

1 **Title Page**

2 **Title:**

3 **Multi-omics Characterization of Epigenetic and Genetic Risk of Alzheimer Disease in Autopsied**
4 **Brains from two Ethnic Groups**

5
6 **Authors:**

7 Yiyi Ma¹⁻³, Dolly Reyes-Dumeyer¹⁻³, Angel Piriz¹, Patricia Recio⁴, Diones Rivera Mejia^{4,5}, Martin
8 Medrano⁶, Rafael A. Lantigua^{1,7}, Jean Paul G. Vonsattel^{1,8}, Giuseppe Tosto^{1,2}, Andrew F. Teich^{1,3,8},
9 Benjamin Ciener^{1,3,8}, Sandra Leskinen^{1,3,8}, Sharanya Sivakumar^{1,3,8}, Michael DeTure⁹, Duara Ranjan⁹,
10 Dennis Dickson⁹, Melissa Murray⁹, Edward Lee¹⁰, David A. Wolk¹⁰, Lee-Way Jin¹¹, Brittany N.
11 Dugger¹¹, Annie Hiniker¹², Robert A. Rissman¹², Richard Mayeux*,¹⁻³ Badri N. Vardarajan.*¹⁻³

12
13 **Affiliations:**

- 14 1. Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Vagelos College of
15 Physicians and Surgeons, Columbia University, New York, NY
- 16 2. G.H. Sergievsky Center, Vagelos College of Physicians and Surgeons, Columbia University, New
17 York, NY
- 18 3. Department of Neurology, Vagelos College of Physicians and Surgeons, Columbia University, and the
19 New York Presbyterian Hospital, New York, NY
- 20 4. CEDIMAT, Santo Domingo, Dominican Republic
- 21 5. Universidad Pedro Henríquez Ureña, Santo Domingo, Dominican Republic
- 22 6. Pontificia Universidad Católica Madre y Maestra (PUCMM), Santiago, Dominican Republic.
- 23 7. Department of Medicine, Vagelos College of Physicians and Surgeons, Columbia University, and the
24 New York Presbyterian Hospital, New York, NY

- 25 8. Department of Pathology and Cell Biology, Vagelos College of Physicians and Surgeons, Columbia
26 University, New York, NY
- 27 9. Department of Neuroscience, Mayo Clinic Florida, Jacksonville, Florida, USA
- 28 10. Department of Neurology and Penn Alzheimer’s Disease Research Center, Perelman School of
29 Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA.
- 30 11. Department of Pathology and Laboratory Medicine, School of Medicine, University of California
31 Davis, Sacramento, CA 95817, USA
- 32 12. Keck School of Medicine of the University of Southern California, Los Angeles, CA

33

34 * **Contributed equally to this work.**

35 * **Correspondence:**

36 Richard Mayeux, MD, MSc,

37 Chair of the Department of Neurology, Vagelos College of Physicians and Surgeons, Columbia

38 University, and the New York Presbyterian Hospital, 630 West 168th street, New York, NY 10032,

39 USA; Tel: 212-305-2391; Email: rpm2@cumc.columbia.edu

40

41 Badri N. Vardarajan, Ph.D. MS

42 Assistant Professor of Neurological Science of the Department of Neurology, Vagelos College of

43 Physicians and Surgeons, Columbia University, and the New York Presbyterian Hospital, 630 West

44 168th street, New York, NY 10032, USA; Tel: 212-342-5449; Email: bnv2103@cumc.columbia.edu

45

46 **Abstract: 334 words; Text: 2,918 words**

47

48

49

50

51

52 **Abstract**

53

54 **Background:**

55 Both genetic variants and epigenetic features contribute to the risk of Alzheimer's disease (AD).

56 We studied the AD association of CpG-related single nucleotide polymorphisms (CGS), which

57 act as the hub of both the genetic and epigenetic effects, in Hispanics decedents and generalized

58 the findings to Non-Hispanic Whites (NHW) decedents.

59

60 **Methods:**

61 First, we derived the dosage of the CpG site-creating allele of multiple CGSes in each 1 KB

62 window across the genome and we conducted a sliding window association test with clinical

63 diagnosis of AD in 7,155 Hispanics (3,194 cases and 3,961 controls) using generalized linear

64 mixed models with the adjustment of age, sex, population structure, genomic relationship matrix,

65 and genotyping batches. Next, using methylation and bulk RNA-sequencing data from the

66 dorsolateral pre-frontal cortex in 150 Hispanics brains, we tested the cis- and trans-effects of AD

67 associated CGS on brain DNA methylation to mRNA expression. For the genes with significant

68 cis- and trans-effects, we checked their enriched pathways.

69

70 **Results:**

71 We identified six genetic loci in Hispanics with CGS dosage associated with AD at genome-wide

72 significance levels: *ADAM20* (Score=55.2, $P=4.06 \times 10^{-8}$), between *VRTN* (Score=-19.6,

73 $P=1.47 \times 10^{-8}$) and *SYNDIGIL* (Score=-37.7, $P=2.25 \times 10^{-9}$), *SPG7* (16q24.3) (Score=40.5,

74 $P=2.23 \times 10^{-8}$), *PVRL2* (Score=125.86, $P=1.64 \times 10^{-9}$), *TOMM40* (Score=-18.58, $P=4.61 \times 10^{-8}$),

75 and *APOE* (Score=75.12, $P=7.26 \times 10^{-26}$). CGSes in *PVRL2* and *APOE* were also genome-wide

76 significant in NHW. Except for *ADAM20*, CGSes in all the other five loci were associated with
77 Hispanic brain methylation levels (mQTLs) and CGSes in *SPG7*, *PVRL2*, and *APOE* were also
78 mQTLs in NHW. Except for *SYNDIGIL* ($P=0.08$), brain methylation levels in all the other five
79 loci affected downstream RNA expression in the Hispanics ($P<0.05$), and methylation at *VRTN*
80 and *TOMM40* were also associated with RNA expression in NHW. Gene expression in these six
81 loci were also regulated by CpG sites in genes that were enriched in the neuron projection and
82 synapse (FDR<0.05).

83

84 **Conclusions:**

85 We identified six CpG associated genetic loci associated with AD in Hispanics, harboring both
86 genetic and epigenetic risks. However, their downstream effects on mRNA expression maybe
87 ethnic specific and different from NHW.

88

89

90 **Keywords:** Alzheimer's disease; Genetics; Epigenetics; Hispanics; Non-Hispanic Whites; CpG-
91 related single nucleotide polymorphism

92

93 **Introduction:**

94 Alzheimer disease (AD) is a chronic and progressive neurodegenerative disorder accompanied
95 by cognitive decline that gradually worsen over years. The etiology of AD is complex involving
96 different molecular mechanisms, which may be the result of not only heritable genetic risks but
97 also by factors that act on the epigenome. The advancement in identifying genetic contributions
98 to AD has also piqued interest in epigenetic contributions. The most recent genome-wide
99 association study (GWAS) of AD reported over 70 genetic loci for AD risk[1]. Candidate gene
100 and genome-wide DNA methylation studies have implicated approximately 21 genetic loci with
101 differential methylation levels associated with AD[2].

102
103 Loci identified in both genetic and epigenetic studies[3] suggest a common molecular hub that
104 captures causal risk factors for AD. *APOE* $\epsilon 4$ is most consistently confirmed genetic risk factor
105 for AD and the methylation levels of the CpG island within *APOE* were found to be lower in AD
106 brains compared to brains from healthy participants[4]. The study of CpG-related single
107 nucleotide polymorphism (CGS) may identify alleles disrupting the CpG dinucleotides leading to
108 the removal of the DNA methylation targeted site. We previously found CGSes in the MS4A
109 region have a dose-dependent effect on AD in persons who identify as non-Hispanic White
110 (NHW) [3]. We identified a statistical association between MS4A CGSes and DNA methylation
111 levels using blood samples and clinical AD but we and others were limited in the downstream
112 functional validation for the top loci.

113
114 In this study, we conducted a systematic analysis based on CGS in persons who identified as
115 Hispanics who enrolled Washington Heights-Inwood Columbia Aging Project (WHICAP) or the

116 Estudio Familiar de Influencia Genética en Alzheimer (EFIGA)[5]. We have studied AD in
117 numerous cohorts of persons who have identified as Hispanics and understand their complex
118 ancestry. We began by using a genome-wide sliding window approach to prioritize genetic loci
119 comprised of CGSes associated with the risk of clinical diagnosis of AD in 7,155 Hispanics
120 decedents. The prioritized loci from the genome-wide analyses were followed by detailed
121 functional studies. We analyzed both the cis- and trans-effects of the molecular mechanisms
122 from genetics to DNA methylation and mRNA gene expression in postmortem brain tissue from
123 Hispanics decedents.

124

125 **Methods:**

126 *Study description:*

127 Cohorts included for genetic studies

128 We included 7,155 Hispanics decedents from the Washington Heights-Inwood Columbia Aging
129 Project (WHICAP)[6] and the Estudio Familiar de Influencia Genética en Alzheimer (EFIGA)[7].
130 The WHICAP study is an ongoing prospective, community-based, multiethnic longitudinal study
131 of Medicare beneficiaries 65 years and older residing in northern Manhattan (Washington
132 Heights, Hamilton Heights, and Inwood). All the participants underwent a comprehensive
133 examination including the assessment of general health and function, standardized physical and
134 neurological examination, and a neuropsychological battery of tests. Follow-up visits were
135 performed every 1.5-2 years, repeating similar examinations. Initiated in 1998, the EFIGA
136 recruited individuals of Caribbean Hispanic ancestry including familial and sporadic AD. The
137 individuals were recruited in New York City using local newspapers, the local Caribbean
138 Hispanic radio station, and postings throughout the Washington Heights-Inwood neighborhood.

139 AD was defined as any individual meeting NINCDS-ADRDA criteria for probable or possible
140 AD[8]. The severity of dementia was rated according to the Clinical Dementia Rating[9].

141
142 We also included 1,283 NHW from the Religious Order Study and the Memory & Aging Project
143 (ROSMAP) study to analyze whether the top loci identified in the Hispanics can be generalized
144 to NHW. ROSMAP recruit older individuals without known dementia and their detailed
145 information of both ante-mortem and post-mortem phenotyping were collected[10]. For this
146 study, we have included in total 1,283 NHW with whole genome sequencing data and clinical
147 diagnosis of AD.

148
149 Cohorts with brain DNA methylation and RNA sequencing (RNA-seq):

150 New York Brain Bank (NYBB): Brain tissue came from The NIA Alzheimer's disease Family
151 Based Study (NIA-AD FBS), WHICAP, EFIGA, and NCRAD. The NIA-AD FBS included
152 9,682 family members, and 1,096 unrelated, nondemented elderly from different race/ethnicity
153 groups from 1,756 families with suspected AD. NCRAD included unaffected individuals from
154 families with a history of AD. A description of the families has been previously detailed in a
155 report [5].

156
157 University of California, Davis Alzheimer's Disease Center (UCD ADC): With the aim to
158 conduct a research on diversity and risk of AD dementia, UCD applied an active community
159 outreach approach to recruit the individuals from the communities of Alameda, Contra Costa,
160 Sacramento, San Joaquin, Solano, and Yolo County. The overall percentage of Hispanic
161 individual 60 years of age and older residing in these counties ranged from 7.1% to 14.3% [11].

162

163 Florida Autopsied Multi-Ethnic (FLAME) cohort: The FLAME cohort is derived from the State
164 of Florida brain bank housed at the Mayo Clinic Florida [12]. The FLAME cohort consists of a
165 total of 2,809 autopsied individuals with a wide range of neurodegenerative diseases, who were
166 self-identified as Hispanic/Latino, black/African, and non-Hispanic white/European. The fixed
167 hemi-brain (typically left hemisphere) was weighed and the frontal cortex was cut and then
168 placed in 10% formalin solution.

169

170 The University of Pennsylvania Integrated Neurodegenerative Disease Biobank: Patients with
171 neurodegenerative disease are recruited into the autopsy program by the different clinical cores.
172 The subjects selected for the autopsy were followed in the clinical centers with detailed clinical
173 information and most of them were also collected with biofluid, neuroimaging, and genetic
174 data/samples. The left hemisphere and brain stem were immersed in 10% neutral buffered
175 formalin for 2 weeks whereas the right hemisphere was sliced coronally and frozen.

176

177 University of California, San Diego Alzheimer's Disease Research Center (UCSD ADC):
178 Postmortem tissue from the center's longitudinally followed cohort was used for this study.
179 Blocks of tissue were provided from autopsy verified cases after fixation in 10% formalin for 4
180 weeks. Cases were selected by using detailed clinical, biomarker and demographics information
181 collected at visits.

182

183 The Religious Order Study and the Memory & Aging Project (ROSMAP): We have included
184 516 NHW who have measurements of both postmortem brain DNA methylation and RNA

185 sequencing (RNA-seq) from their postmortem brain tissues. The details of both datasets were
186 described previously [10]. In brief, the grey matter from the dorsolateral prefrontal cortex
187 (DLPFC) were dissected while still frozen. RNA was extracted for transcriptome library
188 construction following the dUTP protocol and Illumina sequencing. The extracted DNA were
189 processed on the Illumina Infinium HumanMethylation450 BeadChip.

190

191 An informed consent was signed by the participant and/or legal guardian of the individuals
192 included into this study. IRB approval was approved by each institution.

193

194 *Genotype data and annotations of CpG-related SNPs (CGS):*

195 We annotated the CpG-related single nucleotide polymorphisms (CGS) by checking the three
196 base pairs flanking each single nucleotide polymorphism (SNP) and derived the dosage of the
197 CpG site-creating allele of multiple CGSes in each 1 Kb window across the genome. We
198 imputed the genotype according to Haplotype Reference Consortium (HRC) reference panels in
199 7,155 Caribbean Hispanics. We used the whole genome sequencing data of 1,283 NHW.

200

201 *Brain DNA methylation data:*

202 The genome-wide DNA methylation profile was measured by the Infinium MethylationEPIC Kit
203 (Illumina). For each sample we checked the control probes, sex mismatches, contamination, and
204 genotype outliers calling to identify and remove those samples failed quality control (QC). We
205 kept those CpG sites with detection P value < 0.01 across all the qualified samples and masked
206 those sample-specific CpG site with new detection $P > 0.01$ [13]. We further removed those CpG
207 sites reported to have cross-hybridization problems [14, 15] and those polymorphic CpG sites

208 [15, 16]. We further corrected the dye bias for all the qualified CpG probes. Finally, there were
209 179 samples from person of Hispanic descent with 675,583 autosomal probes passed the QC
210 which were included into the current study.

211
212 Brain RNA-seq data: Total RNA was extracted using Qiagen's RNeasy Mini Kit and then was
213 sent to the New York Genome Center for transcriptome library construction, which was
214 sequenced on a NovaSeq 6000 flow cell using 2x100bp cycles, targeting 60 million reads per sample.
215 All the samples included into the analysis passed QC metrics using *FastQC*. Gene counts were calculated
216 using the function *featureCounts*. We applied *ComBat-seq* to correct batch effects. As a result, a total of
217 58,942 genes passed QC metrics and exhibiting non-zero expression across all participants.

218
219 *Statistical analysis:*

220 We scaled the dosage of each window to fit into the value from zero to 2 and then tested the
221 scaled dosage of each window for association with AD using generalized linear mixed models
222 (GLMMs) implemented in the generalized linear mixed model association tests (GMMAT) [17]
223 with the adjustments for age, sex, population substructure, genomic relationship matrix (GRM),
224 and genotyping batches. Genome-wide significance threshold was $p < 5.0 \times 10^{-8}$. For the mQTL
225 analysis, there were 112 samples from persons of Hispanic decent and 571 from persons of
226 NHW decent with both genotype and brain DNA methylation data. We conducted the
227 generalized linear model adjusting for the age at death, sex, and technical covariates of
228 genotyping and methylation batches, and methylation chip and position. We analyzed both the
229 cis- and trans-effect of the DNA methylation on gene expression. For the cis-effect of DNA
230 methylation on gene expression, we conducted a highly adaptive sum of powered score-weighted
231 test (aSPUw) to collapse all the available CpG sites within 100 Kb distance to the gene (from 50

232 Kb upstream of the transcription start site and 50 Kb downstream of the end site of the gene
233 according to GENCODE v44 (GRCh37) annotation) with the adjustment of age, sex, and
234 technical covariates of methylation chip ID and chip position. For the trans-effect of DNA
235 methylation on gene expression, we used the linear mixed model to control for the random effect
236 of methylation array, the fixed covariates of chip position on the methylation array, the batch
237 effects, age at death, and sex.

238

239 *Protein-protein interactive network and pathway analysis:*

240 There were 69 CpG sites annotated to 65 genes associated with the mRNA expression level in
241 human brains. We searched their network enrichment and pathway analysis through STRING
242 (<https://string-db.org>).

243

244 **Results:**

245 **Characteristics of the individuals providing brain samples:** There were brain samples from
246 179 Hispanics and 571 NHW individuals included in the study, and their demographics are
247 shown in **Table 1**. The mean age at death for Hispanics is 80 years while it is 88 years for NHW
248 individuals. 57.54% of Hispanics and 62.9% of NHW were women.

249

250 **CpG identification:** Within the 7,155 Hispanic participants, 1,857,611 windows of 1 Kb
251 genome-wide, with at least two CpG sites, were tested for association with AD. Using the
252 Bonferroni corrected genome-wide significance p-value of $<5.0 \times 10^{-8}$, we identified six genome-
253 wide significant regions: *ADAM20* (Score=55.2, $P=4.06 \times 10^{-8}$), between *VRTN* (Score=-19.6,
254 $P=1.47 \times 10^{-8}$) and *SYNDIG1L* (Score=-37.7, $P=2.25 \times 10^{-9}$), *SPG7* (16q24.3) (Score=40.5,

255 $P=2.23 \times 10^{-8}$), *PVRL2* (Score=125.86, $P=1.64 \times 10^{-9}$), *TOMM40* (Score=-18.58, $P=4.61 \times 10^{-8}$),
256 and *APOE* (Score=75.12, $P=7.26 \times 10^{-26}$). (**Figure 1 & Table 2**). The CGS windows in *PVRL2*
257 (Score=11.25, $P=2.4 \times 10^{-6}$) and *APOE* (Score=14.46, $P=1.53 \times 10^{-11}$) were also significant in the
258 1,283 NHW participants from ROSMAP.

259
260 **Cis-effects of CGS on DNA methylation:** We tested the cis-effects of CGS on molecular
261 phenotypes within 50 Kb flanking the gene. The cis-effects on DNA methylation of the CpG
262 dosage of several CGSes in the windows of AD are presented in **Table 3**. Except for *ADAM20*,
263 all the other five loci have significant associations between the CGSes dosage and the DNA
264 methylation level of the CpG sites within the cis-regions: the intergenic region between *VRTN*
265 and *SYNDIGIL* (cg16837088, $b=-0.04$, $P=2.94 \times 10^{-3}$), *SPG7* (cg26536240, $b=0.02$, $P=5.78 \times 10^{-7}$),
266 *PVRL2* (cg04406254, $b=0.02$, $P=2.49 \times 10^{-3}$), *TOMM40* (cg20051876, $b=-0.04$, $P=0.02$), and
267 *APOE* (cg20090143, $b=-0.01$, $P=0.02$). In NHW, different CpG sites showed statistical
268 significance at *SPG7* (cg02244288, $b=-0.01$, $P=2.76 \times 10^{-63}$), *PVRL2* (cg02613937, $b=-0.01$,
269 $P=2.27 \times 10^{-6}$), and *APOE* (cg02613937, $b=-0.01$, $P=2.36 \times 10^{-6}$). The cis-mQTLs at *ADAM20*
270 (cg04910453, $b=-0.02$, $P=0.05$) reached nominal significance ($P \leq 0.05$) in NHW but not in
271 Hispanics.

272
273 **DNA methylation levels altering downstream cis-mRNA expression:** Next, we tested the
274 methylation sites cis-regulated by AD-associated CGSes (identified above), to determine
275 whether these sites altered downstream mRNA expression in the brain. Because our findings
276 revealed different methylation sites for the same gene in Hispanics and NHW for several AD
277 associated loci, we conducted an aggregate analysis by collapsing all the CpG sites within the

278 cis-region of the targeted gene (**Table 4**). We found that except for *SYNDIGIL*, methylation
279 levels in all the other genes significantly altered the brain RNA expression in Hispanics ($P \leq 0.05$).
280 The significance was replicated for *VRTN* and *TOMM40* in NHW ($P \leq 0.05$).

281
282 **Trans-effects of DNA methylation levels on gene expression.** We tested whether the
283 expression of the genes that harbor AD associated CGS were influenced by genome wide CpG
284 sites in trans. We identified 69 CpG sites across the genome that regulated gene expression of
285 *ADAM20*, *SYNDIGIL*, *SPG7*, *PVRL2*, *TOMM40*, and *APOE* in Hispanics at genome-wide
286 significant levels after Bonferroni correction of the number of CpG sites included into the
287 analysis (**Supplementary Table 2**). At *PVRL2* and *TOMM40*, the same CpG sites regulating the
288 gene expression in Hispanics also regulated the gene expression in NHW ($P < 0.05$).

289
290 These 69 CpG sites which have the trans-effects on the gene expressions were annotated to 65
291 genes. We combined these 65 genes with the target genes of *ADAM20*, *SYNDIGIL*, *SPG7*,
292 *PVRL2*, *TOMM40*, and *APOE*, to be upload to STRINGdb to query the significant biological
293 pathways. The significant pathways ($FDR < 0.05$) were presented in **Table 5**, which involved
294 neuron projection and glutamatergic synapse ($FDR = 0.0189$).

295
296 **Discussion**

297 We have conducted a first multi-omics investigation of CpG related SNPs (CGS) in brain tissue
298 from a group of individuals of Hispanic descent, which confer both genetic and epigenetic
299 effects among individuals. Our study is one of the largest genome-wide association studies with

300 the focus on the special type of genetic variants of CGS in Hispanics. We then assessed the effect
301 cascade from the AD-associated CGSes to brain methylation levels to brain expression levels.
302 This study was unique in terms of its use of Hispanic human brain tissues for AD. Also, the
303 current study provided robust results which survived the most stringent Bonferroni corrections
304 on the multiple testing.

305
306 We identified six genome-wide significant windows in or near *ADAM20*, *VRTN*, *SYNDIGIL*,
307 *SPG7*, *PVRL2*, *TOMM40*, and *APOE*, where the dosage of the CpG dinucleotides (created by the
308 including CGSes) were associated with the risk of clinical diagnosis of AD. In *SPG7*, the AD
309 associated CGS window is associated with increased cis-DNA methylation levels in the frontal
310 cortex, which in turn reduced downstream mRNA expression. We validated the *SPG7* genetic
311 and epigenetic alterations in NHW, but there is no effect on mRNA expression. Similarly, for
312 *SYNDIGIL* and *APOE*, we identified AD-associated CGSes which in turn regulated methylation
313 levels in both Hispanics and NHW brains. However, but the epigenetic modifications had muted
314 effect on downstream gene expression. At *PVRL2*, both the cis effects and trans-effects were
315 statistically significant.

316
317 *SPG7* gene encodes paraplegin, a component of the m-AAA protease, an ATP-dependent
318 proteolytic complex of the mitochondrial inner membrane that degrades misfolded proteins and
319 regulates ribosome assembly. Our finding of its significant effects of its association with AD was
320 consistent with the previous report that the DNA methylation level at *SPG7* was associated with
321 Braak neurofibrillary stages[18]. The *PVRL2* (a.k.a. *NECTIN2*) gene encodes a gene within the
322 nectin subfamily of immunoglobulin-like adhesion molecules that participate in Ca²⁺-

323 independent cell-cell adhesion. It is upstream of *TOMM40* and *APOE* and located within the
324 highly linked genetic cluster of *TOMM40-APOE-APOC2*. *PVRL2* had both cis- and trans-effects
325 between DNA methylation and mRNA gene expression. The CpG island within *APOE* was
326 reported to have lower DNA methylation level in AD patients compared to controls in human
327 postmortem brains[19, 20], which is more profound in glial cells[19]. Lee[21] reported a
328 negative correlation between *APOE* total RNA levels and DNA methylation at the CpG island
329 within *APOE* in human postmortem frontal lobes, and this negative correlation is more obvious
330 in controls compared to AD patients. *SYNDIGIL* (also known as *TMEM90A* or *CAPUCIN*)
331 encodes synapse differentiation-induced gene 1 like. In rodents, memory and motor deficits
332 caused by 1,2-Diacetylbenzene via alteration of the mRNA expression of *Syndig1l* [22] can be
333 improved by prolactin.

334
335 Although this project is currently the largest one with Hispanic brain DNA methylation data, it
336 does have limitations of potential bias by grey vs. white matter composition driven by different
337 protocol used by different sites. Also, the fact that multiple sites contribute to the brain samples
338 may also bring variations into the findings. Since the brain samples from different sites were
339 measured by different methylation batches, we only adjust for the experiment batches not the
340 sites in the regression model to remove the colinear bias.

341
342 This is the first report of robust six genetic loci covering seven genes that act as the hub for both
343 the genetic and epigenetic effects on clinical diagnosis of AD in Hispanic: *ADAM20*, between
344 *VRTN* and *SYNDIGIL*, *SPG7*, *PVRL2*, *TOMM40*, and *APOE*. *PVRL2* and *APOE* were also
345 genetically significant in NHW. Except *ADAM20*, all the other loci have significant mQTL

346 effects in Hispanics, and *SPG7*, *PVRL2*, *APOE* also have significant mQTL in NHW. The DNA
347 methylation levels of all seven genes except for *SYNDIGIL* have significant associations with its
348 mRNA gene expression levels in Hispanic brains, while only *VRTN* and *TOMM40* also showed
349 significant associations on mRNA expression levels in NHW brains. Except for *VRTN*, the
350 mRNA gene expression levels of all the other six genes have significant trans-effects from DNA
351 methylation levels of the CpG sites in Hispanics, while only *PVRL2* and *TOMM40* also showed
352 trans-effects in NHW.

353

354

355 We have identified in total six genetic loci *PVRL2* had both significant cis- and trans-effects from
356 the genetics to epigenetics and then to the mRNA gene expression and the genes for the trans-
357 effects are enriched in the pathways of neuron projection and glutamatergic synapse. Except for
358 *SYNDIGIL* *SPG7* and *APOE* had significant cis-effects while *SYNDIGIL* has significant trans-
359 effects.

360

361

362 **Acknowledgements**

363 We thank all the investigators for this study. We thank Dr. David Bennett for releasing
364 ROSMAP datasets publicly. Data collection and sharing for this project was supported by
365 (WHICAP, R01 AG072474, RF1 AG066107) funded by the National Institute on Aging (NIA)
366 and by the National Center for Advancing Translational Sciences, National Institutes of Health,
367 through Grant Number UL1TR001873. Data collection for this project was supported by the
368 Genetic Studies of Alzheimer disease in Caribbean Hispanics (EFIGA) funded by the National
369 Institute on Aging (NIA) and by the National Institutes of Health (NIH) (R01 AG067501). The
370 project was partially funded by the National Institute on Aging (NIA) of the National Institutes
371 of Health (NIH) under Award Numbers R01AG062517, and P30AG072972, P30AG062429,
372 NIA (P30 AG062677, P01 AG003949), Florida Department of Health, Ed and Ethel Moore
373 Alzheimer Disease Research Program (20A22, 8AZ06), NIH P30AG072979, P01AG066597 and
374 U19AG062418.

375

376 **Tables**

377 Table 1 Characteristics of donors of brain tissues

	Hispanics (N=179)			Non-Hispanic Whites (N=571)
	All (N=179)	With GWAS (N=112)	With RNA-seq (N=150)	
Age at death (y)*	80.29 (11.33)	78.52 (11.24)	81.08 (10.34)	88.32 (6.52)
Female (N) [#]	103 (57.54%)	64 (57.14%)	83 (55.33%)	359 (62.9%)

378 *The mean and standard deviation of the age at death are shown.

379 [#]The number and percentage of female are shown.

380

381

382 Table 2. Top CGS windows associated with Alzheimer's disease.

Chr	Window Position	Gene	Hispanics in WHICAP (N=7155)			Whites in ROSMAP (N=1283)		
			SCORE	VAR	P	SCORE	VAR	P
14	chr14:70994207-70995206	<i>ADAM20</i>	55.19	101.12	4.06E-08	1.06	1.66	0.41
14	chr14:74867207-74868206	<i>VRTN</i> and <i>SYNDIG1L</i>	-37.67	39.70	2.25E-09	-0.20	0.11	0.56
16	chr16:89588052-89589051	<i>SPG7</i>	40.51	52.45	2.23E-08	-3.06	3.51	0.10
19	chr19:45387308-45388307	<i>PVRL2</i>	125.86	435.69	1.64E-09	11.25	5.69	2.40E-06
19	chr19:45402808-45403807	<i>TOMM40</i>	-18.58	11.56	4.61E-08	-0.13	0.009	0.17
19	chr19:45411308-45412307	<i>APOE</i>	75.12	51.03	7.26E-26	14.46	4.60	1.53E-11

383 The dosage of CpG dinucleotides created by multiple CpG-related single nucleotide polymorphisms (CGSs) of each window were scaled into the value from 0 to
 384 2, which was analyzed for its association with Alzheimer's disease using generalized linear mixed models (GLMMs) implemented in the generalized linear
 385 mixed model association tests (GMMAT) with the adjustment of age, sex, and genotyping batches with the random effects of both kinship and genomic
 386 relationship matrix (GRM).

387

388

389

390

391 Table 3. Cis-effects of CGS dosage on DNA methylation.

Chr	Window Position	Gene	Hispanics in WHICAP (N=112)							Whites in ROSMAP (N=571)					Correlation between White top and Hispanic top CpG sites				
			CpG sites (N)	P threshold	CpG	Position	BETA	STDERR	P	Same CpG site in Hispanics			CpG site with minimum P in Whites			Hispanics R	Whites R		
										BETA	STDERR	P	CpG	Position	BETA			STDERR	P
14	chr14:70994207-70995206	ADAM20	7	7.14E-03	cg04910453	71051130	0.01	0.01	0.37	-0.02	0.01	0.05	cg04910453	71051130	-0.02	0.01	0.05	1	1
14	chr14:74867207-74868206	VRTN and SYNDIG1L	95	5.26E-04	cg16837088	74742111	-0.04	0.01	2.94E-03	NA	NA	NA	cg13074788	74720261	-0.01	6.42E-03	0.06	0.17	NA
16	chr16:89588052-89589051	SPG7	186	2.69E-04	cg26536240	89509760	0.02	0.004	5.78E-07	1.75E-02	9.74E-04	6.37E-56	cg02244288	89573955	-0.01	7.49E-04	2.76E-63	-0.48	-0.3
19	chr19:45387308-45388307	PVRL2	67	7.46E-04	cg04406254	45407945	0.02	0.007	2.49E-03	8.88E-03	2.34E-03	1.68E-04	cg02613937	45395297	-0.01	2.87E-03	2.27E-06	-0.12	-0.14
19	chr19:45402808-45403807	TOMM40	61	8.20E-04	cg20051876	45407860	-0.04	0.02	0.02	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
19	chr19:45411308-45412307	APOE	69	7.25E-04	cg20090143	45452003	-0.01	0.004	0.02	-9.05E-04	1.23E-03	0.46	cg02613937	45395297	-0.01	2.42E-03	2.36E-06	-0.31	0.1

392 BETA, STDERR, and P represent the regression coefficient and its corresponding standard error, and P values of the generalized linear mixed regression model
 393 where the dosage of CpG dinucleotides created by multiple CpG-related single nucleotide polymorphisms (CGSs) of each window were exposure variables and
 394 DNA methylation level of each included CpG sites were outcome variable with the adjustment of age, sex, batches of genotyping and methylation, methylation
 395 array position with the random effect of methylation array.
 396 Abbreviations: AD, Alzheimer's disease.
 397 R represent the correlation coefficient between two CpG sites.
 398
 399
 400

401 Table 4. Cis-effects of DNA methylation on gene expression

Chr	Gene	Hispanics in WHICAP (N=112)					Whites in ROSMAP (N=516)				
		SPUw1		aSPUw		P values range	SPUw1		aSPUw		P values range
		T	P	T	P		T	P	T	P	
14	<i>ADAM20</i>	-7.25	0.09	0.01	0.03	[0.012, 0.09]	4.04E-16	0.16	0.10	0.25	[0.1, 0.55]
14	<i>VRTN</i>	-63.76	0.05	0.003	8.00E-03	[0.003, 0.046]	9.99E-16	0	0	9.99E-04	[0, 0.000999]
14	<i>SYNDIG1L</i>	-26.52	0.40	0.03	0.08	[0.03, 0.40]	1.09E-14	0.25	0.03	0.08	[0.03, 0.25]
16	<i>SPG7</i>	-129.93	0.002	0	9.99E-04	[0, 0.005]	5.01E-12	0.47	0.27	0.61	[0.27, 0.62]
19	<i>PVRL2</i>	-75.43	0.002	0	9.99E-04	[0, 0.002]	-4.57E-13	0.19	0.12	0.25	[0.12, 0.25]
19	<i>TOMM40</i>	17.37	0.38	0.003	0.01	[0.003, 0.38]	8.57E-13	0.02	0.02	0.03	[0.02, 0.03]
19	<i>APOE</i>	-83.13	0.001	0	9.99E-04	[0, 0.001]	-2.07E-11	0.96	0.04	0.12	[0.04, 0.97]

402 #We have conducted a highly adaptive sum of powered score-weighted test to collapse all the available CpG sites within 100Kb distance to the gene and analyze
 403 their associations on the gene expression and Braak stages. SPUw1 provides the direction of the score, indicating the effect direction of DNA methylation on
 404 gene expression, where aSPUw test simply combines the results of multiple SPUw tests by taking the minimum P values. The model is gaussian for the
 405 continuous gene expression values and the model is binomial for the binary variable of Braak stage, where score 5 and 6 are coded as 1 and scores from 1 to 4 are
 406 coded as 0.

407 *T and P represent the statistic and its corresponding P values.

408 Abbreviations: SPUw1, sum of powered score-weighted 1 test; aSPUw, adaptive sum of powered score-weighted tests.

409

410

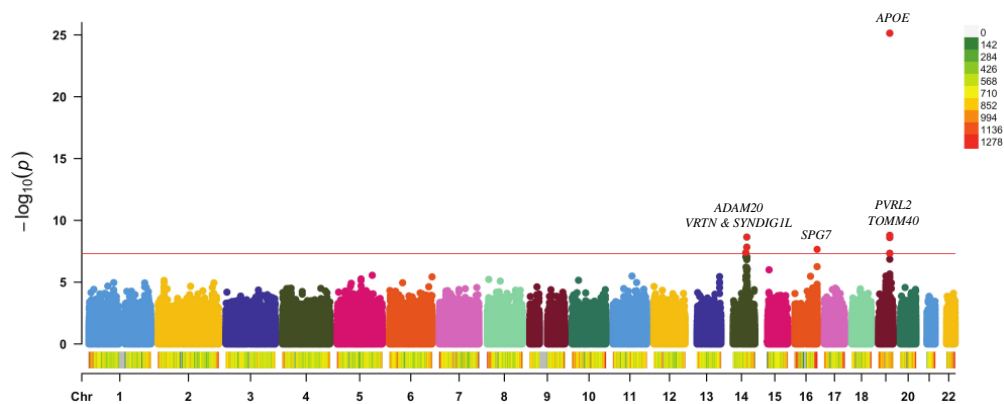
411

412 Table 5. Pathway analysis of the trans-effects of DNA methylation on gene expression.

Category	Term ID	Term Description	Observed Gene Count	Background Gene Count	Strength	False Discovery Rate
GO Component	GO:0030054	Cell junction	19	2115	0.43	0.0189
GO Component	GO:0030424	Axon	11	651	0.71	0.0189
GO Component	GO:0043005	Neuron projection	16	1391	0.54	0.0189
GO Component	GO:0098978	Glutamatergic synapse	8	334	0.86	0.0189
Monarch	EFO:0004612	High density lipoprotein cholesterol measurement	14	740	0.76	0.0015
Monarch	EFO:0004732	Lipoprotein measurement	17	1426	0.56	0.017
Monarch	EFO:0004614	Apolipoprotein A 1 measurement	9	396	0.84	0.0275
Monarch	EFO:0005105	Lipid or lipoprotein measurement	22	2526	0.42	0.0345
Monarch	EFO:0004529	Lipid measurement	21	2400	0.42	0.0365
Monarch	EFO:0004582	Liver enzyme measurement	14	1124	0.58	0.0365
Monarch	EFO:0004747	Protein measurement	36	5856	0.27	0.0365
TISSUES	BTO:0001484	Nervous system	42	6016	0.33	4.03e-05
TISSUES	BTO:0000227	Central nervous system	39	5825	0.31	0.00044
TISSUES	BTO:0000142	Brain	38	5733	0.3	0.00067
TISSUES	BTO:0000282	Head	39	6642	0.25	0.0081
COMPARTMENTS	GOCC:0030054	Cell junction	15	1053	0.64	0.0033
COMPARTMENTS	GOCC:0045202	Synapse	9	493	0.74	0.0426
UniProt Keywords	KW-0025	Alternative splicing	52	10313	0.18	0.0024

413

414 **Figure 1. Manhattan plot of sliding window search across the genome for the risk loci of clinical diagnosis of Alzheimer disease.** Each dot
 415 represents one 1-Kb window, and X and Y axis shows its genomic coordinate and $-\log_{10}$ transformed P value. The horizontal black line shows the
 416 Bonferroni-corrected genome-wide significance threshold ($P \leq 5 \times 10^{-8}$). The density of the windows across the genome were shown as color coded
 417 bars with the color coding is shown on the right.



418
 419
 420
 421

422 **References:**

- 423 1. Holstege H, Hulsman M, Charbonnier C, Grenier-Boley B, Quenez O, Grozeva D, et al. Exome
424 sequencing identifies rare damaging variants in ATP8B4 and ABCA1 as risk factors for Alzheimer's
425 disease. *Nat Genet.* 2022;54(12):1786-94. Epub 2022/11/22. doi: 10.1038/s41588-022-01208-7. PubMed
426 PMID: 36411364; PubMed Central PMCID: PMCPCMC9729101.
- 427 2. Nikolac Perkovic M, Videtic Paska A, Konjevod M, Kouter K, Svob Strac D, Nedic Erjavec G, et
428 al. Epigenetics of Alzheimer's Disease. *Biomolecules.* 2021;11(2). Epub 2021/02/13. doi:
429 10.3390/biom11020195. PubMed PMID: 33573255; PubMed Central PMCID: PMCPCMC7911414.
- 430 3. Ma Y, Jun GR, Chung J, Zhang X, Kunkle BW, Naj AC, et al. CpG-related SNPs in the MS4A
431 region have a dose-dependent effect on risk of late-onset Alzheimer disease. *Aging Cell.*
432 2019;18(4):e12964. Epub 2019/05/31. doi: 10.1111/accel.12964. PubMed PMID: 31144443; PubMed
433 Central PMCID: PMCPCMC6612647.
- 434 4. Shao Y, Shaw M, Todd K, Khrestian M, D'Aleo G, Barnard PJ, et al. DNA methylation of
435 TOMM40-APOE-APOC2 in Alzheimer's disease. *J Hum Genet.* 2018;63(4):459-71. Epub 2018/01/27.
436 doi: 10.1038/s10038-017-0393-8. PubMed PMID: 29371683; PubMed Central PMCID:
437 PMCPCMC6466631.
- 438 5. Reyes-Dumeyer D, Faber K, Vardarajan B, Goate A, Renton A, Chao M, et al. The National
439 Institute on Aging Late-Onset Alzheimer's Disease Family Based Study: A resource for genetic discovery.
440 *Alzheimers Dement.* 2022;18(10):1889-97. Epub 2022/01/04. doi: 10.1002/alz.12514. PubMed PMID:
441 34978149; PubMed Central PMCID: PMCPCMC9250549.
- 442 6. Gu Y, Honig LS, Kang MS, Bahl A, Sanchez D, Reyes-Dumeyer D, et al. Risk of Alzheimer's
443 disease is associated with longitudinal changes in plasma biomarkers in the multi-ethnic Washington
444 Heights-Hamilton Heights-Inwood Columbia Aging Project (WHICAP) cohort. *Alzheimers Dement.*
445 2024. Epub 2024/01/07. doi: 10.1002/alz.13652. PubMed PMID: 38183363.
- 446 7. Vardarajan BN, Faber KM, Bird TD, Bennett DA, Rosenberg R, Boeve BF, et al. Age-specific
447 incidence rates for dementia and Alzheimer disease in NIA-LOAD/NCRAD and EFIGA families:
448 National Institute on Aging Genetics Initiative for Late-Onset Alzheimer Disease/National Cell
449 Repository for Alzheimer Disease (NIA-LOAD/NCRAD) and Estudio Familiar de Influencia Genetica en
450 Alzheimer (EFIGA). *JAMA Neurol.* 2014;71(3):315-23. Epub 2014/01/16. doi:
451 10.1001/jamaneurol.2013.5570. PubMed PMID: 24425039; PubMed Central PMCID: PMCPCMC4000602.
- 452 8. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of
453 Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of
454 Health and Human Services Task Force on Alzheimer's Disease. *Neurology.* 1984;34(7):939-44. doi:
455 10.1212/wnl.34.7.939. PubMed PMID: 6610841.
- 456 9. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of
457 dementia. *Br J Psychiatry.* 1982;140:566-72. Epub 1982/06/01. doi: 10.1192/bjp.140.6.566. PubMed
458 PMID: 7104545.
- 459 10. De Jager PL, Ma Y, McCabe C, Xu J, Vardarajan BN, Felsky D, et al. A multi-omic atlas of the
460 human frontal cortex for aging and Alzheimer's disease research. *Sci Data.* 2018;5:180142. doi:
461 10.1038/sdata.2018.142. PubMed PMID: 30084846; PubMed Central PMCID: PMCPCMC6080491.
- 462 11. Hinton L, Carter K, Reed BR, Beckett L, Lara E, DeCarli C, et al. Recruitment of a community-
463 based cohort for research on diversity and risk of dementia. *Alzheimer Dis Assoc Disord.*
464 2010;24(3):234-41. Epub 2010/07/14. doi: 10.1097/WAD.0b013e3181c1ee01. PubMed PMID: 20625273;
465 PubMed Central PMCID: PMCPCMC2946798.
- 466 12. Santos OA, Pedraza O, Lucas JA, Duara R, Greig-Custo MT, Hanna Al-Shaikh FS, et al.
467 Ethnoracial differences in Alzheimer's disease from the FLorida Autopsied Multi-Ethnic (FLAME) cohort.
468 *Alzheimers Dement.* 2019;15(5):635-43. Epub 2019/02/23. doi: 10.1016/j.jalz.2018.12.013. PubMed
469 PMID: 30792090; PubMed Central PMCID: PMCPCMC6511501.

- 470 13. Heiss JA, Just AC. Improved filtering of DNA methylation microarray data by detection p values
471 and its impact on downstream analyses. *Clin Epigenetics*. 2019;11(1):15. doi: 10.1186/s13148-019-0615-
472 3. PubMed PMID: 30678737; PubMed Central PMCID: PMC6346546.
- 473 14. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of
474 cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray.
475 *Epigenetics*. 2013;8(2):203-9. doi: 10.4161/epi.23470. PubMed PMID: 23314698; PubMed Central
476 PMCID: PMC3592906.
- 477 15. McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL. Identification
478 of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip.
479 *Genom Data*. 2016;9:22-4. doi: 10.1016/j.gdata.2016.05.012. PubMed PMID: 27330998; PubMed Central
480 PMCID: PMC4909830.
- 481 16. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of
482 Infinium DNA methylation BeadChip probes. *Nucleic Acids Res*. 2017;45(4):e22. doi:
483 10.1093/nar/gkw967. PubMed PMID: 27924034; PubMed Central PMCID: PMC5389466.
- 484 17. Chen H, Wang C, Conomos MP, Stilp AM, Li Z, Sofer T, et al. Control for Population Structure
485 and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. *Am J Hum*
486 *Genet*. 2016;98(4):653-66. doi: 10.1016/j.ajhg.2016.02.012. PubMed PMID: 27018471; PubMed Central
487 PMCID: PMC4833218.
- 488 18. Li QS, Sun Y, Wang T. Epigenome-wide association study of Alzheimer's disease replicates 22
489 differentially methylated positions and 30 differentially methylated regions. *Clin Epigenetics*.
490 2020;12(1):149. Epub 2020/10/19. doi: 10.1186/s13148-020-00944-z. PubMed PMID: 33069246;
491 PubMed Central PMCID: PMC7568396.
- 492 19. Tulloch J, Leong L, Thomson Z, Chen S, Lee EG, Keene CD, et al. Glia-specific APOE
493 epigenetic changes in the Alzheimer's disease brain. *Brain Res*. 2018;1698:179-86. Epub 2018/08/07. doi:
494 10.1016/j.brainres.2018.08.006. PubMed PMID: 30081037; PubMed Central PMCID: PMC6388639.
- 495 20. Foraker J, Millard SP, Leong L, Thomson Z, Chen S, Keene CD, et al. The APOE Gene is
496 Differentially Methylated in Alzheimer's Disease. *J Alzheimers Dis*. 2015;48(3):745-55. Epub 2015/09/25.
497 doi: 10.3233/JAD-143060. PubMed PMID: 26402071; PubMed Central PMCID: PMC6469491.
- 498 21. Lee EG, Tulloch J, Chen S, Leong L, Saxton AD, Kraemer B, et al. Redefining transcriptional
499 regulation of the APOE gene and its association with Alzheimer's disease. *PLoS One*.
500 2020;15(1):e0227667. Epub 2020/01/25. doi: 10.1371/journal.pone.0227667. PubMed PMID: 31978088;
501 PubMed Central PMCID: PMC6980611.
- 502 22. Nguyen HD, Jo WH, Hoang NHM, Kim MS. In silico identification of the potential molecular
503 mechanisms involved in protective effects of prolactin on motor and memory deficits induced by 1,2-
504 Diacetylbenzene in young and old rats. *Neurotoxicology*. 2022;93:45-59. Epub 2022/09/14. doi:
505 10.1016/j.neuro.2022.09.002. PubMed PMID: 36100143.
- 506

