Published in final edited form as:

Alzheimers Dement. 2020 December ; 16(Suppl 2): . doi:10.1002/alz.043533.

EPIGENOMIC FEATURES RELATED TO MICROGLIA ARE ASSOCIATED WITH ATTENUATED EFFECT OF APOE ε4 ON ALZHEIMER'S DISEASE RISK IN HUMANS

Yiyi Ma[1], **Lei Yu**[2,3], **Marta Olah**[1], **Rebecca Smith**[4], **Stephanie R. Oatman**[5], **Mariet Allen**[5], **Ehsan Pishva**[4], **Bin Zhang**[6,7], **Vilas Menon**[1], **Nilüfer Ertekin-Taner**[5,8], **Katie Lunnon**[4], **David A. Bennett**[2,3], **Hans-Ulrich Klein**[1], **Philip L. De Jager**[1,9]

[1]Center for Translational & Computational Neuroimmunology, Department of Neurology, Columbia University Medical Center, 630 West 168th street, New York, NY, USA

[2]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA

[3]Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA

[4]University of Exeter Medical School, College of Medicine and Health, Exeter University, Exeter, UK

[5]Mayo Clinic Florida, Department of Neuroscience, Jacksonville, FL 32224, USA

[6]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA

[7]Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA

[8]Mayo Clinic Florida, Department of Neurology, Jacksonville, FL 32224, USA

[9]Cell Circuits Program, Broad Institute, 415 Main street, Cambridge MA, USA

Abstract

Not all *APOE* ε4 carriers who survive to advanced age develop Alzheimer's disease (AD); factors attenuating the risk of ε4 on AD may exist. Guided by the top ε4-attenuating signals from methylome-wide association analyses (N=572, ε4+ and ε4−) of neurofibrillary tangles and neuritic plaques, we conducted a meta-analysis for pathological AD within the ε4+ subgroups (N=235) across four independent collections of brains. Cortical RNA-seq and microglial morphology measurements were used in functional analyses. Three out of the four significant CpG dinucleotides were captured by one principle component (PC1), which interacts with ε4 on AD, and is associated with expression of innate immune genes and activated microglia. In ε4 carriers, reduction in each unit of PC1 attenuated the odds of AD by 58% (OR=2.39, 95%CI=[1.64,3.46],

*Corresponding author: Philip L. De Jager, M.D., Ph.D., Director of the Center for Translational & Computational Neuroimmunology, Department of Neurology, Columbia University Medical Center, 630 West 168th street, New York, NY 10032, USA; Tel: (212) 305-3609; pld2115@cumc.columbia.edu.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

P=7.08x10^{-6}). An epigenomic factor associated with a reduced proportion of activated microglia (microglial epigenomic factor 1) appears to attenuate the risk of ε4 on AD.

Keywords

APOE; Alzheimer's disease; epigenome; microglia; Neurology

1 | NARRATIVE

1.1 | Contextual background

The *APOE* ε4 haplotype contributes the greatest common genetic risk for Alzheimer's disease (AD)[1–4]. However, not all ε4 carriers develop AD. A longitudinal observational study reported that 9 out of the 141 ε4/ε4 individuals remained dementia-free after age 84[5]. Another meta-analysis of cross-sectional studies suggested that the ε4 effect on onset of AD becomes weaker after age 70[1]. These results indicate that this risk haplotype is not fully penetrant and that factors attenuating the genetic risk of ε4 on AD might exist.

Identifying such an *APOE* ε4 attenuator would have significant impacts on both therapeutic development and public health. Genome editing techniques bring us the possibility of editing risk-associated alleles from an individual's genome, but the application of such techniques are still limited to *in vitro* cell cultures. Thus, it is still impossible to change the genotype of *APOE* ε4 in living humans, yielding an ongoing ethical debate relating to whether the physician should release the *APOE* ε4 haplotype information to patients. However, the existence of ε4 attenuators could give us novel and meaningful insights into the development of new, targeted treatments for AD and would enable personalized treatment of ε4 carriers.

Such *APOE* ε4 haplotype attenuators could be genetic – i.e. a genetic variant residing elsewhere in the genome that opposes the effect of ε4 – or it could be environmental, a change in cellular function from an external influence that blocks the functional consequences of ε4. Here, we focused on DNA methylation which can be altered by environmental exposures or life experiences; methylation of selected DNA sequences in the nucleus of cells is a marker for a change in the three-dimensional architecture of DNA in the vicinity of the methylated site. Such conformational changes can be associated with altered patterns of RNA transcription from the DNA[6]. Thus, DNA methylation is considered to be part of the epigenome, the three-dimensional architecture of the genome which can influence the activity of the genomic sequence. DNA methylation is one of the four epigenomic modifications which have the most direct contact with the DNA sequence because a methyl-group is added on top of the DNA nucleotide. In humans, 99% of the DNA methylation occur on cytosine ("C") nucleotide that is next to a guanine ("G") nucleotide, forming a "CpG" dinucleotide in the DNA sequence. Unlike the unmodifiable nature of DNA nucleotide, the DNA methylation can be considered as a "modifiable" feature since the methyl-group can be either added or removed with the help of an enzyme called DNA methyltransferase or other enzymes. The main focus of our study is to identify those modifiable CpG dinucleotides that may act as an *APOE* ε4 attenuator. We report the

discovery of several locations in the human genome that contain such CpG dinucleotides, sites where a change in DNA methylation is associated with a reduced effect of the *APOE* ε4 haplotype on risk of AD.

1.2 | Study conclusions and disease implications

Our study explored the epigenome of the human brain for CpG dinucleotides that attenuate the impact of the strongest but not deterministic[5] genetic risk of AD in the general population: the *APOE* ε4 haplotype. Our staged approach identified 4 CpG dinucleotides on different chromosomes whose level of DNA methylation is associated with AD susceptibility in *APOE* ε4+ individuals; more specifically, it appears that the burden of tau pathology is associated with DNA methylation at these 4 sites. While the 4 CpG dinucleotides are found on different chromosomes, three of them have methylation levels that correlate with one another and represent three independent measurements of the same biological process, so we can collapse them into a single measure, a single principal component (PC1), that is available for each participant. As PC1 gets larger, the risk of AD increases. In downstream analyses, we found that increasing PC1 is also associated with an enlarged proportion of activated microglia in the cortex and that the effect of PC1 is greater in ε4+ than in ε4− participants. Our modeling revealed that each reduction of PC1 in ε4 carriers by 1 unit attenuated risk of AD by 58%, suggesting that we may have found a meaningful ε4 attenuator. Further work is now needed to further characterize the microglia-related biological process that is captured by PC1 to evaluate whether its effect could potentially be mimicked by a therapeutic agent. Given its functional associations, we rename PC1 as “microglial epigenomic factor 1”.

To perform this study, we took advantage of the large and varied sets of molecular and histological data available from the same region of frontal cortex of participants from two longitudinal studies of aging with prospective autopsy: the Religious Order Study (ROS) and the Memory & Aging Project (MAP) which were designed by a single team of investigator to be analyzed together [7]. We therefore refer to them as ROSMAP. We repurposed DNA methylation profiles that we had generated in an earlier study [8] to conduct our staged screen of the epigenome for *APOE* ε4 attenuators. We were able to validate associations with DNA methylation levels at 4 CpG dinucleotides using data generated from independent collections of brain samples, and we then began to characterize the biology captured by our 4 CpG dinucleotides. Our initial unbiased analysis of the cortical transcriptome for RNA transcripts that are associated with these CpG’s clearly pointed towards the innate immune system and particularly myeloid cells as being the source of the altered CpG methylation in cortical tissue. This lead us to focus our attention on this cell type as we dissected the effect of the 4 CpGs further. Namely, they seem to involve effects on two groups of co-expressed neocortical gene transcripts: module 116 which relates to microglial aging and module 5 which is associated with the proportion of activated microglia and contributes to the accumulation of Tau pathology[9]. We previously reported that the proportion of the activated microglia was an independent risk factor for AD that had an effect size comparable to that of ε4 on risk of AD[10]. Our current findings with the same dataset therefore refines our understanding of the relationship between activated microglia and the ε4 allele: while their effects on risk may be largely independent, they may interact to some degree.

Our findings that the top 4 CpG dinucleotides were primarily associated with Tau-related pathologies are in line with previous genetic studies revealing the link between *APOE* and *MAPT*, the gene encoding the Tau protein. A *MAPT* variant enriched in *MAPT* H2 carriers conferred greater protection from AD in *APOE* ε4− individuals[11,12]. Further, a recent genome-wide association study stratified by the *MAPT* haplotype identified an AD susceptibility variant that is enriched in *APOE* ε3 carriers and is protective in individuals with a *MAPT* H1H1 genotype[13]. These findings alongside ours suggest that variation in the products of the *APOE* or *MAPT* loci may interact with one another in different ways to modify AD risk and that this interaction may involve the function of microglial cells. Further, these context-specific factors could establish a link between the *APOE* (a major risk factor for AD) and *MAPT* (major risk factor for other tauopathies but, at best, a weak risk factor for AD) loci.

Our study has several limitations. The sample size is limited even though we have assembled the largest available ε4+ study with both DNA methylation and neuropathology data. Having low power is a common issue for all ε4+ subgroup studies[12]. Trying to circumvent this issue, we utilized a staged approach followed by RNA- and histology-based validation. The inclusion of the same ROSMAP subjects in both stage I and stage II may inflate our results, although we use different traits in each stage and impose a 10 fold more stringent significance threshold to address, in part, this issue. Finally, our cross-sectional design using data from samples collected at autopsy cannot determine the causality of the CpG dinucleotides, which is a general limitation of all postmortem autopsy studies for which we can only have one time point.

Overall, we report that the deleterious effect of the strongest genetic risk (*APOE* ε4) for AD may be attenuated by an epigenomic factor which could work, at least in part, through alterations in the relative proportion of activated microglia (as defined morphologically) and, subsequently, on the accumulation of Tau pathology. This suggests that reduction of microglial activity could be a particularly effective intervention in *APOE* ε4+ individuals who already have evidence of amyloid pathology on imaging or cerebrospinal fluid, and it yields hypotheses that now need to be tested mechanistically to further validate our results and support the proposed sequence of events in which activation of microglia has a strong effect on the accumulation of tau pathology in *APOE* ε4+ individuals.

2 | CONSOLIDATED RESULTS AND STUDY DESIGN

Conducting a DNA methylation genome-wide association study (MWAS) of AD risk within the subgroup of individuals who are ε4+ can provide an unbiased genome-wide scale search for those unknown signals protecting ε4 carriers from having AD. However, such an analysis is very limited in statistical power, as illustrated by our recent whole exome sequencing study with over 3,000 ε4+ subjects[12]. Therefore, we designed a three-stage approach (Figure 1).

In stage I, we took advantage of an existing DNA methylation dataset[8] generated from a random subset of ROSMAP participants. This includes both ε4+ and ε4− individuals with detailed quantitative measures of AD neuropathology to maximize power in performing

an initial MWAS to prioritize a list of CpG dinucleotides which had the potential to be an ε4 attenuator. The resulting prioritized CpG dinucleotides fulfilled two criteria: (1) be associated with AD pathology in all subjects; and (2) reduce the effect of ε4 on AD susceptibility. The latter is determined by comparing the regression coefficients for the ε4 variable before and after adjusting for each CpG dinucleotide and focusing on those CpG dinucleotides which reduce the unadjusted ε4 regression coefficient.

The prioritized list of 25 CpG dinucleotides from stage I were then moved into stage II, where we checked their associations with AD in the ε4+ subgroup. We then assessed for evidence of replication in independent cohorts, and, to increase the statistical power, we conducted a meta-analysis in stage II to combine data from 4 independent cohorts and generate a final summary statistic. At the conclusion of the meta-analysis, we identified 4 CpG sites which might act as the attenuators to the ε4 risk effect on AD; they are cg08706567 at *MPL*, cg26884773 at *TOMM20*, cg12307200 between *LPP* and *TPRG1*, and cg05157625 at *RIN3*. Furthermore, we found that the DNA methylation patterns of these 4 CpG sites and their mutual correlations are similar across different cohorts. For example, the methylation level of the 3 of the CpG dinucleotides - cg08706567, cg26884773, and cg05157625 - are correlated ($r \geq 0.4$), and all three are hypo-methylated levels (<50% methylation) across all cohorts. In order to simplify our downstream analyses, we derived one principal component variable (PC1) to represent the overall effect from these 3 CpG sites. Each participant therefore has a PC1 value, and we found that, in ε4 carriers, a reduction of one unit of PC1 attenuated the odds of AD by 58% (OR=2.39, 95%CI=[1.64,3.46], $P=7.08 \times 10^{-6}$).

In stage III, we conducted validation analyses and explored the possible functional consequences of our 4 CpG dinucleotides. We found that both PC1 and the level of methylation at the single CpG site of cg12307200 are associated with the mRNA expression level of multiple genes in the neocortex of ROSMAP participants, and we replicated this observation in independent datasets. The pathway analysis of these differentially expressed genes suggested the involvement of myeloid cell, which was further supported by our cell type analysis: the altered methylation levels appear to influence microglial gene expression. Finally, we found that both PC1 and cg12307200 have marginal associations with the proportion of histologically-defined activated microglia.

In conclusion, we have identified an epigenomic factor – microglial epigenomic factor 1- which might attenuate the ε4 effect on AD risk through changes in the transcriptome of the neocortex that relate to alterations in the proportion of activated microglia and of their effect on increasing the accumulation of tau pathology.

3 | DETAILED METHODS AND RESULTS

3.1 | Methods

3.1.1 | Study samples and pathological AD measurements—We included whites from the studies of the Religious order Study (ROS) and the Rush Memory and Aging Project (MAP) (www.radc.rush.edu) , the MRC London Neurodegenerative Disease Brain Bank (LBB), and the Mount Sinai Alzheimer's Disease and Schizophrenia Brain Bank

(MSBB) (Supplementary Methods and Table S1)[7,14,15]. ROS and MAP were jointly analyzed as ROSMAP with the adjustment of study variable (ROS and MAP). The LBB (GSE59685) and MSBB (GSE80970) included 68 (ε4+=36) and 129 (ε4+=41) individuals, respectively. Mayo Clinic Brain Bank (MAYO) provided DNA methylation and temporal cortex gene expression data on 45 patients with definite AD[16,17], diagnosed neuropathologically according to NINCDS-ADRDA criteria[18]. Both temporal cortex and prefrontal cortex brain tissues were archived frozenly. The study was approved by IRB of each institute.

3.1.2 | Pathological measurements—The pathological diagnosis of AD is based on the Braak score in LBB and MSBB and the NIA-Reagan score[19,20] in ROSMAP, which relies on both neurofibrillary tangles (Braak) and neuritic plaques (NP). Details of common neuropathologic indices measured in the ROSMAP study were described before[21–24] and in the Supplementary Methods.

3.1.3 | Brain DNA methylation across studies—Details of DNA methylation measurements and data processing of the cortical samples of ROSMAP, LBB, and MSBB were described before[7,8,14,15]. Briefly, the genome-wide DNA methylation was measured by the Illumina 450K methylation array followed by QC and normalization[8,14,15,25]. As a result, β values for 420,132 CpG dinucleotides were included in the ROSMAP MWAS which yielded the 25 sites for subsequent meta-analysis across ROSMAP, LBB and MSBB. In MAYO, only cg05157625 is available out of the 4 top ones which was measured using the reduced representation bisulfite sequencing as described before[26].

3.1.4 | Brain gene expression in ROSMAP, MSBB, and MAYO—In ROSMAP, there were 421 subjects with both data of DNA methylation and RNA sequencing (RNA-seq) (Illumina) from their dorsolateral prefrontal cortex[7] (Supplementary Methods). We included the 17,068 autosomal genes in the unit of normalized log2(cpm) into the transcriptome-wide association study (TWAS). A subset of these subjects (N=413) were previously used to derive the 47 cell-type relevant co-expressed gene module[27].

In MSBB, we downloaded the gene expressions of the transcriptome from Synapse platform (https://www.synapse.org/#!Synapse:syn7391833). Based on the genotype concordance check (Supplementary Methods), we included 50 subjects who have been profiled with both DNA methylation at prefrontal cortex and RNA-seq (Illumina) at BM44 region (closest to the prefrontal cortex).

In MAYO, we included 45 AD cases with both the DNA methylation data at cg05157625 and microarray-based gene expression data from their temporal cortex (Illumina)[16,17].

3.1.5 | Immunohistochemistry (IHC) measurements of cell types and microglia morphology in ROSMAP—A subset of ROSMAP subjects (N=57) with RNA-seq dataset were also profiled with the IHC stainings of markers of different cell types: neurons (NeuN), astrocytes (GFAP), microglia (IBA1), oligodendrocytes (Olig2) and endothelial cells (PECAM-1). The proportion of the microglia cell out of the total number of all different measured cell types were calculated[27]. Another subset (N=136) were

evaluated for their microglia activation based on morphology changes: stage I (thin ramified processes), stage II (plump cytoplasm and thicker processes), and stage III (appearance of macrophages). The percentage or the square-root transformed proportion of the stage III activated microglia out of the sum of all three stages counts (PAM) were derived as before[10]. These measurements are pre-existing and independent from current study.

3.1.6 | Statistical analysis—We used generalized linear regression model with the adjustments of age at death, sex, postmortem interval (if applicable), study (for ROSMAP dataset), technical variables, cell proportion[15] (if applicable) and *APOE* ε4 carrying status (if necessary). We applied the Bonferroni-correction significance threshold. Considering the potential inflation by including the same subjects in stage I and II, we arbitrarily applied 10 times more stringent significance threshold in stage I ($P \leq 1x10^{-8}$ (0.05/420,132/10)) and in stage II ($P \leq 2x10^{-4}$ (0.05/25/10))). The correlations between pairs of the top 4 CpG dinucleotides were presented using R "ggcorrplot" package. The standardized β values (mean=0 and SD=1) of each of the top 4 CpG dinucleotides in ROSMAP were input to derive 4 principal components (PCs) in ROSMAP, which were further projected into LBB and MSBB using the R "factoextra" and "prcomp"

3.1.7 | Pathway analysis—Top TWAS significant genes ($P \leq 2.93x10^{-6}$ (0.05/17,068 autosomal genes)) were followed with pathway enrichment analysis using "STRINGdb" v10 against the functional Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway database (https://www.genome.jp/kegg/pathway.html)[28] with FDR correction[29].

3.1.8 | Fetal brain Hi-C sequencing data downloaded from GEO—We downloaded (04/24/2020) the publically available Hi-C sequencing data of fetal brains[30] (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE77565) to interrogate the inter-chromosomal interactions. For each of the top 4 CpG dinucleotides, we extracted its genomic contacts across the 30,376 regions (100Kb resolution). Using the non-parametric Mann-Whitney-Wilcoxon test, we compared whether the rank of the normalized contact frequency is different between the two types of regions (TWAS or non-TWAS regions).

3.2 | Results

3.2.1 | Demographic characteristics stratified by *APOE ε4*—In ROSMAP, compared to *ε4*−, *ε4*+ participants were younger at the time of death (P=0.02), were more likely to have pathological AD (P=2.48x10^{-9}), and had more abnormally phosphorylated tangles (pTAU) (P=2.61x10^{-8}) and neuritic amyloid plaques (NP) (P=5.6x10^{-11}) (Table 1). In the MRC London Neurodegenerative Disease Brain Bank (LBB) and the Mount Sinai Alzheimer's Disease and Schizophrenia Brain Bank (MSBB) datasets, *ε4*+ subjects were also more likely to have AD (P<0.005) than *ε4*− subjects.

3.2.2 | Identification of CpG dinucleotides attenuating the effect of *APOE ε4* on AD risk

3.2.2.1 | Discovery of CpG dinucleotides: In our stage I analysis with all ROSMAP subjects (Figure 2A & Table S2), 25 CpG dinucleotides (1) were significantly associated with either pTAU or NP ($P \leq 1x10^{-8}$) and (2) had potential to attenuate the effect of

ε4 on pTAU or NP because the regression coefficients of *APOE* ε4, after adjusting for the candidate CpG dinucleotide, were smaller than the unadjusted ones. These 25 CpG were then evaluated in stage II, where we conducted a meta-analysis across 235 *ε4+* individuals assembled from four sample collections: ROS, MAP, LBB and MSBB. We found 4 CpG associated with a pathologic diagnosis of AD (meta-$P \leq 2x10^{-4}$): cg08706567 (*MPL*), cg26884773 (*TOMM20*), cg12307200 (*LPP* and *TPRG1*), and cg05157625 (*RIN3*) (Figure 2B). All of these 4 CpGs have stronger effects on AD susceptibility in *ε4+* than *ε4−*, which is consistently observed across the cohorts (Figure 2C,S1). For the cg08706567 and cg26884773 dinucleotides, the LBB *ε4+* dataset was not included into the meta-analysis since it has the problem of infinite maximum likelihood estimates ($P=1$)[31], and the results using the penalized generalized regression model yielded significant meta-P ($P \leq 0.01$) including only the LBB and MSBB (Table S3). Thus, the independent cohorts offer evidence of replication for the 4 prioritized CpG dinucleotides.

3.2.2.2 | Pathological associations of the 4 CpGs: Aside from its strong associations with pTAU and NP, cg12307200 had weaker associations with diffuse amyloid plaques (BETA=−3.15, SE=1.03, $P=2.34x10^{-3}$), cerebral amyloid angiopathy (BETA=−3.36, SE=1.19, $P=5.07x10^{-3}$), and arteriosclerosis (BETA=−3.47, SE=1.28, $P=6.73x10^{-3}$); and it had no association with Lewy Bodies, hippocampal sclerosis, or TDP-43 proteinopathy ($P>0.05$). The other 3 CpG dinucleotides had weaker associations with NP (BETA=[3.51,4.07], SE=[0.94,0.5], $P=[1.15x10^{-4},2.14x10^{-4}]$) than pTAU (BETA=[8.88,9.81], SE=[1.49,1.66], $P=[1.33x10^{-5},1.64x10^{-4}]$), nominal associations with TDP-43 proteinopathy (BETA=[2.28,3.24], SE=[1.36,1.51], $P=[0.03,0.09]$), and no associations with other neuropathologic indices. Thus, the top 4 CpG dinucleotides were primarily associated with Tau-related pathologies (Table S2).

3.2.2.3 | Derivation of a methylation PC based on the 3 validated CpGs that are correlated: All of the subjects have hypermethylation ($\beta \geq 0.5$) at cg12307200 and hypomethylation ($\beta < 0.5$) at the other 3 CpGs. While they are located on different chromosomes, the methylation level of these 3 CpGs were correlated ($r \geq 0.4$) (Figure 3A,S2). Because such patterns were consistently observed across all the cohorts, we derived a methylation-based principal component 1 (PC1) that captures the effect of those 3 highly-correlated CpGs in a single measure. We developed it using ROSMAP data and then projected it to the samples from LBB and MSBB using eigenvectors. As expected, the projected PC1 in LBB and MSBB captured the 3 CpG dinucleotides in the same way as in ROSMAP (Figure 3B).

3.2.2.4 | Interaction between *APOE* ε4 and the top CpG dinucleotides on AD risk: Here, we evaluated whether our epigenomic factors interacted with *APOE* ε4 to modulate AD risk and compared their effects on AD after stratification by ε4. The cg12307200 site did not display significant evidence of interaction, but the 3 correlated CpG dinucleotides captured by PC1 had nominal significance for ε4 interaction (range of meta-*P* for interaction=[$6.1x10^{-4}$,0.01]) (Figure 3C). This led us to pursue a more comprehensive interaction analysis of PC1 both categorically and continuously. For the categorical interaction, the original continuous PC1 was transformed to a binary variable

based on its median value. For the continuous interaction, the non-scaled continuous value of the PC1 was used. The results were similar for both the non-scaled and scaled values (Table S4), ruling out the potential influences of outliers on the continuous analysis. Both the categorical (meta-P=2.8x10^{-4}) and continuous (meta-P=7.7x10^{-4}) interaction tests were significant. In the categorical analysis, the effect of ε4 on AD was smaller in the subjects with low PC1 (meta-OR=2.57, 95% CI=[1.51,4.39], P=5.13x10^{-4}) than in the subjects with high PC1 (meta-OR=16.16, 95% CI=[6.45,40.51], P=2.93x10^{-9}). In the continuous analysis, the effect of PC1 on AD risk was greater in ε4+ (meta-ln(OR)=0.87, meta-P=7.08x10^{-6}, meta-N=235) than in the ε4− (meta-ln(OR)=0.20, P=2.65x10^{-3}, meta-N=532). In other words, reduction of one unit in PC1 was associated with a 58% ((1-1/exp(0.87))) attenuation in AD risk in ε4+ carriers. Thus, one could potentially influence the magnitude of ε4 risk by modulating PC1, an epigenomic measure that captures the effect of multiple loci.

3.2.2.4 | Functional exploration of neocortical ROSMAP transcriptomes, replication in other independent datasets and validation by Hi-C sequencing data of fetal brains: To understand the functional consequences of our 4 CpG dinucleotides, we conducted a Transcriptome-wide Association Study (TWAS) in the 421 ROSMAP participants who have both DNA methylation and RNA sequence (RNA-seq) data from the same cortical region. There were 71 genes which displayed significant (P≤2.93x10^{-6} (0.05/17,068 autosomal genes)) associations with either PC1 (70 genes) or cg12307200 (3 genes) (Figure 4A&B, Table S5). Except for *RIN3* where there was a cis-effect of the DNA methylation on gene expression, all of the other associations are driven by trans-effects because the TWAS target genes are far from the 4 CpG dinucleotides: they are either on different chromosomes or >5 Mb apart for those found on the same chromosome.

We then attempted to replicate our TWAS findings. In MSBB (50 subjects), the effect of cg12307200 was not replicated. But the majority of the PC1 TWAS genes were replicated. Out of the total of 68 genes also available in MSBB, 94% have the same effect direction as in ROSMAP, and 59% also showed significance (P<0.05) (Figure 4C). In MAYO (45 AD cases), only one of the three PC1 CpGs (cg05157625) was available. Out of the 21 TWAS genes significant with both PC1 and cg05157625 in ROSMAP, 85% have the same effect direction in MAYO as in ROSMAP (Figure 4D). 4 genes are significant in both MAYO and MSBB in relation to PC1, despite the small sample sizes.

To further explore the biological grounding of our observation, we accessed the publicly available Hi-C sequencing data from human fetal brain tissue which captures the 3-dimensional architecture of human cortex, with the caveat that this profiles a stage of brain development. Despite this important limitation, we found evidence that those genomic regions containing the 71 TWAS genes identified in ROSMAP have more contacts with the region covering one of the three PC1 CpGs, cg08706567, than you would expect by chance (P<0.01) (Figure 4E). This suggests that, even at very early stages of brain development, the 71 TWAS genes are physically interacting with at least one element of PC1; cg08706567 may play an important role as a regulator locus whose impact continues into advancing as it influences the impact of ε4.

3.2.3 | The TWAS results implicate microglial activation in the effect on ε4

3.2.3.1 | Pathway analysis of TWAS results suggested the involvement of the myeloid cells: These 71 TWAS significant genes displayed enrichment for 20 KEGG functional pathways (FDR≤0.05, hits>1 and hits%≥3%) (Figure 5A), which were related to multiple aspects of immunity and particularly myeloid cell function, such as osteoclast differentiation (hits=4%, FDR-P=5.5x10^{-5}), phagosome (hits=4%, FDR-P=5.5x10^{-5}), tuberculosis (hits=3%, FDR-P=5.5x10^{-5}), Leishmaniasis (hits=4%, FDR-P=1.3x10^{-3}), and antigen processing and presentation (hits=3%, FDR-P=0.01).

3.2.3.2 | Association with co-expressed gene modules of microglia: We further conducted a complementary association analysis with the 7 modules of co-expressed neocortical genes previously described as being enriched for microglial genes [9,27]: m5, m113, m114, m115, m116, m112, and m117. PC1 was associated with the expression of 5 of these 7 microglia modules (P≤7.14x10^{-3} (0.05/7 microglia modules)) (Figure 5B), but not other non-microglia modules (Table S6). Further, almost all (91%) of the 71 TWAS genes were found in the 5 significant microglia modules, and 58% belonged to m116, the most microglial enriched module; we previously reported[27] this module as being related to microglial aging[9]. It also contains some key AD genes such as *TREM2* and its binding partner *TYROBP*. m5 has been associated the burden of Tau pathology and the proportion of activated microglia (PAM), as defined using data from immunohistochemistry of brain sections and a standard neuropathologic scale based on microglial morphology[9,10]. Our findings do not appear to be driven by changes in microglial cell counts since neither PC1 nor cg12307200 was associated with microglia cell counts estimated using either immunohistochemistry staining for the IBA1 protein or its mRNA expression levels (Table S7). The results remain significant when we account for the proportion of microglial cells. Overall, our 4 epigenomic factors appeared to be related to microglial transcriptional programs captured by the modules of co-expressed genes defined in neocortical data, and the state of the microglia may thus be an important factor in the modulation of the ε4 effect.

3.2.3.3 | Validation using a histology-derived variable of microglia activation: As noted above, the m5 module was associated with PAM, a trait derived from the morphological characterization of microglia in histological sections that we found to be associated with cognitive decline, AD pathologies, and AD dementia[10]. PAM is simply the proportion of microglia that have an activated stage III morphology, and this trait was available in 122 ROSMAP participants that also have DNA methylation data from the same frontal cortex region. PC1 was positively (P=0.02) and cg12307200 was negatively associated with PAM in the midfrontal cortex (P=0.05) (Figure 5C,D). However, the association is modest, and PC1 was not fully explained by PAM. That is, PAM and PC1 capture somewhat different aspects of microglial function that may be partially related to one another. To be thorough, we also evaluated PAM measures in three other brain regions in secondary analyses, and we found that PAM in inferior temporal cortex, posterior putamen and ventral medial caudate also displayed association with PC1 and cg12307200 (Figure S2), suggesting that PC1 and cg12307200 derived from cortical tissue DNA methylation data capture an aspect of microglial state in multiple brain regions, not just the frontal cortex.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

ROSMAP: We are grateful to the participants in the Religious Order Study, the Memory and Aging Project. This work is supported by the US National Institutes of Health [U01AG61356, R01 AG043617, R01 AG042210, R01 AG036042, R01 AG036836, R01 AG032990, R01 AG18023, RC2 AG036547, P50 AG016574, U01 ES017155, KL2 RR024151, K25 AG04190601, R01 AG30146, P30 AG10161, R01 AG17917, R01 AG15819, K08 AG034290, P30 AG10161 and R01 AG11101. Ma Y. is supported by the 2019-AARF-644521.

MSBB: Brain banking and neuropathology assessments for the Mount Sinai cohort was supported by US National Institutes of Health grants AG02219, AG05138, U01 AG046170, RF1 AG057440 and MH064673, and the Department of Veterans Affairs VISN3 MIRECC. Replication work in Boston was supported by US National Institutes of Health grants: R01 AG036042, R01AG036836, R01 AG17917, R01 AG15819, R01 AG032990, R01 AG18023, RC2 AG036547, P30 AG10161, P50 AG016574, U01 ES017155, KL2 RR024151 and K25 AG041906-01.

LBB: Brain banking and neuropathology assessments for the MRC London Neurodegenerative Diseases Brain Bank, which was supported by the Medical Research Council (UK), and Brains for Dementia Research (Alzheimer Brain Bank, UK).

MAYO: Study data[17] were provided by the following sources: The Mayo Clinic Alzheimers Disease Genetic Studies, led by Dr. Nilüfer Ertekin-Taner and Dr. Steven G. Younkin, Mayo Clinic, Jacksonville, FL using samples from the Mayo Clinic Study of Aging, the Mayo Clinic Alzheimers Disease Research Center, and the Mayo Clinic Brain Bank. Data collection was supported through funding by NIA grants P50 AG016574, R01 AG032990, U01 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216, R01 AG003949, NINDS grant R01 NS080820, CurePSP Foundation, and support from Mayo Foundation. Study data includes samples collected through the Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona. The Brain and Body Donation Program is supported by the National Institute of Neurological Disorders and Stroke (U24 NS072026 National Brain and Tissue Resource for Parkinsons Disease and Related Disorders), the National Institute on Aging (P30 AG19610 Arizona Alzheimers Disease Core Center), the Arizona Department of Health Services (contract 211002, Arizona Alzheimers Research Center), the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901 and 1001 to the Arizona Parkinson's Disease Consortium) and the Michael J. Fox Foundation for Parkinsons Research.

We gratefully acknowledged Dr. Hyejung Won for his guidance to interpret the HiC sequencing data from the human fetal brains.

DATA AVAILABILITY

All the data and analysis output are available via the AD Knowledge Portal (https://adknowledgeportal.synapse.org). The AD Knowledge Portal is a platform for accessing data, analyses, and tools generated by the Accelerating Medicines Partnership (AMP-AD) Target Discovery Program and other National Institute on Aging (NIA)-supported programs to enable open-science practices and accelerate translational learning. The data, analyses and tools are shared early in the research cycle without a publication embargo on secondary use. Data is available for general research use according to the following requirements for data access and data attribution (https://adknowledgeportal.synapse.org/DataAccess/Instructions). The link to the data and analysis output for this manuscript is https://www.synapse.org/#!Synapse:syn22240706.

REFERENCE

1. Farrer LAet 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. JAMA278, 1349–56 (1997). [PubMed: 9343467]
2. Sherva R & Farrer LA Power and pitfalls of the genome-wide association study approach to identify genes for Alzheimer's disease. Curr Psychiatry Rep 13, 138–46 (2011). [PubMed: 21312009]
3. Lambert JCet al.Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet45, 1452–8 (2013). [PubMed: 24162737]
4. Kunkle BWet al.Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. Nat Genet51, 414–430 (2019). [PubMed: 30820047]
5. Meyer MRet al.APOE genotype predicts when--not whether--one is predisposed to develop Alzheimer disease. Nat Genet19, 321–2 (1998). [PubMed: 9697689]
6. Ma Y & Ordovas JM The integration of epigenetics and genetics in nutrition research for CVD risk factors. Proc Nutr Soc 76, 333–346 (2017). [PubMed: 27919301]
7. De Jager PLet al.A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. Sci Data5, 180142 (2018). [PubMed: 30084846]
8. De Jager PLet al.Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. Nat Neurosci17, 1156–63 (2014). [PubMed: 25129075]
9. Patrick ET, M.; Ergun A; Ng B; Casazza W; Cimpean M; Yung C; Schneider JA; Bennett DA; Gaiteri C; De Jager PL; Bradshaw EM; Mostafavi S. Deconvolving the contributions of cell-type heterogeneity on cortical gene expression. bioRxiv (2019).
10. Felsky Det al.Neuropathological correlates and genetic architecture of microglial activation in elderly human brain. Nat Commun10, 409 (2019). [PubMed: 30679421]
11. Jun Get al.A novel Alzheimer disease locus located near the gene encoding tau protein. Mol Psychiatry21, 108–17 (2016). [PubMed: 25778476]
12. Ma Yet al.Analysis of Whole-Exome Sequencing Data for Alzheimer Disease Stratified by APOE Genotype. JAMA Neurol (2019).
13. Strickland SLet al.MAPT haplotype-stratified GWAS reveals differential association for AD risk variants. Alzheimers Dement (2020).
14. Lunnon Ket al.Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. Nat Neurosci17, 1164–70 (2014). [PubMed: 25129077]
15. Smith RGet al.Elevated DNA methylation across a 48-kb region spanning the HOXA gene cluster is associated with Alzheimer's disease neuropathology. Alzheimers Dement14, 1580–1588 (2018). [PubMed: 29550519]
16. Zou Fet al.Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. PLoS Genet8, e1002707 (2012). [PubMed: 22685416]
17. Allen Met al.Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. Sci Data3, 160089 (2016). [PubMed: 27727239]
18. McKhann Get al.Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology34, 939–44 (1984). [PubMed: 6610841]
19. Consensus recommendations for the postmortem diagnosis of Alzheimer's disease. The National Institute on Aging, and Reagan Institute Working Group on Diagnostic Criteria for the Neuropathological Assessment of Alzheimer's Disease. Neurobiol Aging18, S1–2 (1997). [PubMed: 9330978]
20. Bennett DAet al.Neuropathology of older persons without cognitive impairment from two community-based studies. Neurology66, 1837–44 (2006). [PubMed: 16801647]
21. Yu L, Boyle PA, Leurgans S, Schneider JA & Bennett DA Disentangling the effects of age and APOE on neuropathology and late life cognitive decline. Neurobiol Aging 35, 819–26 (2014). [PubMed: 24199961]

22. Bennett DA, Schneider JA, Arvanitakis Z & Wilson RS Overview and findings from the religious orders study. Curr Alzheimer Res 9, 628–45 (2012). [PubMed: 22471860]

23. Bennett DAet al.Overview and findings from the rush Memory and Aging Project. Curr Alzheimer Res9, 646–63 (2012). [PubMed: 22471867]

24. White CCet al.Identification of genes associated with dissociation of cognitive performance and neuropathological burden: Multistep analysis of genetic, epigenetic, and transcriptional data. PLoS Med14, e1002287 (2017). [PubMed: 28441426]

25. Pidsley Ret al.A data-driven approach to preprocessing Illumina 450K methylation array data. BMC Genomics14, 293 (2013). [PubMed: 23631413]

26. Allen Met al.Gene expression, methylation and neuropathology correlations at progressive supranuclear palsy risk loci. Acta Neuropathol132, 197–211 (2016). [PubMed: 27115769]

27. Mostafavi Set al.A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. Nat Neurosci21, 811–819 (2018). [PubMed: 29802388]

28. Szklarczyk Det al.STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res43, D447–52 (2015). [PubMed: 25352553]

29. Benjamini Y HYControlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Statist. Soc.57, 289–300 (1995).

30. Won Het al.Chromosome conformation elucidates regulatory relationships in developing human brain. Nature538, 523–527 (2016). [PubMed: 27760116]

31. Lesaffre EA, Partial Aseparation in logistic discrimination. Journal of the Royal Statistical Society. Series B. Methodological1, 109–116 (1989).

RESEARCH IN CONTEXT

1. **Systematic review:** The authors reviewed the published literature (e.g. PubMed) on the associations between *APOE* ε4 and Alzheimer's disease (AD) which indicates the presence of potential suppressor of the ε4 risk on AD.
2. **Interpretation:** The alterations of methylation values of 7 CpG sites reduce the ε4 risk on pathological AD by 9% and these sites regulate the mRNA expressions of immune genes related to microglia, which was confirmed by the immuno-histochemistry experiments. Having low proportion of activated microglia may act as a suppressor of ε4 effect on AD by epigenomic mechanisms.
3. **Future directions:** Larger scale human and mechanistic studies are necessary to confirm our findings and the identifications of the upstream environmental factors and their working mechanisms are of particular interest.

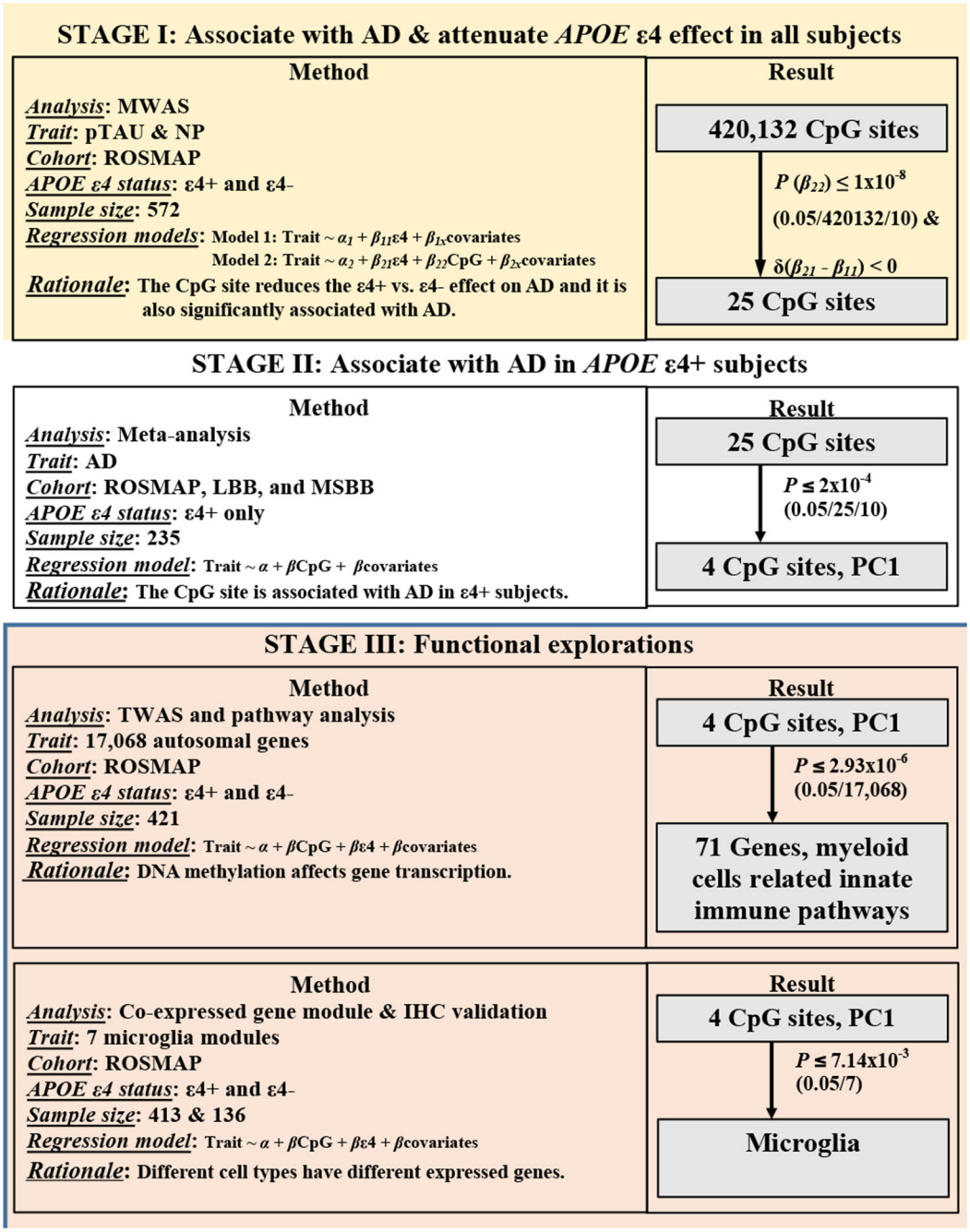

Figure 1. Flow chart of analysis plan and results.

Abbreviations: MWAS, methylome-wide association analysis; NP, neuritic plaque; pTAU, abnormally phosphorylated Tau protein, AT8; TWAS, transcriptome-wide association analysis; IHC, immuno-histochemistry measurements.

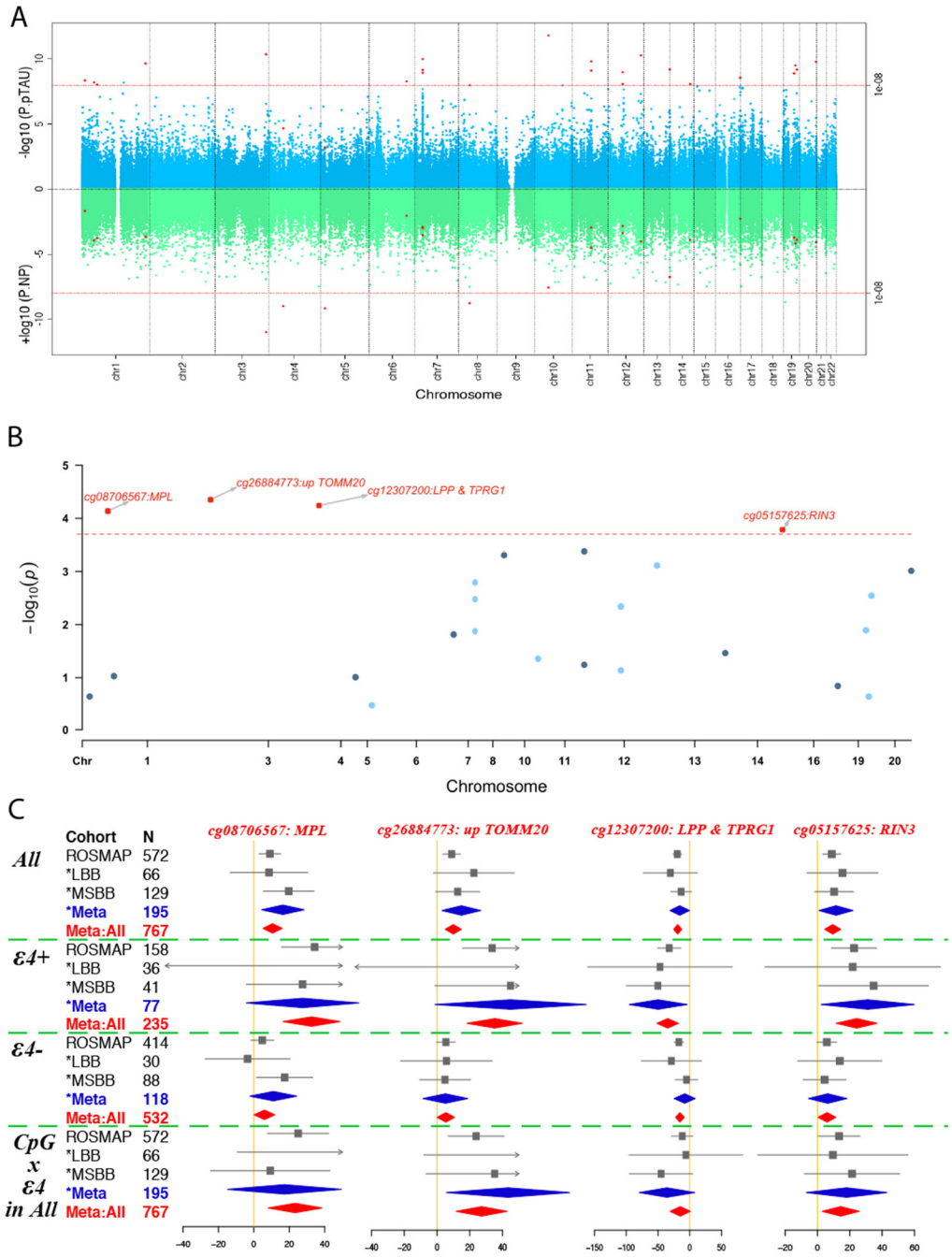

Figure 2. Identification of the top 4 CpG dinucleotides.

(A) Miami plot shows the results of the methylation genome-wide association study (MWAS) on pTAU (upper panel in blue) and NP (lower panel in green) in ROSMAP (N=572). The Y axis show the −log10 transformed *P* value of each of the genome-wide CpG dinucleotides which are shown in dots and ordered according to their genomic coordinates on the shared X axis. The red dashed line represents the significance threshold (*P*=1E-8=0.05/420132/10) and those CpG dinucleotides passing the genome-wide significance threshold for either pTAU or NP are shown in red dots. (B) Candidate

manhattan plot shows the meta-analyzed associations of the above selected 25 CpG dinucleotides with pathological AD in ε4+ subjects across ROSMAP, LBB, and MSBB cohorts (N=235). Y axis represent the −log10 transformed meta analyzed *P* values of the regression coefficient estimates of the logistic regression model, in which the outcome variable is the pathological diagnosis of Alzheimer's disease (AD) (no=0 and yes=1), the exposure variable is the methylation status of each CpG dinucleotides (0 to 1) and the covariates include age at death, postmortem interval, sex, and study, ethnicity principle components, methylation experiment batches in ROSMAP, and cell proportion in LBB and MSBB. The horizontal red dashed line represents the Bonferroni corrected *P* value threshold of ($0.05/(25*10) = 2x10^{-4}$) and the top 4 significant CpG dinucleotides are represented in red squares with the annotation of their closest genes. (C) Forest plots for the top 4 CpG dinucleotides across different cohorts for their associations with AD within: 1) all the subjects; 2) subjects carrying; or 3) not carrying the *APOE* ε4 allele; and 4) the interaction test between the methylation level and *APOE* ε4 allele carrying status within all the subjects. The filled square and horizontal line for each population or the filled summary diamonds (blue for the meta-analysis across the replication cohorts while red for the joint meta-analysis across all the cohorts) denote the estimated regression coefficient (BETA) and its 95% CI per unit increase in the methylation level of each CpG dinucleotide or the interaction term of the methylation times the ε4-carrying status (yes=1 and no=0) for the binary outcome variable of the pathological diagnosis of AD (no=0 and yes=1) with the adjustment of the covariates of age at death, postmortem interval, sex, study, ethnicity principle components, methylation experiment batches in ROSMAP, and cell proportion in LBB and MSBB. The arrows indicate the estimates are out of the boundaries.
Abbreviations: NP, neuritic plaque; pTAU, abnormally phosphorylated Tau protein, AT8.

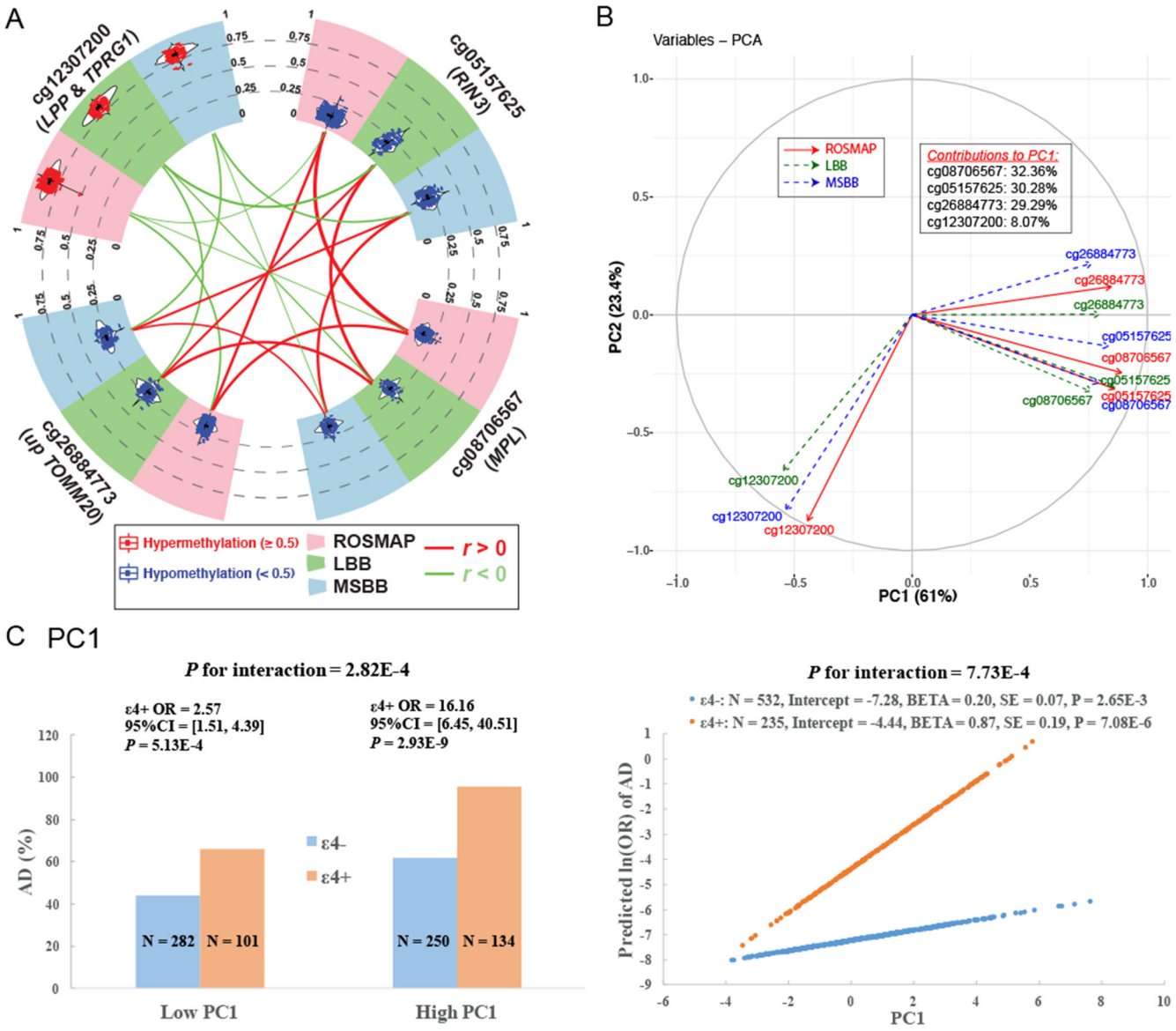

Figure 3. Consistent correlations among the 4 CpG dinucleotides across cohorts and the derivation of the PC1.

(A) Circos plot shows the consistent distributions of each of the top 4 CpG dinucleotides and their mutual correlations across ROSMAP (pink sector), LBB (green sector) and MSBB (blue sector). In the outer layer, the distribution of each CpG dinucleotide within each cohort is shown in a violin plot where each dot represents a subject and the black horizontal and vertical lines denote the mean and standard deviation. Red dots represent those subjects with hypermethylation level (≥0.5) while blue ones represent those subjects with hypomethylation level (<0.5). In the inner layer, the correlations between each pair of 2 CpG dinucleotides within each cohort were represented by their connected lines, where the color denote the direction (red for r>0 and green for r<0) and the thickness denote the strength of their correlations (thicker lines represent stronger correlations). (B) Variance plot of the PC1 vs. PC2 and their contributions to each of the top 4 CpG dinucleotides across

ROSMAP (red solid arrow), LBB (blue dashed arrow), and MSBB (blue dashed arrow). (C) Interaction effect on AD between *APOE* ε4 and PC1. The subgroup of ε4+ and ε4− are represented as blue and orange. Interaction with categorical variables are shown on the left panel where the continuous variable of PC1 were transformed to the binary variable based on the median value. The Y axis represent the percentage of AD cases across 4 subgroups of subjects. The right panel show the interactions with the untransformed continuous PC1. the Y axis represent the predicted ln(OR) of AD calculated based on the statistics within ε4+ and ε4− subgroup using the logistic regression model with binary AD status (case=1 and control=0) as the outcome variable adjusting the covariates of age at death, sex, study, postmortem interval, methylation experimental batches and two major ethnic principles. Abbreviations: OR, odds ratio.

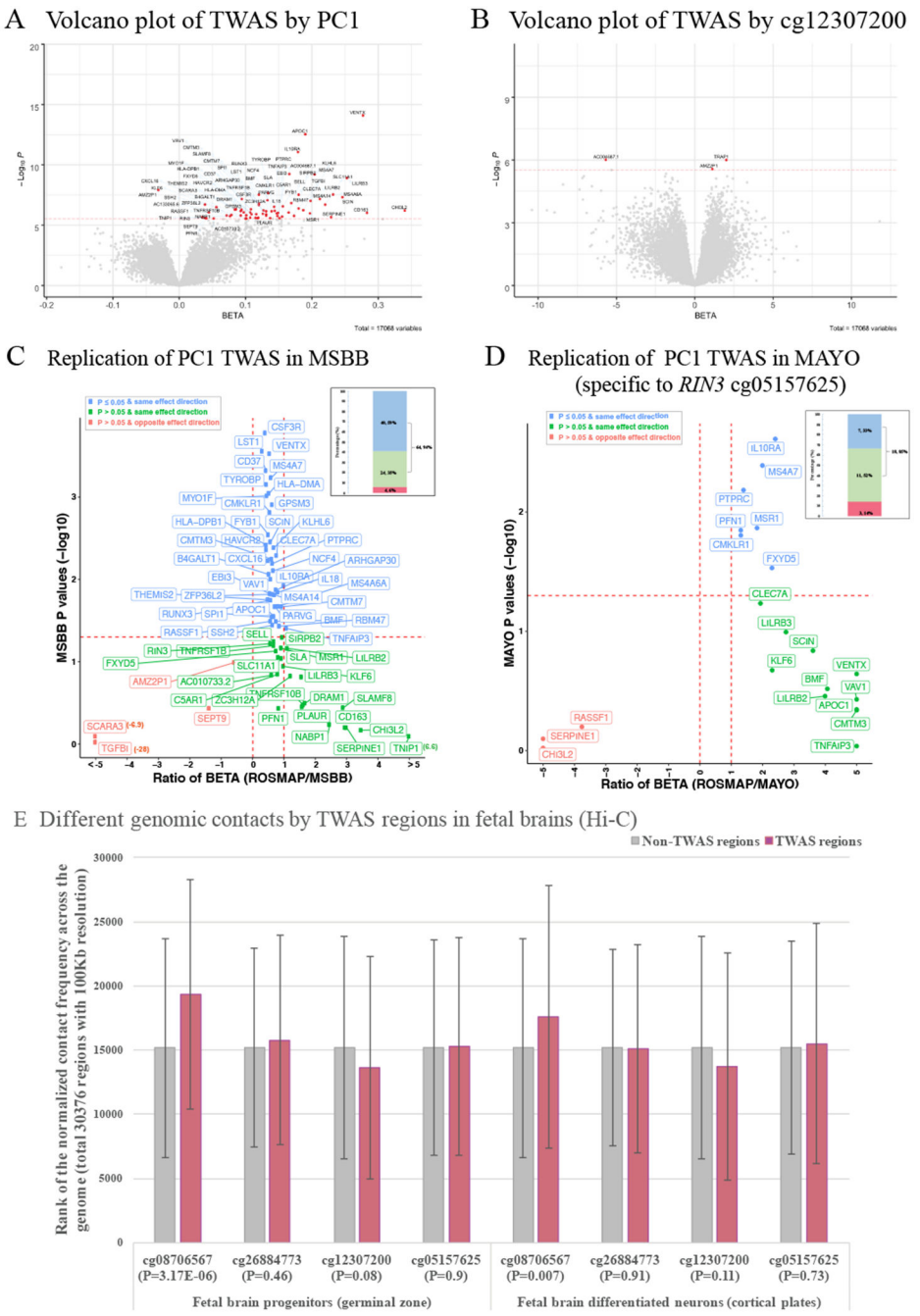

Figure 4. Discovery and replication of TWAS results of the PC1 and cg12307200.

(A) and (B) Volcano plot shows the TWAS results of PC1 and cg12307200. The X axis shows the BETA and the Y axis shows the −log10 transformed P values for the exposure variable of PC1 or cg12307200 on the mRNA expression levels of each gene. The 71 unique genes (70 for PC1 and 3 for cg12307200) passed the Bonferroni corrected significance threshold of $P \leq 2.93 \times 10^{-6}$ (0.05/17,068 autosomal genes) are shown in red dots with their gene name connected through blue arrows. (C) and (D) Replication of the PC1 TWAS results in MSBB and MAYO. ROSMAP has identified 70 TWAS genes for PC1 and 68 of

them are available in MSBB. There are 25 (out of 70) genes are significant for cg05157625 (the only CpG dinucleotide available in MAYO) and 21 of them are available in MAYO. The scatter plot shows the comparison of each gene between ROSMAP and MSBB (or MAYO) where the X axis represents the ratio of the regression coefficient obtained in the two cohorts (ROSMAP over MSBB (or MAYO) and the red dashed vertical lines for the ratio=0 or 1) and the Y axis shows the *P* values in MSBB (or MAYO) (red dashed horizontal line for the P=0.05). The blue dots are those genes with nominal significance ($P \leq 0.05$) in MSBB (or MAYO) and with the same effect direction in both ROSMAP and MSBB (or MAYO), the green dots are those non-significant genes ($P > 0.05$) in MSBB (or MAYO) but with the same effect direction, and the red dots are those non-significant genes in MSBB (or MAYO) and also with the opposite effect direction. The number and their percentage of these three groups of genes are presented in the bar plot (upper right corner) with the same color codings. (E) Different genomic contacts by TWAS regions in the fetal brains analyzed with the published Hi-C sequencing data (Won et al., 2016) downloaded from GEO (GSE77565). Silver and rose gold bars represent those regions outside and within the 70 significant TWAS genes in ROSMAP and the *P* value for the differences between these two groups for each CpG dinucleotide are shown in X axis.

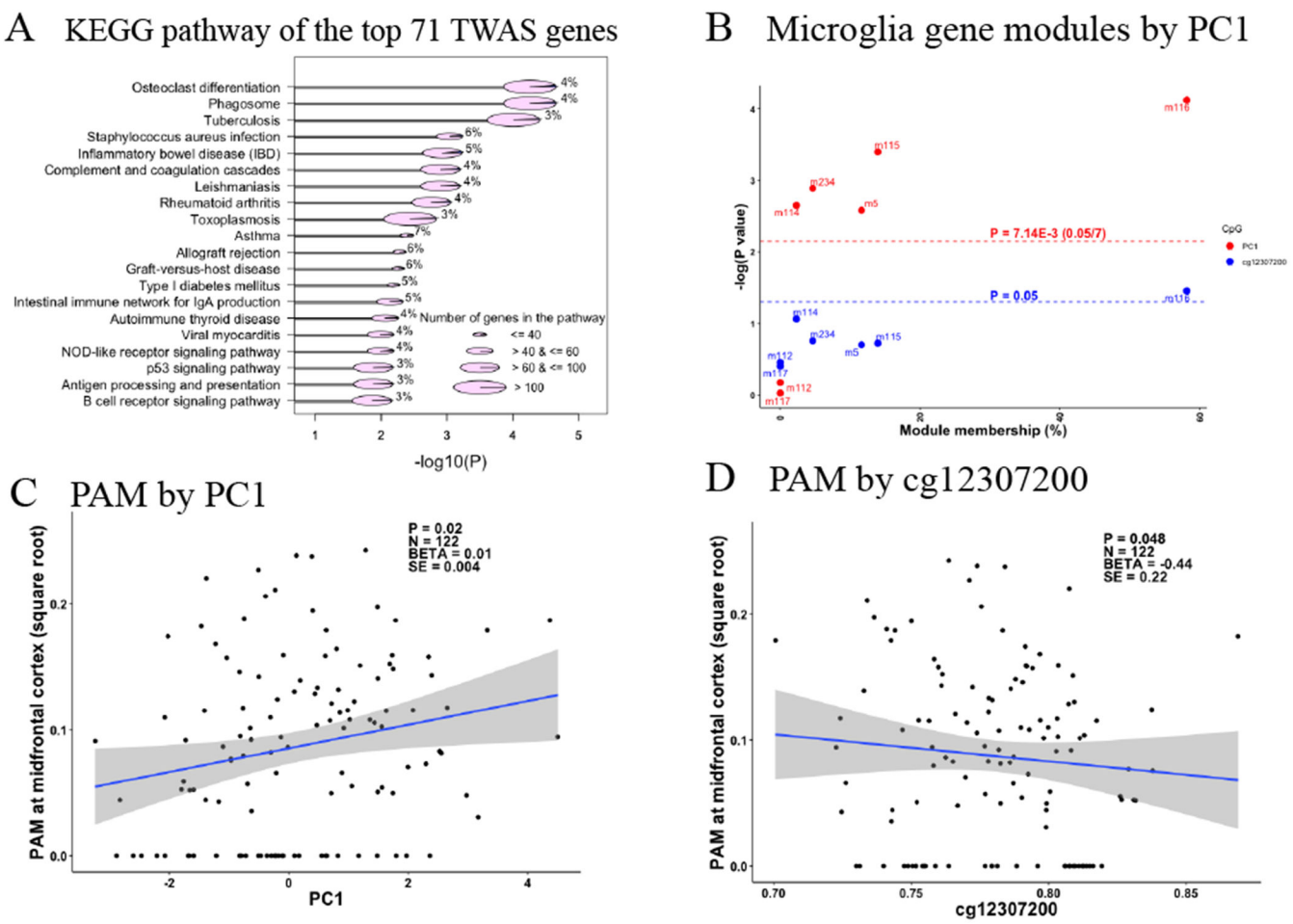

Figure 5. Microglia relevance to PC1 and cg12307200.

(A) Plot of the KEGG pathway analysis of the top 71 TWAS genes which have significant associations with the PC1 and cg12307200. Only those pathways with FDR adjusted *P* value ≤ 0.05 and the number (percentage) of the hit genes > 1 (≥ 3%) were shown in the plot. Y axis list the name of these KEGG functional pathways and X axis shows their corresponding −log10 transformed FDR adjusted *P* values. The size of the pie represents the number of the member genes of each functional pathway, which are categorized into four types ranging from the smallest to the largest containing ≤40, >40 & ≤60, >60 & ≤100, and >100 member genes. Each pie was split into 2 slices with the area of the blue slice represents the percentage of the hit genes out of the total member genes for each pathway and their number are shown for each pathway. (B) Scatter plot of the associations with the methylation factors (PC1 in red and cg12307200 in blue) and expression levels of the 7 reported gene modules for microglia. Y axis represent the −log10 transformation of the *P* values of the associations between the exposure variables of PC1 or cg12307200 on the outcome variable of the expression values of each gene module, and the X axis represent the module membership, which is defined as the percentage of the module member genes out of our identified 43 gene list (28 genes are not mapped to any of the reported gene modules). The horizontal red and blue dashed lines represent the Bonferroni corrected significance threshold of $P \leq 7.14 \times 10^{-3}$ (0.05/7 gene modules) and the nominal significance threshold

of $P \leq 0.05$. (C) and (D) show the scatter plot of the associations between proportion of activated microglia (PAM) at midfrontal cortex as Y axis and the methylation factors (PC1 and cg12307200) as X axis. Each dot is one observation and the regression line and its 95% CI are represented by the blue line and grey area. The estimated statistics are shown on the upper right corner of the plot. The BETA, SE, and P represent the estimated regression coefficient and its standard error, P value of the exposure variable of PC1 or cg12307200 on the outcome of PAM. The N represent the sample size within the analysis. For all the analysis, we used the generalized linear regression model with the adjustments of the age at death, sex, postmortem interval, study, *APOE* ε4 binary status, two major ethnic principles and methylation experimental batches.

Table 1.

Demographics of subjects carrying or not carrying *APOE ε4* allele across ROSMAP, LBB and MSBB

	ROSMAP			LBB			MSBB		
	ε4+	**ε4−**	***P***	**ε4+**	**ε4−**	***P***	**ε4+**	**ε4−**	***P***
N[#]	158	414		36	32		41	88	
Age at death[*]	87.31 (6.01)	88.70 (6.65)	0.02	84.58 (9.00)	84.72 (10.13)	0.95	86.27 (7.22)	85.33 (7.80)	0.5
Female[$]	93 (59%)	268 (65%)	0.23	25 (69%)	21 (66%)	0.94	27 (66%)	53 (60%)	0.68
AD cases[$]	128 (81%)	221 (53%)	$2.48x10^{-9}$	33 (92%)	19 (59%)	0.004	34 (83%)	40 (45%)	0.0001
pTAU (sqrt)[*]	2.75 (1.67)	1.92 (1.16)	$2.61x10^{-8}$	NA	NA	NA	NA	NA	NA
NP[*]	1.20 (0.92)	0.64 (0.73)	$5.60x10^{-11}$	NA	NA	NA	NA	NA	NA

[#]N represent the total number of subjects with measurements of any of the pathological traits listed in the table.

[*]represent the mean and standard deviation of each trait in subgroups of subjects carrying or not carrying *APOE4* allele and the P values of their differences using t test.

[$]represent the count and percentage of each trait in subgroups of subjects carrying or not carrying *APOE4* allele and the *P* values of their differences using chi-square test.

Abbreviations: pTAU, abnormally phosphorylated Tau protein, AT8.

Alzheimers Dement. Author manuscript; available in PMC 2021 December 01.