; $P=1.18 \times 10^{-6}$) and cg16740586 (effect size = 0.004; effect size standard error = 0.001;
282 $P=4.93 \times 10^{-5}$), and the *DHCR24* promoter DMP, cg17901584 (effect size = -0.007; effect size standard
283 error = 0.001; $P=6.04 \times 10^{-6}$), for which it mediated 25.2%, 11.5%, and 18.2% of the relationship,
284 respectively (Supplementary Table 5). For some sites, inspection of the *P*-values indicated total and
285 non-HDL cholesterol to be significant mediators but the proportion of the relationship between
286 *APOE* ε4 vs. ε2 carrier status and methylation attributable to the mediator was negative
287 (Supplementary Table 5). This indicates that, at these sites, the direction of the association between

288 the cholesterol phenotypes and methylation is the opposite to the direction of the association
289 between *APOE* ε4 vs. ε2 carrier status and methylation.

290 **3.4. Assessment of meQTLs associated with loci that are differentially methylated between**
291 ***APOE* ε4 and ε2 carriers**

292 To explore the DMP and DMR CpGs further, meQTL analyses were performed. Whilst it was
293 expected that meQTLs for the DMP and DMR CpGs would be identified at the *APOE* locus, the
294 identification of meQTLs outside of this locus would be of particular interest. Should meQTLs outside
295 of the *APOE* locus be found to be show non-random segregation with *APOE* ε4 vs. ε2 carrier status,
296 these meQTL SNPs might contribute to the methylation differences observed in this study and *APOE*
297 genotype effects more generally.

298 It was possible to assess meQTLs for 26 of the 31 CpGs of interest (from the DMP and DMR
299 analyses); amongst these CpGs, 23 were associated with a meQTL. In total, 3727 significant CpG-SNP
300 associations were identified for the 23 CpGs, involving 1654 unique SNPs (Figure 3; Supplementary
301 Table 6). Unsurprisingly, more than half of the meQTLs (n=947) were located in a ~719kb region
302 (chr19: 45,004,645– 45,723,446; GRCh37/hg19) spanning *APOE*. The *APOE* region meQTLs are
303 associated with 16 CpGs, of which 14 are located at the *APOE* locus. None of these meQTLs is
304 associated with all 16 CpGs: two are each associated with nine CpGs: rs7412, one of the *APOE*
305 ε2/ε3/ε4-defining SNPs; and rs41290120, an intronic *PVRL2* SNP that is in high linkage disequilibrium
306 with rs7412 with D' = 0.85 in the British population (46). The two CpGs associated in *trans* with SNPs
307 in the *APOE* region are cg16000331 in *SREBF2* and cg19751789 in *LDLR*.

308 Outside of the *APOE* locus, the remaining 707 meQTLs, which are associated with 10 CpGs, are
309 located in 11 genomic regions (Figure 3; Supplementary Table 7), with each region containing
310 meQTLs associated with between one and eight CpGs-of-interest. To assess whether these meQTLs
311 might contribute to *APOE* ε4 vs. ε2-associated methylation differences, their association with *APOE*

ε4 vs. ε2 carrier status was assessed. No significant associations were observed, suggesting that the *APOE* ε4 vs. ε2-associated methylation differences are predominantly driven by genotype at the *APOE* locus.

To investigate potential trait/disease associations with variation in methylation levels at the CpGs-of-interest, the GWAS catalog was queried(43). This identified 234/1654 meQTLs as having genome-wide significant associations with 316 traits (Supplementary Table 8). More than one third of the associations are with a lipid-related traits, including LDL, HDL and total cholesterol levels. As expected, many of the meQTL SNPs within the *APOE* locus have previously been associated with AD and related traits, such as "Cerebrospinal fluid p-tau levels", "Cerebral amyloid deposition (PET imaging)" and "Cognitive decline". Interestingly, five SNPs located outside of the *APOE* locus have also been associated with traits related cognitive ability ("Cognitive ability, years of educational attainment or schizophrenia (pleiotropy) ", "General cognitive ability", "Intelligence" and "Self-reported math ability"). Four of these SNPs encompass the 3' end of *CCDC134* and most of the neighbouring *SREBF2*. Between them, these four SNPs are associated in *cis* with methylation at the four CpGs forming the two *SREBF2* DMRs. The fifth SNP, which is located on chromosome 6 in the pseudogene *CCDC162P*, is associated with methylation at CpGs in *SREBF2* and *LDLR*. Three meQTL SNPs have been associated with several age-related disorders (e.g. heart failure, stroke, and cancer) and endophenotypes of these disorders (including cholesterol levels, blood pressure and blood glucose) in a pleiotropic GWAS meta-analysis (47).

4. Discussion

We performed the first epigenome-wide comparison of DNA methylation between carriers of the *APOE* ε4 and ε2 haplotypes, which confer risk for and protection from AD, respectively. In large discovery and replication samples, we confirm the presence of *APOE* haplotype-associated methylation differences in *APOE*, demonstrate that differences in methylation at the *APOE* locus

336 span a broad genomic locus encompassing several genes, and find evidence for altered methylation
337 at sites unlinked to the *APOE* locus. The observed methylation differences are located in a network
338 of genes involved in lipid metabolism and homeostasis.

339 Methylation differences were identified using discovery, replication and meta-analysis EWASs and
340 DMR analysis. Eight DMPs located on chromosome 19 in a ~169kb region spanning from upstream of
341 *BCL3* to the *APOE*'s fourth exon showed replicated association. An additional twelve DMPs, eight of
342 which are located in a ~233kb region at the *APOE* locus, were identified by meta-analysing the
343 discovery and replication samples. DMR analysis identified six regions of differential methylation,
344 both within and outside of the *APOE* locus.

345 Within the *APOE* gene, two DMPs, cg06750524, in the second intron, and cg16471933, in the fourth
346 exon, were identified. *APOE* ε4 carriers showed higher methylation levels at both. This observation
347 directly replicates a previous study (21) and is in line with Foraker et al.'s observation of increased
348 methylation of the *APOE* exon 4 CpG island in ε4 carriers (20). Moreover, we have previously
349 demonstrated (48) that the pattern of methylation in *APOE* in our sample is consistent with that
350 described by Ma et al. (2015) (21). Pairwise comparisons revealed differences in *APOE* methylation
351 to be driven both by differences between ε2 carriers and ε3/ε3 homozygotes and ε4 carriers and
352 ε3/ε3 homozygotes. One interpretation of this observation is that the spectrum of methylation at
353 the *APOE* DMPs reflects the spectrum of AD risk conferred by different AD genotypes. It is clear,
354 however, that additional, likely experimental, studies are required to assess the implications of the
355 observed methylation pattern.

356 The differentially methylated CpGs at the *APOE* locus span a broad region that encompasses several
357 genes containing AD-associated variants (49). Long-ranging linkage disequilibrium in the region
358 complicates the interpretation of association signals; however, conditional analysis and fine-
359 mapping studies suggest the presence of multiple independent AD risk loci across the region (3, 49).

As such, the methylation differences observed in this study may be associated with variants that, whilst being in LD with the *APOE* ε2/ε4-defining SNPs, confer risk via different pathways to these SNPs. This notion is supported by the observation that SNPs that define an *APOE* ε4-independent AD-risk haplotype in *PVRL2* (49) are highly significant meQTLs for the most significant DMP identified in this study.

Beyond the *APOE* locus, DMPs were identified in an *ABCG1* intron, and upstream of *DHCR24* and *LDLR*. Comparisons with ε3/ε3 homozygotes suggested the ε2 allele to be the primary driver of these differences, suggesting the possibility that altered methylation of genes involved in lipid metabolism might contribute to this allele's protective effects. DMRs were identified in the gene body of *SREBF2* and in the putative promoter region of *LDLR*. The CpGs involved in the DMPs and DMRs located outside of the *APOE* locus are associated with several meQTLs, with all of the CpGs except those involved in the *LDLR* DMR being associated with meQTLs in *cis* as well as in *trans*. Our findings did not, however, support a role for *cis* meQTLs for these CpGs driving associations with *APOE* ε4 vs. ε2 carrier status.

The genes outside of the *APOE* locus that harbour differentially methylated CpGs are implicated in lipid metabolism or homeostasis. *ABCG1*, which is highly expressed in the brain, encodes a cholesterol and phospholipid transporter and is involved in regulating the sterol biosynthetic pathway (50). *DHCR24*, which encodes the cholesterol biosynthesis enzyme 3ß-hydroxysterol-Δ24 reductase, also known as seladin-1, plays a neuroprotective role in AD-related stress conditions, including Aβ toxicity, oxidative stress and inflammation (51, 52). The alteration of seladin-1 expression in mouse brain and human neuroblastoma cell cultures has been shown to affect β-secretase processing of amyloid precursor protein, with reduced seladin-1 being associated with an increased rate of Aβ production (53). Future studies should assess whether methylation-associated differences in the brain expression of seladin-1 (6) might mediate the established associations between *APOE* ε4 vs. ε2 haplotype and Aβ production. The *LDLR* gene encodes the LDL receptor, one

385 of the neuronal receptors capable of mediating the endocytosis of ApoE, thus, maintaining brain
386 cholesterol homeostasis. *LDLR* expression is regulated, in part, by *SREBF2*, a transcriptional regulator
387 of sterol-regulated genes, which contains a SNP that is associated both with *SREBF2* expression and
388 CSF levels of the AD biomarkers Aβ and tau (54).

389 The link between *APOE* ε4 vs. ε2-associated methylation differences and lipid-related processes and
390 pathways was further supported by GO and KEGG analyses, the identification of meQTLs for the
391 differentially methylated CpGs, which were clustered in genomic regions that contain several lipid-
392 related genes, and their GWAS-associated phenotypes. It would be of interest to investigate the
393 mechanisms underlying the clustering of meQTLs in these genomic regions. Future studies might
394 assess, for example, the extent to which meQTLs associated with the differentially methylated CpGs
395 are enriched in these regions and whether they disproportionately affect certain sequence motifs.
396 Previous EWASs have also identified associations between some of the *APOE* ε4 vs. ε2-associated
397 CpGs and cholesterol levels: the *DHCR24* (cg17901584), *ABCG1* (cg06500161) and *SREBF2*
398 (cg16000331) DMPs have been associated with HDL cholesterol, total cholesterol and triglyceride
399 levels (15, 55-57). Comparisons with previous EWASs are, however, limited by the fact that the
400 majority of previous EWASs used the 450K array, which, does not contain 10 of the *APOE* ε4 vs. ε2-
401 associated CpGs.

402 As differences in lipid metabolism between carriers of the *APOE* ε4 and ε2 haplotypes are well-
403 documented (6), we assessed whether variation in blood cholesterol levels might mediate the
404 observed *APOE* ε4 vs. ε2-associated methylation differences. HDL cholesterol was found to be a
405 partial mediator of the relationship between *APOE* ε4 vs. ε2 carrier status and methylation at three
406 loci located outside of the *APOE* locus (two within *ABCG1* and one in the promoter of *DHCR24*), thus
407 suggesting one mechanism that might underlie these *trans* effects. Consistent with our observation
408 that methylation differences at these loci appear to be predominantly driven by *APOE* ε2 carriers
409 (when compared to *APOE* ε3/ε3 homozygotes), higher HDL cholesterol levels have been reported in

carriers of *APOE* ε2 (58). The effect of HDL cholesterol on methylation varied between the three loci, with *APOE* ε2 carriers showing increased methylation at the site located in the *DHCR24* promoter and decreased methylation at the two *ABCG1* sites. This suggests that increased HDL cholesterol levels do not exert a general effect on methylation but rather that methylation varies in a locus-specific manner in response to variation in HDL levels. It should be noted that an assumption of this analysis is that reverse causation does not exist between the outcome, methylation, and the mediator, cholesterol. Previous Mendelian Randomisation studies have predominantly supported this premise (59, 60); however, the ability to identify robust genetic instruments has limited both the number of methylation sites assessed and the ability to assess reverse causation. Limitations to the GS:SFHS cholesterol data should also be noted when interpreting these findings: triglyceride levels were not measured, preventing LDL cholesterol assessment; and blood samples were not taken at a consistent time of day or after fasting.

The cross-sectional nature of this study precludes the observed methylation differences being interpreted as conferring risk, protection or compensation. Comparison of methylation at these loci in *APOE* ε4 and ε2 carriers with AD would be useful in addressing this question; however, the optimum study design would involve the longitudinal assessment of the trajectory of ε4 vs. ε2-associated methylation differences in AD-free individuals in midlife who either do or do not later develop AD. These analyses are currently not feasible due to the small sizes of existing AD patient blood-based DNA methylation samples and insufficient follow-up time of large population-based samples.

Studies assessing the association of neuropathological hallmarks (neuritic plaque burden and/or neurofibrillary tangles) of AD with DNA methylation in the brain have not identified the loci identified in the present study (10, 12, 61). Although the phenotypes assessed differ, the existence of *APOE* haplotype-associated differences in Aβ metabolism and tau phosphorylation (6) suggest that some degree of overlap might be expected. The neuropathological hallmarks of AD are,

435 however, complex phenotypes and *APOE* haplotype will be one of many contributing factors (De
436 Jager et al. (10) reported that *APOE* ε4 could account for 13.9% of the variance in NP burden
437 observed in their participants). In addition, the smaller samples assessed by De Jager et al. (10),
438 Lunnon et al. (12) and Smith et al. (61) may have been inadequately powered to detect any
439 methylation differences driven by *APOE* haplotype. Differences in age and methylation profiling
440 platform are also likely to limit comparability: the participants assessed in these studies were much
441 older (mean age >75 years) than those assessed in our study (mean age ~ 50 years) and array
442 differences mean that only two thirds of our DMP/DMR probes were assessed. Two important
443 corollaries of the age difference are that brain-based studies are more likely to (i) suffer from
444 survivor bias and (ii) be better suited to investigating end-of-disease processes. It is also important
445 to note that *APOE* is involved in multiple processes, with *APOE* ε4 conferring risk for AD, at least in
446 part, via mechanisms that are not related to Aβ or tau pathology. A recent study has indicated that
447 *APOE* ε4-associated breakdown of the blood-brain barrier in the hippocampus and medial temporal
448 lobe contributes to *APOE* ε4-associated cognitive decline independently of Aβ or tau (62).

449 The blood provides an easily accessible tissue that can be repeatedly sampled to characterise pre-
450 morbid markers of risk. The extent to which it can provide mechanistic insights into diseases that are
451 considered predominantly brain-based, however, is a perennial subject of debate. *Cis* meQTL effects
452 tend to be highly correlated (r = 0.78) between the blood and the brain (63), supporting the use of
453 the blood to study the effects of genetic risk factors for brain-based diseases. It is also important to
454 note the increasing recognition of the role of peripheral processes in conferring risk for AD (64). As
455 the blood provides a conduit by which many circulating factors (e.g. plasma proteins and microbial
456 metabolites) reach the brain and affect brain ageing (65), assessing DNA methylation in the blood is
457 likely to be informative regarding systemic factors contributing to AD pathogenesis. Although *APOE*
458 is synthesised separately in the blood and the brain and neither *APOE* nor cholesterol can cross the
459 blood-brain barrier (66, 67), there is cross-talk between brain and blood cholesterol via oxysterols

(67), levels of which vary by *APOE* ε2/ε3/ε4 haplotype (68). Peripheral hypercholesterolemia has been associated with increased oxysterol levels in the brain, which have been implicated in with production and accumulation of Aβ, increased neuroinflammation and neuronal death (67).

The association between *APOE* genotype and AD varies between populations (5), with studies in populations of Hispanic and African ancestry often reporting attenuated effect sizes for the ε4 allele compared to studies involving European and Asian participants (69, 70). Moreover, Rajabli et al. (70) have shown that genetic variation local to *APOE* is likely to confer protection from the effects of the ε4 allele in individuals of African ancestry. As the participants in the present study were of European ancestry, it should be noted that these findings are likely to be European-specific and future studies should assess their generalisability and relevance to AD pathogenesis in other populations.

5. Conclusions

This is the first study to characterise epigenome-wide DNA methylation differences between carriers of *APOE* ε4 and ε2. In AD-free individuals, we identified several methylation differences both at the *APOE* locus and in the rest of the genome, which converge on lipid-related pathways. Strengths of the study include the large samples available for EWAS analysis, the epigenome-wide approach, the use of a well-phenotyped cohort with genotype data, and the avoidance of reverse causation by studying AD-free participants. Future studies should investigate the causal relationship between *APOE* genotype, DNA methylation and lipid-related processes and their role in AD pathogenesis.

Abbreviations

Aβ: amyloid beta; AD: Alzheimer's disease; CI: confidence interval; DMP: differentially methylated position; DMR: differentially methylated region; EWAS: epigenome-wide association study; GS:SFHS: Generation Scotland: Scottish Family Health Study; HDL: high density lipoprotein; LDL: low density lipoprotein; PC: principal component; SNP: single nucleotide polymorphism

Declarations

Ethics approval and consent to participate

GS:SFHS obtained ethical approval from the NHS Tayside Committee on Medical Research Ethics, on behalf of the National Health Service (reference: 05/S1401/89) and has Research Tissue Bank Status (reference: 15/ES/0040), providing generic ethical approval for a wide range of uses within medical research. All experimental methods were in accordance with the Helsinki declaration.

Consent for publication

Not applicable

Availability of data and materials

According to the terms of consent for GS:SFHS, access to data must be reviewed by the GS Access Committee (access@generationscotland.org).

Competing interests

AMM has received grant support from Pfizer, Eli Lilly, Janssen and The Sackler Trust. These sources are not connected to the current investigation. AMM has also received speaker fees from Janssen and Illumina. The remaining authors report no conflicts of interest.

Funding

This work was supported by a Wellcome Trust Strategic Award "STratifying Resilience and Depression Longitudinally" (STRADL) [104036/Z/14/Z] to AMM, KLE, CSH, DJP and others, and an MRC Mental Health Data Pathfinder Grant [MC_PC_17209] to AMM and DJP. REM is supported by an Alzheimer's Research UK major project grant [ARUK-PG2017B-10]. KV is funded by the Wellcome Trust Translational Neuroscience PhD Programme at the University of Edinburgh [108890/Z/15/Z].

ADB would like to acknowledge funding from the Wellcome PhD training fellowship for clinicians [204979/Z/16/Z], the Edinburgh Clinical Academic Track (ECAT) programme. Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Clinical Research Facility, Edinburgh, Scotland and was funded by the UK's Medical Research Council and the Wellcome Trust [104036/Z/14/Z]. DNA methylation profiling of the GS:SFHS samples was funded by the Wellcome Trust Strategic Award [10436/Z/14/Z] with additional funding from a 2018 NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation [27404].

The funding bodies did not play any role in the design of the study, the collection, analysis, or interpretation of the data, or in the writing of the manuscript.

Authors' contributions

Conception and design: RMW, KLE, REM; data analysis: RMW, KV, MLB, ADB, CH. Drafting the article: RMW and KLE; data preparation: RMW, MLB, SWM, KR, AC, ADB, YZ, CA; data collection: AMM, KLE, CSH, DJP. Revision of the article: RMW, KV, MLB, ADB, YZ, CA, AC, CSH, DJP, REM, KLE; all authors read and approved the final manuscript.

Acknowledgements

We are grateful to all the families who took part, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, healthcare assistants and nurses.

References

1. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet. 2013;45(12):1452-8.

2. Kunkle BW, Grenier-Boley B, Sims R, Bis JC, Damotte V, Naj AC, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. Nat Genet. 2019;51(3):414-30.

3. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat Genet. 2019;51(3):404-13.

4. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science. 1993;261(5123):921-3.

5. Farrer LA, Cupples LA, Haines JL, Hyman B, Kukull WA, Mayeux R, et al. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. JAMA. 1997;278(16):1349-56.

6. Safieh M, Korczyn AD, Michaelson DM. ApoE4: an emerging therapeutic target for Alzheimer's disease. BMC Med. 2019;17(1):64.

7. Tzioras M, Davies C, Newman A, Jackson R, Spires-Jones T. Invited Review: APOE at the interface of inflammation, neurodegeneration and pathological protein spread in Alzheimer's disease. Neuropathol Appl Neurobiol. 2019;45(4):327-46.

8. Qian J, Wolters FJ, Beiser A, Haan M, Ikram MA, Karlawish J, et al. APOE-related risk of mild cognitive impairment and dementia for prevention trials: An analysis of four cohorts. PLoS Med. 2017;14(3):e1002254.

9. Babenko VN, Afonnikov DA, Ignatieva EV, Klimov AV, Gusev FE, Rogaev EI. Haplotype analysis of APOE intragenic SNPs. BMC Neurosci. 2018;19(Suppl 1):16.

10. De Jager PL, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L, et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. Nat Neurosci. 2014;17(9):1156-63.

11. Gasparoni G, Bultmann S, Lutsik P, Kraus TFJ, Sordon S, Vlcek J, et al. DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex. Epigenetics Chromatin. 2018;11(1):41.

12. Lunnon K, Smith R, Hannon E, De Jager PL, Srivastava G, Volta M, et al. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. Nat Neurosci. 2014;17(9):1164-70.

13. Slieker RC, Relton CL, Gaunt TR, Slagboom PE, Heijmans BT. Age-related DNA methylation changes are tissue-specific with ELOVL2 promoter methylation as exception. Epigenetics Chromatin. 2018;11(1):25.

14. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature. 2017;541(7635):81-6.

564 15. Hedman AK, Mendelson MM, Marioni RE, Gustafsson S, Joehanes R, Irvin MR, et al.
565 Epigenetic Patterns in Blood Associated With Lipid Traits Predict Incident Coronary Heart Disease
566 Events and Are Enriched for Results From Genome-Wide Association Studies. Circ Cardiovasc Genet.
567 2017;10(1).

568 16. Singmann P, Shem-Tov D, Wahl S, Grallert H, Fiorito G, Shin SY, et al. Characterization of
569 whole-genome autosomal differences of DNA methylation between men and women. Epigenetics
570 Chromatin. 2015;8:43.

571 17. Galanter JM, Gignoux CR, Oh SS, Torgerson D, Pino-Yanes M, Thakur N, et al. Differential
572 methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental
573 exposures. Elife. 2017;6.

574 18. Yuan V, Price EM, Del Gobbo G, Mostafavi S, Cox B, Binder AM, et al. Accurate ethnicity
575 prediction from placental DNA methylation data. Epigenetics Chromatin. 2019;12(1):51.

576 19. Yu CE, Cudaback E, Foraker J, Thomson Z, Leong L, Lutz F, et al. Epigenetic signature and
577 enhancer activity of the human APOE gene. Hum Mol Genet. 2013;22(24):5036-47.

578 20. Foraker J, Millard SP, Leong L, Thomson Z, Chen S, Keene CD, et al. The APOE Gene is
579 Differentially Methylated in Alzheimer's Disease. J Alzheimers Dis. 2015;48(3):745-55.

580 21. Ma Y, Smith CE, Lai CQ, Irvin MR, Parnell LD, Lee YC, et al. Genetic variants modify the effect
581 of age on APOE methylation in the Genetics of Lipid Lowering Drugs and Diet Network study. Aging
582 Cell. 2015;14(1):49-59.

583 22. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. Cohort Profile:
584 Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their
585 potential for genetic research on health and illness. Int J Epidemiol. 2013;42(3):689-700.

586 23. Smith BH, Campbell H, Blackwood D, Connell J, Connor M, Deary IJ, et al. Generation
587 Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability.
588 BMC medical genetics. 2006;7:74.

589 24. Amador C, Huffman J, Trochet H, Campbell A, Porteous D, Generation S, et al. Recent
590 genomic heritage in Scotland. BMC Genomics. 2015;16:437.

591 25. Kerr SM, Campbell A, Murphy L, Hayward C, Jackson C, Wain LV, et al. Pedigree and
592 genotyping quality analyses of over 10,000 DNA samples from the Generation Scotland: Scottish
593 Family Health Study. BMC medical genetics. 2013;14:38.

594 26. Barbu MC, Walker RM, Howard DM, Evans KL, Whalley HC, Porteous DJ, et al. Epigenetic
595 prediction of major depressive disorder. medRxiv. 2019:19001123.

596 27. Madden RA, McCartney DL, Walker RM, Hillary RF, Bermingham ML, Rawlik K, et al. Birth
597 weight predicts psychiatric and physical health, cognitive function, and DNA methylation differences
598 in an adult population. bioRxiv. 2019:664045.

599 28. Bermingham ML, Walker RM, Marioni RE, Morris SW, Rawlik K, Zeng Y, et al. Identification of
600 novel differentially methylated sites with potential as clinical predictors of impaired respiratory
601 function and COPD. EBioMedicine. 2019;43:576-86.

602 29. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R
603 Foundation for Statistical Computing; 2019.

30. Team RC. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2020.

31. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.

32. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88(1):76-82.

33. Saffari A, Silver MJ, Zavattari P, Moi L, Columbano A, Meaburn EL, et al. Estimation of a significance threshold for epigenome-wide association studies. Genet Epidemiol. 2018;42(1):20-33.

34. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-1.

35. Lenth RV. Least-Squares Means: The R Package lsmeans. Journal of Statistical Software. 2016;69(1):1-33.

36. Suderman M, Staley JR, French R, Arathimos R, Simpkin A, Tilling K. dmrff: identifying differentially methylated regions efficiently with power and control. bioRxiv. 2018:508556.

37. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics. 2016;32(2):286-8.

38. Tingley D, Yamamoto T, Hirose K, Keele L, Imai K. mediation: R Package for Causal Mediation Analysis. Journal of Statistical Software. 2014;59(5).

39. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48(10):1279-83.

40. Nagy R, Boutin TS, Marten J, Huffman JE, Kerr SM, Campbell A, et al. Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 Generation Scotland participants. Genome Med. 2017;9(1):23.

41. Zeng Y, Amador C, Xia C, Marioni R, Sproul D, Walker RM, et al. Parent of origin genetic effects on methylation in humans are common and influence complex trait variation. Nat Commun. 2019;10(1):1383.

42. Haller T, Kals M, Esko T, Magi R, Fischer K. RegScan: a GWAS tool for quick estimation of allele effects on continuous traits and their combinations. Brief Bioinform. 2015;16(1):39-44.

43. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;47(D1):D1005-D12.

44. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjalmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet. 2015;47(3):284-90.

45. Maas SCE, Vidaki A, Wilson R, Teumer A, Liu F, van Meurs JBJ, et al. Validated inference of smoking habits from blood with a finite DNA methylation marker set. Eur J Epidemiol. 2019;34(11):1055-74.

641 46. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific
642 haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics.
643 2015;31(21):3555-7.

644 47. He L, Kernogitski Y, Kulminskaya I, Loika Y, Arbeev KG, Loiko E, et al. Pleiotropic Meta-
645 Analyses of Longitudinal Studies Discover Novel Genetic Variants Associated with Age-Related
646 Diseases. Front Genet. 2016;7:179.

647 48. Mur J, McCartney DL, Walker RM, Campbell A, Bermingham ML, Morris SW, et al. DNA
648 methylation in the APOE gene: its link with Alzheimer's and cardiovascular health. bioRxiv.
649 2019:811224.

650 49. Zhou X, Chen Y, Mok KY, Kwok TCY, Mok VCT, Guo Q, et al. Non-coding variability at the
651 APOE locus contributes to the Alzheimer's risk. Nat Commun. 2019;10(1):3310.

652 50. Burgess BL, Parkinson PF, Racke MM, Hirsch-Reinshagen V, Fan J, Wong C, et al. ABCG1
653 influences the brain cholesterol biosynthetic pathway but does not affect amyloid precursor protein
654 or apolipoprotein E metabolism in vivo. J Lipid Res. 2008;49(6):1254-67.

655 51. Greeve I, Hermans-Borgmeyer I, Brellinger C, Kasper D, Gomez-Isla T, Behl C, et al. The
656 human DIMINUTO/DWARF1 homolog seladin-1 confers resistance to Alzheimer's disease-associated
657 neurodegeneration and oxidative stress. J Neurosci. 2000;20(19):7345-52.

658 52. Martiskainen H, Paldanius KMA, Natunen T, Takalo M, Marttinen M, Leskela S, et al. DHCR24
659 exerts neuroprotection upon inflammation-induced neuronal death. J Neuroinflammation.
660 2017;14(1):215.

661 53. Crameri A, Biondi E, Kuehnle K, Lutjohann D, Thelen KM, Perga S, et al. The role of seladin-
662 1/DHCR24 in cholesterol biosynthesis, APP processing and Abeta generation in vivo. EMBO J.
663 2006;25(2):432-43.

664 54. Picard C, Julien C, Frappier J, Miron J, Theroux L, Dea D, et al. Alterations in cholesterol
665 metabolism-related genes in sporadic Alzheimer's disease. Neurobiol Aging. 2018;66:180 e1- e9.

666 55. Braun KVE, Dhana K, de Vries PS, Voortman T, van Meurs JBJ, Uitterlinden AG, et al.
667 Epigenome-wide association study (EWAS) on lipids: the Rotterdam Study. Clin Epigenetics.
668 2017;9:15.

669 56. Pfeiffer L, Wahl S, Pilling LC, Reischl E, Sandling JK, Kunze S, et al. DNA methylation of lipid-
670 related genes affects blood lipid levels. Circ Cardiovasc Genet. 2015;8(2):334-42.

671 57. Sayols-Baixeras S, Subirana I, Lluis-Ganella C, Civeira F, Roquer J, Do AN, et al. Identification
672 and validation of seven new loci showing differential DNA methylation related to serum lipid profile:
673 an epigenome-wide approach. The REGICOR study. Hum Mol Genet. 2016;25(20):4556-65.

674 58. Tan CE, Tai ES, Tan CS, Chia KS, Lee J, Chew SK, et al. APOE polymorphism and lipid profile in
675 three ethnic groups in the Singapore population. Atherosclerosis. 2003;170(2):253-60.

676 59. Dekkers KF, van Iterson M, Slieker RC, Moed MH, Bonder MJ, van Galen M, et al. Blood lipids
677 influence DNA methylation in circulating cells. Genome Biol. 2016;17(1):138.

678 60. Sayols-Baixeras S, Tiwari HK, Aslibekyan SW. Disentangling associations between DNA
679 methylation and blood lipids: a Mendelian randomization approach. BMC Proc. 2018;12(Suppl 9):23.

680 61. Smith RG, Pishva E, Shireby G, Smith AR, Roubroeks JAY, Hannon E, et al. Meta-analysis of
681 epigenome-wide association studies in Alzheimer's disease highlights 220 differentially methylated
682 loci across cortex. bioRxiv. 2020:2020.02.28.957894.

683 62. Montagne A, Nation DA, Sagare AP, Barisano G, Sweeney MD, Chakhoyan A, et al. APOE4
684 leads to blood-brain barrier dysfunction predicting cognitive decline. Nature. 2020;581(7806):71-6.

685 63. Qi T, Wu Y, Zeng J, Zhang F, Xue A, Jiang L, et al. Identifying gene targets for brain-related
686 traits using transcriptomic and methylomic data from blood. Nat Commun. 2018;9(1):2282.

687 64. Morris G, Berk M, Maes M, Puri BK. Could Alzheimer's Disease Originate in the Periphery and
688 If So How So? Mol Neurobiol. 2019;56(1):406-34.

689 65. Pluvinage JV, Wyss-Coray T. Systemic factors as mediators of brain homeostasis, ageing and
690 neurodegeneration. Nat Rev Neurosci. 2020.

691 66. Liu M, Kuhel DG, Shen L, Hui DY, Woods SC. Apolipoprotein E does not cross the blood-
692 cerebrospinal fluid barrier, as revealed by an improved technique for sampling CSF from mice. Am J
693 Physiol Regul Integr Comp Physiol. 2012;303(9):R903-8.

694 67. Gamba P, Testa G, Gargiulo S, Staurenghi E, Poli G, Leonarduzzi G. Oxidized cholesterol as
695 the driving force behind the development of Alzheimer's disease. Front Aging Neurosci. 2015;7:119.

696 68. Jenner AM, Lim WL, Ng MP, Wenk MR, Shui G, Sharman MJ, et al. The effect of APOE
697 genotype on brain levels of oxysterols in young and old human APOE epsilon2, epsilon3 and epsilon4
698 knock-in mice. Neuroscience. 2010;169(1):109-15.

699 69. Blue EE, Horimoto A, Mukherjee S, Wijsman EM, Thornton TA. Local ancestry at APOE
700 modifies Alzheimer's disease risk in Caribbean Hispanics. Alzheimers Dement. 2019;15(12):1524-32.

701 70. Rajabli F, Feliciano BE, Celis K, Hamilton-Nelson KL, Whitehead PL, Adams LD, et al. Ancestral
702 origin of ApoE epsilon4 Alzheimer disease risk in Puerto Rican and African American populations.
703 PLoS Genet. 2018;14(12):e1007791.

704 **Table 1.** *APOE* ε4 vs. ε2-associated DMPs identified by meta-analysis of the discovery and replication

705 EWASs

Probe ID	Gene symbol	Gene feature*	Chr.	BP†	Effect‡	SE	*P* value
cg13375295			19	45344725	-0.1031	0.0049	2.80×10^{-100}
cg06750524	*APOE*	Body	19	45409955	0.1122	0.008	1.05×10^{-44}
cg16094954	*BCL3*	TSS1500	19	45251180	-0.0994	0.0081	8.18×10^{-35}
cg10762466			19	45347693	-0.0463	0.004	1.37×10^{-30}
cg16471933	*APOE*	Body	19	45411802	0.0606	0.0055	7.17×10^{-28}
cg10178308	*PVRL2*	TSS200	19	45349383	0.1075	0.0103	2.04×10^{-25}
cg27087650	*BCL3*	Body	19	45255796	0.0455	0.0044	3.77×10^{-25}
cg04488858			19	45242346	-0.0514	0.0065	2.25×10^{-15}
cg11643040	*PVRL2*	Body	19	45361327	-0.0278	0.0038	1.46×10^{-13}
cg26631131			19	45240591	0.0298	0.0042	2.45×10^{-12}
cg17901584	*DHCR24;RP11-67L3.4*	TSS1500	1	55353706	-0.0403	0.0058	3.58×10^{-12}
cg06198803	*PVRL2*	Body	19	45371896	-0.041	0.006	1.04×10^{-11}
cg16740586	*ABCG1*	Body	21	43655919	0.0332	0.005	3.58×10^{-11}
cg03793277	*APOC1*	TSS1500	19	45416910	-0.0304	0.0049	5.99×10^{-10}
cg06500161	*ABCG1*	Body	21	43656587	0.0247	0.0042	2.67×10^{-9}
cg09555818	*APOC2;APOC4*	5' UTR; 1st exon	19	45449301	-0.0531	0.0091	5.77×10^{-9}
cg13119609	*APOC2;APOC4*	5' UTR; 1st exon	19	45449297	-0.0464	0.008	5.84×10^{-9}

cg15233575			19	45221584	-0.0223	0.0039	7.17×10^{-9}
cg14645843			19	45454752	-0.0346	0.0062	2.31×10^{-8}
cg19751789	*LDLR*		19	11199944	-0.0338	0.0061	2.43×10^{-8}

706

707 Abbreviations: BP, base position; Chr., chromosome; SE, standard error; TSS, transcription start site;

708 UTR, untranslated region

709 *Gene feature: 5' UTR: between the TSS and the ATG; Body: between the ATG and the stop codon;

710 TSS200: within 200 bases 5' of the TSS; TSS1500: within 1500 bases 5' of the TSS.

711 †Base position in genome assembly hg19/GRCh37

712 ‡Effect direction is relative to carriers of the ε2 allele

713 **Table 2.** Significant DMRs identified through DMR meta-analysis of the discovery and replication

714 sample EWAS results

Chr.	Coordinates*	Gene symbol	Effect†	SE	Adj. *P* value‡	CpGs
19	45449297-45449301	*APOC2;APOC4*	-0.231	0.0364	1.63×10^{-4}	cg13119609; cg09555818
19	45449099-45449150	*APOC4-APOC2;APOC2;APOC4*	-0.212	0.0356	0.00203	cg01958934; cg10872931
19	11199851-11199903	*LDLR*	-0.135	0.0245	0.0290	cg07960944; cg05249393; cg22381454;
22	42230879-42230899	*SREBF2*	-0.189	0.0329	0.00755	cg15128785; cg12403973
22	42229983-42230138	*SREBF2*	-0.176	0.0312	0.0118	cg09978077; cg16000331
19	45429771-45429870	*APOC1P1*	0.148	0.0269	0.0301	cg23184690; cg08121984

715

716 Abbreviations: Chr., chromosome; SE, standard error; Adj., adjusted; CpGs, cytosine and guanine

717 nucleotides linked by a phosphate bond

718 *DMR start and end coordinates in genome assembly hg19/GRCh37

719 †Effect direction is relative to carriers of the ε2 allele

720 ‡Bonferroni-adjusted *P* value

721 **Table 3.** Top 20 GO terms showing significant enrichment for *APOE* ε4 vs. ε2-associated differentially

722 methylated loci

Ontology category	Term	Proportion *	*P*-value	Genes
BP	cholesterol metabolic process	7/146	2.00×10^{-11}	*DHCR24, APOC1, APOE, LDLR, SQLE, SREBF2, ABCG1*
BP	secondary alcohol metabolic process	7/156	3.59×10^{-11}	*DHCR24, APOC1, APOE, LDLR, SQLE, SREBF2, ABCG1*
BP	chylomicron remnant clearance	4/9	4.29×10^{-11}	*APOC1, APOC2, APOE, LDLR*
BP	triglyceride-rich lipoprotein particle clearance	4/9	4.29×10^{-11}	*APOC1, APOC2, APOE, LDLR*
BP	sterol metabolic process	7/162	4.78×10^{-11}	*DHCR24, APOC1, APOE, LDLR, SQLE, SREBF2, ABCG1*
BP	cholesterol homeostasis	6/94	7.57×10^{-11}	*MYLIP, APOC2, APOE, LDLR, SREBF2, ABCG1*
BP	sterol homeostasis	6/94	7.57×10^{-11}	*MYLIP, APOC2, APOE, LDLR, SREBF2, ABCG1*
BP	cholesterol transport	6/98	1.04×10^{-10}	*APOC1, APOC2, APOE, LDLR, SREBF2, ABCG1*
BP	regulation of plasma lipoprotein particle levels	6/100	1.17×10^{-10}	*MYLIP, APOC1, APOC2, APOE, LDLR, ABCG1*
BP	phospholipid transport	6/90	1.27×10^{-10}	*APOC1, APOC2, APOE, KCNN4, LDLR, ABCG1*
BP	regulation of lipid	7/190	$1.50 \times 10^{-}$	*APOC1, APOC2, APOE,*

	biosynthetic process		10^{-10}	*LDLR, SQLE, SREBF2, ABCG1*
BP	phospholipid efflux	4/12	1.50×10^{-10}	*APOC1, APOC2, APOE, ABCG1*
BP	sterol transport	6/111	2.65×10^{-10}	*APOC1, APOC2, APOE, LDLR, SREBF2, ABCG1*
BP	cholesterol efflux	5/56	4.60×10^{-10}	*APOC1, APOC2, APOE, SREBF2, ABCG1*
BP	organophosphate ester transport	6/116	6.18×10^{-10}	*APOC1, APOC2, APOE, KCNN4, LDLR, ABCG1*
BP	regulation of cholesterol transport	5/61	6.55×10^{-10}	*APOC1, APOC2, APOE, SREBF2, ABCG1*
BP	regulation of steroid metabolic process	6/123	6.67×10^{-10}	*APOC1, APOE, LDLR, SQLE, SREBF2, ABCG1*
BP	regulation of sterol transport	5/62	7.34×10^{-10}	*APOC1, APOC2, APOE, SREBF2, ABCG1*
BP	lipid localization	8/407	7.54×10^{-10}	*APOC1, APOC2, APOE, KCNN4, LDLR, SQLE, SREBF2, ABCG1*
BP	high-density lipoprotein particle remodeling	4/18	1.14×10^{-9}	*APOC1, APOC2, APOE, ABCG1*

723

724 Abbreviations: BP, biological process; CC, cellular component; GO, gene ontology; MF, molecular

725 function

726 *Number of significant target list-associated Entrez IDs associated with the gene ontology term /total

727 number of Entrez IDs associated with the GO term. The target list comprised probes that met a

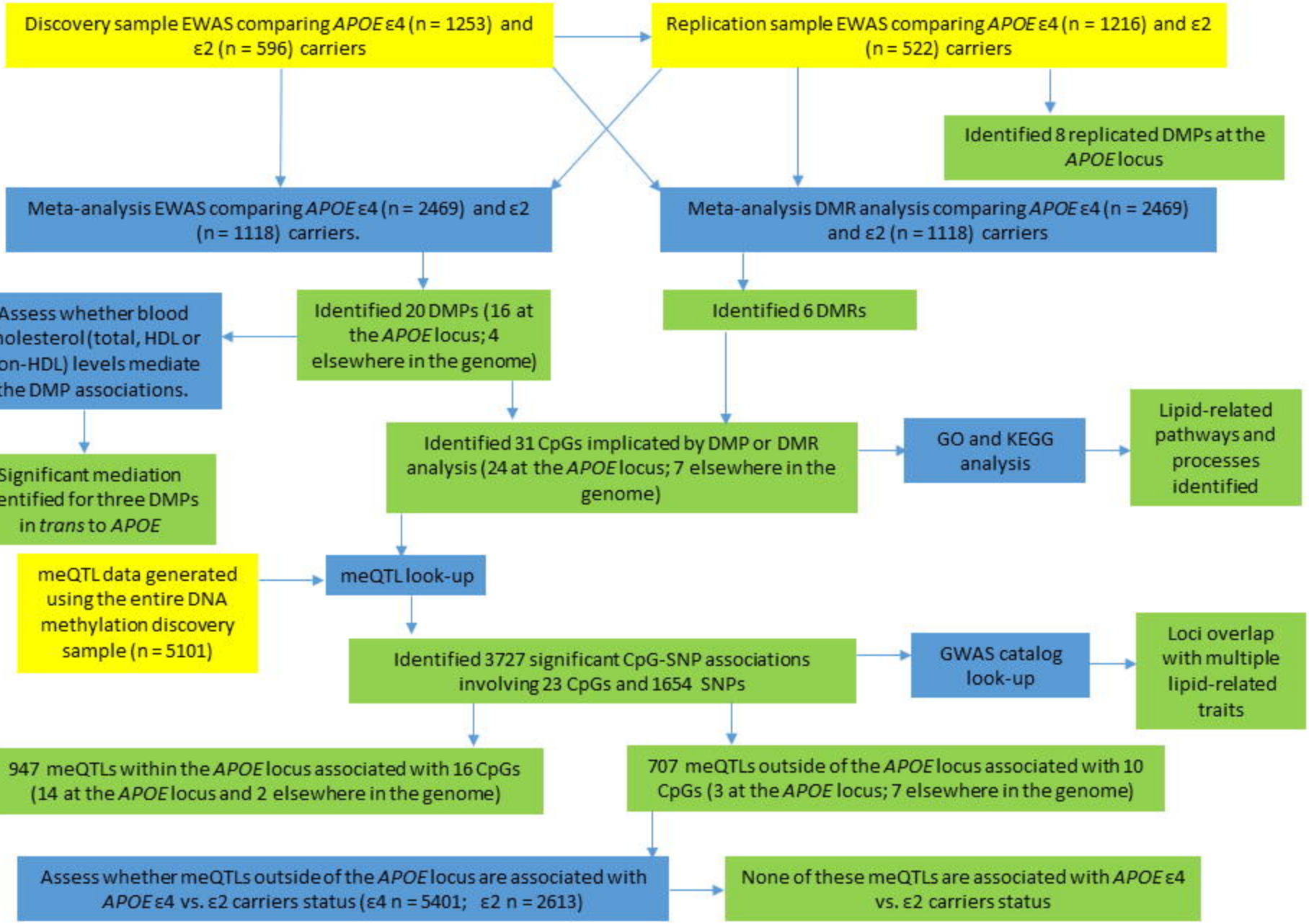
728 nominal threshold for association with *APOE* ε4 vs. ε2 carrier status of $P \leq 1 \times 10^{-5}$.

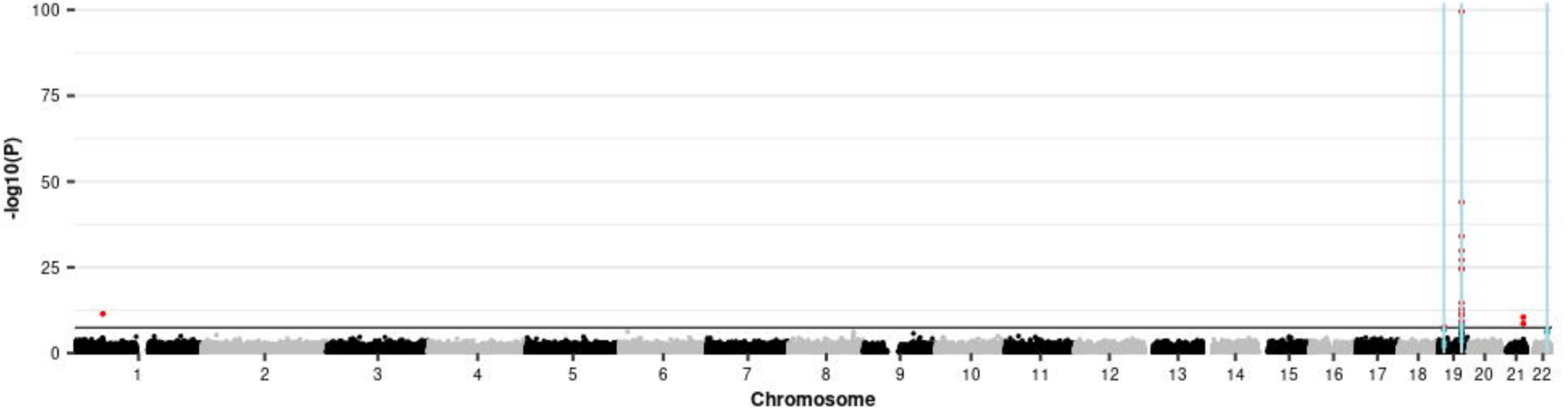
Figure legends

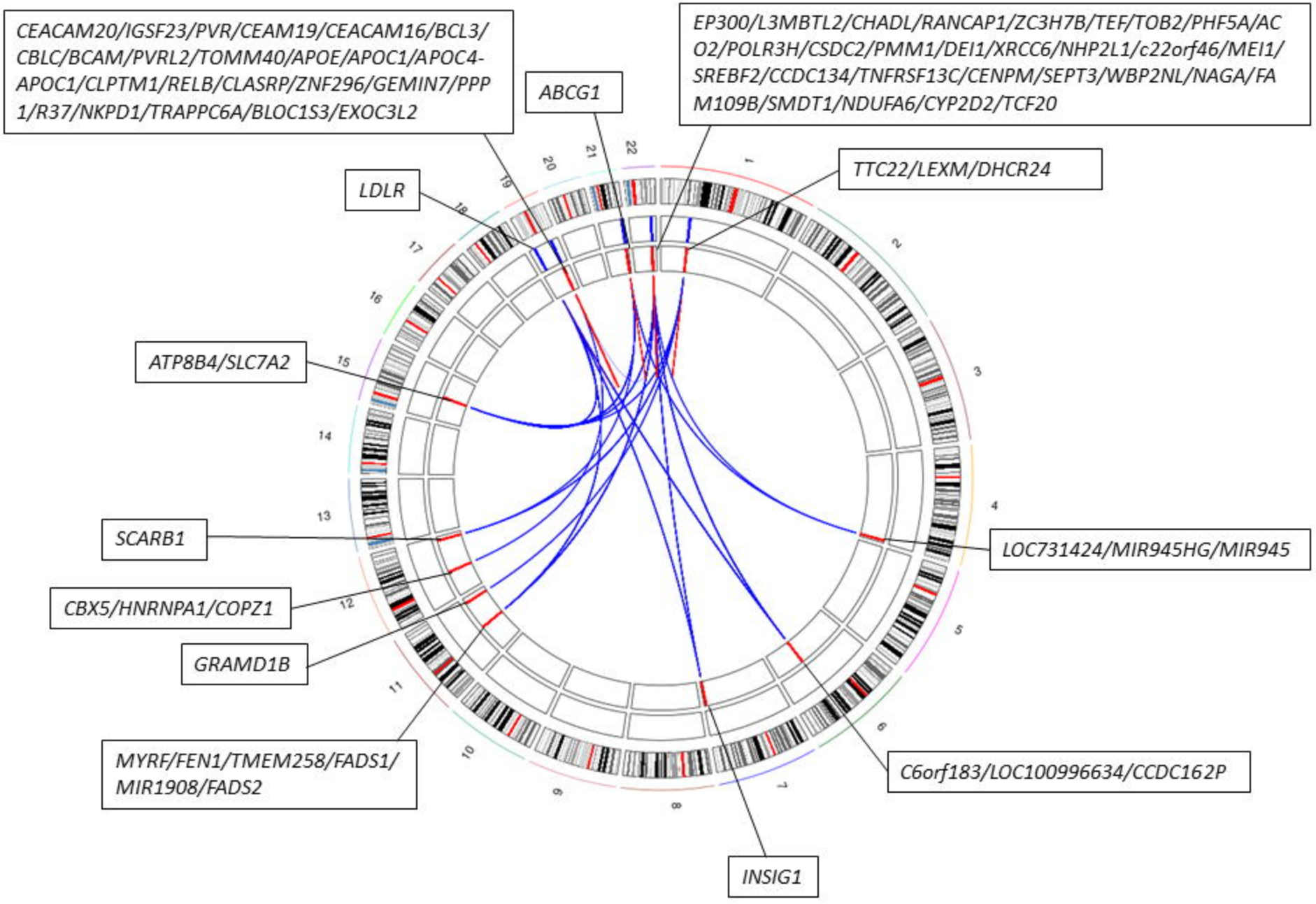
Figure 1. Flow chart indicating the analyses carried out in this study. Yellow boxes indicate datasets used for the analysis, blue boxes describe the analysis performed and green boxes contain the results of the analysis. Arrows indicate for which analyses the datasets were used, the order of the analyses and the results from each analysis.

Figure 2. Manhattan plot showing the *APOE* ε4 vs. ε2 carrier EWAS and DMR meta-analyses results. Each point represents one of the 772,453 loci included in the EWAS meta-analysis, with the point's position being determined by genomic position (x-axis) and significance in the EWAS meta-analysis ($-\log_{10}$ *P* value; y-axis). Sites attaining genome-wide significance ($P \leq 3.6 \times 10^{-8}$) are indicated in red and those that are involved in a significant DMR (Bonferroni-correct $P \leq 0.05$) are indicated in blue. The locations of DMRs are further indicated by vertical blue lines. The solid horizontal line is the threshold for genome-wide significance ($P \leq 3.6 \times 10^{-8}$).

Figure 3. Circular plot indicating the locations of *APOE* ε4 vs. ε2 carrier-associated DMP and DMR CpGs. The first track shows a chromosome ideogram (hg19/GRCh37). The genomic locations of CpGs identified as being DMPs or in DMRs identified in *APOE* ε4 vs. ε2 carriers are indicated by blue lines on the second track and the meQTLs associated with these CpGs are indicated by the red lines on the third track. The connections between CpGs and meQTLs indicate regulatory relationships (*cis* interactions in red; *trans* interactions in blue). Gene symbols for genes located in each CpG/meQTL-harbouring region are indicated.


Discovery sample EWAS comparing *APOE* ε4 (n = 1253) and ε2 (n = 596) carriers
Replication sample EWAS comparing *APOE* ε4 (n = 1216) and ε2 (n = 522) carriers
Identified 8 replicated DMPs at the *APOE* locus
Meta-analysis EWAS comparing *APOE* ε4 (n = 2469) and ε2 (n = 1118) carriers.
Meta-analysis DMR analysis comparing *APOE* ε4 (n = 2469) and ε2 (n = 1118) carriers
Assess whether blood cholesterol (total, HDL or non-HDL) levels mediate the DMP associations.
Identified 20 DMPs (16 at the *APOE* locus; 4 elsewhere in the genome)
Identified 6 DMRs
Significant mediation identified for three DMPs in *trans* to *APOE*
Identified 31 CpGs implicated by DMP or DMR analysis (24 at the *APOE* locus; 7 elsewhere in the genome)
GO and KEGG analysis
Lipid-related pathways and processes identified
meQTL data generated using the entire DNA methylation discovery sample (n = 5101)
meQTL look-up
Identified 3727 significant CpG-SNP associations involving 23 CpGs and 1654 SNPs
GWAS catalog look-up
Loci overlap with multiple lipid-related traits
947 meQTLs within the *APOE* locus associated with 16 CpGs (14 at the *APOE* locus and 2 elsewhere in the genome)
707 meQTLs outside of the *APOE* locus associated with 10 CpGs (3 at the *APOE* locus; 7 elsewhere in the genome)
Assess whether meQTLs outside of the *APOE* locus are associated with *APOE* ε4 vs. ε2 carriers status (ε4 n = 5401; ε2 n = 2613)
None of these meQTLs are associated with *APOE* ε4 vs. ε2 carriers status


100
75
50
25
0
-log10(P)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome


CEACAM20/IGSF23/PVR/CEAM19/CEACAM16/BCL3/CBLC/BCAM/PVRL2/TOMM40/APOE/APOC1/APOC4-APOC1/CLPTM1/RELB/CLASRP/ZNF296/GEMIN7/PPP1/R37/NKPD1/TRAPPC6A/BLOC1S3/EXOC3L2
ABCG1
EP300/L3MBTL2/CHADL/RANCAP1/ZC3H7B/TEF/TOB2/PHF5A/ACO2/POLR3H/CSDC2/PMM1/DEI1/XRCC6/NHP2L1/c22orf46/MEI1/SREBF2/CCDC134/TNFRSF13C/CENPM/SEPT3/WBP2NL/NAGA/FAM109B/SMDT1/NDUFA6/CYP2D2/TCF20
TTC22/LEXM/DHCR24
LDLR
ATP8B4/SLC7A2
SCARB1
LOC731424/MIR945HG/MIR945
CBX5/HNRNPA1/COPZ1
GRAMD1B
MYRF/FEN1/TMEM258/FADS1/MIR1908/FADS2
C6orf183/LOC100996634/CCDC162P
INSIG1