



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

Yiyi Ma^{1,2} | Gyungah R. Jun^{1,3,4} | Jaeyoon Chung¹ | Xiaoling Zhang^{1,3} | Brian W. Kunkle⁵ | Adam C. Naj^{6,7} | Charles C. White^{8,9} | David A Bennett¹⁰ | Philip L. De Jager^{2,8,9} | Alzheimer's Disease Genetics Consortium | Richard Mayeux¹¹ | Jonathan L. Haines¹² | Margaret A. Pericak-Vance⁵ | Gerard D. Schellenberg⁷ | Lindsay A. Farrer^{1,3,4,13,14} | 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 PH 19-406.
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 RC2AG036528. Data for this study were prepared, archived, and distributed by the National Institute on

Abstract

CpG-related single nucleotide polymorphisms (CGS) have the potential to perturb DNA methylation; however, their effects on Alzheimer disease (AD) risk have not been evaluated systematically. We conducted a genome-wide association study using a sliding-window approach to measure the combined effects of CGSes on AD risk in a discovery sample of 24 European ancestry cohorts (12,181 cases, 12,601 controls) from the Alzheimer's Disease Genetics Consortium (ADGC) and replication sample of seven European ancestry cohorts (7,554 cases, 27,382 controls) from the International Genomics of Alzheimer's Project (IGAP). The potential functional relevance of significant associations was evaluated by analysis of methylation and expression levels in brain tissue of the Religious Orders Study and the Rush Memory and

Alzheimer's Disease Genetics Consortium group members are shown in Appendix 1

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Aging Cell* published by the Anatomical Society and John Wiley & Sons Ltd.

Aging Alzheimer's Disease Data Storage Site (NIAGADS) at the University of Pennsylvania and funded by NIA grant U24-AG041689-01. Samples from the National Centralized Repository for Alzheimer's Disease and Related Dementias (NCRAD), which receives government support under a cooperative agreement grant (U24-AG21886) awarded by the National Institute on Aging (NIA), were used in this study. We thank contributors who collected samples used in this study, as well as patients and their families, whose help and participation made this work possible. The NACC database is funded by NIA/NIH Grant U01-AG016976. NACC data are contributed by the NIA-funded ADCs: P30 AG019610 (PI Eric Reiman, MD), P30 AG013846 (PI Neil Kowall, MD), P50 AG008702 (PI Scott Small, MD), P50 AG025688 (PI Allan Levey, MD, PhD), P50 AG047266 (PI Todd Golde, MD, PhD), P30 AG010133 (PI Andrew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P50 AG005134 (PI Bradley Hyman, MD, PhD), P50 AG016574 (PI Ronald Petersen, MD, PhD), P50 AG005138 (PI Mary Sano, PhD), P30 AG008051 (PI Thomas Wisniewski, MD), P30 AG013854 (PI M. Marsel Mesulam, MD), P30 AG008017 (PI Jeffrey Kaye, MD), P30 AG010161 (PI David Bennett, MD), P50 AG047366 (PI Victor Henderson, MD, MS), P30 AG010129 (PI Charles DeCarli, MD), P50 AG016573 (PI Frank LaFerla, PhD), P50 AG005131 (PI James Brewer, MD, PhD), P50 AG023501 (PI Bruce Miller, MD), P30 AG035982 (PI Russell Swerdlow, MD), P30 AG028383 (PI Linda Van Eldik, PhD), P30 AG053760 (PI Henry Paulson, MD, PhD), P30 AG010124 (PI John Trojanowski, MD, PhD), P50 AG005133 (PI Oscar Lopez, MD), P50 AG005142 (PI Helena Chui, MD), P30 AG012300 (PI Roger Rosenberg, MD), P30 AG049638 (PI Suzanne Craft, PhD), P50 AG005136 (PI Thomas Grabowski, MD), P50 AG033514 (PI Sanjay Asthana, MD, FRCP), P50 AG005681 (PI John Morris, MD), P50 AG047270 (PI Stephen Strittmatter, MD, PhD). ROSMAP data were generated with support from NIA grants P30-AG10161, R01-AG17917, R01-AG36042, and U01-AG46152. This work was also supported by NIA grants R01-AG048927 and RF1-AG057519 (to LAF).

Aging Project (ROSMAP), and in whole blood of Framingham Heart Study participants (FHS). Genome-wide significant ($p < 5 \times 10^{-8}$) associations were identified with 171 1.0 kb-length windows spanning 932 kb in the *APOE* region (top $p < 2.2 \times 10^{-308}$), five windows at *BIN1* (top $p = 1.3 \times 10^{-13}$), two windows at *MS4A6A* (top $p = 2.7 \times 10^{-10}$), two windows near *MS4A4A* (top $p = 6.4 \times 10^{-10}$), and one window at *PICALM* ($p = 6.3 \times 10^{-9}$). The total number of CGS-derived CpG dinucleotides in the window near *MS4A4A* was associated with AD risk ($p = 2.67 \times 10^{-10}$), brain DNA methylation ($p = 2.15 \times 10^{-10}$), and gene expression in brain ($p = 0.03$) and blood ($p = 2.53 \times 10^{-4}$). Pathway analysis of the genes responsive to changes in the methylation quantitative trait locus signal at *MS4A4A* (cg14750746) showed an enrichment of methyltransferase functions. We confirm the importance of CGS in AD and the potential for creating a functional CpG dosage-derived genetic score to predict AD risk.

KEYWORDS

Alzheimer disease, DNA methylation, epigenetics, eQTL, genetics, mQTL

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

methylation (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 a haplotype of multiple CGSes is associated with DNA methylation patterns on a genome-wide scale (Ma et al., 2016). DNA methylation has been shown to influence risk of age-related diseases (Hunter et al., 2012; De Jager et al., 2014). For example, a genome-wide DNA methylation study reported association of AD pathological features with methylation changes at several loci (De Jager et al., 2014). Also, levels of DNA methylation of *GSTM1* and *GSTM5* have been associated with risk of age-related macular degeneration (Hunter et al., 2012).

In this study, we evaluated the association of AD with CGSes genome-wide and validated significant findings by expression quantitative trait locus (eQTL) and methylation QTL (mQTL) analyses.

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 ($\lambda = 1.099$, 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 $p = 1.27 \times 10^{-13}$, CGS dosage $p = 4.74 \times 10^{-18}$), *MS4A6A* (SKAT-O $p = 2.66 \times 10^{-10}$, CGS dosage $p = 3.40 \times 10^{-10}$), *MS4A4A* (SKAT-O $p = 6.36 \times 10^{-10}$, CGS dosage $p = 2.67 \times 10^{-10}$), *PICALM* (SKAT-O $p = 6.34 \times 10^{-9}$, CGS dosage $p = 1.42 \times 10^{-9}$), and *APOE* (SKAT-O $p = 2.99 \times 10^{-46}$, CGS dosage $p = 2.77 \times 10^{-556}$) (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, $p = 2.67 \times 10^{-10}$ 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, $p = 2.77 \times 10^{-556}$).

TABLE 1 Top-ranked windows associated with AD by SKAT-O and CG dosage methodologies in discovery stage

| Chr | Gene | Start | End | N of CGSes | P range of CGSes (min, max) | Window P ^a | Window P ^b | Beta (SE) |
|---------------------------------------|---------------------|-------------|-------------|------------|-----------------------------|-----------------------|-----------------------|---------------|
| Common loci identified by two methods | | | | | | | | |
| 2 | <i>BIN1</i> | 127,847,001 | 127,848,000 | 2 | (1.48E-13, 5.88E-06) | 1.27E-13 | 2.14E-03 | -0.02 (0.005) |
| 2 | <i>BIN1</i> | 127,881,001 | 127,882,000 | 2 | (2.5E-12, 3.67E-03) | 1.09E-03 | 4.74E-18 | 0.18 (0.02) |
| 11 | <i>MS4A6A</i> | 59,923,001 | 59,924,000 | 2 | (1.41E-10, 1.25E-09) | 2.66E-10 | 3.40E-10 | -0.01 (0.002) |
| 11 | <i>MS4A4A</i> | 60,087,501 | 60,088,500 | 2 | (8.44E-12, 1.23E-05) | 6.36E-10 | 2.67E-10 | -0.02 (0.003) |
| 11 | <i>PICALM</i> | 85,759,501 | 85,760,500 | 2 | (3.77E-05, 0.11) | 6.34E-09 | 9.28E-05 | 0.01 (0.002) |
| 11 | <i>PICALM</i> | 85,845,001 | 85,846,000 | 2 | (3.77E-05, 0.11) | 5.10E-02 | 1.42E-09 | 0.13 (0.02) |
| 19 | <i>APOE</i> | 45,411,501 | 45,412,500 | 2 | (<2.23E-308, 3.56E-28) | 2.99E-46 | 2.77E-556 | 0.2 (0.004) |
| Top loci identified by either method | | | | | | | | |
| 1 | <i>CR1</i> | 207,737,501 | 207,738,500 | 3 | (1.49E-10, 0.06) | 7.01E-04 | 8.57E-11 | -0.15 (0.02) |
| 6 | <i>LRFN2-UNC5CL</i> | 40,825,501 | 40,826,500 | 3 | (1.38E-06, 0.90) | 1.21E-08 | 8.00E-02 | -0.01 (0.005) |
| 6 | <i>TREM2</i> | 41,128,501 | 41,129,500 | 5 | (1.34E-06, 0.92) | 1.73E-08 | 6.80E-06 | -0.08 (0.002) |

^aP values obtained by SKAT-O test.

^bP values obtained by CGSes 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.

TABLE 2 Top-ranked windows associated with AD in replication stage

| Window | Chr | Start | End | Region or closest gene | N of CGSes | rsID | Effect allele | MAF | Top CGS in the window | | | Discovery stage | | | Replication stage | | |
|--------|-----|-------------|-------------|------------------------|------------|------------|---------------|------|-----------------------|-------------------|------------|-----------------|-------------------|-------------------------|-------------------|--------------|-------------------------|
| | | | | | | | | | rsID | Effect allele | MAF | P for top CGS | P for window | OR (95% CI) for top CGS | P for top CGS | P for window | OR (95% CI) for top CGS |
| 1 | 2 | 127,847,001 | 127,848,000 | BIN1 | 2 | rs35114168 | A | 0.39 | 1.27E-13 | 1.16 (1.12, 1.21) | 1.48E-13 | 2.96E-07 | 1.13 (1.08, 1.18) | 2.96E-07 | 1.13 (1.08, 1.18) | 2.96E-07 | |
| 2 | 11 | 59,923,001 | 59,924,000 | MS4A6A | 2 | rs983392 | G | 0.39 | 2.66E-10 | 0.88 (0.85, 0.91) | 1.41E-10 | 2.83E-05 | 0.91 (0.87, 0.95) | 2.53E-05 | 0.91 (0.87, 0.95) | 2.53E-05 | |
| 3 | 11 | 60,087,501 | 60,088,500 | MS4A4A/MS4A6E | 2 | rs4354705 | C | 0.36 | 6.36E-10 | 0.87 (0.83, 0.90) | 8.44E-12 | 6.93E-03 | 0.94 (0.90, 0.98) | 3.97E-03 | 0.94 (0.90, 0.98) | 3.97E-03 | |
| 4 | 11 | 85,759,501 | 85,760,500 | PICALM | 2 | rs694011 | T | 0.32 | 6.34E-09 | 0.90 (0.86, 0.95) | 3.77E-05 | 1.65E-05 | 0.90 (0.86, 0.94) | 9.30E-06 | 0.90 (0.86, 0.94) | 9.30E-06 | |
| 5 | 19 | 45,411,501 | 45,412,500 | APOE | 2 | rs429358 | C | 0.25 | 2.99E-46 | 3.73 (3.53, 3.94) | <2.23E-308 | <2.23E-308 | 3.47 (3.26, 3.69) | 1.09E-345 | 3.47 (3.26, 3.69) | 1.09E-345 | |

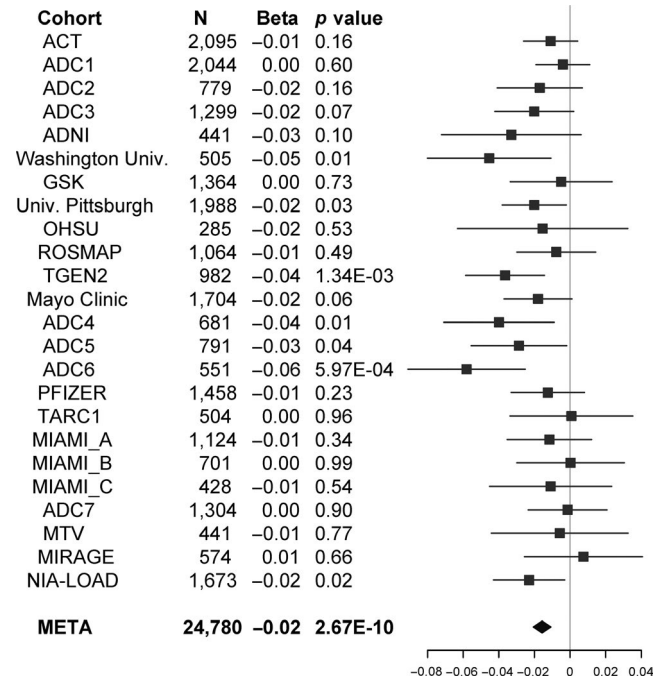


FIGURE 1 Forest plot of dose–response effect of the number of CpG dinucleotides created by the CGSes in the intergenic window close to *MS4A4A* on the logged odds ratio of AD. The filled square and horizontal line for each population or the filled diamond for the summary denote the estimated logged odds ratio and its 95% CI per unit increase in the number of CpG dinucleotides 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* ($R^2 = 0.98$) which is 165 kb from the top window at *MS4A4A* ($R^2 = 0.37$). 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 ($p \leq 0.05$) 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 $p = 2.15 \times 10^{-9}$ at cg14750746). This window was also nominally associated with increased methylation in blood after correcting for the same 176 methylation probes (nominal

TABLE 3 Comparisons of the top windows containing CGSes versus non-CGSes

| Chr | Start | End | Gene | P of window | | N of variants | | LD between CGS and non-CGS (R^2) | |
|-----|-------------|-------------|--------|-------------|----------|---------------|---------|--------------------------------------|----------------------|
| | | | | CGS | Non-CGS | CGS | Non-CGS | Top CGS and non-CGS | Any pairs (min, max) |
| 2 | 127,847,001 | 127,848,000 | BIN1 | 1.27E-13 | 3.30E-05 | 2 | 2 | 0.01 | (2.38E-05, 0.01) |
| 11 | 59,923,001 | 59,924,000 | MS4A6A | 2.66E-10 | 0.786 | 2 | 4 | 0.98 | (8.73E-03, 0.98) |
| 11 | 60,087,501 | 60,088,500 | MS4A4A | 6.36E-10 | 0.029 | 2 | 9 | 0.37 | (5.99E-04, 1) |
| 11 | 85,759,501 | 85,760,500 | PICALM | 6.34E-09 | 0.035 | 2 | 6 | 0.15 | (1.33E-03, 0.24) |
| 19 | 45,411,501 | 45,412,500 | APOE | 2.99E-46 | NA | 2 | 0 | NA | NA |

$p = 3.34 \times 10^{-4}$ 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 ($p = 2.53 \times 10^{-4}$). The *MS4A6A* window was associated with DNA methylation (adjusted $p = 1.47 \times 10^{-7}$) and gene expression ($p = 5.89 \times 10^{-26}$) 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 $p = 3.2 \times 10^{-6}$), there were 34 nominally associated genes ($p < 5 \times 10^{-3}$) (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 | | | |
| | | | Beta (SE) ^a | P1 ^a | P2 ^a | P ^b |
| BIN1 | 127,800,646 | cg00436254 | -1.87E-04 (2.28E-03) | 0.93 | 1.00 | 2.08E-03 |
| MS4A6A | 59,824,541 | cg01917716 | NA | NA | NA | NA |
| MS4A4A | 60,101,475 | cg14750746 | 5.60E-03 (8.11E-04) | 1.22E-11 | 2.15E-09 | 0.03 |
| PICALM | 85,566,560 | cg15822411 | -3.42E-03 (1.77E-03) | 0.05 | 1.00 | 1.00 |
| APOE | 45,395,297 | cg02613937 | -9.96E-04 (4.19E-04) | 0.02 | 1.00 | 0.42 |

^aStatistics obtained from CGSes dosage tests. P1 represents uncorrected *p*-values, and P2 represents Bonferroni corrected *p*-values calculated by multiplying the number of methylation probes included in the test which are within 1Mb distance to the window.

^bStatistics obtained from SKAT-O tests.

2005). Genes in the *MS4A* family have been shown to activate T cells and trigger production of inflammatory cytokines (Yan et al., 2013). Expression of *PICALM* was reduced in subjects who underwent gastric bypass surgery to reverse their pro-inflammatory state of obesity (Ghanim et al., 2012), and *PICALM* overexpression in vitro was found to reduce the endosomal localization of the mannose-6-phosphate receptor (M6PR) which binds to the herpes virus (Brunetti, Dingwell, Wale, Graham, & Johnson, 1998). It is still controversial whether *APOE-ε4* causes anti- or pro-inflammatory effects, but it is generally accepted that *APOE* is related to inflammation (Dorey, Chang, Liu, Yang, & Zhang, 2014). Our collective findings suggest that DNA methylation may be a molecular mechanism underlying aberrant inflammatory responses related to AD.

Our findings also suggest that the sliding-window approach focused on CGSes is useful for identifying loci whose influence on disease risk may involve clinically relevant epigenetic mechanisms. In the large GWAS conducted by the International Genomics of Alzheimer's Project (Lambert et al., 2013), approximately 44% of the top AD-associated SNPs are CGSes. However, not all of these CGSes were significantly associated with AD in our analysis (e.g., CGSes in *CR1*, *CD2AP*, and *CLU*; Table S1). Interestingly, none of these three loci were reported to have significant brain methylation changes related to AD pathology (De Jager et al., 2014), indicating that their effects on AD may not involve DNA methylation.

Our study has several limitations. All the identified top windows for AD were previously reported loci associated with AD (Guerreiro et al., 2013; Jonsson et al., 2013; Lambert et al., 2013; Naj et al., 2011). This was expected because the samples of our and previously published GWAS are highly overlapping. However, our study ascribes potential function to some of these results, especially those occurring in noncoding regions. In order to identify the relative importance of the CGSes in the top windows compared to non-CGSes, we performed conditional analysis adjusting for the top GWAS SNP. For all windows, the association signal for both the GWAS SNP and CpG dosage was attenuated when both were included in the model. In particular, for the intergenic window between *MS4A4A* and *MS4A6E*, the *p*-values for both CpG dosage and the GWAS variant had similar

reduction in significance (Table S7). The squared correlation (r^2) between the GWAS variant and the CGS with the largest influence on the dosage effect in *MS4A4A* window is 0.56. Thus, it is not possible to conclude from the conditional analysis whether the GWAS variant, the window CpG dosage, or another variant in the region that is correlated with both of these markers, is responsible for the association. We did not remove CGSes in high LD, which may inflate the number of significant findings. However, some of these associations may be independent because multiple adjacent methylated CpG sites can serve as the platform for chromatin binding proteins that lead to changes in chromatin state (Bartke et al., 2010). Another concern is that despite experimental evidence suggesting an optimal window size of 1kb, it is unknown whether other window sizes may increase power. Also, our selection of the default weights of variants has bias toward rare variants. Finally, we observed that the CGS most significantly associated with AD risk also has significant mQTL and eQTL effects that survive regional multiple test correction but do not achieve genome-wide significance.

In conclusion, we confirmed the importance of CGS in AD and the potential for creating a functional genetic score based on CpG dosage to predict disease risk. However, it is unknown whether these CGS signals act as causative mechanisms in AD progression. Further replication and mechanistic studies are necessary to validate these findings. Future genome-wide mQTL and eQTL analyses may extend our findings.

4 | EXPERIMENTAL PROCEDURES

4.1 | Genome-wide association analysis of CGSes with AD

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.

| | | | Gene expression | | | | |
|------------------------|-----------------|-----------------|------------------------|----------------|----------------|------------------------|----------------|
| Blood | | | Brain | | | Blood | |
| Beta (SE) ^a | P1 ^a | P2 ^a | Beta (SE) ^a | p ^a | P ^b | Beta (SE) ^a | p ^a |
| -5.48E-03 (6.40E-04) | 1.97E-17 | 5.56E-15 | 2.40 (2.27) | 0.29 | 0.41 | -8.40E-03 (4.15E-03) | 0.04 |
| -2.91E-03 (4.72E-04) | 8.13E-10 | 1.47E-07 | NA | NA | NA | -0.04 (3.55E-03) | 5.89E-26 |
| -2.49E-03 (6.93E-04) | 3.34E-04 | 0.06 | 0.14 (0.07) | 0.03 | 0.07 | 0.09 (0.02) | 2.53E-04 |
| -6.97E-04 (4.10E-04) | 0.09 | 1.00 | 3.14E-03 (0.41) | 0.99 | 0.76 | 7.68E-03 (5.11E-03) | 0.13 |
| -4.25E-03 (2.90E-03) | 0.14 | 1.00 | -2.93 (6.19) | 0.64 | 0.60 | 6.94E-03 (6.13E-03) | 0.26 |

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:0016864 | 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) ≤ 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-corrected 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., Younkin, C. S., Crook, J., Pankratz, V. S., ... Woltjer, 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 methylation. *Cell*, *143*(3), 470–484. <https://doi.org/10.1016/j.cell.2010.10.012>
- Bellenguez, C., Charbonnier, C., Grenier-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*(B), 289–300.
- Brunetti, 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., Boulden, 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, BIN1, RHBDF2 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>
- Goecks, 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 II 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., Rogava, 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 variant 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.jalz.2016.12.012>
- Jun, G., Ibrahim-Verbaas, C. A., Vronska, 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 DNA 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-Filippini, J., ... Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *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). Transcriptomics 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 SLC10A2, 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., Vardarajan, B. N., Buross, J., ... Schellenberg, G. D. (2011). Common variants at MS4A4/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., Badarinarayan, N., Jakobsdottir, J., ... Schellenberg, G. D. (2017). Rare coding variants in PLCG2, ABI3, 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>
- Tang, R., Feng, T., Sha, Q., & Zhang, S. (2009). A variable-sized sliding-window approach for genetic association studies via principal component analysis. *Annals of Human Genetics*, 73(Pt 6), 631–637. <https://doi.org/10.1111/j.1469-1809.2009.00543.x>
- Walshe, C. A., Beers, S. A., French, R. R., Chan, C. H., Johnson, P. W., Packham, G. K., ... Cragg, M. S. (2008). Induction of cytosolic calcium flux by CD20 is dependent upon B Cell antigen receptor signaling. *Journal of Biological Chemistry*, 283(25), 16971–16984. <https://doi.org/10.1074/jbc.M708459200>
- Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, 26(17), 2190–2191. <https://doi.org/10.1093/bioinformatics/btq340>
- Yan, Y., Li, Z., Zhang, G. X., Williams, M. S., Carey, G. B., Zhang, J., ... Xu, H. (2013). Anti-MS4a4B treatment abrogates MS4a4B-mediated protection in T cells and ameliorates experimental autoimmune encephalomyelitis. *Apoptosis*, 18(9), 1106–1119. <https://doi.org/10.1007/s10495-013-0870-2>
- Zhi, D., Aslibekyan, S., Irvin, M. R., Claas, S. A., Borecki, I. B., Ordovas, J. M., ... Arnett, D. K. (2013). SNPs located at CpG sites modulate genome-epigenome interaction. *Epigenetics*, 8(8), 802–806. <https://doi.org/10.4161/epi.25501>
- Zuccolo, J., Deng, L., Unruh, T. L., Sanyal, R., Bau, J. A., Storek, J., ... Deans, J. P. (2013). Expression of MS4A and TMEM176 Genes in Human B Lymphocytes. *Frontiers in Immunology*, 4, 195. <https://doi.org/10.3389/fimmu.2013.00195>

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/acel.12964>

APPENDIX 1

MEMBERS OF THE ALZHEIMER DISEASE GENETICS CONSORTIUM

Marilyn S. Albert¹, Roger L. Albin^{2–4}, Liana G. Apostolova⁵, Steven E. Arnold⁶, Sanjay Asthana^{7–9}, Craig S. Atwood^{7,9}, Clinton T. Baldwin¹⁰, Michael M. Barmada¹¹, Lisa L. Barnes^{12,13}, Thomas G. Beach¹⁴, James T. Becker¹⁵, Duane Beekly¹⁶, David A. Benönett¹², Eileen H. Bigio^{17,18}, Thomas D. Bird^{19,20}, Deborah Blacker^{21,22}, Bradley F. Boeve²³, James D. Bowen²⁴, Adam Boxer²⁵, James R. Burke²⁶, Joseph Buxbaum¹⁰⁸, Nigel J. Cairns²⁷, Laura B. Cantwell²⁸, Chuanhai Cao²⁹, Chris S. Carlson³⁰, Cynthia M. Carlsson⁸, Regina M. Carney³¹, Minerva M. Carrasquillo³², Steven L. Carroll³³, Helena C. Chui³⁴, Clark³⁵, Paul Crane⁸⁰, David H. Cribbs³⁶, Elizabeth A. Crocco³¹,

Carlos Cruchaga³⁷, Charles DeCarli³⁸, F. Yesim Demirci¹¹, Philip De Jager¹⁰⁹, Malcolm Dick³⁹, Dennis Dickson³², Ranjan Duara⁴⁰, Nilufer Ertekin-Taner^{32,41}, Kelley M. Faber⁴², Kenneth B. Fallon³³, Martin R. Farlow⁴³, Steven Ferris⁴⁴, Tatiana M. Foroud⁴², Matthew Frosch¹¹⁰, Douglas R. Galasko⁴⁵, Marla Gearing^{46,47}, Daniel H. Geschwind⁴⁸, Bernardino Ghetti⁸⁵, John Gilbert⁷⁰, Jonathan D. Glass⁴⁹, Alison M. Goate³⁷, Neill R. Graff-Radford^{32,41}, Robert C. Green⁵⁰, John H. Growdon⁵¹, Hakon Hakonarson⁵², Ronald L. Hamilton⁵³, Kara L. Hamilton-Nelson⁷⁰, John Hardy¹¹¹, Lindy E. Harrell³⁵, Elizabeth Head⁵⁴, Lawrence S. Honig⁵⁵, Matthew J. Huentelman¹¹², Christine M. Hulette⁵⁶, Bradley Hyman⁵¹, Gail P. Jarvik^{57,58}, Gregory A. Jicha⁵⁹, Lee-Way Jin⁶⁰, Gyungah Jun^{10,61,62}, M. Ilyas Kamboh^{11,75}, Anna Karydas²⁵, John S. K. Kauwe⁶³, Jeffrey A. Kaye^{64,65}, Ronald Kim⁶⁶, Neil W. Kowall^{67,68}, Joel H. Kramer⁶⁹, Patricia Kramer⁶⁴, Walter Kukull^{16,113}, Brian W. Kunkle⁷⁰, Frank M. LaFerla⁷¹, James J. Lah⁴⁹, Eric Larson⁸⁰, James B. Leverenz⁷², Allan I. Levey⁴⁹, Ge Li⁷³, Andrew P. Lieberman⁷⁴, Chiao-Feng Lin²⁸, Oscar L. Lopez⁷⁵, Kathryn L. Lunetta⁶¹, Constantine G. Lyketsos⁷⁶, Wendy J. Mack⁷⁷, Daniel C. Marson³⁵, Eden R. Martin¹⁷, Frank Martiniuk⁷⁸, Deborah Mash¹⁰⁷, Eliezer Masliah^{45,79}, Wayne C. McCormick⁸⁰, Susan M. McCurry⁸¹, Andrew N. McDavid³⁰, Ann C. McKee^{67,68}, Marsel Mesulam^{18,82}, Bruce L. Miller²⁵, Carol A. Miller⁸³, Joshua W. Miller⁶⁰, John C. Morris^{27,84}, Shubhabrata Mukherjee⁸⁰, Jill R. Murrell^{42,85}, Amanda J. Myers³¹, Adam C. Naj¹¹⁴, John M. Olichney³⁸, Vernon S. Pankratz⁸⁶, Joseph E. Parisi⁸⁷, Amanda Partch²⁸, Henry L. Paulson⁸⁸, William Perry⁷⁰, Elaine Peskind⁷³, Ronald C. Petersen²³, Aimee Pierce³⁶, Wayne W. Poon³⁹, Huntington Potter⁸⁹, Joseph F. Quinn⁶⁴, Ashok Raj²⁹, Murray Raskind⁷³, Eric Reiman^{115,116}, Barry Reisberg^{44,90}, Christiane Reitz^{55,91,92}, John M. Ringman⁵, Erik D. Roberson³⁵, Ekaterina Rogaeva⁹³, Howard J. Rosen²⁵, Roger N. Rosenberg⁹⁴, Mark A. Sager⁸, Mary Sano⁹⁵, Andrew J. Saykin^{42,96}, Julie Schneider^{12,117}, Lon S. Schneider^{34,97}, William W. Seeley²⁵, Amanda G. Smith²⁹, Joshua A. Sonnen⁹⁸, Salvatore Spina⁸⁵, Peter St George Hyslop^{93,99}, Robert A. Stern⁶⁷, Rudolph E. Tanzi⁵¹, Tricia A. Thornton-Wells¹⁰⁰, John Trojanowski²⁸, Juan C. Troncoso¹⁰¹, Debby W. Tsuang^{20,73}, Otto Valladares²⁸, Vivianna M. Van Deerlin²⁸, Linda J. Van Eldik¹⁰², Badri N. Vardarajan^{55,91,92}, Harry V. Vinters^{5,103}, Jean-Paul Vonsattel¹¹⁸, Li-San Wang²⁸, Sandra Weintraub^{18,104}, Kathleen A. Welsh-Bohmer^{26,105}, Jennifer Williamson⁵⁵, Sarah Wishnek⁷⁰, Randall L. Woltjer¹⁰⁶, Clinton B. Wright¹⁰⁷, Steven Younkin³², Chang-En Yu⁸⁰, and Lei Yu¹²

¹Department of Neurology, Johns Hopkins University, Baltimore, Maryland; ²Department of Neurology, University of Michigan, Ann Arbor, Michigan; ³Geriatric Research, Education and Clinical Center (GRECC), VA Ann Arbor Healthcare System (VAAHS), Ann Arbor, Michigan; ⁴Michigan Alzheimer Disease Center, Ann Arbor, Michigan; ⁵Department of Neurology, University of California, Los Angeles, Los Angeles, California; ⁶Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania; ⁷Geriatric Research, Education and Clinical Center (GRECC), University of Wisconsin, Madison, Wisconsin; ⁸Department of Medicine, University of Wisconsin, Madison, Wisconsin; ⁹Wisconsin Alzheimer's

Institute, Madison, Wisconsin; ¹⁰Department of Medicine (Biomedical Genetics), Boston University, Boston, Massachusetts; ¹¹Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania; ¹²Department of Neurological Sciences, Rush University Medical Center, Chicago, Illinois; ¹³Department of Behavioral Sciences, Rush University Medical Center, Chicago, Illinois; ¹⁴Civin Laboratory for Neuropathology, Banner Sun Health Research Institute, Phoenix, Arizona; ¹⁵Departments of Psychiatry, Neurology, and Psychology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; ¹⁶National Alzheimer's Coordinating Center, University of Washington, Seattle, Washington; ¹⁷Department of Pathology, Northwestern University Feinberg School of Medicine, Chicago, Illinois; ¹⁸Cognitive Neurology and Alzheimer's Disease Center, Northwestern University Feinberg School of Medicine, Chicago, Illinois; ¹⁹Department of Neurology, University of Washington, Seattle, Washington; ²⁰VA Puget Sound Health Care System/GRECC, Seattle, Washington; ²¹Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts; ²²Department of Psychiatry, Massachusetts General Hospital/Harvard Medical School, Boston, Massachusetts; ²³Department of Neurology, Mayo Clinic, Rochester, Minnesota; ²⁴Swedish Medical Center, Seattle, Washington; ²⁵Department of Neurology, University of California, San Francisco, San Francisco, California; ²⁶Department of Medicine, Duke University, Durham, North Carolina; ²⁷Department of Pathology and Immunology, Washington University, St. Louis, Missouri; ²⁸Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania; ²⁹USF Health Byrd Alzheimer's Institute, University of South Florida, Tampa, Florida; ³⁰Fred Hutchinson Cancer Research Center, Seattle, Washington; ³¹Department of Psychiatry and Behavioral Sciences, Miller School of Medicine, University of Miami, Miami, Florida; ³²Department of Neuroscience, Mayo Clinic, Jacksonville, Florida; ³³Department of Pathology, University of Alabama at Birmingham, Birmingham, Alabama; ³⁴Department of Neurology, University of Southern California, Los Angeles, Los Angeles, California; ³⁵Department of Neurology, University of Alabama at Birmingham, Birmingham, Alabama; ³⁶Department of Neurology, University of California, Irvine, Irvine, California; ³⁷Department of Psychiatry and Hope Center Program on Protein Aggregation and Neurodegeneration, Washington University School of Medicine, St. Louis, Missouri; ³⁸Department of Neurology, University of California, Davis, Sacramento, California; ³⁹Institute for Memory Impairments and Neurological Disorders, University of California, Irvine, Irvine, California; ⁴⁰Wien Center for Alzheimer's Disease and Memory Disorders, Mount Sinai Medical Center, Miami Beach, Florida; ⁴¹Department of Neurology, Mayo Clinic, Jacksonville, Florida; ⁴²Department of Medical and Molecular Genetics, Indiana University, Indianapolis, Indiana; ⁴³Department of Neurology, Indiana University, Indianapolis, Indiana; ⁴⁴Department of Psychiatry, New York University, New York, New York; ⁴⁵Department of Neurosciences, University of California, San Diego, La Jolla, California; ⁴⁶Department of Pathology and Laboratory Medicine, Emory University, Atlanta, Georgia; ⁴⁷Emory Alzheimer's Disease

Center, Emory University, Atlanta, Georgia; ⁴⁸Neurogenetics 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 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 Neuroimmunology, 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.