CpG-related SNPs in the MS4A region have a dose-dependent effect on risk of late–onset Alzheimer disease

Yiyi Ma1,2 | Gyungah R. Jun1,3,4 | Jaeyoon Chung1 | Xiaoling Zhang1,3 | Brian W. Kunkle5 | Adam C. Naj6,7 | Charles C. White8,9 | David A Bennett10 | Philip L. De Jager2,8,9 | Alzheimer's Disease Genetics Consortium | Richard Mayeux11 | Jonathan L. Haines12 | Margaret A. Pericak-Vance5 | Gerard D. Schellenberg7 | Lindsay A. Farrer1,3,4,13,14 | Kathryn L. Lunetta3

1Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, Massachusetts
2Center for Translational & Computational Neuroimmunology, Multiple Sclerosis Clinical Care and Research Center, Division of Neuroimmunology, Department of Neurology, Columbia University Medical Center, New York, New York
3Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts
4Department of Ophthalmology, Boston University School of Medicine, Boston, Massachusetts
5John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, Florida
6Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania
7Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania
8Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Departments of Neurology and Psychiatry, Brigham and Women's, Boston, Massachusetts
9Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts
10Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois
11Department of Neurology and Sergievsky Center, Columbia University, New York, New York
12Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio
13Department of Neurology, Boston University School of Medicine, Boston, Massachusetts
14Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts

Correspondence
Yiyi Ma, Center for Translational & Computational Neuroimmunology, Multiple Sclerosis Clinical Care and Research Center, Division of Neuroimmunology, Department of Neurology, Columbia University Medical Center, Columbia University, 630 W 168th Street, New York, NY 10032. Email: ym2666@cumc.columbia.edu

Funding information
The Alzheimer's Disease Genetics Consortium supported the collection of samples used in this study through National Institute on Aging (NIA) grants U01-AG032984 and RC2-AG036528. Data for this study were prepared, archived, and distributed by the National Institute on...
1 | INTRODUCTION

Much has been learned about the genetic basis of Alzheimer disease (AD), the most common cause of dementia in the elderly. Genome-wide association studies (GWAS) have identified common and rare variants in more than 30 loci that contribute to AD risk (Bellenguez et al., 2017; Hollingworth et al., 2011; Jakobsdottir et al., 2016; Jun et al., 2017, 2016; Lambert et al., 2013; Mez et al., 2017; Naj et al., 2011; Sims et al., 2017). However, these associations explain only a fraction of the heritability of AD, and their functional consequence also remains unclear (Lambert et al., 2013; Ridge, Mukherjee, Crane, & Kauwe, 2013). Thus, here we investigate AD risk from a different perspective.

Epigenetic phenomena such as DNA methylation may be involved but have not been studied extensively in AD. DNA methylation is intimately associated with genetic variation because of frequent attachment of a methyl group directly to a DNA nucleotide, particularly a dinucleotide comprising a cytosine and guanine (CpG). CpG-related SNPs (CGS) alter the sequence of the primary target sites for DNA
methylated regions (Lister et al., 2009) and account for a significant fraction (~38%-88%) of allele-specific methylation (ASM) regions in the human genome (Shoemaker, Deng, Wang, & Zhang, 2010). It has been demonstrated that more than 80% of CGSes have a regulatory role in DNA methylation (Zhi et al., 2013). Recently, we found that almost more than 80% of CGSes have a regulatory role in human genome (Shoemaker, Deng, Wang, & Zhang, 2010). It has been demonstrated that almost all CGSes in a window and weighted rare variants more strongly than common variants. Because the SKAT-O window-based test does not consider the effect direction of the variants in each window, we also tested a model including CGS dosage which was calculated as the total number of CpG dinucleotides created by the CGSes in the window. Genome-wide analysis of 2,288,371 overlapping windows each containing at least two CGSes showed little evidence of inflation (λ = 1.099, Figure S1). SKAT-O and CGS dosage approaches provided similar results across the genome (Figure S2) including five distinct genome-wide significant loci with windows at BIN1 (SKAT-O p = 1.27 × 10^{-13}, CGS dosage p = 4.74 × 10^{-10}), MS4A4A (SKAT-O p = 2.66 × 10^{-10}, CGS dosage p = 3.40 × 10^{-10}), MS4A4A (SKAT-O p = 6.36 × 10^{-10}, CGS dosage p = 2.67 × 10^{-10}), PICALM (SKAT-O p = 6.34 × 10^{-9}, CGS dosage p = 1.42 × 10^{-5}), and APOE (SKAT-O p = 2.99 × 10^{-44}, CGS dosage p = 2.77 × 10^{-536}) (Table 1). Although the top windows at BIN1 and PICALM identified by SKAT-O do not reach genome-wide significance in the CGS dosage test, the CGS dosage test identified significant associations with other windows at these loci. Windows at LRFN2-UNC5CL and TREM2 are genome-wide significant with only the SKAT-O test, whereas the windows at CR1 are genome-wide significant with only the CGS dosage test. All genome-wide significant windows identified by SKAT-O were replicated (Table 2).

Windows in MS4A4A and MS4A6A showed a strong negative dosage effect of CpG dinucleotides on AD risk (change in log odds of AD = −0.01 and −0.02 per one unit dinucleotide increase, p = 2.67 × 10^{-10} and 3.4 × 10^{-10}, respectively). This effect was evident in 18 out of 24 cohorts (Figure 1). The dosage of CpG dinucleotides created by the two CGSes in the APOE window has significantly positive association with AD risk (change in log odds of AD = 0.2 per one unit dinucleotide increase, p = 2.77 × 10^{-526}).

**2 | RESULTS**

**2.1 Sliding Window Association of CGSes with AD**

Association of AD with CGSes was tested genome-wide using sliding windows that were 1 kb in length, overlapping by 0.5 kb and contained at least two CGSes. These analyses, which were performed using SKAT-O (Lee, Wu, & Lin, 2012), considered the combined effects of all CGSes in a window and weighted rare variants more heavily than common variants. Because the SKAT-O window-based test does not consider the effect direction of the variants in each window, we also tested a model including CGS dosage which was calculated as the total number of CpG dinucleotides created by the CGSes in the window. Genome-wide analysis of 2,288,371 overlapping windows each containing at least two CGSes showed little evidence of inflation (λ = 1.099, Figure S1). SKAT-O and CGS dosage approaches provided similar results across the genome (Figure S2) including five distinct genome-wide significant loci with windows at BIN1 (SKAT-O p = 1.27 × 10^{-13}, CGS dosage p = 4.74 × 10^{-10}), MS4A4A (SKAT-O p = 2.66 × 10^{-10}, CGS dosage p = 3.40 × 10^{-10}), MS4A4A (SKAT-O p = 6.36 × 10^{-10}, CGS dosage p = 2.67 × 10^{-10}), PICALM (SKAT-O p = 6.34 × 10^{-9}, CGS dosage p = 1.42 × 10^{-5}), and APOE (SKAT-O p = 2.99 × 10^{-44}, CGS dosage p = 2.77 × 10^{-536}) (Table 1). Although the top windows at BIN1 and PICALM identified by SKAT-O do not reach genome-wide significance in the CGS dosage test, the CGS dosage test identified significant associations with other windows at these loci. Windows at LRFN2-UNC5CL and TREM2 are genome-wide significant with only the SKAT-O test, whereas the windows at CR1 are genome-wide significant with only the CGS dosage test. All genome-wide significant windows identified by SKAT-O were replicated (Table 2).

Windows in MS4A4A and MS4A6A showed a strong negative dosage effect of CpG dinucleotides on AD risk (change in log odds of AD = −0.01 and −0.02 per one unit dinucleotide increase, p = 2.67 × 10^{-10} and 3.4 × 10^{-10}, respectively). This effect was evident in 18 out of 24 cohorts (Figure 1). The dosage of CpG dinucleotides created by the two CGSes in the APOE window has significantly positive association with AD risk (change in log odds of AD = 0.2 per one unit dinucleotide increase, p = 2.77 × 10^{-526}).

| Chr | Gene | Start | End | N of CGSes | P range of CGSes (min, max) | Window P^a | Window P^b | Beta (SE) |
|-----|------|-------|-----|-----------|-----------------------------|------------|------------|-----------|
| 2   | BIN1 | 127,847,001 | 127,848,000 | 2 | (1.48E-13, 5.88E-06) | 1.27E-13 | 2.14E-03 | -0.02 (0.005) |
| 2   | BIN1 | 127,881,001 | 127,882,000 | 2 | (2.5E-12, 3.67E-03) | 1.09E-03 | 4.74E-18 | 0.18 (0.02) |
| 11  | MS4A6A | 59,923,001 | 59,924,000 | 2 | (1.41E-10, 1.25E-09) | 2.66E-10 | 3.40E-10 | -0.01 (0.002) |
| 11  | MS4A4A | 60,087,501 | 60,088,500 | 2 | (8.44E-12, 1.23E-05) | 6.36E-10 | 2.67E-10 | -0.02 (0.003) |
| 11  | PICALM | 85,759,501 | 85,760,500 | 2 | (3.77E-05, 0.11) | 6.34E-09 | 9.28E-05 | 0.01 (0.002) |
| 11  | PICALM | 85,845,001 | 85,846,000 | 2 | (3.77E-05, 0.11) | 5.10E-02 | 1.42E-09 | 0.13 (0.02) |
| 19  | APOE | 45,411,501 | 45,412,500 | 2 | (>2.23E-308, 3.56E-28) | 2.99E-46 | 2.77E-556 | 0.2 (0.004) |

^aP values obtained by SKAT-O test.

^bP values obtained by CGS dosage test and beta represent the change in log odds of AD per 1-unit increase in dosage of CpG dinucleotides comprising the CpG-related SNPs in the window.
In order to show the unique role of CGSes in these windows, we compared the significance level for the windows under two conditions, (a) including only CGSes and (b) including only non-CGSes, which are the SNPs that do not disrupt CpG dinucleotide formation. As shown in Table 3, the p values of all the identified AD-associated windows in Table 1 were attenuated when only non-CGSes were included in the test. The number of CGSes and non-CGSes in each top window differed by no more than two except for the windows at MS4A4A and PICALM. There is very modest LD between the top CGS and non-CGS for the most of the windows except for a window at MS4A6A (R² = 0.98) which is 165 kb from the top window at MS4A4A (R² = 0.37). The attenuation of the significance level was also observed at the individual SNP level for the comparisons of the two types of SNPs in each window, noting that the APOE region did not contain any non-CGSes (Table S4).

2.2 | Association of CGSes with DNA methylation and gene expression

Windows containing CGSes located in MA4A4A, PICALM and APOE were associated (p ≤ 0.05) with the degree of DNA methylation in brains (Table 4); however, only the MS4A4A window was significantly associated in brains after correction for the 176 methylation probes tested for association (adjusted p = 2.15 × 10⁻⁷ at cg14750746). This window was also nominally associated with increased methylation in blood after correcting for the same 176 methylation probes (nominal
p = 3.34 × 10^{-4} and adjusted p = 0.06). In addition, the number of CpG dinucleotides created by the CGSes in the intergenic window between MS4A4A and MS4A6E was associated with increased expression of MS4A4A in both brain (p = 0.03) and blood (p = 2.53 × 10^{-8}). The MS4A6A window was associated with DNA methylation (adjusted p = 1.47 × 10^{-7}) and gene expression (p = 5.89 × 10^{-26}) in blood, but rs12226022 was not well imputed in the ROSMAP dataset to test this association in brain.

2.3 | Pathway analysis at the MS4A4A window

Transcriptome analysis using RNAseq data from the Religious Order Study and Rush Memory and Aging Project (ROSMAP) brain samples was performed to identify the set of genes whose expression is influenced by methylation of CpG site cg14750746 that was associated with the dosage of MS4A4A CGSes (Table 4). In total, 15,508 protein-coding genes remained in the analysis after removing genes expressed in less than 10% subjects. Although no genes remained significant after correcting for the number tests (threshold p = 3.2 × 10^{-6}), there were 34 nominally associated genes (p < 5×10^{-5}) (Table S5) and pathway analysis showed enrichment in methyltransferase activity (Table 5).

3 | DISCUSSION

Our study using a sliding-window approach confirmed the importance of CGS in AD and is the first to report dosage effects of CpG dinucleotides created by CGSes on AD risk. In particular, we identified six windows with a significant effect of the number of CpG dinucleotides on AD risk, including a novel and robust dose-dependent effect in an intergenic window located between MS4A4A and MS4A6E. The number of CpG dinucleotides created by the CGSes within this window is inversely associated with the risk of AD. The potential functional importance of CGSes in AD is supported by evidence showing that the significance of almost all of the top windows was attenuated when non-CGSes were included instead of CGSes.

This observation does not seem to be related to the differences in the number of variants or LD between CGSes and non-CGSes.

The MS4A gene cluster encodes a family of proteins spanning the cellular membrane four times which share similar polypeptide sequence and predicted topological structure. MS4A6A expression in brain is positively associated with AD-related neurofibrillary tangles and neuritic plaques (Karch et al., 2012; Martiskainen et al., 2015). AD risk alleles at these loci were reported to be associated with higher expression in brain (Allen et al., 2012; Karch et al., 2012; Martiskainen et al., 2015). The underlying mechanism for the effects of MS4A genes on AD may be related to their regulation of calcium channels (Walshe et al., 2008), immune system (Zuccolo et al., 2013). Our findings of an association of the CpG dinucleotide dosage in this region with AD risk suggest a potential novel AD-related mechanism involving MS4A genes. Further experiments examining DNA methylation in the MS4A region are necessary to clarify the exact mechanism.

All of the loci identified in our study using a sliding-window approach were previously reported to be associated with AD through DNA methylation analyses, indicating an overlap between genetic and epigenetic mechanisms. For example, brain DNA methylation levels of CpG sites located in the top-ranked loci have been associated with clinical and pathological diagnoses of AD in a sample of 740 ROSMAP participants, many of whom are included in the ADGC GWAS dataset (De Jager et al., 2014). The mQTL CpG sites identified in our study are correlated with the previously reported (De Jager et al., 2014) AD-associated CpG sites in both brain and blood (all p < 0.05) (Table S6), but it is unclear why the pairs of methylation probes in MS4A region and APOE are inversely correlated in brain and blood.

All of the genes identified by our analyses have been implicated in inflammation and the immune system. BIN1 knock-out mice were shown to have higher incidence of inflammation during aging (Chang et al., 2007). BIN1 was also reported to be related to inflammation and immunity by its participation in the phagocytic pathway (Gold et al., 2004) and regulation of critical enzymes against pathogens (Muller, DuHadaway, Donover, Sutanto-Ward, & Prendergast,
TABLE 4 Association between CGSes and methylation and gene expression

| Gene  | Position | Name    | Methylation of CpG site (Brain) | Methylation of CpG site (AD) |
|-------|----------|---------|--------------------------------|----------------------------|
| BIN1  | 127,800,646 | cg000436254 | β (SE) = −1.87E−04 (2.28E−03) | P1 = 0.93, P2 = 1.00, p = 2.08E−03 |
| MS4A6A | 59,824,541 | cg01971716 | NA                              | NA                         |
| MS4A4A | 60,101,475 | cg14750746 | β (SE) = 5.60E−03 (8.11E−04)    | P1 = 1.22E−11, P2 = 2.15E−09, p = 0.03 |
| PICALM | 85,566,560 | cg15822411 | β (SE) = −3.42E−03 (1.77E−03)  | P1 = 0.05, P2 = 1.00, p = 1.00 |
| APOE  | 45,395,297 | cg02613937 | β (SE) = −9.96E−04 (4.19E−04)  | P1 = 0.02, P2 = 1.00, p = 0.42 |

Statistical significance was assessed using CGSes dosage and the association was significant at the genome-wide level (p < 5.0E−8) or 1.00E−04 (adj-p < 1.00E−04).

4.1.1 | CGS annotation

CGSes were annotated as described previously (Ma et al., 2016). In brief, CGS information was retrieved by Galaxy (Goecks, Nekrutenko, Taylor, & Galaxy, 2010) from UCSC human genome browser based on SNP141 and human hg19 sequence data.
4.1.2 | Discovery stage subjects

The discovery stage included 12,181 unrelated cases and 12,601 controls from 22 cohorts with European ancestry participating in the Alzheimer’s Disease Genetic Consortium (ADGC) (Table S2).

Characteristics of the ADC7 cohort are provided in the Appendix S1, and details of other study cohorts were previously described (Jun et al., 2016; Lambert et al., 2013). Studies of the individual cohorts were approved by the appropriate Institutional Review Boards, and written informed consent for all subjects was provided on behalf of

TABLE 5 Enrichment of methyltransferase activities in the regulatory network of MS4A cluster-associated CpG site (cg14750746) in brain using Gene Ontology (GO) terms

| GO term ID    | GO term description                                      | P         | FDR   |
|---------------|----------------------------------------------------------|-----------|-------|
| GO:0050313    | sulfur dioxygenase activity                              | 8.31E−04  | 0.03  |
| GO:0008276    | protein methyltransferase activity                       | 1.19E−03  | 0.03  |
| GO:0008170    | N-methyltransferase activity                             | 1.23E−03  | 0.03  |
| GO:0070905    | serine binding                                           | 1.66E−03  | 0.03  |
| GO:0003713    | transcription coactivator activity                       | 1.69E−03  | 0.03  |
| GO:0004843    | ubiquitin-specific protease activity                     | 2.44E−03  | 0.03  |
| GO:0042799    | histone methyltransferase activity (H4-K20 specific)     | 2.49E−03  | 0.03  |
| GO:0019783    | ubiquitin-like protein-specific protease activity        | 2.95E−03  | 0.03  |
| GO:0036459    | ubiquitinyl hydrolase activity                           | 3.01E−03  | 0.03  |
| GO:0008234    | cysteine-type peptidase activity                         | 8.70E−03  | 0.06  |
| GO:0008139    | nuclear localization sequence binding                    | 9.11E−03  | 0.06  |
| GO:0008168    | methyltransferase activity                               | 9.91E−03  | 0.06  |
| GO:0016741    | transferase activity, transferring one-carbon groups     | 0.01      | 0.06  |
| GO:0003756    | protein disulfide isomerase activity                     | 0.02      | 0.08  |
| GO:0016846    | intramolecular oxidoreductase activity, transposing S-S bonds | 0.02   | 0.08  |
| GO:0005096    | GTPase activator activity                                | 0.02      | 0.08  |
| GO:0016702    | oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen | 0.02  | 0.08  |
| GO:0016701    | oxidoreductase activity, acting on single donors with incorporation of molecular oxygen | 0.02  | 0.08  |
| GO:0030695    | GTPase regulator activity                                | 0.02      | 0.08  |
| GO:0005048    | signal sequence binding                                  | 0.02      | 0.08  |
| GO:0060589    | nucleoside-triphosphatase regulator activity             | 0.02      | 0.09  |
| GO:0018024    | histone-lysine N-methyltransferase activity              | 0.03      | 0.09  |
| GO:0016278    | lysine N-methyltransferase activity                      | 0.03      | 0.10  |
| GO:0016279    | protein-lysine N-methyltransferase activity              | 0.03      | 0.10  |
themselves or for substantially cognitively impaired subjects, by a caregiver, legal guardian, or other proxy.

4.2 | Statistical analysis

Details of SNP genotyping and quality control are described elsewhere (Jun et al., 2016; Lambert et al., 2013). SNP genotype imputation was performed using IMPUTE2 with reference haplotypes from the March 2012 release of 1,000 Genomes. Principal component (PC) analysis was conducted using the smartpca program in EIGENSOFT (Patterson, Price, & Reich, 2006; Price et al., 2006) to evaluate population substructure within each dataset. Association of AD risk with CGSes was tested using a sliding-window approach (Tang, Feng, Sha, & Zhang, 2009). Windows spanning 1kb were constructed based on evidence suggesting that sequence variants within 1 kb can affect the methylation status of a gene (Lienert et al., 2011). Consecutive windows with a 500 bp overlap were tested to optimize power for detection of associations and ensure a sufficient number of SNPs in each window. Thus, for example, each unique 2000 bp region contains three overlapping windows. CGSes with imputation quality ($r^2 \leq 0.4$ or genotype data available for less than half of the cohorts were removed. Windows with fewer than two CGSes were omitted from the analysis. After these filtering steps, 2,288,371 windows remained for association analyses.

The association of AD with the combined effects of multiple CGSes in each window on the risk of AD was evaluated by logistic regression using the optimal sequence kernel association test (SKAT-O) (Lee et al., 2012) using R package seqMeta (https://cran.r-project.org/web/packages/seqMeta/index.html) as implemented in Universal Genome Analyst (UGA) software (https://github.com/rmkoesterer/uga). The fast $P$ value calculation "integration" method was used as a screening tool. Windows with $p \leq 5 \times 10^{-4}$ or no reported $p$ value were re-analyzed using the "saddlepoint" method (Duchesne & de Micheaux, 2010). We used the default weights of the seqMeta package to up weight the contributions from rare variants with the aim to identify potential novel loci. The same methodology was applied to the analysis of non-CGSes. SKAT-O is not sensitive to effect direction of the individual variants included in the test and thus does not produce effect estimates. Thus, we also conducted the dose–response effect of the multiple CGSes in the window on AD risk using logistic regression. The allele that creates a CpG dinucleotide was considered as the effect allele and the allele that disrupts the CpG dinucleotide as the reference allele. The sum of the imputed dosages for multiple CGSes in each window was calculated and used as the exposure variable for the logistic regression model with AD status as the outcome. The summary statistics for regression coefficients and robust standard errors from each cohort were meta-analyzed using an inverse variance-weighted, fixed-effects approach implemented in METAL (Willer, Li, & Abecasis, 2010). Both SKAT-O and dosage analyses were adjusted for age, sex, and PCs. Windows surviving Bonferroni-correction genome-wide significance level ($p \leq 5 \times 10^{-8}$) from both methodologies were considered. The genome-wide summary statistics from the two methodologies are provided in Table S8.

4.2.1 | Replication testing

Cohort-specific GWAS summary statistics were obtained from a prior AD GWAS conducted by the IGAP consortium, which includes 7,554 unrelated cases and 27,382 controls from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium, the European Alzheimer’s Disease Initiative (EADI), and the Genetic and Environmental Risk in Alzheimer’s Disease (GERAD) consortium (Lambert et al., 2013). The protocols and participant consent forms were approved by each institution. The combined effects of multiple CGSes in each window on AD were determined using the GATES method, implemented in the GATES R package (Li, Gui, Kwan, & Sham, 2011). This method extends the Simes test to combine the $p$-values of the SNPs within a region into an overall regional $p$ value.

4.3 | mQTL Analysis

Brain mQTL was obtained for 740 subjects (mean age at death = 88 years, 63.6% female) from the Religious Order Study and Rush Memory and Aging Project (ROSMAP), and blood mQTL data obtained from 2,405 participants (mean age = 66 years, 54% female) of the Framingham Heart Study (FHS) Offspring cohort at examination 8 were downloaded from dbGAP (Table S3). DNA methylation profiles for both studies were measured by the Illumina Infinium HumanMethylation450 BeadChip. Analyses of FHS data were conducted in two stages. A linear mixed model was used to derive the residuals of the DNA methylation of the probe adjusted for the imputed cell types (CD8T, CD4T, NK, B-cell, monocyte), row and column as fixed-effects, chip ID as a random effect at first. Then, each residual was regressed on the CGSes dosage in models including age and sex as fixed-effects and kinship matrix as random effect to account for familial correlation. Analyses of ROSMAP data were conducted with the linear model by adjusting the methylation batch, age at death, sex, post-mortem interval, and study group (ROS or MAP), which was test to be the most appropriate model for the data as reported by De Jager et al. (2014). $p$-values were adjusted using a Bonferroni correction for the total number of probes tested within each window.

4.4 | eQTL analysis

Brain RNAseq data were obtained for 580 ROSMAP subjects (mean age at death = 89 years, 63.3% female), and whole blood array-based expression data for 5,252 FHS Offspring cohort (examination 8) and Generation 3 (examination 2) participants (mean age = 55 years, 54% female) were obtained from dbGAP (Supplementary Table 3). Normalized gene expression level was regressed on the sum of dosages of CpG dinucleotides in each window with covariates for age, sex, and the first three PCs of ancestry using a linear mixed model for analyses of FHS data and a general linear model for analyses of ROSMAP data. $p$-Values were corrected for the seven tests (i.e., 7 genes) performed.
4.5 | Pathway analysis

Using the ROSMAP brain methylation and RNAseq data, we performed a genome-wide expression-methylation scan using a general linear model with the methylation of CpG site cg14750476 as the exposure variable and the normalized gene expression levels of all the protein-coding genes as outcomes (n = 15,508), including the same covariates as in the mQTL and eQTL analyses. Genes with p < 0.005 were included in the pathway enrichment analysis implemented in the software of STRINGdb (Szklarczyk et al., 2015), which conducted a hypergeometric test, using the false discovery rate (FDR) to correct for multiple tests (Benjamini, 1995), to query the enrichment of the input gene sets against the background gene list in Gene Ontology database classified as "molecular function".

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

YM, KLL, and LAF wrote the manuscript. YM performed the data analysis. XZ, BWK, ACN, CCW, PLDJ, and DAB interpreted genetic association, mQTL and eQTL analyses. GRJ and JC provided technical support. RM, JLH, MAP-V, GDS, and LAF obtained the funding for this study. KLL and LAF supervised the project. All authors read and approved the final manuscript.

ORCID

Yiyi Ma https://orcid.org/0000-0002-3609-8877

REFERENCES

Allen, M., Zou, F., Chai, H. S., Younklin, C. S., Crook, J., Pankratz, V. S., ... Wolter, R. L. (2012). Novel late-onset Alzheimer disease loci variants associate with brain gene expression. Neurology, 79(3), 221–228. https://doi.org/10.1212/WNL.0b013e3182605801

Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S. C., Mann, M., & Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone modification. Cell, 143(3), 470–484. https://doi.org/10.1016/j.cell.2010.10.012

Bellenguez, C., Charbonnier, C., Gnerier-Boley, B., Quenez, O., Le Guennec, K., Nicolas, G., ... collaborators, C. M. (2017). Contribution to Alzheimer’s disease risk of rare variants in TREM2, SORL1, and ABCA7 in 1779 cases and 1273 controls. Neurobiology of Aging, 59, 220 e221–220 e229. https://doi.org/10.1016/j.neurobiolaging.2017.07.001

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Royal Statistical Society: Series B (Methodological), 57(1B), 289–300

Brunet, C. R., Dingwell, K. S., Wale, C., Graham, F. L., & Johnson, D. C. (1998). Herpes simplex virus gD and virions accumulate in endosomes by mannose 6-phosphate-dependent and -independent mechanisms. Journal of Virology, 72(4), 3330–3339.

Chang, M. Y., Bouliden, J., Katz, J. B., Wang, L., Meyer, T. J., Soler, A. P., ... Prendergast, G. C. (2007). Bin1 ablation increases susceptibility to cancer during aging, particularly lung cancer. Cancer Research, 67(16), 7605–7612. https://doi.org/10.1158/0008-5472.CAN-07-1100

De Jager, P. L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L. C., Yu, L., ... Bennett, D. A. (2014). Alzheimer’s disease: Early alterations in brain DNA methylation at ANK1, BDNF, RHBDL2 and other loci. Nature Neuroscience, 17(9), 1156–1163. https://doi.org/10.1038/nn.3786

Dorey, E., Chang, N., Liu, Q. Y., Yang, Z., & Zhang, W. (2014). Apolipoprotein E, amyloid-beta, and neuroinflammation in Alzheimer’s disease. Neurosci Bull, 30(2), 317–330. https://doi.org/10.1007/s12264-013-1422-z

Duchesne, P., & de Micheaux, P. L. (2010). Computing the distribution of quadratic forms: Further comparisons between the Liu-Tang-Zhang approximation and exact methods. Computational Statistics and Data Analysis, 54, 858–862.

Ghanim, H., Monte, S. V., Sia, C. L., Abuaysheh, S., Green, K., Caruana, J. A., & Dandona, P. (2012). Reduction in inflammation and the expression of amyloid precursor protein and other proteins related to Alzheimer’s disease following gastric bypass surgery. Journal of Clinical Endocrinology and Metabolism, 97(7), E1197–1201. https://doi.org/10.1210/jc.2011-3284

Goeccks, J., Nekrutenko, A., Taylor, J., & Galaxy, T. (2010). Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biology, 11(8), R86. https://doi.org/10.1186/gb-2010-11-8-r86

Gold, E. S., Simmons, R. M., Petersen, T. W., Campbell, L. A., Kuo, C. C., & Aderem, A. (2004). Amphiphysin IIm is required for survival of Chlamydia pneumoniae in macrophages. Journal of Experimental Medicine, 200(5), 581–586. https://doi.org/10.1084/jem.20040546

Guerreiro, R., Wojtas, A., Bras, J., Carrasquillo, M., Rogaea, E., Majounie, E., ... Analysis, A. G. (2013). TREM2 variants in Alzheimer’s disease. New England Journal of Medicine, 368(2), 117–127. https://doi.org/10.1056/NEJMoa1211851

Hollingworth, P., Harold, D., Sims, R., Gerrish, A., Lambert, J. C., Carrasquillo, M. M., ... Williams, J. (2011). Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer’s disease. Nature Genetics, 43(5), 429–435. https://doi.org/10.1038/ng.803

Hunter, A., Spechler, P. A., Cwanger, A., Song, Y., Zhang, Z., Ying, G. S., ... Dunaief, J. L. (2012). DNA methylation is associated with altered gene expression in AMD. Investigative Ophthalmology & Visual Science, 53(4), 2089–2105. https://doi.org/10.1167/iovs.11-8449

Jakobsdottir, J., van der Lee, S. J., Bis, J. C., Chouraki, V., Li-Kroeger, D., Yamamoto, S., ... van Duijn, C. M. (2016). Rare functional variants in TM2D3 is associated with Late-Onset Alzheimer’s disease. PLoS Genetics, 12(10), e1006327. https://doi.org/10.1371/journal.pgen.1006327

Jonsson, T., Stefansson, H., Steinberg, S., Jonsdottir, I., Jonsson, P. V., Snaedal, J., ... Stefansson, K. (2013). Variant of TREM2 associated with the risk of Alzheimer’s disease. New England Journal of Medicine, 368(2), 107–116. https://doi.org/10.1056/NEJMoa1211103

Jun, G. R., Chung, J., Mez, J., Barber, R., Beecham, G. W., Bennett, D. A., ... Farrer, L. A. (2017). Transethnic genome-wide scan identifies novel Alzheimer’s disease loci. Alzheimer’s & Dementia, 13(7), 727–738. https://doi.org/10.1016/j.j-alz.2016.12.012

Jun, G., Ibrahim-Verbaas, C. A., Vronskaya, M., Lambert, J. C., Chung, J., Naj, A. C., ... Farrer, L. A. (2016). A novel Alzheimer disease locus located near the gene encoding tau protein. Molecular Psychiatry, 21(1), 108–117. https://doi.org/10.1038/mp.2015.23

Karch, C. M., Jeng, A. T., Nowotny, P., Cady, J., Cruchaga, C., & Goate, A. M. (2012). Expression of novel Alzheimer’s disease risk genes in control and Alzheimer’s disease brains. PLoS ONE, 7(11), e50976. https://doi.org/10.1371/journal.pone.0050976
Lambert, J. C., Ibrahim-Verbaas, C. A., Harold, D., Naj, A. C., Sims, R., Bellenguez, C., ... Amouyel, P. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease. *Nature Genetics*, 45(12), 1452–1458. https://doi.org/10.1038/ng.2802

Lee, S., Wu, M. C., & Lin, X. (2012). Optimal tests for rare variant effects in sequencing association studies. *Biostatistics*, 13(4), 762–775. https://doi.org/10.1093/biostatistics/kxs014

Li, M. X., Gui, H. S., Kwan, J. S., & Sham, P. C. (2011). GATES: A rapid and powerful gene-based association test using extended Simes procedure. *American Journal of Human Genetics*, 88(3), 283–293. https://doi.org/10.1016/j.ajhg.2011.01.019

Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., & Schubeler, D. (2011). Identification of genetic elements that autonomously determine methylation states. *Nature Genetics*, 43(11), 1091–1097. https://doi.org/10.1038/ng.946

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Liebert, P., ... Schubeler, D. (2011). Identification of genomic DNA methylation marks that define mammalian gene CpG islands. *Nature*, 462(7271), 315–322. https://doi.org/10.1038/nature08514

Ma, Y., Smith, C. E., Lee, Y. C., Parnell, L. D., Lai, C. Q., & Ordovas, J. M. (2016). Haplotypes of CpG-related SNPs and associations with DNA methylation patterns. *In Genome-wide Association Studies: From Polymorphism to Personalized Medicine* (pp. 193–207). Cambridge, UK: Cambridge University Press.

Martiskainen, H., Viswanathan, J., Nykanen, N. P., Kurki, M., Helisalmi, S., Natunen, T., Hiltunen, M. (2015). Transcriptional and mechanistic elucidation of Alzheimer’s disease risk genes in the brain and in vitro models. *Neurobiology of Aging*, 36(2), 1221.e1215–1228. https://doi.org/10.1016/j.neurobiolaging.2014.09.003

Mez, J., Chung, J., Jun, G., Kriegel, J., Bourlas, A. P., Sherva, R., ... Farrer, L. A. (2017). Two novel loci, COBL and SLCO1A2, for Alzheimer’s disease in African Americans. *Alzheimer’s & Dementia*, 13(2), 119–129. https://doi.org/10.1016/j.jalz.2016.09.002

Muller, A. J., DuHadaway, J. B., Donover, P. S., Sutanto-Ward, E., & Prendergast, G. C. (2005). Inhibition of indoleamine 2,3-dioxygenase, an immunoregulatory target of the cancer suppression gene Bin1, potentiates cancer chemotherapy. *Nature Medicine*, 11(3), 312–319. https://doi.org/10.1038/nm1196

Naj, A. C., Jun, G., Beecham, G. W., Wang, L. S., Vardaraj, B. N., Buros, J., ... Schellenberg, G. D. (2011). Common variants at MS4A4A/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer’s disease. *Nature Genetics*, 43(5), 436–441. https://doi.org/10.1038/ng.801

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, 2(12), e190. https://doi.org/10.1371/journal.pgen.0020190

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, 38(8), 904–909. https://doi.org/10.1038/ng1847

Ridge, P. G., Mukherjee, S., Crane, P. K., Kauwe, J. S., & Alzheimer’s Disease Genetics, C (2013). Alzheimer’s disease: Analyzing the missing heritability. *PLoS ONE*, 8(11), e79771. https://doi.org/10.1371/journal.pone.0079771

Shoemaker, R., Deng, J., Wang, W., & Zhang, K. (2010). Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Research*, 20(7), 883–889. https://doi.org/10.1101/gr.104695.109

Sims, R., van der Lee, S. J., Naj, A. C., Bellenguez, C., Badarinarayanan, N., Jakobsdottir, J., ... Schellenberg, G. D. (2017). Rare coding variants in PLCG2, AB13, and TREM2 implicate microglial-mediated innate immunity in Alzheimer’s disease. *Nature Genetics*, 49(9), 1373–1384. https://doi.org/10.1038/ng.3916

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., von Mering, C. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, 43(Database issue), D447–452. https://doi.org/10.1093/nar/gku1003.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Ma Y, Jun GR, Chung J, et al. CpG-related SNPs in the MS4A region have a dose-dependent effect on risk of late-onset Alzheimer disease. *Aging Cell*. 2019;18:e12964. https://doi.org/10.1111/ace1.12964

APPENDIX 1

MEMBERS OF THE ALZHEIMER DISEASE GENETICS CONSORTIUM

Marilyn S. Albert1, Roger L. Albin2,4, Liana G. Apostolova6, Steven E. Arnold6, Sanjay Asthana7,9, Craig S. Atwood17,9, Clinton T. Baldwin10, Michael M. Barnard11, Lisa L. Barnes12,13, Thomas G. Beach14, James T. Becker15, Duane Beekly16, Adam Boxer17, Sanjay A. Breen18, James R. Burke26, Joseph Buxbaum23, David A. Benöe18, Eileen H. Bigio17,18, Thomas D. Bird19,20, Deborah Blacker21,22, Bradley F. Boone23, James D. Bowen24, Adam Boxer25, James R. Burke26, Joseph Buxbaum108, Nigel J. Cairns27, Laura B. Cantweil28, Chuanhai Cao29, Chris S. Carlson30, Cynthia M. Carlson5, Regina M. Carney31, Minerva M. Carrasquillo32, Steven L. Carroll33, Helen C. Chui34, Clark,35 Paul Crane36, David H. Cribs36, Elizabeth A. Crocco31,
Center, Emory University, Atlanta, Georgia; 49Neurogenetics Program, University of California, Los Angeles, Los Angeles, California; 49Department of Neurology, Emory University, Atlanta, Georgia; 50Division of Genetics, Department of Medicine and Partners Center for Personalized Genetic Medicine, Brigham and Women’s Hospital and Harvard Medical School, Boston, Massachusetts; 51Department of Neurology, Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts; 52Center for Applied Genomics, Children’s Hospital of Philadelphia, Philadelphia, Pennsylvania; 53Department of Pathology (Neuropathology), University of Pittsburgh, Pittsburgh, Pennsylvania; 54Sanders-Brown Center on Aging, Department of Molecular and Biomedical Pharmacology, University of Kentucky, Lexington, Kentucky; 55Taub Institute on Alzheimer’s Disease and the Aging Brain, Department of Neurology, Columbia University, New York, New York; 56Department of Pathology, Duke University, Durham, North Carolina; 57Department of Genome Sciences, University of Washington, Seattle, Washington; 58Department of Medicine (Medical Genetics), University of Washington, Seattle, Washington; 59Sanders-Brown Center on Aging, Department of Neurology, University of Kentucky, Lexington, Kentucky; 60Department of Pathology and Laboratory Medicine, University of California, Davis, Sacramento, California; 61Department of Biostatistics, Boston University, Boston, Massachusetts; 62Department of Ophthalmology, Boston University, Boston, Massachusetts; 63Department of Biology, Brigham Young University, Provo, Utah; 64Department of Neurology, Oregon Health & Science University, Portland, Oregon; 65Department of Neurology, Portland Veterans Affairs Medical Center, Portland, Oregon; 66Department of Pathology and Laboratory Medicine, University of California, Irvine, Irvine, California; 67Department of Neurology, Boston University, Boston, Massachusetts; 68Department of Pathology, Boston University, Boston, Massachusetts; 69Department of Neuropsychology, University of California, San Francisco, San Francisco, California; 70The John P. Hussman Institute for Human Genomics, University of Miami, Miami, Florida; 71Department of Neurobiology and Behavior, University of California, Irvine, Irvine, California; 72Cleveland Clinic Lou Ruvo Center for Brain Health, Cleveland Clinic, Cleveland, Ohio; 73Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, Washington; 74Department of Pathology, University of Michigan, Ann Arbor, Michigan; 75University of Pittsburgh Alzheimer’s Disease Research Center, Pittsburgh, Pennsylvania; 76Department of Psychiatry, Johns Hopkins University, Baltimore, Maryland; 77Department of Preventive Medicine, University of Southern California, Los Angeles, California; 78Department of Medicine–Pulmonary, New York University School of Medicine, New York, New York; 79Department of Pathology, University of California, San Diego, La Jolla, California; 80Department of Medicine, University of Washington, Seattle, Washington; 81Northwest Research Group on Aging, University of Washington School of Nursing, Seattle, Washington; 82Department of Neurology, Northwestern University Feinberg School of Medicine, Chicago, Illinois; 83Department of Pathology, University of Southern California, Los Angeles, Los Angeles, California; 84Department of Neurology, Washington University, St. Louis, Missouri; 85Department of Pathology and Laboratory Medicine, Indiana University, Indianapolis, Indiana; 86Department of Biostatistics, Mayo Clinic, Rochester, Minnesota; 87Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota; 88Michigan Alzheimer’s Disease Center, Department of Neurology, University of Michigan, Ann Arbor, Michigan; 89Department of Neurology, University of Colorado School of Medicine, Aurora, Colorado; 90Alzheimer’s Disease Center, New York University, New York, New York; 91Gertrude H. Sergievsky Center, Columbia University, New York, New York; 92Department of Neurology, Columbia University, New York, New York; 93Tanz Centre for Research in Neurodegenerative Disease, University of Toronto, Toronto, Ontario, Canada; 94Department of Neurology, University of Texas Southwestern, Dallas, Texas; 95Department of Psychiatry, Mount Sinai School of Medicine, New York, New York; 96Department of Radiology and Imaging Sciences, Indiana University, Indianapolis, Indiana; 97Department of Psychiatry, University of Southern California, Los Angeles, California; 98Department of Pathology, University of Washington, Seattle, Washington; 99Cambridge Institute for Medical Research and Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom; 100Center for Human Genetics and Research, Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee; 101Department of Pathology, Johns Hopkins University, Baltimore, Maryland; 102Sanders-Brown Center on Aging, Department of Anatomy and Neurobiology, University of Kentucky, Lexington, Kentucky; 103Department of Pathology & Laboratory Medicine, University of California, Los Angeles, Los Angeles, California; 104Department of Psychiatry, Northwestern University Feinberg School of Medicine, Chicago, Illinois; 105Department of Psychiatry & Behavioral Sciences, Duke University, Durham, North Carolina; 106Department of Pathology, Oregon Health & Science University, Portland, Oregon; 107Evelyn F. McKnight Brain Institute, Department of Neurology, Miller School of Medicine, University of Miami, Miami, Florida; 108Department of Psychiatry, Mount Sinai Hospital, New York, New York; 109Center for Translational & Computational Neuromunology, Multiple Sclerosis Clinical Care and Research Center, Division of Neuroimmunology, Department of Neurology, Columbia University Medical Center, New York, New York; 110C.S. Kubik Laboratory for Neuropathology, Massachusetts General Hospital, Charlestown, Massachusetts; 111Department of Molecular Neuroscience, University College London, London, United Kingdom; 112Neurogenomics Division, Translational Genomics Research Institute, Phoenix, Arizona; 113Department of Epidemiology, University of Washington, Seattle, Washington; 114Division of Epidemiology, University of Pennsylvania, Philadelphia, Pennsylvania; 115Arizona Alzheimer’s Consortium, Banner Alzheimer’s Institute, Phoenix, Arizona; 116Department of Psychiatry, University of Arizona, Phoenix, Arizona; 117Department of Pathology (Neuropathology), Rush University Medical Center, Chicago, Illinois; 118New York Brain Bank, Columbia University, New York, New York.