

1 EPIGENOMIC FEATURES RELATED TO MICROGLIA ARE ASSOCIATED WITH
2 ATTENUATED EFFECT OF APOE ϵ 4 ON ALZHEIMER'S DISEASE RISK IN HUMANS

3
4 Yiyi Ma¹, Lei Yu^{2,3}, Marta Olah¹, Rebecca Smith⁴, Stephanie R. Oatman⁵, Mariet Allen⁵, Ehsan
5 Pishva⁴, Bin Zhang^{6,7}, Vilas Menon¹, Nilüfer Ertekin-Taner^{5,8}, Katie Lunnon⁴, David A.
6 Bennett^{2,3}, Hans-Ulrich Klein¹, Philip L. De Jager^{1,9}

7
8 ¹ Center for Translational & Computational Neuroimmunology, Department of Neurology,
9 Columbia University Medical Center, 630 West 168th street, New York, NY, USA.

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

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

12 ⁴ University of Exeter Medical School, College of Medicine and Health, Exeter University,
13 Exeter, UK.

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

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

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

19 ⁸ Mayo Clinic Florida, Department of Neurology, Jacksonville, FL 32224, USA.

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

21

22

23 * Corresponding author: Philip L. De Jager, M.D., Ph.D.,

24 Director of the Center for Translational & Computational Neuroimmunology, Department of

25 Neurology, Columbia University Medical Center, 630 West 168th street, New York, NY 10032,

26 USA; Tel: (212) 305-3609; Email: pld2115@cumc.columbia.edu

27

28 **Words count (counts: 3,499 and limits: \leq 3,500 words, excluding abstract, references, figures
29 and tables):**

30 **Number of references: (counts: 31 and limits: \leq 50)**

31 **Number of tables and figures: (counts: 6 and limits: \leq 6)**

32 **Keywords: APOE; Alzheimer's disease; epigenome; microglia**

33 **Classification: Neurology**

34

35

36 **Abstract (149 words / <150 words)**

37 **INTRODUCTION:** Not all *APOE* $\epsilon 4$ carriers who survive to advanced age develop Alzheimer's
38 disease (AD); factors attenuating the risk of $\epsilon 4$ on AD may exist.

39
40 **METHODS:** Guided by the top $\epsilon 4$ -attenuating signals from methylome-wide association
41 analyses (N=572, $\epsilon 4+$ and $\epsilon 4-$) of neurofibrillary tangles and neuritic plaques, we conducted a
42 meta-analysis for pathological AD within the $\epsilon 4+$ subgroups (N=235) across four independent
43 collections of brains. Cortical RNA-seq and microglial morphology measurements were used in
44 functional analyses.

45
46 **RESULTS:** Three out of the four significant CpG dinucleotides were captured by one principle
47 component (PC1), which interacts with $\epsilon 4$ on AD, and is associated with expression of innate
48 immune genes and activated microglia. In $\epsilon 4$ carriers, reduction in each unit of PC1 attenuated
49 the odds of AD by 58% (OR=2.39, 95% CI=[1.64,3.46], $P=7.08 \times 10^{-6}$).

50
51 **DISCUSSION:** An epigenomic factor associated with a reduced proportion of activated
52 microglia appears to attenuate the risk of $\epsilon 4$ on AD.

53

54

55 **1. Introduction**

56 The *APOE* $\epsilon 4$ haplotype contributes the greatest common genetic risk for Alzheimer's
57 disease (AD)¹⁻⁴. However, not all $\epsilon 4$ carriers develop AD. A longitudinal observational study
58 reported that 9 out of the 141 $\epsilon 4/\epsilon 4$ individuals remained dementia-free after age 84⁵.

59 Another meta-analysis of cross-sectional studies suggested that the $\epsilon 4$ effect on AD becomes
60 weaker after age 70¹. These results indicate that factors attenuating the genetic risk of $\epsilon 4$ on
61 AD might exist. Such attenuators could be either genetic or environmental or both; here, we
62 focused on DNA methylation features which might act as a modifiable mechanism on the
63 human genome⁶.

64 Conducting a DNA methylation genome-wide association study (MWAS) of AD risk
65 within the $\epsilon 4+$ subgroup can provide an unbiased search for those unknown signals
66 protecting $\epsilon 4$ carriers from having AD. However, such an analysis is very limited by
67 statistical power, as illustrated by our recent whole exome sequencing study with over 3,000

68 $\epsilon 4+$ subjects⁷. Therefore, we designed a three-stage approach (**Figure 1**). In stage I, we took
69 advantage of an existing DNA methylation dataset⁸ generated from a random subset of
70 participants in the Religious order (ROS) and Memory & Aging Projects (MAP). This
71 includes both $\epsilon 4+$ and $\epsilon 4-$ individuals with detailed quantitative measures of AD
72 neuropathology to maximize power in performing an initial MWAS to prioritize a list of CpG
73 dinucleotides which had the potential to be an $\epsilon 4$ attenuator. The resulting prioritized CpG
74 dinucleotides fulfilled two criteria: (1) be associated with AD pathology in all subjects; and
75 (2) reduce the effect of $\epsilon 4$ on AD susceptibility. The latter is determined by comparing the
76 regression coefficients for the $\epsilon 4$ variable before and after adjusting for each CpG
77 dinucleotide and focusing on those CpG dinucleotides which reduce the unadjusted $\epsilon 4$
78 regression coefficient. The prioritized list of CpG dinucleotides from stage I were then
79 moved into stage II, where we checked their associations with AD in the $\epsilon 4+$ subgroup. We
80 then assessed for evidence of replication in independent cohorts, and, to increase the
81 statistical power, we conducted a meta-analysis in stage II to combine data from 4
82 independent cohorts and generate a final summary statistic. In stage III, we conducted
83 validation analyses and explored the relevant functions. As a result, we have identified an
84 epigenomic factor which might attenuate the $\epsilon 4$ effect on AD risk through changes in the
85 transcriptome of the neocortex that relate to alterations in the proportion of activated
86 microglia and their effect on the accumulation of Tau pathologies.

87

88 **2. Methods**

89 *2.1 Study samples and pathological AD measurements*

90 We included whites from the studies of the Religious order Study (ROS) and the
91 Rush Memory and Aging Project (MAP) (www.radc.rush.edu) , the MRC London
92 Neurodegenerative Disease Brain Bank (LBB), and the Mount Sinai Alzheimer's Disease and
93 Schizophrenia Brain Bank (MSBB) (**Supplementary Methods and Table S1**)⁹⁻¹¹. ROS and
94 MAP were jointly analyzed as ROSMAP with the adjustment of study variable (ROS and
95 MAP). The LBB (GSE59685) and MSBB (GSE80970) included 68 ($\epsilon 4+=36$) and 129
96 ($\epsilon 4+=41$) individuals, respectively. Mayo Clinic Brain Bank (MAYO) provided DNA
97 methylation and temporal cortex gene expression data on 45 patients with definite AD^{12,13},
98 diagnosed neuropathologically according to NINCDS-ADRDA criteria¹⁴. Both temporal
99 cortex and prefrontal cortex brain tissues were archived frozenly. The study was approved by
100 IRB of each institute.

101

102 *2.2 Pathological measurements*

103 The pathological diagnosis of AD is based on the Braak score in LBB and MSBB and the
104 NIA-Reagan score^{15,16} in ROSMAP, which relies on both neurofibrillary tangles (Braak) and
105 neuritic plaques (NP). Details of common neuropathologic indices measured in the ROSMAP
106 study were described before¹⁷⁻²⁰ and in the **Supplementary Methods**.

107

108 *2.3 Brain DNA methylation across studies*

109 Details of DNA methylation measurements and data processing of the cortical samples of
110 ROSMAP, LBB, and MSBB were described before⁸⁻¹¹. Briefly, the genome-wide DNA
111 methylation was measured by the Illumina 450K methylation array followed by QC and
112 normalization^{8,10,11,21}. As a result, β values for 420,132 CpG dinucleotides were included in
113 the ROSMAP MWAS which yielded the 25 sites for subsequent meta-analysis across

114 ROSMAP, LBB and MSBB. In MAYO, only cg05157625 is available out of the 4 top ones
115 which was measured using the reduced representation bisulfite sequencing as described
116 before²².

117

118 *2.4 Brain gene expression in ROSMAP, MSBB, and MAYO*

119 In ROSMAP, there were 421 subjects with both data of DNA methylation and RNA
120 sequencing (RNA-seq) (Illumina) from their dorsolateral prefrontal cortex⁹ (Supplementary
121 Methods). We included the 17,068 autosomal genes in the unit of normalized log₂(cpm) into
122 the transcriptome-wide association study (TWAS). A subset of these subjects (N=413) were
123 previously used to derive the 47 cell-type relevant co-expressed gene module²³.

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

129 In MAYO, we included 45 AD cases with both the DNA methylation data at cg05157625
130 and microarray-based gene expression data from their temporal cortex (Illumina)^{12,13}.

131

132 *2.5 Immunohistochemistry (IHC) measurements of cell types and microglia morphology in*

133 *ROSMAP*

134 A subset of ROSMAP subjects (N=57) with RNA-seq dataset were also profiled with the
135 IHC stainings of markers of different cell types: neurons (NeuN), astrocytes (GFAP),
136 microglia (IBA1), oligodendrocytes (Olig2) and endothelial cells (PECAM-1). The

137 proportion of the microglia cell out of the total number of all different measured cell types
138 were calculated²³. Another subset (N=136) were evaluated for their microglia activation
139 based on morphology changes: stage I (thin ramified processes), stage II (plump cytoplasm
140 and thicker processes), and stage III (appearance of macrophages). The percentage or the
141 square-root transformed proportion of the stage III activated microglia out of the sum of all
142 three stages counts (PAM) were derived as before²⁴. These measurements are pre-existing
143 and independent from current study.

144

145 *2.6 Statistical analysis*

146 We used generalized linear regression model with the adjustments of age at death, sex,
147 postmortem interval (if applicable), study (for ROSMAP dataset), technical variables, cell
148 proportion¹¹ (if applicable) and *APOE* $\epsilon 4$ carrying status (if necessary). We applied the
149 Bonferroni-correction significance threshold. Considering the potential inflation by including
150 the same subjects in stage I and II, we arbitrarily applied 10 times more stringent significance
151 threshold in stage I ($P \leq 1 \times 10^{-8}$ (0.05/420,132/10)) and in stage II ($P \leq 2 \times 10^{-4}$ (0.05/25/10)).

152 The correlations between pairs of the top 4 CpG dinucleotides were presented using R
153 "ggcorrplot" package. The standardized β values (mean=0 and SD=1) of each of the top 4
154 CpG dinucleotides in ROSMAP were input to derive 4 principal components (PCs) in
155 ROSMAP, which were further projected into LBB and MSBB using the R "factoextra" and
156 "prcomp"

157

158 *2.7 Pathway analysis*

159 Top TWAS significant genes ($P \leq 2.93 \times 10^{-6}$ (0.05/17,068 autosomal genes)) were
160 followed with pathway enrichment analysis using “STRINGdb” v10 against the functional
161 Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway database
162 (<https://www.genome.jp/kegg/pathway.html>)²⁵ with FDR correction²⁶.

163

164 *2.8 Fetal brain Hi-C sequencing data downloaded from GEO*

165 We downloaded (04/24/2020) the publically available Hi-C sequencing data of fetal
166 brains²⁷ (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE77565>) to interrogate the
167 inter-chromosomal interactions. For each of the top 4 CpG dinucleotides, we extracted its
168 genomic contacts across the 30,376 regions (100Kb resolution). Using the non-parametric
169 Mann-Whitney-Wilcoxon test, we compared whether the rank of the normalized contact
170 frequency is different between the two types of regions (TWAS or non-TWAS regions).

171

172 **3. Results**

173 *1.1. Demographic characteristics by APOE $\epsilon 4$*

174 In ROSMAP, compared to $\epsilon 4^-$, $\epsilon 4^+$ participants were younger at the time of death
175 ($P=0.02$), more likely to have pathological AD ($P=2.48 \times 10^{-9}$), and had more abnormally
176 phosphorylated tangles (pTAU) ($P=2.61 \times 10^{-8}$) and neuritic amyloid plaques (NP) ($P=5.6 \times 10^{-11}$) (**Table 1**). In the MRC London Neurodegenerative Disease Brain Bank (LBB) and the
177 Mount Sinai Alzheimer's Disease and Schizophrenia Brain Bank (MSBB) datasets, $\epsilon 4^+$
178 subjects were also more likely to have AD ($P < 0.005$) than $\epsilon 4^-$ subjects.

180

181 *1.2. Identification of CpG dinucleotides attenuating the effect of APOE $\epsilon 4$ on AD*

182 *3.2.1 Discovery of CpG dinucleotides*

183 In our stage I analysis with all ROSMAP subjects (**Figure 2A & Table S2**), 25 CpG
184 dinucleotides (1) were significantly associated with either pTAU or NP ($P \leq 1 \times 10^{-8}$) and (2)
185 had potential to attenuate the effect of $\epsilon 4$ on pTAU or NP because the regression coefficients
186 of *APOE* $\epsilon 4$, after adjusting for the candidate CpG dinucleotide, were smaller than the
187 unadjusted ones. These 25 CpG were then evaluated in stage II, where we conducted a meta-
188 analysis across 235 $\epsilon 4+$ individuals assembled from four sample collections: ROS, MAP,
189 LBB and MSBB. We found 4 CpG associated with a pathologic diagnosis of AD (meta-
190 $P \leq 2 \times 10^{-4}$): cg08706567 (*MPL*), cg26884773 (*TOMM20*), cg12307200 (*LPP* and *TPRG1*),
191 and cg05157625 (*RIN3*) (**Figure 2B**). All of these 4 CpGs have stronger effects on AD
192 susceptibility in $\epsilon 4+$ than $\epsilon 4-$, which is consistently observed across cohorts (**Figure 2C,S1**).
193 For the cg08706567 and cg26884773, the LBB $\epsilon 4+$ dataset was not included into the meta-
194 analysis since it has the problem of infinite maximum likelihood estimates ($P=1$)²⁸, and the
195 results using the penalized generalized regression model yielded significant meta-P ($P \leq 0.01$)
196 including only the LBB and MSBB (**Table S3**). Thus, the independent cohorts offer evidence
197 of replication.

198 *3.2.2 Pathological associations of the 4 CpGs*

199 Aside from its strong associations with pTAU and NP, cg12307200 had weaker
200 associations with diffuse plaques (BETA=-3.15, SE=1.03, $P=2.34 \times 10^{-3}$), cerebral amyloid
201 angiopathy (BETA=-3.36, SE=1.19, $P=5.07 \times 10^{-3}$), and arteriosclerosis (BETA=-3.47,
202 SE=1.28, $P=6.73 \times 10^{-3}$); and it had no association with Lewy Bodies, hippocampal sclerosis,
203 or TDP-43 proteinopathy ($P > 0.05$). The other 3 CpG dinucleotides had weaker associations
204 with NP (BETA=[3.51,4.07], SE=[0.94,0.5], $P=[1.15 \times 10^{-4}, 2.14 \times 10^{-4}]$) than pTAU

205 (BETA=[8.88,9.81], SE=[1.49,1.66], P =[1.33×10^{-5} , 1.64×10^{-4}]), nominal associations with
206 TDP-43 proteinopathy (BETA=[2.28,3.24], SE=[1.36,1.51], P =[0.03,0.09]), and no
207 associations with other neuropathologic indices. Thus, the top 4 CpG dinucleotides were
208 primarily associated with Tau-related pathologies (Table S2).

209

210 *1.3. Derivation of the methylation PC based on the identified top 4 CpGs*

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

219

220 *1.4. Interaction between APOE $\epsilon 4$ and the top CpG dinucleotides on AD risk*

221 Here, we evaluated whether our epigenomic factors interacted with *APOE* $\epsilon 4$ to modulate
222 AD risk and compared their stratified effects on AD by $\epsilon 4$. The cg12307200 site did not
223 display significant evidence of interaction, but the 3 correlated CpG dinucleotides captured
224 by PC1 had nominal significance for $\epsilon 4$ interaction (range of meta- P for interaction=[6.1×10^{-4} ,
225 0.01]) (**Figure 3C**). This led us to pursue a more comprehensive interaction analysis of PC1
226 both categorically and continuously. For the categorical interaction, the original continuous
227 PC1 was transformed to a binary variable based on its median value. For the continuous

228 interaction, the non-scaled continuous value of the PC1 was used. The results were similar
229 for both the non-scaled and scaled values (**Table S4**), ruling out the potential influences of
230 outliers on the continuous analysis. Both the categorical (meta- $P=2.8 \times 10^{-4}$) and continuous
231 (meta- $P=7.7 \times 10^{-4}$) interaction tests were significant. In the categorical analysis, the effect of
232 $\epsilon 4$ on AD was smaller in the subjects with low PC1 (meta-OR=2.57, 95% CI=[1.51,4.39],
233 $P=5.13 \times 10^{-4}$) than the subjects with high PC1 (meta-OR=16.16, 95% CI=[6.45,40.51],
234 $P=2.93 \times 10^{-9}$). In the continuous analysis, the effect of PC1 on AD risk was greater in $\epsilon 4+$
235 (meta- $\ln(\text{OR})=0.87$, meta- $P=7.08 \times 10^{-6}$, meta- $N=235$) than in the $\epsilon 4-$ (meta- $\ln(\text{OR})=0.20$,
236 $P=2.65 \times 10^{-3}$, meta- $N=532$). In other words, reduction of one unit in PC1 was associated with
237 a 58% $((1-1/\exp(0.87)))$ attenuation in AD risk in $\epsilon 4+$ carriers. Thus, one could potentially
238 influence the magnitude of $\epsilon 4$ risk by modulating PC1, an epigenomic measure that captures
239 the effect of multiple loci.

240

241 *1.5. Functional exploration neocortical ROSMAP transcriptomes, replications in other*
242 *independent studies and validations by Hi-C sequencing data of fetal brains*

243 To understand the functional consequences of our 4 CpG dinucleotides, we then
244 conducted a Transcriptome-wide Association Study (TWAS) in the 421 ROSMAP
245 participants who have both DNA methylation and RNA sequence (RNA-seq) data from the
246 same cortical region. There were 71 genes which displayed significant ($P \leq 2.93 \times 10^{-6}$
247 (0.05/17,068 autosomal genes)) associations with either PC1 (70 genes) or cg12307200 (3
248 genes) (**Figure 4A&B, Table S5**). Except for *RIN3* where there was a cis-effect of the DNA
249 methylation on the gene expression, all the other associations are driven by the trans-effects

250 because the TWAS target genes are far from the 4 CpG dinucleotides: they are either on
251 different chromosomes or >5 Mb apart for those on the same chromosome.

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

260 To further explore the biological grounding of our observation, we accessed the
261 publically available Hi-C sequencing data from human fetal brain tissue which captures the
262 3-dimensional architecture of human cortex, with the caveat that this profiles a stage of brain
263 development. Despite this important limitation, we found evidence that those genomic
264 regions containing the 71 TWAS genes identified in ROSMAP have more contacts with the
265 region covering one of the three PC1 CpGs, cg08706567, than you would expect by chance
266 ($P<0.01$) (**Figure 4E**). This suggests that, even at very early stages of brain development, the
267 71 TWAS genes are physically interacting with at least one element of PC1; cg08706567
268 may play an important role as a regulator locus whose impact continues into advancing as it
269 influences the impact of $\epsilon 4$.

270

271 *1.6. The TWAS results implicate microglial activation in the effect on $\epsilon 4$*

272 *3.6.1 Pathway analysis of TWAS results suggested the involvement of the myeloid cells*

273 These 71 TWAS significant genes displayed enrichment for 20 KEGG functional
274 pathways (FDR \leq 0.05, hits $>$ 1 and hits% \geq 3%) (**Figure 5A**), which were related to multiple
275 aspects of immunity and particularly myeloid cell function, such as osteoclast differentiation
276 (hits=4%, FDR- $P=5.5\times 10^{-5}$), phagosome (hits=4%, FDR- $P=5.5\times 10^{-5}$), tuberculosis (hits=3%,
277 FDR- $P=5.5\times 10^{-5}$), Leishmaniasis (hits=4%, FDR- $P=1.3\times 10^{-3}$), and antigen processing and
278 presentation (hits=3%, FDR- $P=0.01$).

279 3.6.2 Association with co-expressed gene modules of microglia

280 We further conducted a complementary association analysis with the 7 modules of co-
281 expressed genes previously described as being enriched for microglial genes^{23,29}: m5, m113,
282 m114, m115, m116, m112, and m117. PC1 was associated with the expression of 5 of these 7
283 microglia modules ($P\leq 7.14\times 10^{-3}$ (0.05/7 microglia modules)) (**Figure 5B**), but not other non-
284 microglia modules (**Table S6**). Further, almost all (91%) of the 71 TWAS genes were found
285 in the 5 significant microglia modules, and 58% belonged to m116, the most microglial
286 enriched module; we previously reported²³ this module as being related to microglial aging²⁹.
287 It also contains some key AD genes such as *TREM2* and its binding partner *TYROBP*. m5 has
288 been associated the burden of Tau pathology and the proportion of activated microglia
289 (PAM), as defined using data from immunohistochemistry of brain sections and a standard
290 neuropathologic scale^{24,29}. Our findings do not appear to be driven by changes in microglial
291 cell counts since neither PC1 nor cg12307200 was associated with the microglia cell counts
292 estimated by either the immunohistochemistry staining for IBA1 protein or its mRNA
293 expression levels (**Table S7**). The results remain significant when we account for the
294 proportion of microglial cells. Overall, our 4 epigenomic factors appeared to be related to
295 microglial transcriptional programs captured by the modules of co-expressed genes defined

296 in neocortical data, and the state of the microglia may thus be an important factor in the
297 modulation of the $\epsilon 4$ effect.

298 *3.6.3 Validation using a histology-derived variable of microglia activation*

299 As noted above, the m5 module was associated with PAM, a trait derived from the
300 morphological characterization of microglia in histological sections that we found to be
301 associated with cognitive decline, AD pathologies, and AD dementia²⁴. PAM is simply the
302 proportion of microglia that have an activated stage III morphology, and this trait was
303 available in 122 ROSMAP participants that also have DNA methylation data from the same
304 frontal cortex region. PC1 was positively ($P=0.02$) and cg12307200 was negatively
305 associated with PAM in the midfrontal cortex ($P=0.05$) (**Figure 5C,D**). However, the
306 association is modest, and PC1 was not fully explained by PAM. That is, PAM and PC1
307 capture somewhat different aspects of microglial function that may be partially related to one
308 another. To be thorough, we also evaluated PAM measures in three other brain regions in
309 secondary analyses, and we found that PAM in inferior temporal cortex, posterior putamen
310 and ventral medial caudate also displayed association with PC1 and cg12307200 (**Figure S2**),
311 suggesting that PC1 and cg12307200 derived from cortical tissue DNA methylation data
312 capture an aspect of microglial state in multiple brain regions, not just the frontal cortex.

313

314 **4. Discussion**

315 Our study explored the epigenome of the human brain for CpG dinucleotides that
316 attenuate the impact of the strongest but not deterministic⁵ genetic risk of AD in the general
317 population: the *APOE* $\epsilon 4$ haplotype. Our staged approach identified 4 primarily Tau-related
318 CpGs and three of them were captured by one principal component (PC1). PC1 is positively

319 associated with the proportion of activated microglia and had a significant interaction with $\epsilon 4$
320 showing stronger effect in $\epsilon 4+$ than $\epsilon 4-$ subjects. Each unit reduction of PC1 in $\epsilon 4$ carriers
321 attenuated AD risk by 58%, suggesting that we may have found a meaningful $\epsilon 4$ attenuator.
322 Further work is now needed to further characterize PC1 to evaluate whether its effect could
323 potentially be mimicked by a therapeutic agent.

324 We took advantage of the large and varied sets of molecular and histological data
325 available in the same brain region of ROSMAP participants from which the DNA
326 methylation profile were obtained to begin to characterize the biology captured by our 4 CpG
327 dinucleotides. Our initial unbiased analysis clearly pointed towards the innate immune
328 system and particularly myeloid cells, leading us to focus our attention on this cell type as we
329 dissected the effect of the 4 CpGs. Namely, they seem to involve effects on two modules of
330 co-expressed neocortical genes: module 116 which relates to microglial aging and module 5
331 which is associated with proportion of activated microglia and contributes to the
332 accumulation of Tau pathology²⁹. We previously reported that the proportion of the activated
333 microglia was an independent risk factor for AD that had an effect size comparable to that of
334 $\epsilon 4$ on AD²⁴. Our current findings with the same dataset therefore refine our understanding of
335 the relationship between activated microglia and the $\epsilon 4$ allele: while their effects on risk may
336 be largely independent, they may interact to some degree.

337 Our findings that the top 4 CpG dinucleotides were primarily associated with Tau-related
338 pathologies are in line with previous genetic studies revealing the link between *APOE* and
339 *MAPT*, the gene encoding the Tau protein. A *MAPT* variant enriched in *MAPT* H2 carriers
340 conferred greater protection from AD in *APOE* $\epsilon 4-$ individuals^{7,30}. A recent *MAPT* haplotype
341 stratified GWAS identified a variant enriched in *APOE* $\epsilon 3$ -carriers and protective in *MAPT*

342 H1H1 individuals³¹. These findings alongside ours suggest the presence of either genetic or
343 epigenetic factors which behave in a differential manner either on the *APOE* or *MAPT*
344 genetic background to modify AD risk. Further, these context-specific factors could establish
345 a link between *APOE* (major risk factor for AD) and *MAPT* (major risk factor for
346 tauopathies).

347 Our study has several limitations. The sample size is limited even though we have
348 assembled the largest $\epsilon 4+$ study with both DNA methylation and neuropathology data.
349 Having low power is a common issue for all $\epsilon 4+$ subgroup studies⁷. Trying to circumvent
350 this issue, we utilized a staged approach followed by RNA- and histology-based validation.
351 The inclusion of the same ROSMAP subjects in both stage I and stage II may inflate our
352 results, although we use different traits in each stage and impose a 10 fold more stringent
353 significance threshold to address, in part, this issue. Finally, our cross-sectional design cannot
354 to determine the causality, which is a general limitation of all postmortem autopsy studies for
355 which we can only have one time point.

356

357 **5. Conclusions**

358 We reported that the deleterious effect of the strongest genetic risk (*APOE* $\epsilon 4$) for AD
359 may be attenuated by an epigenomic factor which could work, at least in part, through
360 alterations in the relative proportion of activated microglia and, subsequently, on the
361 accumulation of Tau pathology. Further mechanistic studies are necessary to validate our
362 results and demonstrate the sequence of events outlined in the hypothesis suggested by our
363 current findings.

364

365 **SUPPLEMENTAL MATERIAL CONTENTS**

366

367 **SUPPLEMENTAL METHODS (Text)**

368

369 **SUPPLEMENTAL TABLES**

370 **Table S1.** Population demographics included in each analysis across all studies.

371 **Table S2.** Summary statistics of the 25 CpG dinucleotides for their associations with the 11
372 common neuropathologies in all subjects of ROSMAP, and meta-interactions with APOE ϵ 4 on
373 pathological AD in all subjects across ROSMAP, LBB, and MSBB, and meta-associations with
374 pathological AD in subgroups of subjects carrying or not carrying ϵ 4 allele across ROSMAP,
375 LBB, and MSBB (in excel spreadsheet).

376 **Table S3.** Meta-analysis of cg08706567 and cg26884773 using penalized generalized regression
377 model.

378 **Table S4.** Interaction and association tests with scaled continuous variables.

379 **Table S5.** Summary statistics of the top 71 genes from the transcriptome wide association for the
380 PC1 and cg12370200 in ROSMAP with the adjustment of APOE ϵ 4 status MSBB (in excel
381 spreadsheet).

382 **Table S6.** Summary statistics of the association between the 47 gene modules and the PC1 and
383 cg12370200 in ROSMAP with the adjustment of APOE ϵ 4 status MSBB (in excel spreadsheet).

384 **Table S7.** Associations of PC1 and cg12307200 with microglia cell type proportion in subset of
385 subjects with immuno-histochemistry measurements (N=57).

386

387 **SUPPLEMENTAL FIGURES**

388 **Figure S1.** Forest plots of the association between the pathological diagnosis of Alzheimer's
389 disease (AD) and methylation level at the 4 top CpG dinucleotides.

390 **Figure S2.** Distributions of the scaled values of each of the top 4 CpG dinucleotides within each
391 cohort before the derivation of PC1 and their pairwise correlations.

392 **Figure S3.** Associations of the proportion of activated microglia (PAM) with the methylation
393 PC1 and cg12307200.

394

395 **DATA AVAILABILITY**

396 All the data and analysis output are available via the AD Knowledge Portal

397 (<https://adknowledgeportal.synapse.org>). The AD Knowledge Portal is a platform for accessing

398 data, analyses, and tools generated by the Accelerating Medicines Partnership (AMP-AD) Target

399 Discovery Program and other National Institute on Aging (NIA)-supported programs to enable

400 open-science practices and accelerate translational learning. The data, analyses and tools are

401 shared early in the research cycle without a publication embargo on secondary use. Data is

402 available for general research use according to the following requirements for data access and

403 data attribution (<https://adknowledgeportal.synapse.org/DataAccess/Instructions>). The link to the
404 data and analysis output for this manuscript is <https://www.synapse.org/#!Synapse:syn22240706>.

405

406

407 **ACKNOWLEDGMENTS**

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

414 **MSBB:** Brain banking and neuropathology assessments for the Mount Sinai cohort was
415 supported by US National Institutes of Health grants AG02219, AG05138, U01 AG046170, RF1
416 AG057440 and MH064673, and the Department of Veterans Affairs VISN3 MIRECC.

417 Replication work in Boston was supported by US National Institutes of Health grants: R01
418 AG036042, R01AG036836, R01 AG17917, R01 AG15819, R01 AG032990, R01 AG18023,
419 RC2 AG036547, P30 AG10161, P50 AG016574, U01 ES017155, KL2 RR024151 and K25
420 AG041906-01.

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

424 **MAYO:** Study data¹³ were provided by the following sources: The Mayo Clinic Alzheimers
425 Disease Genetic Studies, led by Dr. Nilüfer Ertekin-Taner and Dr. Steven G. Younkin, Mayo

426 Clinic, Jacksonville, FL using samples from the Mayo Clinic Study of Aging, the Mayo Clinic
427 Alzheimers Disease Research Center, and the Mayo Clinic Brain Bank. Data collection was
428 supported through funding by NIA grants P50 AG016574, R01 AG032990, U01 AG046139,
429 R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216, R01
430 AG003949, NINDS grant R01 NS080820, CurePSP Foundation, and support from Mayo
431 Foundation. Study data includes samples collected through the Sun Health Research Institute
432 Brain and Body Donation Program of Sun City, Arizona. The Brain and Body Donation Program
433 is supported by the National Institute of Neurological Disorders and Stroke (U24 NS072026
434 National Brain and Tissue Resource for Parkinsons Disease and Related Disorders), the National
435 Institute on Aging (P30 AG19610 Arizona Alzheimers Disease Core Center), the Arizona
436 Department of Health Services (contract 211002, Arizona Alzheimers Research Center), the
437 Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901 and 1001 to the
438 Arizona Parkinson's Disease Consortium) and the Michael J. Fox Foundation for Parkinsons
439 Research.

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

442

443 **DATA AVAILABILITY**

444 All the data and analysis output are available via the AD Knowledge Portal
445 (<https://adknowledgeportal.synapse.org>). The AD Knowledge Portal is a platform for accessing
446 data, analyses, and tools generated by the Accelerating Medicines Partnership (AMP-AD) Target
447 Discovery Program and other National Institute on Aging (NIA)-supported programs to enable
448 open-science practices and accelerate translational learning. The data, analyses and tools are

449 shared early in the research cycle without a publication embargo on secondary use. Data is
450 available for general research use according to the following requirements for data access and
451 data attribution (<https://adknowledgeportal.synapse.org/DataAccess/Instructions>). See data and
452 supporting information: <https://doi.org/10.7303/syn22240706>

453

454 **CONFLICTS OF INTEREST**

455 The authors declare no conflicts of interest.

456

457 **REFERENCE**

- 458 1. Farrer, L.A. *et al.* Effects of age, sex, and ethnicity on the association between apolipoprotein E
459 genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis
460 Consortium. *JAMA* **278**, 1349-56 (1997).
- 461 2. Sherva, R. & Farrer, L.A. Power and pitfalls of the genome-wide association study approach to
462 identify genes for Alzheimer's disease. *Curr Psychiatry Rep* **13**, 138-46 (2011).
- 463 3. Lambert, J.C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for
464 Alzheimer's disease. *Nat Genet* **45**, 1452-8 (2013).
- 465 4. Kunkle, B.W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk
466 loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* **51**, 414-430 (2019).
- 467 5. Meyer, M.R. *et al.* APOE genotype predicts when--not whether--one is predisposed to develop
468 Alzheimer disease. *Nat Genet* **19**, 321-2 (1998).
- 469 6. Ma, Y. & Ordovas, J.M. The integration of epigenetics and genetics in nutrition research for CVD
470 risk factors. *Proc Nutr Soc* **76**, 333-346 (2017).
- 471 7. Ma, Y. *et al.* Analysis of Whole-Exome Sequencing Data for Alzheimer Disease Stratified by APOE
472 Genotype. *JAMA Neurol* (2019).
- 473 8. De Jager, P.L. *et al.* Alzheimer's disease: early alterations in brain DNA methylation at ANK1,
474 BIN1, RHBDF2 and other loci. *Nat Neurosci* **17**, 1156-63 (2014).
- 475 9. De Jager, P.L. *et al.* A multi-omic atlas of the human frontal cortex for aging and Alzheimer's
476 disease research. *Sci Data* **5**, 180142 (2018).
- 477 10. Lunnon, K. *et al.* Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's
478 disease. *Nat Neurosci* **17**, 1164-70 (2014).
- 479 11. Smith, R.G. *et al.* Elevated DNA methylation across a 48-kb region spanning the HOXA gene
480 cluster is associated with Alzheimer's disease neuropathology. *Alzheimers Dement* **14**, 1580-
481 1588 (2018).
- 482 12. Zou, F. *et al.* Brain expression genome-wide association study (eGWAS) identifies human
483 disease-associated variants. *PLoS Genet* **8**, e1002707 (2012).
- 484 13. Allen, M. *et al.* Human whole genome genotype and transcriptome data for Alzheimer's and
485 other neurodegenerative diseases. *Sci Data* **3**, 160089 (2016).

- 486 14. McKhann, G. *et al.* Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work
487 Group under the auspices of Department of Health and Human Services Task Force on
488 Alzheimer's Disease. *Neurology* **34**, 939-44 (1984).
- 489 15. Consensus recommendations for the postmortem diagnosis of Alzheimer's disease. The National
490 Institute on Aging, and Reagan Institute Working Group on Diagnostic Criteria for the
491 Neuropathological Assessment of Alzheimer's Disease. *Neurobiol Aging* **18**, S1-2 (1997).
- 492 16. Bennett, D.A. *et al.* Neuropathology of older persons without cognitive impairment from two
493 community-based studies. *Neurology* **66**, 1837-44 (2006).
- 494 17. Yu, L., Boyle, P.A., Leurgans, S., Schneider, J.A. & Bennett, D.A. Disentangling the effects of age
495 and APOE on neuropathology and late life cognitive decline. *Neurobiol Aging* **35**, 819-26 (2014).
- 496 18. Bennett, D.A., Schneider, J.A., Arvanitakis, Z. & Wilson, R.S. Overview and findings from the
497 religious orders study. *Curr Alzheimer Res* **9**, 628-45 (2012).
- 498 19. Bennett, D.A. *et al.* Overview and findings from the rush Memory and Aging Project. *Curr*
499 *Alzheimer Res* **9**, 646-63 (2012).
- 500 20. White, C.C. *et al.* Identification of genes associated with dissociation of cognitive performance
501 and neuropathological burden: Multistep analysis of genetic, epigenetic, and transcriptional
502 data. *PLoS Med* **14**, e1002287 (2017).
- 503 21. Pidsley, R. *et al.* A data-driven approach to preprocessing Illumina 450K methylation array data.
504 *BMC Genomics* **14**, 293 (2013).
- 505 22. Allen, M. *et al.* Gene expression, methylation and neuropathology correlations at progressive
506 supranuclear palsy risk loci. *Acta Neuropathol* **132**, 197-211 (2016).
- 507 23. Mostafavi, S. *et al.* A molecular network of the aging human brain provides insights into the
508 pathology and cognitive decline of Alzheimer's disease. *Nat Neurosci* **21**, 811-819 (2018).
- 509 24. Felsky, D. *et al.* Neuropathological correlates and genetic architecture of microglial activation in
510 elderly human brain. *Nat Commun* **10**, 409 (2019).
- 511 25. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree
512 of life. *Nucleic Acids Res* **43**, D447-52 (2015).
- 513 26. Benjamini Y, H.Y. Controlling the false discovery rate: a practical and powerful approach to
514 multiple testing. *J. Roy. Statist. Soc.* **57**, 289-300 (1995).
- 515 27. Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing
516 human brain. *Nature* **538**, 523-527 (2016).
- 517 28. Lesaffre, E.A., A. Partial separation in logistic discrimination. *Journal of the Royal Statistical*
518 *Society. Series B. Methodological* **1**, 109-116 (1989).
- 519 29. Patrick, E.T., M.; Ergun, A.; Ng, B.; Casazza, W.; Cimpean, M.; Yung, C.; Schneider, J.A.; Bennett,
520 D.A.; Gaiteri, C.; De Jager P.L.; Bradshaw, E.M.; Mostafavi, S. Deconvolving the contributions of
521 cell-type heterogeneity on cortical gene expression. *bioRxiv* (2019).
- 522 30. Jun, G. *et al.* A novel Alzheimer disease locus located near the gene encoding tau protein. *Mol*
523 *Psychiatry* **21**, 108-17 (2016).
- 524 31. Strickland, S.L. *et al.* MAPT haplotype-stratified GWAS reveals differential association for AD risk
525 variants. *Alzheimers Dement* (2020).

526

527 **Tables:**
 528 **Table 1. Demographics of subjects carrying or not carrying *APOE ε4* allele across ROSMAP, LBB and MSBB**

	ROSMAP			LBB			MSBB	
	ε4+	ε4-	<i>P</i>	ε4+	ε4-	<i>P</i>	ε4+	ε4-
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)
Female [§]	93 (59%)	268 (65%)	0.23	25 (69%)	21 (66%)	0.94	27 (66%)	53 (60%)
AD cases [§]	128 (81%)	221 (53%)	2.48x10 ⁻⁹	33 (92%)	19 (59%)	0.004	34 (83%)	40 (45%)
pTAU (sqrt)*	2.75 (1.67)	1.92 (1.16)	2.61x10 ⁻⁸	NA	NA	NA	NA	NA
NP*	1.20 (0.92)	0.64 (0.73)	5.60x10 ⁻¹¹	NA	NA	NA	NA	NA

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

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

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

534 Abbreviations: pTAU, abnormally phosphorylated Tau protein, AT8.

535

536 **Figure 1. Flow chart of analysis plan and results.** Abbreviations: MWAS, methylome-wide
537 association analysis; NP, neuritic plaque; pTAU, abnormally phosphorylated Tau protein, AT8;
538 TWAS, transcriptome-wide association analysis; IHC, immuno-histochemistry measurements.

539
540 **Figure 2. Identification of the top 4 CpG dinucleotides.** (A) Miami plot shows the results of
541 the methylation genome-wide association study (MWAS) on pTAU (upper panel in blue) and NP
542 (lower panel in green) in ROSMAP (N=572). The Y axis show the $-\log_{10}$ transformed P value of
543 each of the genome-wide CpG dinucleotides which are shown in dots and ordered according to
544 their genomic coordinates on the shared X axis. The red dashed line represents the significance
545 threshold ($P=1E-8=0.05/420132/10$) and those CpG dinucleotides passing the genome-wide
546 significance threshold for either pTAU or NP are shown in red dots. (B) Candidate manhattan
547 plot shows the meta-analyzed associations of the above selected 25 CpG dinucleotides with
548 pathological AD in $\epsilon 4+$ subjects across ROSMAP, LBB, and MSBB cohorts (N=235). Y axis
549 represent the $-\log_{10}$ transformed meta analyzed P values of the regression coefficient estimates
550 of the logistic regression model, in which the outcome variable is the pathological diagnosis of
551 Alzheimer's disease (AD) (no=0 and yes=1), the exposure variable is the methylation status of
552 each CpG dinucleotides (0 to 1) and the covariates include age at death, postmortem interval, sex,
553 and study, ethnicity principle components, methylation experiment batches in ROSMAP, and
554 cell proportion in LBB and MSBB. The horizontal red dashed line represents the Bonferroni
555 corrected P value threshold of $(0.05/(25*10) = 2 \times 10^{-4})$ and the top 4 significant CpG
556 dinucleotides are represented in red squares with the annotation of their closest genes. (C) Forest
557 plots for the top 4 CpG dinucleotides across different cohorts for their associations with AD
558 within: 1) all the subjects; 2) subjects carrying; or 3) not carrying the *APOE* $\epsilon 4$ allele; and 4) the
559 interaction test between the methylation level and *APOE* $\epsilon 4$ allele carrying status within all the
560 subjects. The filled square and horizontal line for each population or the filled summary
561 diamonds (blue for the meta-analysis across the replication cohorts while red for the joint meta-
562 analysis across all the cohorts) denote the estimated regression coefficient (BETA) and its 95%
563 CI per unit increase in the methylation level of each CpG dinucleotide or the interaction term of
564 the methylation times the $\epsilon 4$ -carrying status (yes=1 and no=0) for the binary outcome variable of
565 the pathological diagnosis of AD (no=0 and yes=1) with the adjustment of the covariates of age
566 at death, postmortem interval, sex, study, ethnicity principle components, methylation
567 experiment batches in ROSMAP, and cell proportion in LBB and MSBB. The arrows indicate
568 the estimates are out of the boundaries.
569 Abbreviations: NP, neuritic plaque; pTAU, abnormally phosphorylated Tau protein, AT8.

570 **Figure 3. Consistent correlations among the 4 CpG dinucleotides across cohorts and the**
571 **derivation of the PC1.** (A) Circos plot shows the consistent distributions of each of the top 4
572 CpG dinucleotides and their mutual correlations across ROSMAP (pink sector), LBB (green
573 sector) and MSBB (blue sector). In the outer layer, the distribution of each CpG dinucleotide
574 within each cohort is shown in a violin plot where each dot represents a subject and the black
575 horizontal and vertical lines denote the mean and standard deviation. Red dots represent those
576 subjects with hypermethylation level (≥ 0.5) while blue ones represent those subjects with
577 hypomethylation level (< 0.5). In the inner layer, the correlations between each pair of 2
578 CpG dinucleotides within each cohort were represented by their connected lines, where the color
579 denote the direction (red for $r > 0$ and green for $r < 0$) and the thickness denote the strength of their
580 correlations (thicker lines represent stronger correlations). (B) Variance plot of the PC1 vs. PC2
581 and their contributions to each of the top 4 CpG dinucleotides across ROSMAP (red solid arrow),

582 LBB (blue dashed arrow), and MSBB (blue dashed arrow). (C) Interaction effect on AD between
583 *APOE* $\epsilon 4$ and PC1. The subgroup of $\epsilon 4+$ and $\epsilon 4-$ are represented as blue and orange. Interaction
584 with categorical variables are shown on the left panel where the continuous variable of PC1 were
585 transformed to the binary variable based on the median value. The Y axis represent the
586 percentage of AD cases across 4 subgroups of subjects. The right panel show the interactions
587 with the untransformed continuous PC1. the Y axis represent the predicted ln(OR) of AD
588 calculated based on the statistics within $\epsilon 4+$ and $\epsilon 4-$ subgroup using the logistic regression model
589 with binary AD status (case=1 and control=0) as the outcome variable adjusting the covariates of
590 age at death, sex, study, postmortem interval, methylation experimental batches and two major
591 ethnic principles.
592 Abbreviations: OR, odds ratio.
593

594 **Figure 4. Discovery and replication of TWAS results of the PC1 and cg12307200.** (A) and
595 (B) Volcano plot shows the TWAS results of PC1 and cg12307200. The X axis shows the BETA
596 and the Y axis shows the $-\log_{10}$ transformed P values for the exposure variable of PC1 or
597 cg12307200 on the mRNA expression levels of each gene. The 71 unique genes (70 for PC1 and
598 3 for cg12307200) passed the Bonferroni corrected significance threshold of $P \leq 2.93 \times 10^{-6}$
599 ($0.05/17,068$ autosomal genes) are shown in red dots with their gene name connected through
600 blue arrows. (C) and (D) Replication of the PC1 TWAS results in MSBB and MAYO. ROSMAP
601 has identified 70 TWAS genes for PC1 and 68 of them are available in MSBB. There are 25 (out
602 of 70) genes are significant for cg05157625 (the only CpG dinucleotide available in MAYO) and
603 21 of them are available in MAYO. The scatter plot shows the comparison of each gene between
604 ROSMAP and MSBB (or MAYO) where the X axis represents the ratio of the regression
605 coefficient obtained in the two cohorts (ROSMAP over MSBB (or MAYO) and the red dashed
606 vertical lines for the ratio=0 or 1) and the Y axis shows the P values in MSBB (or MAYO) (red
607 dashed horizontal line for the $P=0.05$). The blue dots are those genes with nominal significance
608 ($P \leq 0.05$) in MSBB (or MAYO) and with the same effect direction in both ROSMAP and MSBB
609 (or MAYO), the green dots are those non-significant genes ($P > 0.05$) in MSBB (or MAYO) but
610 with the same effect direction, and the red dots are those non-significant genes in MSBB (or
611 MAYO) and also with the opposite effect direction. The number and their percentage of these
612 three groups of genes are presented in the bar plot (upper right corner) with the same color
613 codings. (E) Different genomic contacts by TWAS regions in the fetal brains analyzed with the
614 published Hi-C sequencing data (Won et al., 2016) downloaded from GEO (GSE77565). Silver
615 and rose gold bars represent those regions outside and within the 70 significant TWAS genes in
616 ROSMAP and the P value for the differences between these two groups for each CpG
617 dinucleotide are shown in X axis.
618

619 **Figure 5. Microglia relevance to PC1 and cg12307200.** (A) Plot of the KEGG pathway
620 analysis of the top 71 TWAS genes which have significant associations with the PC1 and
621 cg12307200. Only those pathways with FDR adjusted P value ≤ 0.05 and the number
622 (percentage) of the hit genes > 1 ($\geq 3\%$) were shown in the plot. Y axis list the name of these
623 KEGG functional pathways and X axis shows their corresponding $-\log_{10}$ transformed FDR
624 adjusted P values. The size of the pie represents the number of the member genes of each
625 functional pathway, which are categorized into four types ranging from the smallest to the largest
626 containing ≤ 40 , >40 & ≤ 60 , >60 & ≤ 100 , and >100 member genes. Each pie was split into 2
627 slices with the area of the blue slice represents the percentage of the hit genes out of the total

628 member genes for each pathway and their number are shown for each pathway. (B) Scatter plot
629 of the associations with the methylation factors (PC1 in red and cg12307200 in blue) and
630 expression levels of the 7 reported gene modules for microglia. Y axis represent the $-\log_{10}$
631 transformation of the P values of the associations between the exposure variables of PC1 or
632 cg12307200 on the outcome variable of the expression values of each gene module, and the X
633 axis represent the module membership, which is defined as the percentage of the module
634 member genes out of our identified 43 gene list (28 genes are not mapped to any of the reported
635 gene modules). The horizontal red and blue dashed lines represent the Bonferroni corrected
636 significance threshold of $P \leq 7.14 \times 10^{-3}$ (0.05/7 gene modules) and the nominal significance
637 threshold of $P \leq 0.05$. (C) and (D) show the scatter plot of the associations between proportion of
638 activated microglia (PAM) at midfrontal cortex as Y axis and the methylation factors (PC1 and
639 cg12307200) as X axis. Each dot is one observation and the regression line and its 95% CI are
640 represented by the blue line and grey area. The estimated statistics are shown on the upper right
641 corner of the plot. The BETA, SE, and P represent the estimated regression coefficient and its
642 standard error, P value of the exposure variable of PC1 or cg12307200 on the outcome of PAM.
643 The N represent the sample size within the analysis. For all the analysis, we used the generalized
644 linear regression model with the adjustments of the age at death, sex, postmortem interval, study,
645 *APOE* $\epsilon 4$ binary status, two major ethnic principles and methylation experimental batches.
646

STAGE I: Associate with AD & attenuate *APOE* $\epsilon 4$ effect in all subjects

Method

Analysis: MWAS

Trait: pTAU & NP

Cohort: ROSMAP

***APOE* $\epsilon 4$ status:** $\epsilon 4+$ and $\epsilon 4-$

Sample size: 572

Regression models: Model 1: Trait $\sim \alpha_1 + \beta_{11}\epsilon 4 + \beta_{1s}\text{covariates}$
Model 2: Trait $\sim \alpha_2 + \beta_{21}\epsilon 4 + \beta_{22}\text{CpG} + \beta_{2s}\text{covariates}$

Rationale: The CpG site reduces the $\epsilon 4+$ vs. $\epsilon 4-$ effect on AD and it is also significantly associated with AD.

Result

420,132 CpG sites

$P(\beta_{22}) \leq 1 \times 10^{-8}$
(0.05/420132/10) &

$\delta(\beta_{21} - \beta_{11}) < 0$

25 CpG sites

STAGE II: Associate with AD in *APOE* $\epsilon 4+$ subjects

Method

Analysis: Meta-analysis

Trait: AD

Cohort: ROSMAP, LBB, and MSBB

***APOE* $\epsilon 4$ status:** $\epsilon 4+$ only

Sample size: 235

Regression model: Trait $\sim \alpha + \beta\text{CpG} + \beta\text{covariates}$

Rationale: The CpG site is associated with AD in $\epsilon 4+$ subjects.

Result

25 CpG sites

$P \leq 2 \times 10^{-4}$
(0.05/25/10)

4 CpG sites, PC1

STAGE III: Functional explorations

Method

Analysis: TWAS and pathway analysis

Trait: 17,068 autosomal genes

Cohort: ROSMAP

***APOE* $\epsilon 4$ status:** $\epsilon 4+$ and $\epsilon 4-$

Sample size: 421

Regression model: Trait $\sim \alpha + \beta\text{CpG} + \beta\epsilon 4 + \beta\text{covariates}$

Rationale: DNA methylation affects gene transcription.

Result

4 CpG sites, PC1

$P \leq 2.93 \times 10^{-6}$
(0.05/17,068)

71 Genes, myeloid cells related innate immune pathways

Method

Analysis: Co-expressed gene module & IHC validation

Trait: 7 microglia modules

Cohort: ROSMAP

***APOE* $\epsilon 4$ status:** $\epsilon 4+$ and $\epsilon 4-$

Sample size: 413 & 136

Regression model: Trait $\sim \alpha + \beta\text{CpG} + \beta\epsilon 4 + \beta\text{covariates}$

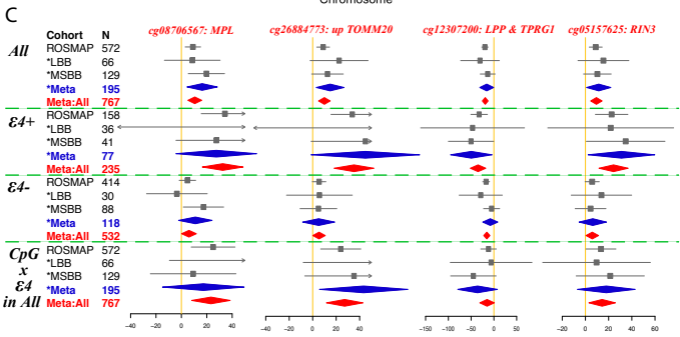
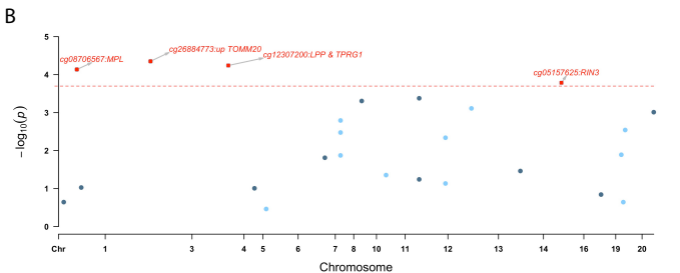
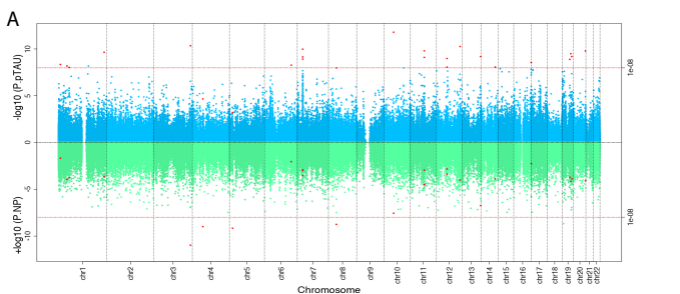
Rationale: Different cell types have different expressed genes.

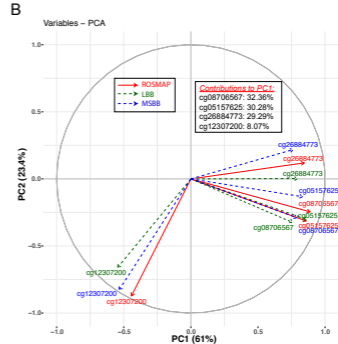
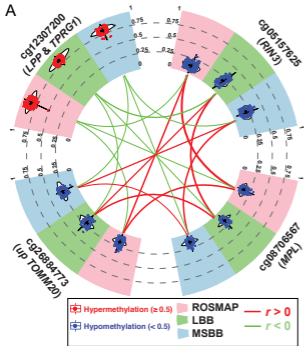
Result

4 CpG sites, PC1

$P \leq 7.14 \times 10^{-3}$
(0.05/7)

Microglia



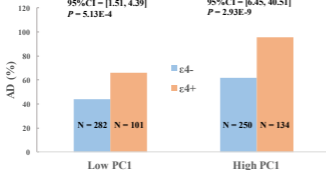


C PC1

P for interaction = $2.82E-4$

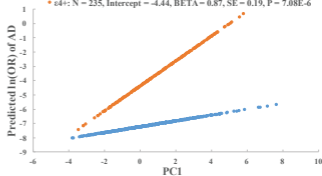
$\epsilon 4^-$ OR = 1.57
 95% CI = [1.51, 4.39]
 $P = 5.13E-4$

$\epsilon 4^+$ OR = 16.16
 95% CI = [6.45, 40.51]
 $P = 2.93E-9$

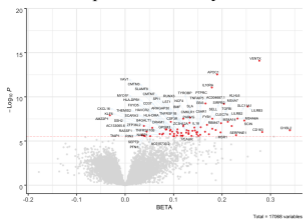


P for interaction = $7.73E-4$

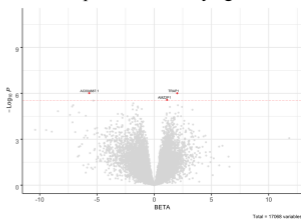
$\epsilon 4^-$: N = 532, Intercept = -7.28, BETA = 0.20, SE = 0.07, $P = 2.65E-3$
 $\epsilon 4^+$: N = 235, Intercept = -4.44, BETA = 0.87, SE = 0.19, $P = 7.08E-6$



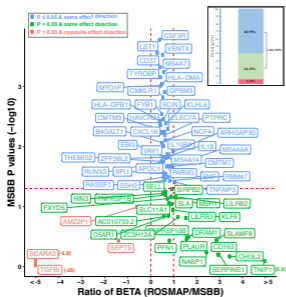
A Volcano plot of TWAS by PC1



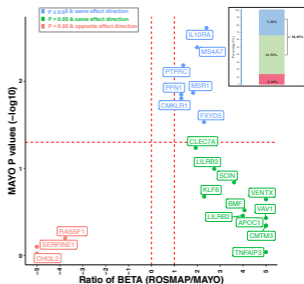
B Volcano plot of TWAS by cg12307200



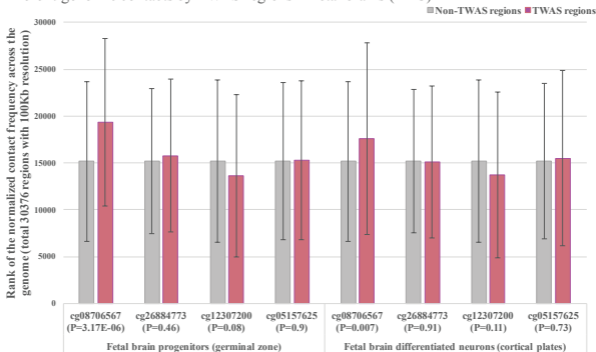
C Replication of PC1 TWAS in MSBB



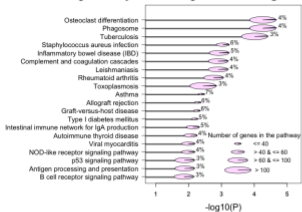
D Replication of PC1 TWAS in MAYO (specific to *RIN3* cg05157625)



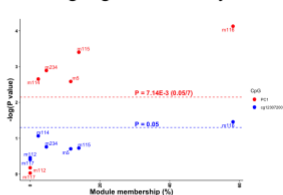
E Different genomic contacts by TWAS regions in fetal brains (Hi-C)



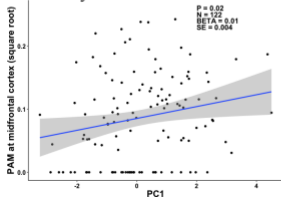
A KEGG pathway of the top 71 TWAS genes



B Microglia gene modules by PC1



C PAM by PC1



D PAM by cg12307200

