nature communications

Article

https://doi.org/10.1038/s41467-025-58159-w

Brain 5-hydroxymethylcytosine alterations are associated with Alzheimer's disease neuropathology

Received: 1 July 2024

Accepted: 11 March 2025

Published online: 22 March 2025

Check for updates

Jinying Zhao[1] ✉, Tongjun Gu[2,9], Cheng Gao[1,9], Guanhong Miao[1,9], Helena Palma-Gudiel[2], Lei Yu[3], Jingyun Yang[3], Yanling Wang[3], Yujing Li[4], Junghwa Lim[4], Ronghua Li[4], Bing Yao[4], Hao Wu[5], Julie A. Schneider[3], Nicholas Seyfried[6], Francine Grodstein[7], Philip L. De Jager[8], Peng Jin[4] ✉ & David A. Bennett[3] ✉

5-hydroxymethylcytosine, also known as the sixth DNA base of the genome, plays an important role in brain aging and neurological disorders such as Alzheimer's disease. However, little is known about its genome-wide distribution and its association with Alzheimer's disease pathology. Here, we report a genome-wide profiling of 5-hydroxymethylcytosine in 1079 autopsied brains (dorsolateral prefrontal cortex) of older individuals and assess its association with multiple measures of Alzheimer's disease pathologies, including pathological diagnosis of Alzheimer's disease, amyloid-β load, and PHFtau tangle density. Of 197,765 5-hydroxymethylcytosine regions detected, we identified 2821 differentially hydroxymethylated regions associated with Alzheimer's disease neuropathology after controlling for multiple testing and covariates. Many differentially hydroxymethylated regions are located within known Alzheimer's disease loci, such as *RIN3, PLCG2, ITGA2B*, and *USP6NL*. Integrative multi-omics analyses support a potential mechanistic role of 5-hydroxymethylcytosine alterations in Alzheimer's disease. Our study presents a large-scale genome-wide atlas of 5-hydroxymethylcytosine in Alzheimer's brain and offers insight into the mechanism underlying Alzheimer's disease pathogenesis.

Alzheimer's disease (AD) is a devastating neurodegenerative disorder affecting over 35 million people worldwide. The pathological hallmarks of AD include extracellular amyloid plaques, intracellular neurofibrillary tangles, and progressive loss of neurons in the brain. Both genetic and environmental factors are involved in the pathogenesis of AD. Large-scale genome-wide association studies (GWAS) have identified over 70 AD-associated loci, which collectively explain about 40% of the genetic variance in AD[1]. Accumulated evidence suggests that

[1]Health Informatics Institute, University of South Florida, Tampa, FL, USA. [2]Department of Epidemiology, College of Public Health and Health Professions, University of Florida, Gainesville, FL, USA. [3]Rush Alzheimer's Disease Center & Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA. [4]Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA. [5]Department of Biostatistics and Bioinformatics, Emory University School of Public Health, Atlanta, GA, USA. [6]Department of Biochemistry, Emory University School of Medicine, Atlanta, GA, USA. [7]Rush Alzheimer's Disease Center & Department of Internal Medicine, Rush University Medical Center, Chicago, IL, USA. [8]Center for Translational & Computational Neuroimmunology, Department of Neurology and the Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Medical Center, New York, NY, USA. [9]These authors contributed equally: Tongjun Gu, Cheng Gao, Guanhong Miao. ✉e-mail: jinyingz@usf.edu; peng.jin@emory.edu; David_A_Bennett@rush.edu

epigenetic alterations are also implicated in and provide additional insights into AD pathology.

DNA methylation (DNAm) at the fifth position of cytosine (5-methylcytosine, 5mC) is the best-studied epigenetic mechanism involved in gene expression and neurodegenerative disorders. 5-hydroxymethylcytosine (5hmC), also known as the sixth DNA base, is an oxidative product of 5mC catalyzed by the ten-eleven translocation (TET) family of proteins via the process of active demethylation. We and others have shown that 5hmC is particularly enriched in brain[2–4], dynamically regulated during neurodevelopment and aging[3–5], and responsive to stress[6]. The level of 5hmC varies substantially between different cell types and tissues[7,8], and altered 5hmC affects gene expression[9,10]. These findings suggest that 5hmC represents another layer of epigenetic regulation that may play an important role in brain aging and AD. However, only a few studies examined the association of 5hmC alterations with AD in human brain tissue samples. Among these, we[11,12] and others[13,14] have previously reported genome-wide associations of brain 5hmC alterations with AD, but these studies were small and results were conflicting. Other studies[15–20] reported an association of global 5hmC level with AD using immunohistochemistry, which does not provide site/region-specific information for 5hmC, and thus cannot evaluate the effect of specific 5hmC alterations on AD. To date, a detailed genome-wide profiling of brain 5hmC, its associations with AD neuropathologies as well as its link with other brain omics is lacking, especially in a large collection of human brain tissue samples.

In this study, we conducted a genome-wide mapping of brain 5hmC by selective chemical labeling sequencing (hMe-Seal) in 1,079 autopsied brains. We investigated the association of 5hmC alterations with AD neuropathologies in two well-characterized community-based cohorts of aging and dementia that include a brain donation at the time of death. To examine the potential functional role of 5hmC alterations in AD pathology, we integrated 5hmC with other brain omics data including DNA methylation, gene expression, protein expression, and histone modification in the same participants and further performed functional annotation and colocalization analyses.

Table 1 | Characteristics of ROSMAP study participants

Stratified by cohort			
Characteristics	**MAP (discovery, *n* = 600)**	**ROS (replication, *n* = 450)**	**All (*n* = 1050)**
Age at death (years)	90.0 ± 6.2	88.8 ± 6.8	89.5 ± 6.5
Female, *n* (%)	417 (69.5%)	313 (69.6%)	730 (69.5%)
Education (years)	14.5 ± 2.9	18.1 ± 3.4	16.1 ± 3.6
AD, *n* (%)	391 (65.2%)	288 (64.0%)	679 (64.7%)
Amyloid-β (Aβ) load	1.8 ± 1.2	1.5 ± 1.1	1.7 ± 1.2
Neurofibrillary tangles (PHFtau)	7.5 ± 8.7	6.5 ± 7.5	7.1 ± 8.2
PMI (hours)	7.9 ± 4.6	9.1 ± 7.1	8.4 ± 5.9
Stratified by pathological diagnosis of AD (yes/no)			
	AD (*n* = 679)	Non-AD (*n* = 371)	
Age at death (years)	90.4 ± 6.1	87.8 ± 6.9	
Female, *n* (%)	494 (72.8%)	236 (63.6%)	
Education (years)	16.0 ± 3.6	16.1 ± 3.5	
Amyloid-β (Aβ) load	2.3 ± 0.8	0.5 ± 0.7	
Neurofibrillary tangles (PHFtau)	9.5 ± 9.2	2.6 ± 2.6	
PMI (hours)	8.4 ± 5.4	8.5 ± 6.6	

Continuous variables were expressed as mean ± standard deviation and qualitative variables were expressed as *n* (%).
AD pathologic diagnosis of Alzheimer's disease, *PMI* postmortem interval.

Results

Description of study participants and data

We generated 5hmC profiles in 1117 postmortem brain tissue samples who were part of the Religious Order Study (ROS)[21] or the Rush Memory and Aging Project (MAP)[22], collectively called ROSMAP. The procedures for sample selection are shown in Supplementary Fig. 1. All participants were without known dementia at enrollment and followed annually until death. The 5hmC data was generated by hMe-Seal[3,4] in a sample of dorsolateral prefrontal cortex (DLPFC) from each participant (Methods). After data pre-processing and quality control, our dataset included 197,765 5hmC regions (Supplementary Data 1) in 1050 older individuals (mean age: 89.5, age range: 65.9–108.3, Table 1). A median of 33.14 million reads were sequenced per sample. Batch effects and other unwanted variations in the 5hmC data were removed using the program RUVr (Methods), and residuals were used in downstream statistical analyses. More details for the distribution of brain 5hmC data can be found in Supplementary Fig. 2–3.

Our analyses focused on identifying differentially hydroxymethylated regions (DhMRs) associated with three AD neuropathological phenotypes: a binary variable for neuropathologic diagnosis of AD (yes/no) and two quantitative measures of molecular specificity for AD: amyloid-β (Aβ) load and paired helical filaments of tau protein density (PHFtau). Multi-omics data in the dorsolateral prefrontal cortex from the same individuals were also available for most participants[23], e.g., genetics (whole-genome sequencing), transcriptomics (RNA-seq), histone acetylation (H3K9ac) by ChIP-Seq[24], and proteomics data by tandem mass tags-mass spectrometry (TMT-MS)[25].

Differentially hydroxymethylated regions (DhMRs) are associated with AD neuropathology. We employed a multi-stage approach to identify and validate AD-associated DhMRs (Fig. 1). In stage I (discovery), we performed a genome-wide scan to identify DhMRs associated with each quantitative neuropathological phenotype in 600 participants from MAP. In stage II (replication), we replicated the top hits in 450 participants from ROS, and then meta-analyzed the results from both cohorts. In stage III (functional annotation), we carried out integrative multi-omics analysis to annotate and functionally validate the putative DhMRs (i.e., those replicated in stage II) using gene expression, histone acetylation, and proteomics data generated in the same brain cortex. This analysis seeks to determine whether alterations in DNA hydroxymethylome meaningfully affect gene expression and other molecular processes and helps identify potential functional genes in the putative regions.

Our discovery stage included 600 participants from MAP. We identified differentially hydroxymethylated regions (DhMRs) associated with three neuropathological phenotypes. Given the large number of 5hmC regions and their inter-correlations, we used elastic net (EN) implemented in *glmnet*[26] to select 5hmC regions associated with each neuropathological measure (Methods). The model adjusted for age, sex, postmortem interval (PMI), and education. This analysis resulted in 2939 DhMRs associated with AD diagnosis (Supplementary Data 2), 1,770 DhMRs associated with Aβ load (Supplementary Data 3), and 3110 DhMRs associated with PHFtau tangles (Supplementary Data 4). Since elastic net shrinks regression coefficients (i.e., effect size) towards zero, we then estimated the effect size of the selected 5hmC regions by linear regression, adjusting for the same covariates. At the Benjamin-Hochberg (BH) corrected false discovery rate (FDR) < 0.05, 2705 regions (92% of EN-selected) were associated with AD diagnosis (Supplementary Data 5), 1750 regions (98.9% of EN-selected) were associated with Aβ load (Supplementary Data 6) and 2626 regions (84.4% of EN-selected) were associated with PHFtau tangles (Supplementary Data 7) in the regression analysis. Of these, higher levels of 5hmC at 1362, 1260, and 1644 regions were associated with a lower risk of AD or lower levels of Aβ and PHFtau. In comparison,

https://doi.org/10.1038/s41467-025-58159-w

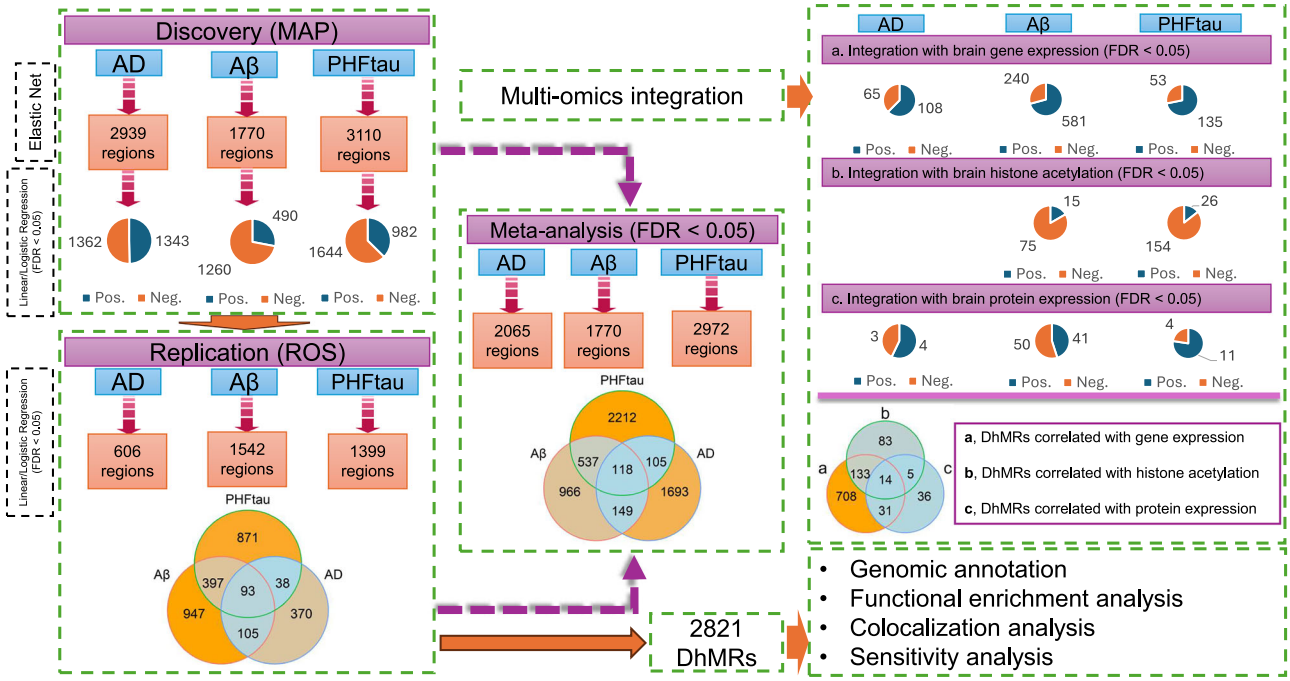

Fig. 1 | Schematic illustration of our study design and major findings. DhMR: differentially hydroxymethylated region; FDR: false discovery rate.

higher levels of 5hmC at 1343, 490 and 982 regions were associated with a higher risk of AD diagnosis or higher levels of Aβ and PHFtau.

We replicated these DhMRs in an independent sample comprising 450 participants from the ROS cohort. At FDR < 0.05, a total of 2821 DhMRs, including 606 (22.4% of those identified in the discovery sample) associated with AD diagnosis, 1542 (88.1% of those identified in the discovery sample) associated with Aβ load, and 1399 (53.3% of those identified in the discovery sample) associated with PHFtau tangles, remained statistically significant with the same direction of association (Supplementary Data 8–10). Figure 2 illustrates the AD-associated 5hmC regions (i.e., DhMRs) identified in either MAP (discovery) or ROS (replication) or both. The top 15 ranked DhMRs associated with AD diagnosis in both cohorts are shown in Table 2. Majority (>74%) of the identified DhMRs showed consistent effect sizes and direction of association across the two cohorts (Fig. 3), indicating the robustness of our findings. Some DhMRs appeared to be clustered in the same chromosomal segment and their hydroxymethylation levels were highly correlated (Fig. 4), suggesting that these regions might capture the same effect on AD pathology. Moreover, some DhMRs harbor known AD GWAS loci (Fig. 5).

Of the 2821 DhMRs replicated in ROS, 93 regions were significantly associated with all three neuropathological phenotypes: AD diagnosis, Aβ,and PHFtau (Supplementary Data 11). Notably, 43 DhMRs (out of 2821) also harbor known AD variants (Supplementary Data 12). However, the association between 5hmC and AD pathology appeared not to be mediated by genetic variants because additionally adjusting for the genotype of lead SNPs in the candidate regions did not substantially attenuate the observed associations (Supplementary Data 13). Of note, among the 43 DhMRs harboring known AD variants, seven DhMRs (C19P081372, C19P082146, C19P083148, C19P083158, C19P083160, C19P083167, and C19P083192) are located on chromosome 19 where the *APOE* gene resides. To examine whether the associations of these 5hmC regions with AD neuropathology are driven by *APOE*, we further adjusted for *APOE* genotype in a model that included covariates (age at death, sex, education, PMI) and the lead SNP genotype in each candidate region. It showed that the observed associations on chromosome 19 remained significant after further controlling for *APOE* genotype (Supplementary Data 14). These results suggest that 5hmC alterations and their proximal genetic variants may represent different genomic variations that independently affect AD pathologies.

The effect of an individual 5hmC region on AD pathology was in general small to moderate (Supplementary Data 15). On average, a single DhMR explained 1.95% (range: 0–8.86%) and 0.60% (range: 0–2.93%) of the variance in Aβ and PHFtau, respectively. The 93 DhMRs in common collectively explained 30.9% and 13.7% of the variance in Aβ load and PHFtau tangles, respectively. In comparison, the 71 CpGs detected in previous EWAS (Illumina 450 K BeadChips) explained 28.7% of the variation in neuritic plaques in ROSMAP[27]. Interestingly, majority of the identified DhMRs showed lower DNA hydroxymethylation in participants with AD diagnosis (Fig. 6). Similarly, for most of the DhMRs associated with Aβ and PHFtau, a higher level of DNA hydroxymethylation was associated with a lower level of Aβ load and PHFtau tangles. These findings suggest that hypo-hydroxymethylation in human brain is associated with AD neuropathology.

To increase statistical power and identify additional disease-related 5hmC alterations, we combined the results from both cohorts by fixed-effect meta-analysis (Methods). At FDR < 0.05, 5hmC alterations in 2065, 1770, and 2972 regions were significantly associated with AD diagnosis (Supplementary Data 16), Aβ (Supplementary Data 17), and PHFtau (Supplementary Data 18), respectively. Of these, 118 regions were associated with all three phenotypes (Supplementary Data 19). These 118 regions in common collectively explained 34.9% and 17.6% of variance in Aβ load and PHFtau tangles, respectively.

Genomic features of the identified DhMRs. To better understand the functional consequences of the identified DhMRs, we examined their distributions in relation to genomic features. Of the 2821 DhMRs (606 associated with AD diagnosis, 1542 associated with Aβ load, 1399 associated with PHFtau tangles, annotated to 2642 distinct genes) identified in stage II (replication), about 69% reside in exons, ~15% in intergenic regions, and ~17% in introns (Supplementary Fig. 4a). We further annotated the AD-associated 5hmC regions to chromatin states based on a reference chromatin map of dorsolateral prefrontal cortex (E073_15_coreMarks_hg38lift_stateno.bed) generated by the NIH

https://doi.org/10.1038/s41467-025-58159-w

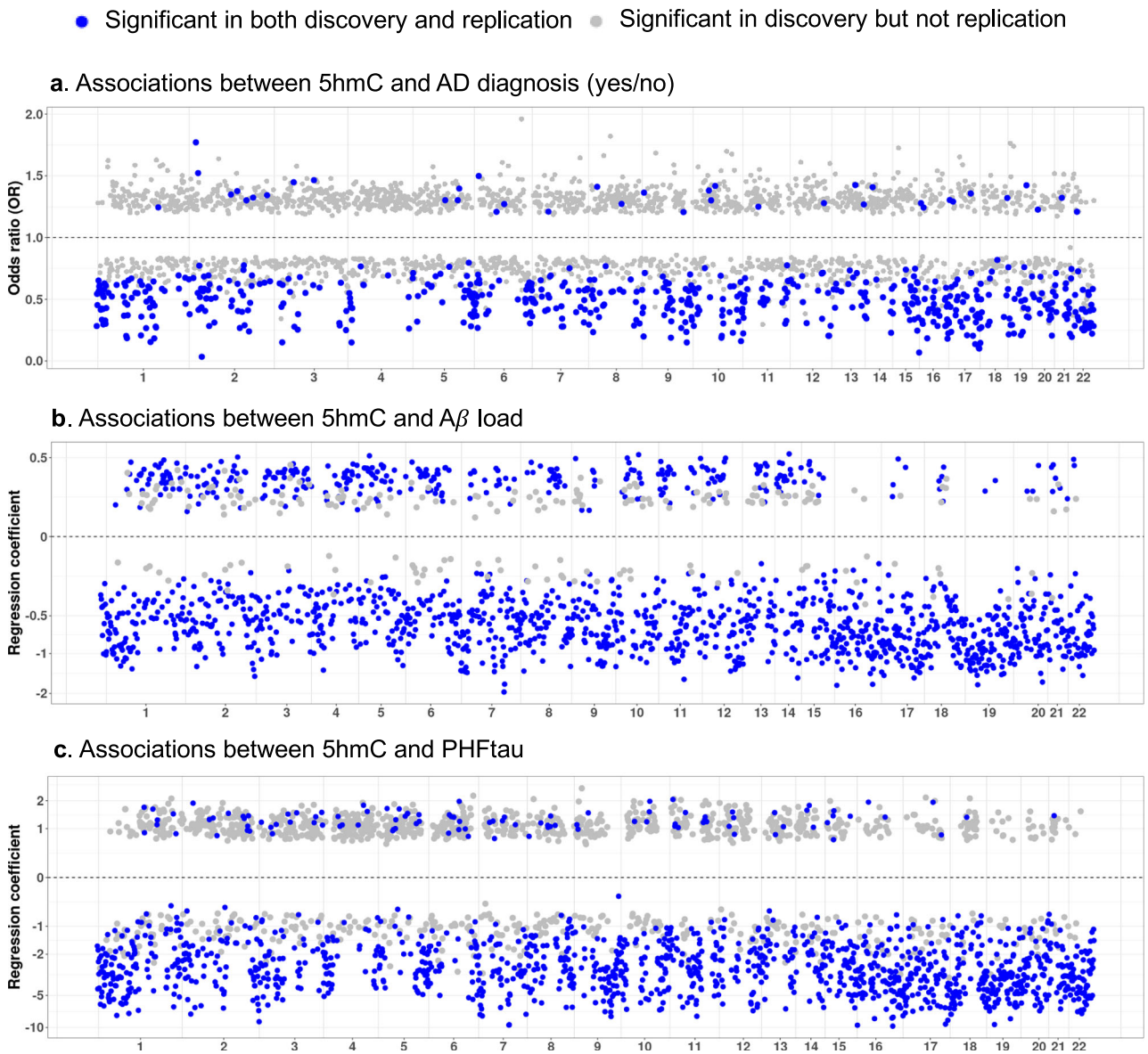

Fig. 2 | Miami plots displaying the association between 5hmC alterations and AD pathology. Each dot represents a 5hmC region, and all dots shown in the figure were selected by elastic net. Regions that were significant in the discovery (MAP) but not the replication (ROS) cohort are marked in gray. Regions that were significant in both the discovery and replication cohorts are marked in blue. **a** Association of 5hmC with AD diagnosis. Odds ratios (ORs) for AD diagnosis were derived from logistic regression; **b** Association of 5hmC with Aβ load. Regression coefficients were derived from linear regression; **c** Association of 5hmC with PHFtau. Regression coefficients were derived from linear regression. All models adjusted for age at death, sex, education, and PMI.

Roadmap Epigenomics Project (https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html#core_15state). AD-associated 5hmC regions are more frequently found at genomic positions corresponding to chromatin states. Compared to non-AD associated 5hmC regions (total 194,944), AD-associated regions (total 2821) were significantly more likely to coincide with regions related to transcription start site (TSS), transcription states, and enhancer states, but less likely to co-position with regions corresponding to heterochromatin and quiescent states (Supplementary Fig. 4b, Supplementary Data 20). These observations suggest that altered 5hmC may contribute to AD through influencing chromatin states and downstream gene transcription.

Integrative multi-omics analyses revealed functional implication of 5hmC alterations in AD pathology. To better understand the functional consequences of the identified DhMRs, we next performed several integrative multi-omics analyses using matched DNA hydroxymethylation and other brain omics data (DLPFC) measured on the same individuals, including gene expression, histone acetylation, and proteomics. Below we briefly described these findings.

a). AD-associated 5hmC alterations correlated with brain gene expression. Focusing on the 2821 DhMRs associated with AD pathologies (606 regions associated with AD diagnosis, 1542 regions associated with Aβ, 1399 regions associated with PHFtau, annotated to 2642 distinct genes), we evaluated whether altered DNA hydroxymethylation affects the expression of genes located in the vicinity (±50 kb) of each DhMR using brain RNA-seq data (DLPFC) from the same participants (Supplementary Data 21). At FDR < 0.05, the levels of 886 DhMRs (annotated to 818 distinct genes) were correlated with gene expression (Supplementary Data 22). Of these, 108 AD-associated, 581 Aβ-associated and 135 PHFtau-associated 5hmC regions were positively correlated with gene expression, whereas 65 AD-associated, 240 Aβ-associated and 53 PHFtau-associated 5hmC regions

https://doi.org/10.1038/s41467-025-58159-w

Table 2 | Top fifteen differentially hydroxymethylated 5hmC regions (DhMRs) associated with pathological diagnosis of AD

DhMR	Chr	Position (bp, start-end)	Discovery (MAP)		Replication (ROS)		[a]Selected known gene(s) within 50 kb of DhMR
			OR (95% CI)	*p*-value	OR (95% CI)	*p*-value	
C01P001166	1	22843393–22935266	0.15 (0.10, 0.24)	5.88×10^{-17}	0.22 (0.14, 0.34)	4.48×10^{-11}	*MIR4253, EPHB2, LACTBL1*
C02P101050	2	235901491–236150321	0.03 (0.02, 0.07)	1.97×10^{-24}	0.18 (0.10, 0.30)	1.19×10^{-10}	*GBX2, ASB18, GBX2-AS1, AGAP1*
C03P110503	3	13288402–13407287	0.15 (0.10, 0.23)	1.58×10^{-17}	0.30 (0.21, 0.45)	2.22×10^{-9}	*NUP210, IQSEC1*
C06P149062	6	5090194–5160804	0.27 (0.19, 0.38)	1.83×10^{-13}	0.31 (0.21, 0.46)	8.06×10^{-9}	*MIR3691, PPP1R3G, LYRM4*
C07P167237	7	98939735–98984755	0.28 (0.20, 0.38)	4.44×10^{-15}	0.32 (0.22, 0.47)	1.65×10^{-9}	*TRRAP, SMURF1*
C12P043553	12	109508902–109572682	0.20 (0.14, 0.29)	1.18×10^{-18}	0.36 (0.26, 0.51)	2.30×10^{-9}	*MMAB, UBE3B, MVK, KCTD10*
C13P051680	13	110209697–110252420	0.36 (0.27, 0.48)	4.78×10^{-12}	0.34 (0.24, 0.48)	2.77×10^{-10}	*COL4A1*
C16P064804	16	3944997–4060779	0.07 (0.04, 0.13)	3.40×10^{-17}	0.10 (0.05, 0.20)	2.87×10^{-11}	*ADCY9*
C17P075085	17	78546049–78652324	0.10 (0.06, 0.17)	4.03×10^{-18}	0.15 (0.09, 0.27)	7.39×10^{-11}	*SCAT1, DNAH17, CYTH1*
C19P082282	19	35070278–35106894	0.21 (0.14, 0.31)	3.91×10^{-16}	0.36 (0.25, 0.53)	1.28×10^{-7}	*HPN-AS1, FXYD3, MIR6887, HPN, LGI4, FXYD1, SCN1B, FXYD7, FXYD5, GRAMD1A*
C21P107042	21	36802083–36828245	0.46 (0.37, 0.59)	1.01×10^{-10}	0.42 (0.32, 0.56)	6.24×10^{-10}	*HLCS-IT1, HLCS*
C22P109014	22	39010977–39058414	0.25 (0.17, 0.37)	3.44×10^{-12}	0.22 (0.14, 0.34)	3.18×10^{-11}	*APOBEC3D, APOBEC3F, APOBEC3C, APOBEC3B, APOBEC3G, APOBEC3H, APOBEC3A*
Overlapped with known AD loci							
C05P137398	5	14695819–14753803	0.32 (0.23, 0.43)	3.96×10^{-13}	0.43 (0.32, 0.59)	8.07×10^{-8}	*OTULIN, ANKH*
C08P181044	8	144148659–144205346	0.20 (0.13, 0.31)	3.77×10^{-12}	0.32 (0.20, 0.52)	5.56×10^{-6}	*MROH1, TSSK5P, HGH1, WDR97, MAF1, SHARPIN*
C10P017404	10	11428830–11467102	0.38 (0.29, 0.49)	9.07×10^{-14}	0.41 (0.31, 0.54)	1.29×10^{-10}	*USP6NL*

Odds ratios (ORs) were obtained from logistic regression, adjusting for age at death, sex, education, PMI, and unwanted variations. All DhMRs in the table remained significant after multiple testing corrections (i.e., two-sided FDR < 0.05). The bolded regions were also significantly associated with Aβ load and PHFtau tangles.
Chr chromosome, *DhMRs* differentially hydroxymethylated regions, *OR* odds ratio, *CI* confidence interval.
[a]Selected known genes within 50 kb of the DhMR, with the first gene being the closest to DhMR.

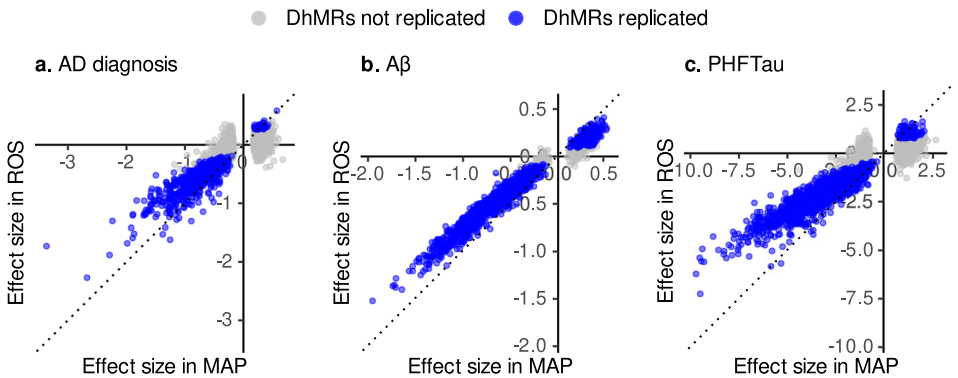

Fig. 3 | Comparison of the effect sizes (regression coefficients) and direction of association between the discovery (MAP, *n* = 600) and replication (ROS, *n* = 450) cohorts. Effect sizes for AD diagnosis were derived from logistic regression, while effect sizes for Aβ load and PHFtau were derived from linear regression. All models adjusted for age at death, sex, education, and PMI. **a** AD diagnosis: Spearman's correlation of effect sizes between cohorts: $r = 0.70$ ($P < 2.2\times10^{-16}$); concordance rate for the direction of association across cohorts: 74.3% (sign test $P < 2.2\times10^{-16}$). **b** Aβ load: Spearman's correlation of effect sizes between cohorts: $r = 0.98$ ($P < 2.2\times10^{-16}$); concordance rate for the direction of association across cohorts: 98.3% (sign test $P < 2.2\times10^{-16}$). **c** PHFtau: Spearman's correlation of effect sizes between cohorts: $r = 0.91$ ($P < 2.2\times10^{-16}$); concordance rate for the direction of association across cohorts: 88.6% (sign test $P < 2.2\times10^{-16}$). The significant correlations of effect sizes and high concordance rates for direction of association across cohorts signify the robustness of our findings. All statistical tests were two-sided. Source data are provided as a Source Data file.

were negatively correlated with gene expression (Supplementary Fig. 5, Supplementary Data 22). Moreover, the expression level for most of these genes (83%) was also significantly associated with AD diagnosis, Aβ or PHFtau (Supplementary Data 23). These genes were clustered into multiple protein-protein interaction (PPI) networks (Supplementary Fig. 6). The most significant biological processes and pathways of these functional clusters were associated with mTOR signaling, mediator complex and transcriptional regulation of RNA polymerase transcripts (II and III), SWI/SNF and npBAF complex, γ-secretase complex, DNA replication and repair, etc. (Supplementary Fig. 6, Supplementary Data 24).

b). AD-associated 5hmC alterations correlated with H3K9 acetylation (H3K9ac). To further functionally annotate the AD-associated 5hmC regions, we calculated the correlation between the level of 5hmC and the level of H3K9ac generated in the dorsolateral prefrontal cortex from the same individuals (Supplementary Data 21) in ROSMAP[24]. At FDR < 0.05, none of the 5hmC regions associated with AD diagnosis (out of 606) was correlated with H3K9ac, despite that eight regions did show nominal correlation with H3K9ac at FDR < 0.1 (Supplementary Data 25). In comparison, the levels of 90 Aβ-associated (out of 1542) and 180 PHFtau-associated (out of 1399) 5hmC regions were correlated with the level of H3K9ac at FDR < 0.05 (Supplementary Data 26–27).

https://doi.org/10.1038/s41467-025-58159-w

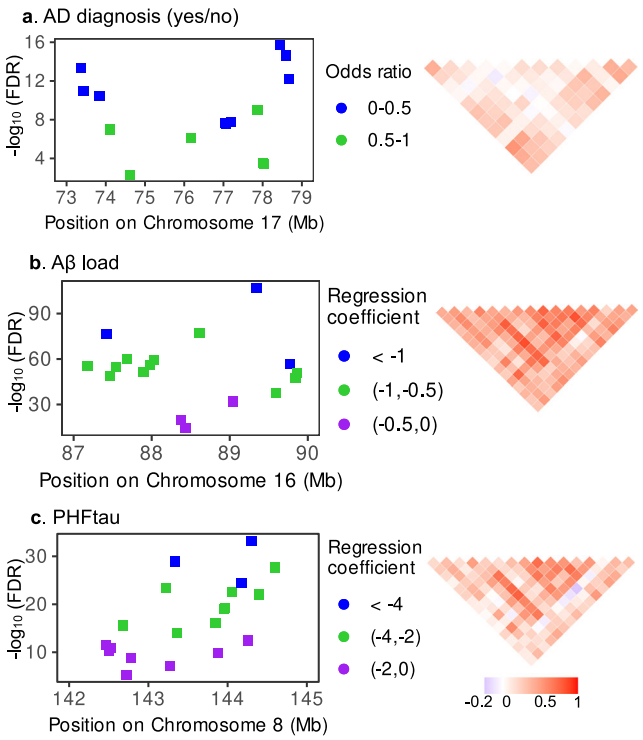

Fig. 4 | Regional plots depicting the associations of 5hmC regions with AD neuropathology (i.e., DhMRs) in three candidate regions on chromosomes 17, 16, and 8. Each dot represents a DhMR. Dot colors reflect the magnitude of association effect size (i.e., odds ratio (OR) or regression coefficient) between altered 5hmC and AD neuropathology. ORs for AD diagnosis were derived from logistic regression, while regression coefficients for Aβ load and PHFtau were derived from linear regression. All models adjusted for age at death, sex, education, and PMI. Multiple testing was performed using the Benjamini-Hochberg false discovery rate (FDR) method. Heatmap (right) depicts the partial correlations (i.e., Spearman correlation of DhMR residuals after regressing out age at death, sex, education, PMI) between selected DhMRs in the candidate region. X-axis: genomic position (Mb). Y-axis: -$\log_{10}$(two-sided FDR). **a** DhMRs associated with AD diagnosis (yes/no) in a candidate region on chromosome 17 (73-79 Mb); **b** DhMRs associated with Aβ load in a candidate region on chromosome 16 (87-90 Mb); **c** DhMRs associated with PHFtau tangles in a candidate region on chromosome 8 (142-145 Mb). Source data are provided as a Source Data file.

Interestingly, majority (~85%) of these regions were negatively correlated with H3K9ac expression. These results suggest that 5hmC alterations may affect AD through regulation of gene transcription.

c). AD-associated 5hmC alterations correlated with brain protein expression. In a subset of individuals (Supplementary Data 21) with both 5hmC and brain proteomics data (DLPFC)[28], we calculated the correlation of AD-associated 5hmC with protein abundances (Methods). At FDR < 0.05, the expression levels of 91 proteins were significantly correlated with the abundance of 5hmC regions that were significantly associated with AD pathologies (Supplementary Data 28). These proteins are significantly enriched in biological processes related to amyloid precursor protein catabolic process, regulation of Aβ formation, regulation of signal transduction, neurogenesis, phosphorylation, synapse, protein binding, proteolysis, neuronal ion channel clustering, regulation of protein modification process, etc. (Supplementary Data 29).

AD-associated 5hmC alterations are enriched in annotated functional regions. To examine whether the identified DhMRs were randomly distributed across various regulatory and cell-type specific elements in the genome, we performed enrichment analysis of the AD-associated 5hmC regions using GARFIELD v2.0[29] (Methods). This method leverages GWAS findings and various functional (e.g., regulatory or protein-coding) annotations to identify features relevant to a trait of interest at various GWAS *p*-value cutoffs, and then tests the significance of the enrichment by a generalized linear model. GARFIELD accounts for major sources of confounding factors by incorporating high-LD proxies ($r^2 > 0.8$), minor allele frequency (MAF), and transcription start site (TSS) distance as categorical covariates. In this analysis, we used summary statistics of susceptibility variants located in an extended region of the DhMRs (±500 kb) based on three large-scale GWAS meta-analysis for AD[30–32]. At GWAS threshold of 1×10^{-5}, genetic variants residing in regions harboring AD-associated 5hmC alterations were significantly enriched in DNase I hypersensitive sites (DHSs), transcribed histone modifications, TSS, and transcribed chromatin states, but depleted in repressed chromatin states and H3K27me3 in various tissues such as blood, blood vessel, embryonic stem cells (ESCs), colon, liver, and fetal intestine(Supplementary Figs. 7–9; Supplementary Data 30–32). Together, these results suggest that altered 5hmC may play a functional role in AD pathology.

Colocalization of AD-associated 5hmC regions with known GWAS loci and eQTLs for AD. To investigate the potential causal role of the identified 5hmC regions in AD pathology, we performed colocalization analysis by *coloc* (Methods) using summary statistics from three large-scale GWAS meta-analysis for AD[30–32]. At PP4 > 0.8 (i.e., a high posterior probability that a single shared variant is responsible for two signals), three 5hmC regions, including one associated with Aβ (C12P044736) and two associated with PHFtau (C12P044736, C14P057817), colocalize with known AD genes (*SCARB1* and *SLC24A4/RIN3*) near the variants rs12582221 and rs12590654 (Supplementary Data 33). We also performed colocalization analysis using summary statistics from a cell-type specific *cis*-eQTL analysis of eight human brain cell types[33] and found that one of the colocalized genes, *SLC24A4/RIN3*, is expressed in both astrocytes (Supplementary Fig. 10) and microglia (Supplementary Fig. 11). Collectively, these findings suggest that the identified AD-associated 5hmC regions may play a causal role in AD pathology.

Results of sensitivity analysis. The observed associations of 5hmC regions on chromosome 19 with AD pathology are unlikely to be driven by *APOE* because additional adjustment for *APOE* genotype did not change the results (Supplementary Data 14). In addition, further adjustment for cell type proportions (estimated using brain snRNA-seq data in ROSMAP participants, see Methods) did not attenuate the observed associations of 5hmC with any of the three phenotypes (i.e., AD diagnosis, Aβ, PHFtau), suggesting that cell type proportions did not confound our results. Finally, the observed associations of 5hmC with Aβ or PHFtau slightly attenuated but remained significant after controlling for each other (Supplementary Data 34–35), suggesting that 5hmC alterations may have an independent effect on these two hallmarks of AD pathologies.

Discussion

In this large-scale genome-wide profiling of brain cortex 5hmC, we identified extensive differences of DNA hydroxymethylation in relation to AD pathological traits. Specifically, we identified nearly 3000 differentially hydroxymethylated regions (DhMRs) associated with AD neuropathology after adjusting for covariates and multiple testing. Of these, 118 DhMRs were associated with all three neuropathological measurements, including a pathological diagnosis of AD, amyloid-β load, and tau tangle density. In addition, many identified regions harbor known GWAS genes or loci for AD. Integrative multi-omics analyses suggest a potential mechanistic role of altered DNA hydroxymethylation in AD pathology. For example, integration of AD-

https://doi.org/10.1038/s41467-025-58159-w

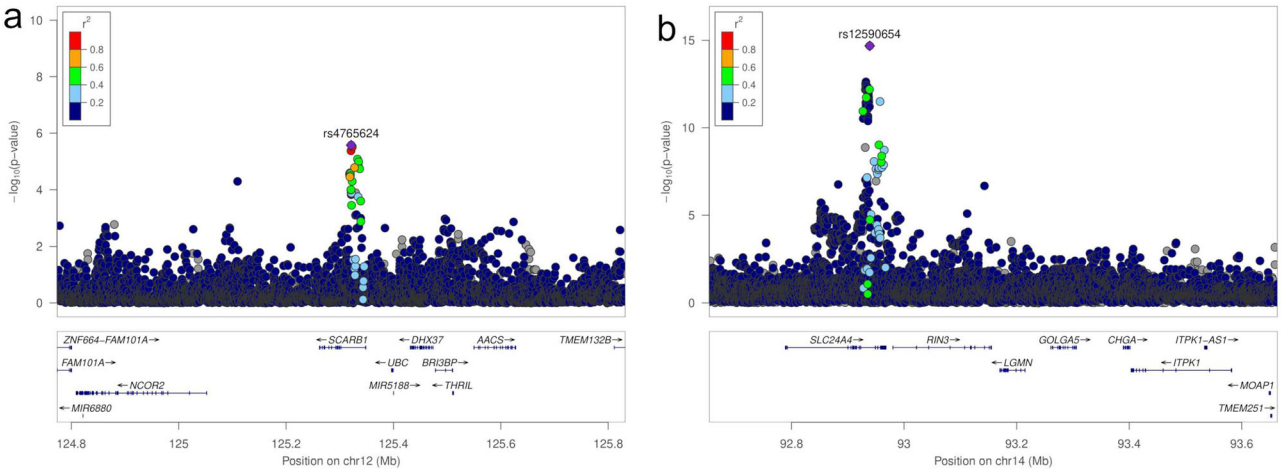

Fig. 5 | AD-associated 5hmC regions are located within known AD GWAS loci. **a** An AD-associated 5hmC region on chromosome 12 (C12P044736, 124,788,270-124,847,332 bp) contains SNPs (± 500 kb) known to be associated with AD in previous GWAS (peak SNP: rs4765624; $P = 2.66 \times 10^{-6}$). **b** An AD-associated 5hmC region on chromosome 14 (C14P057817: 92687207-92697256 bp) harbors SNPs known to be associated with AD in previous GWAS (peak SNP: rs12590654; $P = 2.08 \times 10^{-15}$). X-axis: genomic positions; Y-axis: $-\log_{10}$(*P*-value). The GWAS nominal P-values were obtained by two-sided z-score test. Source data are provided as a Source Data file.

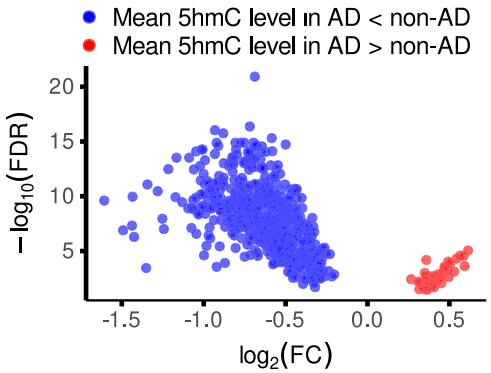

Fig. 6 | Loss of DNA hydroxymethylation (i.e., hypo-hydroxymethylation) associated with AD neuropathology. Fold change (FC) was defined as the ratio of mean 5hmC level at each DhMR in AD divided by the mean 5hmC level at the same DhMR in non-AD. $\log_2(FC) < 0$ indicates a lower 5hmC level in AD than non-AD (i.e., loss of DNA hydroxymethylation), while $\log_2(FC) > 0$ indicates a higher 5hmC level in AD than non-AD (i.e., gain of DNA hydroxymethylation). In our study, majority (93.6%) of the identified DhMRs exhibited lower 5hmC levels in AD compared to non-AD, indicating that loss of DNA hydroxymethylation is associated with AD pathology. X-axis: $\log_2$(FC); Y-axis: $-\log_{10}$(two-sided FDR) where multiple testing was performed using the Benjamini-Hochberg false discovery rate (FDR) method. Source data are provided as a Source Data file.

associated 5hmC with brain gene expression data revealed that DNA hydroxymethylation was predominantly positively correlated with *cis*-acting gene expression, though negative correlations were also observed. Integration of 5hmC with brain histone acetylation (H3K9ac) data demonstrated that the identified DhMRs were largely negatively correlated with H3K9ac expression, and more likely to co-position with active transcription start site (TSS), transcription states, and enhancer states, suggesting that altered 5hmC may contribute to AD through altering chromatin states and gene transcription. Integration of 5hmC with brain proteomic data showed that the identified DhMRs were correlated with the abundance of some proteins that are enriched in Aβ formation, signal transduction, neurogenesis, and neuronal ion channel activity. Enrichment analysis indicated that the identified DhMRs were significantly enriched in functional regulatory elements such as DHSs and transcribed chromatin states. In addition, some AD-associated 5hmC regions colocalize with known AD loci (e.g., *SLC24A4/RIN3*) and expression quantitative trait loci (eQTLs). Protein-protein interaction network analysis revealed that genes showing both differential hydroxymethylation and expression were significantly enriched in pathways involved in neurological pathogenesis, such as mTOR signaling[34], γ-secretase complex[35], mediator complex, and transcriptional regulation[36], SWI/SNF complex[37], and DNA replication and repair[38]. Together, these results demonstrate extensive alterations of DNA hydroxymethylome associated with neuropathology and highlight the critical role of active DNA demethylation in the pathogenesis of Alzheimer's dementia.

Notably, of the 2821 DhMRs associated with AD neuropathologies, the genes annotated to 886 5hmC regions (total 818 unique genes) also showed significant differential expression in relation to AD, suggesting that altered 5hmC may affect AD through regulating gene transcription. However, the quantitative relationship between 5hmC and gene expression varies across tissues and cell types[39]. In addition, previous evidence suggests that DNA methylation/hydroxymethylation may also affect disease through other pathways, such as regulating immune responses and oxidative stress[40–42], altering chromatin remodeling and genomic integrity[43–45], influencing microRNA expression[46–49], affecting neuronal myelination/remyelination or ion channel activity[11,50], etc. Of the 886 genes showing both differential hydroxymethylation and expression, many are known to be involved in neuropathy (e.g., *CACNA1A, ATXN1, RFX1, HIPK2, HSF1, ITGA2B, IGF1R, PLCG2*), cognitive performance (e.g., *CAMTA1, CC2D1A, CHST3, DPP6, GAL, HCN2, KCNH2, KMT2C*), synaptic plasticity (e.g., *PRKCZ, CC2D1A, CTBP2, GNA11, IGF1R, EIF3B, DOCK1, ADORA3*), immune function (e.g., *AGAP3, AP3D1, CARS2, CEP192, COTL1, TFCP2L1*), mitochondrial function (*e.g., CARS2, LETM1, MIGA2, TOP1MT*), and oxidative stress (e.g., *ATP11A, CPT1A, GSTP1, TIMM44*). In addition, many annotated genes were previously associated with AD (e.g., *ANK1, FOXK1, NKX6-3, TNIP1, SLC24A4/RIN3, PLCG2, ACE, BCAM*) or neuropsychiatric disorders (e.g., *CRHBP, PTPN21, AGAP1, CAMTA1, CNGB1, CPT1A, FGFR2, PADI2*). Many identified genes are drug targets for AD (e.g., *CACNA1A, DLGAP1, DOCK1, DPP6, GAD1, IGF1R, PLCG2, ACE*) based on the Agora database[51]. Moreover, consistent with our previous findings[11], we identified alterations of 5hmC in genes involved in ion channels (e.g., *CACNA1A, KCNJ11, KCNT1, ATP11A, LETM1, SLC8B1, TMEM175, USP36*), a mechanism known to be implicated in AD pathogenesis[52]. Additionally, many identified DhMRs harbor known AD GWAS genes or loci (e.g., *ANKH, BCAM, ITGA2B, INPP5D, PLCG2, RIN3, SHARPIN, USP6NL*). Furthermore, our analyses led to the

identification of genes and pathways associated with AD. For example, multiple genes encoding zinc finger proteins (e.g., *ZDHHC1, ZHX2, ZBED4, ZC3H12C, ZC3H18, ZC3H4, ZCCHC14, ZDHHC14*) exhibited both differential hydroxymethylation and expression in our sample. Other genes showing differential hydroxymethylation are involved in mTOR signaling (e.g., *NPRL3*, *RPTOR, ULK1*), *Wnt* signaling (e.g., *LRP5, LETM1, RNF213*), and cardiometabolic processes (e.g., *CPT1A, KCNJ11, TCF7L2*), etc. Together, our results unraveled a critical role of altered DNA hydroxymethylation in AD neuropathology. If validated, these candidate genes may serve as potential therapeutic targets for AD treatment.

Several findings are worthy of discussion. First, the majority of the identified DhMRs showed lower DNA hydroxymethylation in participants with AD diagnosis, indicating loss of DNA hydroxymethylation (i.e., hypo-hydroxymethylation) in human brains with AD compared to no AD. This is consistent with previous studies[11,13,17,20] showing loss of active DNA demethylation in AD. Second, genetic polymorphisms are important modifiers of DNA methylation, but the observed associations of 5hmC with AD neuropathology appeared not to be driven by genetic polymorphisms, because additional controlling for genotypes of nearby genes including *APOE* did not change our results. Third, in line with a recent study showing that 5hmC preferentially decorates gene bodies and outperforms gene body 5mC in reflecting gene expression[53], we found that 5hmC regions are mainly located in gene bodies (~ 70% among all identified regions, ~85% among AD-associated regions), where their abundances are largely positively correlated with *cis*-acting gene expression. Fourth, consistent with a previous study showing that 5mC may affect amyloid load and tau tangles independently[54], we found that 5hmC alterations may also have an independent effect on these two molecularly-specific hallmarks of AD pathology as additionally adjusting for each other in the statistical models did not affect the observed associations. Fifth, the effect of an individual 5hmC region on AD pathology was in general small to moderate (2.74% for AD diagnosis, 1.95% for Aβ load, 0.60% for PHFtau tangles). A total of 118 DhMRs (associated with all three pathological phenotypes in meta-analysis) collectively explain about 34.9% and 17.6% of the variance in Aβ load and PHFtau tangles, respectively. In comparison, a total of 71 CpG sites reported in a previous EWAS in ROSMAP accounted for 28.7% of the variation in neuritic plaques[27].

Strengths of our study include the use of hMe-Seal technology applied to a large collection of human brain tissue samples (> 1000) from two community-based cohorts of aging and dementia, the rich clinical data and detailed postmortem pathological evaluations, the use of multiple pathological phenotypes, including a binary outcome variable for pathological diagnosis of AD and two quantitative measures of molecular specificity for AD pathology (Aβ load and PHFtau tangles), and the multi-omics data in the same brain cortex from the same individuals. To the best of our knowledge, the current study represents by far the largest sequencing-based genome-wide profiling of brain 5hmC in relation to AD pathology in the field.

Several limitations should also be noted. First, our 5hmC and other omics data were generated in bulk brain tissue which contains a mixture of different cell types. In addition, we only investigated genome-wide patterns of DNA hydroxymethylome in the prefrontal cortex, but brain hydroxymethylation is likely to be cell-type and region-specific. Future research should use cell-type specific or ideally, single-cell technologies and include other brain regions. Second, the resolution of the hMe-Seal method used for 5hmC quantification is relatively low (~ 100 bp). Single-base resolution technique, such as Tet-assisted bisulfite sequencing (TAB-seq)[55], is required to elucidate the relationship between 5hmC and 5mC at a genome-scale. Third, our study sample comprises primarily highly educated older individuals of European ancestry, mostly older women. Additional work will need to be done in diverse populations with different demographic characteristics. Finally, it is unclear whether the differences in DNA hydroxymethylation seen in AD brains are the cause or consequence of neuropathological changes.

Methods

Participants and cohorts

The current study included deceased participants from two longitudinal studies of aging and dementia: Religious Orders Study (ROS) and Rush Memory and Aging Project (MAP). Detailed study design and methods have been described previously[21,22,56]. Briefly, ROS recruited elderly Catholic priests, nuns, and brothers free of known dementia at time of enrollment (since 1994) across the USA. MAP recruited elderly men and women who were free of known dementia at the time of enrollment (since 1997) in the Chicago metropolitan area. In both cohorts, participants underwent detailed annual clinical evaluations and agreed to donate their brains at the time of death. The follow-up rate for survivors exceeds 90% in both cohorts, while the autopsy rate exceeds 90% in ROS, and 80% in MAP. Majority of the ROSMAP participants (96%) were self-reported non-Latinx Whites. The two studies share a large common core of identical clinical and pathological data at the item level collected by the same people with the same trainers. Thus, they can be studied either individually or together. The ROSMAP studies were approved by an Institutional Review Board of Rush University Medical Center. Each participant signed an informed consent, Anatomic Gift Act, and repository consent allowing their data and biospecimens to be repurposed. The current study was approved by the Institutional Review Board of University of Florida.

Phenotypes

Our primary phenotypes of interest included a pathological diagnosis of AD (yes/no) and two quantitative measures of AD neuropathologies including amyloid-β (Aβ) load and paired helical filament (PHF) tau tangle density. Brain autopsies were performed as previously described[21,22,56]. Neuropathological examinations were performed by data collectors blinded to the clinical data. Briefly, Aβ and PHFtau tangles were quantified via immunohistochemistry across eight brain regions: the CA1/subiculum of the hippocampus, angular gyrus, and entorhinal, superior frontal, mid prefrontal, inferior temporal, anterior cingulate and calcarine cortices. Mean percentage of the area positive for Aβ was averaged across brain regions. PHFtau tangle density (per mm^2) was determined using systematic sampling; PHFtau tangle score was then averaged across brain regions. The neuropathological diagnosis of AD was made following a modified NIA-Reagan Institute criterion. Details of the pathology collection have been described elsewhere[57,58]. All clinical data were reviewed by a neurologist blinded to postmortem data at the time of death. Presence of either mild cognitive impairment (MCI) or Alzheimer's dementia was evaluated at every visit as previously reported[59–61]. Participants with no MCI or Alzheimer's dementia were rendered as having no cognitive impairment (NCI). Diagnostic procedures have been described previously[59–61].

Brain genomic DNA isolation

Genomic DNA from ~100 mg sections of frozen dorsolateral prefrontal cortex (DLPFC) was isolated by homogenizing the samples in 600 µl digestion buffer (100 mM Tris-HCl, pH 8.5, 5 mM EDTA, 0.2% SDS, 200 mM NaCl), Proteinase K (Thermo Fisher, cat# EO0491) treatment at 55 °C for overnight. Following digestion, 600 µl of Phenol:-Chloroform:Isoamyl Alcohol (25:24:1 Saturated with 10 mM Tris, pH 8.0, 1 mM EDTA) (Sigma-Aldrich, cat# P-3803) was added and mixed thoroughly prior to centrifugation at 16,000 × *g* for 10 min. PicoGreen was used to quantitate the DNA.

5hmC profiling by 5hmC-selective chemical labeling method (hMe-Seal)

5hmC enrichment was performed according to the previously described 5hmC-specific chemical labeling and affinity purification

protocol[4]. Briefly, DNA libraries were prepared using 2 µg of captured DNA or input DNA and performed following the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® manufacturers protocol. Prior to sequencing, the quality of prepared libraries was checked using Bioanalyzer and TapeStation. Sequencing was performed on the HiseqX 2 x150 platform.

5hmC data pre-processing, quality control, and normalization

Raw sequencing reads were trimmed by Trimmomatic v0.39 to remove adapter sequences and poor-quality reads. High-quality reads were mapped to the reference human genome GRCh38 by Bowtie2-2.3.5.1 with uniquely mapped reads retained for the following analyses. PCR duplicates were removed by Picard v2.21.2. We used MACS2 (v2.2.7.1) to identify DhMRs with default parameters, except for enabling the --broad option to call broad peaks at a threshold of q-value < 0.01. Additionally, twelve input samples were combined and used as control samples during peak calling for all samples. After peak calling, all peaks with at least one base pair (bp) overlapping across all samples were merged using DiffBind v3.0 (https://bioconductor.org/packages/devel/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf). The read counts were then re-calculated for each merged peak. Consensus regions were normalized by trimmed mean of M-values (*TMM*) implemented in the *edgeR* Bioconductor package[62]. There was no constraint on the size of the 5hmC peak regions. Moreover, since there has been no established standard for brain 5hmC peak calling, we tried different thresholds of samples (e.g., 50, 100, 150, 200) to identify an optimal cutoff to select high-confident 5hmC regions. We found that the results were largely consistent under different scenarios. We chose a threshold of 200 samples—the highest threshold used by far based on our knowledge—to select high-confident 5hmC regions (i.e., regions detected in ≥ 200 samples). This threshold accounts for approximately 19% of the samples analyzed (200 out of 1050). In comparison, a previous study[24] in ROSMAP used 15% of the samples analyzed (100 out of 669 samples) as a cutoff to filter peaks of brain histone acetylation. A total of 197,765 high-confident 5hmC regions (size ranging from 392 bp to 248,830 bp) from 1050 unique individuals were included in the statistical analysis.

Controlling for batch-specific variance in the 5hmC data

We used the program RUVr (remove unwanted variation using residuals)[63] implemented in the R package *RUVSeq* v1.36.0 (https://bioconductor.org/packages/release/bioc/html/RUVSeq.html) to detect and remove unwanted variations, e.g., technical batches, library preparation, cellular heterogeneity, and other nuisance effects, without requiring direct knowledge of the sources of unwanted variations. RUVr uses the residuals from a first-pass generalized linear model (GLM) of the read counts[63] and considers the least differentially expressed genes across samples. The default parameters were used in this analysis. We identified nine unwanted variations for each phenotype of interest (i.e., AD diagnosis, Aβ load, and PHFtau tangles). These unwanted variations were regressed out from the 5hmC data, and the residuals were used in the statistical analysis. Principal component analysis (PCA) was used to detect outliers and evaluated the performance of batch removal.

To avoid potential over-correction, we examined the associations of the unwanted variations (derived by RUVr) with technical batch variables (e.g., sequencing lab, library size, sequencing depth, etc.), cell type proportions, and clinical covariates (e.g., age at death, sex, education, and PMI), respectively. It shows that these unwanted variations were significantly associated with technical batches (e.g., sequencing lab, library size, sequencing depth) and heterogeneity of cell type proportions, but not clinical variables (age at death, sex, PMI, and education).

5hmC data annotation

To facilitate the comparison of 5hmC with other omics data, we annotated the 5hmC regions to their nearby genes (±500 kb) by ChIPpeakAnno (https://bioconductor.org/packages/release/bioc/html/ChIPpeakAnno.html), UROPA[64], and Annotatr (https://bioconductor.org/packages/release/bioc/html/annotatr.html). Chromatin states were obtained from a DLPFC sample with minimal neuropathology (E073) included in the NIH Roadmap Epigenomics Project.

Brain RNA-seq data

Bulk brain (DLPFC) RNA-seq data were downloaded from Synapse (https://www.synapse.org/#!Synapse:syn3388564). Detailed methods for RNA-seq data generation and pre-processing have been described previously[23]. Briefly, 50 mg frozen brain tissue was dissected and homogenized in DNA/RNA shield buffer (Zymo, R1100) with 3 mm beads using a bead homogenizer. RNA was subsequently extracted using Chemagic RNA tissue kit (Perkin Elmer, CMG-1212) on a Chemagic 360 instrument. RNA was concentrated (Zymo, R1080), and RQN values calculated with a Fragment Analyzer total RNA assay (Agilent, DNF-471). RNA concentration was determined using Qubit broad-range RNA assay (Invitrogen, Q10211) according to the manufacturer's instructions. 500 ng total RNA was used as input for sequencing library generation and rRNA was depleted with RiboGold (Illumina, 20020599). A Zephyr G3 NGS workstation (Perkin Elmer) was utilized to generate TruSeq stranded sequencing libraries (Illumina, 20020599) with custom unique dual indexes (IDT) according to the manufacturer's instructions with the following modifications. RNA was fragmented for 4 min at 85 °C. First-strand synthesis was extended to 50 min. Size selection post adapter ligation was modified to select for larger fragments. Library size and concentrations were determined using an NGS fragment assay (Agilent, DNF-473) and Qubit ds DNA assay (Invitrogen, Q10211) respectively, according to the manufacturer's instructions. The modified protocol yielded libraries with an average insert size of around 330-370 bp. Libraries were normalized for molarity and sequenced on a NovaSeq 6000 (Illumina) at 40–50 M reads, 2 × 150 bp paired end. The RNA-Seq data were then processed by the parallelized and automatic pipeline. These pipelines include trimming the beginning and ending bases from each read, identifying, and trimming adapter sequences from reads, detecting, and removing rRNA reads, aligning reads to reference genome. We used the non-gapped aligner Bowtie to align reads to transcriptome reference and then applied RSEM v1.3.3 to estimate expression levels for all transcripts. The FPKM values were the outcome of our data RNA-Seq pipeline. We used the *edgeR* Bioconductor package to normalize the data, and the $\log_2$-transformed data were used in downstream analysis. Potential batch effect was detected and removed by RUVr. Outliers were detected by PCA. A total of 819 samples have complete data for both 5hmC and RNA-seq data (Supplementary Data 22).

Brain H3K9ac (ChIP-seq) data

To examine the relationship between AD-associated 5hmC and histone modification, we downloaded the histone 3 lysine 9 acetylation (H3K9ac) data from synapse (https://www.synapse.org/Synapse:syn4896408). Detailed methods for H3K9ac data generation, processing, and quality control have been described previously[24]. Briefly, 50 mg of gray matter was dissected on ice from biopsies of the dorsolateral prefrontal cortex of ROSMAP participants. The tissue was minced and crosslinked with 1% formaldehyde at room temperature for 15 min and quenched with 0.125 M Glycine. The tissue was then homogenized in cell lysis buffer using the Tissue Lyser and a 5 mm stainless steel bead. Then the nuclei were lysed in nuclei lysis buffer and chromatin was sheared by sonication. Chromatin was incubated overnight at 4 C with the H3K9ac antibody (Millipore anti-H3K9ac Ab, catalog # 06-942, lot: 31636) and purified with protein A sepharose beads. The final DNA was extracted and used for Illumina library construction following usual methods of end repair, adapter ligation, and gel size selection. Samples were pooled and sequenced with 36 bp single-end reads on the Illumina HiSeq. For data processing, short

single-end reads were aligned against the human reference GRCh37/hg19 using the Burrows-wheeler Aligner (BWA 0.7.4). Peaks were detected for each sample individually by MACS2 using the broad peak option, a stringent q-value cutoff of 0.001, and pooled genomic DNA of seven samples as a negative control library. Subsequently, H3K9ac domains were defined as genomic regions that were detected as a peak in at least 100 (15%) of our 669 samples. Regions neighboring within 100 bp were merged and very small regions of less than 100 bp were removed, resulting in 26,384 H3K9ac domains. H3K9ac levels were quantified for each sample and domain by counting the number of reads in the domains after extending reads towards the 3'-end by the estimated DNA fragment length. The *edgeR* Bioconductor package was further applied to perform TMM and FPKM normalization. Potential batches and outliers were detected by PCA. No batch effect or outliers were identified. In the current analysis, we converted the genomic coordinates to GRCh38/hg38 and matched each of the identified DhMRs to their nearby H3K9ac (±50 kb). A total of 513 ROSMAP participants with both 5hmC and H3K9ac data were included in this analysis (Supplementary Data 22).

Brain proteomic data

To examine the relationship between AD-associated 5hmC and protein expression in the same brain cortex, we downloaded the brain proteomic data from synapse (https://www.synapse.org/Synapse:syn17015098). A detailed description of the brain proteomic data ($n = 980$) including batch correction and data normalization ($\log_2$-transformed) has been reported elsewhere[25,65,66]. Briefly, brain tissue samples (DLPFC) were homogenized, and protein concentration was determined. After protein digestion, isobaric tandem mass tag (TMT) peptide labeling and high pH fractionation were performed. Fractions were then analyzed by liquid chromatography-mass spectrometry. The resulting mass spectrometry spectra were searched against the UniProt human protein database, with individual protein abundance checked against the global internal standard. Technical confounders were regressed out from the proteomic data. A total of 8356 proteins passed the final quality control. We further performed PCA and removed two outlier samples. Using a gene-centric approach, we matched the identified DhMRs to their nearby proteins (±50 kb) and then calculated the correlation between the levels of AD-associated 5hmC and protein expression. A total of 731 unique individuals with both 5hmC and proteomic data were included in this analysis (Supplementary Data 21).

Accounting for cell type proportions in the 5hmC data

Our 5hmC data were generated using bulk brain tissue which consists of many different cell types. To examine whether cell type proportions affect our results, we downloaded the bulk RNA-seq data ($n = 834$) from Synapse (https://www.synapse.org/#!Synapse:syn3388564) and the single-nucleus RNA-seq (snRNA) data ($n = 48$) from synapse (https://www.synapse.org/#!Synapse:syn18485175). Both datasets were generated in the prefrontal cortex of ROSMAP participants, and detailed methods for the brain RNA-seq data[67] and snRNA-seq data[68] have been reported elsewhere. We used Bisque to estimate cell type proportions in the bulk brain tissue, as previously described[69]. Bisque implements a regression-based approach that uses snRNA-seq as a reference and has been shown to be highly efficient for estimating cell type proportions in bulk RNA-seq data[69]. The program identified eight major cell types, including excitatory neurons, inhibitory neurons, astrocytes, microglia, oligodendrocytes, oligodendrocyte progenitor cells (OPC), endothelial cells, and pericytes. Based on the correlation among these cell types, we further grouped them into four cellular groups, including neurons (inhibitory and excitatory neurons), glia (microglia, oligodendrocytes, and OPC), vascular cells (pericytes and endothelial cells) and astrocytes. The proportions of these four cellular groups were included as covariates in the statistical analysis.

Statistical modeling and data analysis

Identification of differentially hydroxymethylated regions (DhMRs) associated with AD neuropathology. Given the high dimension of the 5hmC data and the strong intercorrelations between 5hmC regions, we used an elastic net model implemented in the R package *glmnet* to identify DhMRs associated with each of the three neuropathological phenotypes, including neuropathological diagnosis of AD, Aβ load, and PHFtau tangles. Briefly, we used ten-fold cross-validation to select 5hmC regions associated with each phenotype. As a hybrid of ridge and LASSO regression, elastic net model is controlled by two parameters: α and λ[70]. We used elastic net model with $\alpha = 0.05$, a common choice for methylation data, based on the performance of the model after testing different values in the range between 0 and 1. Another parameter λ was chosen by cross-validation. The elastic net model included all 197,765 5hmC regions (residuals after regressing out unwanted variations derived from RUVr) and clinical factors, including age at death, sex, education, and postmortem interval (PMI). We repeated this process 100 times with different data partitions and obtained the frequency of selection for each 5hmC over the 100 iterations. Regions selected ≥ 60 times over 100 runs were considered top candidates associated with AD pathology.

For replication, we performed the elastic net regression in MAP cohort (discovery, $n = 600$), and then confirmed the results in ROS cohort (replication, $n = 450$). Fixed-effect meta-analysis was performed to combine the results from both cohorts. A Benjamin-Hochberg (BH) adjusted false discovery rate (FDR) < 0.05 was used to determine statistical significance in the replication stage and meta-analysis. In addition, since elastic net shrinks the coefficients towards zero, we estimated the effect of each selected 5hmC by logistic or linear regression, adjusting for age at death, sex, education, and PMI.

We calculated the variance in Aβ or PHFtau that is explained by each of the identified DhMRs using the R-squared (i.e., coefficient of determination in a regression model) method. Briefly, R-squared is a statistical measure that determines the proportion of variance in the dependent variable that can be explained by the independent variable. In our case, the dependent variable was Aβ load or PHFtau tangles, and the independent variable was each of the identified DhMRs (identified and replicated in both cohorts). We calculated the difference in R-squared by building an initial model consisting of clinical variables only (age at death, sex, education, PMI) and then a second model that additionally included each of the identified DhMRs. The proportion of variance explained by each DhMR was obtained by subtracting the R-squared values of the initial models from those of the second models. Results for the variance explained in Aβ load and PHFtau tangles are shown in Supplementary Data 15.

Integrative multi-omics in the same brain cortex. To examine the functional consequence of 5hmC alterations on other molecular phenotypes (e.g., gene expression, protein expression, histone acetylation), we annotated the identified DhMRs to their nearby genes (±50 kb) and then calculated the correlation of each DhMR with its cognate gene expression, protein expression, or histone acetylation (H3K9ac). We also performed linear or logistic regression to examine whether the differentially hydroxymethylated genes are also differentially expressed in relation to AD pathology, adjusting for age at death, sex, education, and PMI. FDR < 0.05 was used to determine statistical significance for all analyses.

Functional enrichment analysis. To assess whether the identified DhMRs were randomly distributed among functional or regulatory elements in the genome, we performed functional enrichment analysis using GARFIELD (v2)[29]. Briefly, this method leverages GWAS summary statistics at various p-value thresholds and calculates the fold enrichment of various regulatory elements based on an annotation panel that includes 1005 annotations (e.g., genetic elements, chromatin states,

histone modifications, DNase I hypersensitive sites, and transcription factor binding sites in different cell lines) from ENCODE, GENCODE, and the Roadmap Epigenomics Project[71–73]. Statistical significance of each regulatory element at one or more GWAS *p*-value thresholds was tested by permutation while accounting for minor allele frequency (MAF), the number of LD proxies ($r^2 > 0.8$), and the distance to the nearest transcription start site (TSS)[29]. In this analysis, we used summary statistics from three large-scale GWAS meta-analysis for AD[30–32]. For each DhMR, we included genetic variants located within 500 kb of the DhMR and calculated the odds ratios (ORs) of enrichment at various GWAS significance thresholds. *P*-values of the enrichment analysis were adjusted for multiple testing based on the effective number of annotations.

Protein-protein interaction (PPI) networks of differentially hydroxymethylated genes. To examine the direct and functional connections between the differentially hydroxymethylated and differentially expressed genes, we constructed protein-protein interaction (PPI) networks using the Search Tool for the Retrieval of Interacting Genes/proteins (STRING) v12.0[74]. Protein interactions with a probabilistic confidence score ≥ 0.7 (high confidence) were considered significant. Enrichment analysis was also conducted using STRING and regulatory elements with FDR < 0.05 were deemed to be significant. A total of 740 unique genes showing both differentially hydroxymethylated and differentially expressed (FDR < 0.05) in relation to AD were included in this analysis. All genes in the database were used as a background set in the enrichment analysis.

Colocalization analysis. To investigate the potential functional impact of the identified DhMRs on AD pathology, we performed colocalization analysis using the R package *coloc*[75]. The program estimates the posterior probability of each DhMR containing a single variant affecting both 5hmC and AD phenotype or gene expression. We used the default prior probabilities that a variant is equally associated with each phenotype ($p_1 = 1 \times 10^{-4}$; $p_2 = 1 \times 10^{-4}$) and both phenotypes jointly ($p_{12} = 1 \times 10^{-5}$). A posterior probability ≥ 80% was used to indicate evidence of colocalization between DhMR and variants influencing AD or cell-type specific gene expression.

In this analysis, we used summary statistics from three previously published AD GWAS meta-analysis[30–32]. Data from Jansen et al.[32] were downloaded at https://ctg.cncr.nl/software/summary_statistics (71,880 AD cases, 383,378 controls). Data from Wightman et al.[31] were obtained directly from the authors (39,918 AD cases, 358,140 controls). Data from Bellenguez et al.[30] were downloaded from the European Bioinformatics Institute GWAS Catalog at https://www.ebi.ac.uk/gwas/ under accession number GCST90027158 (106,398 AD, 423,821 controls). Similar analysis was performed to examine whether the identified DhMRs colocalize with gene expression quantitative trait loci (eQTLs) for psychiatric and neurological disorders. The *cis*-eQTL data were based on gene expression data in eight human brain cell types derived from prefrontal cortex, temporal cortex, and white matter[33]. The eQTL summary statistics were downloaded at https://zenodo.org/records/7276971. The colocalization analysis tested genetic variants located within 500 kb for each of the 2821 DhMRs associated with AD pathologies.

Sensitivity analysis. To examine whether the observed associations of 5hmC regions on chromosome 19 with AD pathology are driven by *APOE*, we additionally adjusted for *APOE* genotypes in the model. To investigate whether cell type proportions confound our results, we further adjusted for cell type proportions derived from snRNA-seq data (DLPFC) in 48 ROSMAP participants. To assess whether tau tangles mediate the association of 5hmC alteration with amyloid load, we repeated the analysis of Aβ by further controlling for PHFtau. Similar analysis was performed to examine the potential mediation effect of amyloid load on the association of 5hmC with tau tangles. An attenuated association would be evidence of mediation. This analysis focused on the 490 5hmC regions associated with both Aβ and PHFtau.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Brain 5hmC data have been deposited into the Gene Expression Omnibus (GEO) under the accession number GSE291402. The following data used in this study are publicly available: ROSMAP genotype data are available on Synapse under accession code [syn11724057]; ROSMAP RNA-seq data are available on Synapse under accession code [syn3388564]; ROSMAP proteomics data are available on Synapse under accession code [syn17015098]; ROSMAP histone modification data are available on Synapse under accession code [syn4896408]; summary statistics from Jansen et al. are available on https://ctg.cncr.nl/software/summary_statistics; summary statistics from Bellenguez et al. were accessed through the National Human Genome Research Institute-European Bioinformatics Institute GWAS catalog under accession number GCST90027158; summary statistics from Wightman et al.[31] were obtained directly from the authors; reference chromatin map data are available on https://egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html#core_15state. Raw data for figures are available in the Source Data file. Source data are available with this paper. Source data are provided with this paper.

Code availability

The code used to perform the analyses described in this study is available on GitHub at https://github.com/Zhao-research/5hmC and Zenodo[76].

References

1. Nazarian, A. & Kulminski, A. M. Evaluation of the genetic variance of Alzheimer's disease explained by the disease-associated chromosomal regions. *J. Alzheimer's Dis.* **70**, 907–915 (2019).
2. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
3. Szulwach, K. E. et al. 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.* **14**, 1607–1616 (2011).
4. Song, C.-X. et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* **29**, 68–72 (2011).
5. Spiers, H., Hannon, E., Schalkwyk, L. C., Bray, N. J. & Mill, J. 5-hydroxymethylcytosine is highly dynamic across human fetal brain development. *BMC Genomics* **18**, 738 (2017).
6. Papale, L. A., Madrid, A., Li, S. & Alisch, R. S. Early-life stress links 5-hydroxymethylcytosine to anxiety-related behaviors. *Epigenetics* **12**, 264–276 (2017).
7. Nestor, C. E. et al. Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Res.* **22**, 467–477 (2012).
8. Pastor, W. A. et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**, 394–397 (2011).
9. Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S. & Jacobsen, S. E. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol.* **12**, R54 (2011).
10. Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151**, 1417–1430 (2012).

11. Zhao, J. et al. A genome-wide profiling of brain DNA hydroxymethylation in Alzheimer's disease. *Alzheimer's Dement. J. Alzheimer's Assoc.* **13**, 674–688 (2017).
12. Bernstein, A. I. et al. 5-Hydroxymethylation-associated epigenetic modifiers of Alzheimer's disease modulate Tau-induced neurotoxicity. *Hum. Mol. Genet.* **25**, 2437–2450 (2016).
13. Smith, A. R. et al. Parallel profiling of DNA methylation and hydroxymethylation highlights neuropathology-associated epigenetic variation in Alzheimer's disease. *Clin. Epigenet.* **11**, 52 (2019).
14. Ellison, E. M., Bradley-Whitman, M. A. & Lovell, M. A. Single-base resolution mapping of 5-hydroxymethylcytosine modifications in hippocampus of Alzheimer's disease subjects. *J. Mol. Neurosci.* **63**, 185–197 (2017).
15. Phipps, A. J., Vickers, J. C., Taberlay, P. C. & Woodhouse, A. Neurofilament-labeled pyramidal neurons and astrocytes are deficient in DNA methylation marks in Alzheimer's disease. *Neurobiol. Aging* **45**, 30–42 (2016).
16. Lashley, T. et al. Alterations in global DNA methylation and hydroxymethylation are not detected in Alzheimer's disease. *Neuropathol. Appl. Neurobiol.* **41**, 497–506 (2015).
17. Condliffe, D. et al. Cross-region reduction in 5-hydroxymethylcytosine in Alzheimer's disease brain. *Neurobiol. aging* **35**, 1850–1854 (2014).
18. Coppieters, N. et al. Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. *Neurobiol. aging* **35**, 1334–1344 (2014).
19. Bradley-Whitman, M. & Lovell, M. Epigenetic changes in the progression of Alzheimer's disease. *Mechan. Ageing Dev.* **134**, 486–495 (2013).
20. Chouliaras, L. et al. Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiol. Aging* **34**, 2091–2099 (2013).
21. Bennett, D. A., Schneider, J. A., Arvanitakis, Z. & Wilson, R. S. Overview and findings from the religious orders study. *Curr. Alzheimer Res.* **9**, 628–645 (2012).
22. Bennett, D. A. et al. Overview and findings from the rush Memory and Aging Project. *Curr. Alzheimer Res.* **9**, 646–663 (2012).
23. De Jager, P. L. et al. A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci. Data* **5**, 180142 (2018).
24. Klein, H.-U. et al. Epigenome-wide study uncovers large-scale changes in histone acetylation driven by tau pathology in aging and Alzheimer's human brains. *Nat. Neurosci.* **22**, 37–46 (2019).
25. Wingo, A. P. et al. Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer's disease pathogenesis. *Nat. Genet.* **53**, 143–146 (2021).
26. Tay, J. K., Narasimhan, B. & Hastie, T. Elastic net regularization paths for all generalized linear models. *J. Stat. Softw.* **106**, 1–31 (2023).
27. De Jager, P. L. et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat. Neurosci.* **17**, 1156–1163 (2014).
28. Wingo, A. P. et al. Shared proteomic effects of cerebral atherosclerosis and Alzheimer's disease on the human brain. *Nat. Neurosci.* **23**, 696–700 (2020).
29. Iotchkova, V. et al. GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat. Genet.* **51**, 343–353 (2019).
30. Bellenguez, C. et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436 (2022).
31. Wightman, D. P. et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat. Genet.* **53**, 1276–1282 (2021).
32. Jansen, I. E. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).
33. Bryois, J. et al. Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. *Nat. Neurosci.* **25**, 1104–1112 (2022).
34. Querfurth, H. & Lee, H.-K. Mammalian/mechanistic target of rapamycin (mTOR) complexes in neurodegeneration. *Mol. Neurodegener.* **16**, 44 (2021).
35. Hur, J. Y. γ-Secretase in Alzheimer's disease. *Exp. Mol. Med.* **54**, 433–446 (2022).
36. Schiano, C., Luongo, L., Maione, S. & Napoli, C. Mediator complex in neurological disease. *Life Sci.* **329**, 121986 (2023).
37. Tibshirani, M. et al. Dysregulation of chromatin remodelling complexes in amyotrophic lateral sclerosis. *Hum. Mol. Genet.* **26**, 4142–4152 (2017).
38. Dileep, V. et al. Neuronal DNA double-strand breaks lead to genome structural variations and 3D genome disruption in neurodegeneration. *Cell* **186**, 4404–4421.e4420 (2023).
39. Feng, B. et al. An epigenetic manifestation of Alzheimer's disease: DNA methylation. *Actas Esp. Psiquiatr* **52**, 365–374 (2024).
40. Millan, M. J. The epigenetic dimension of Alzheimer's disease: causal, consequence, or curiosity? *Dialogues Clin. Neurosci.* **16**, 373–393 (2014).
41. Johnson, N. D. & Conneely, K. N. The role of DNA methylation and hydroxymethylation in immunosenescence. *Ageing Res. Rev.* **51**, 11–23 (2019).
42. Wei, H. et al. Role of oxidative stress and DNA hydroxymethylation in the neurotoxicity of fine particulate matter. *Toxicology* **380**, 94–103 (2017).
43. Zhang, H. et al. Distinct roles of nucleosome sliding and histone modifications in controlling the fidelity of transcription initiation. *RNA Biol.* **18**, 1642–1652 (2021).
44. Rose, N. R. & Klose, R. J. Understanding the relationship between DNA methylation and histone lysine methylation. *Biochim. Biophys. Acta* **1839**, 1362–1372 (2014).
45. Sriraman, A., Debnath, T. K., Xhemalce, B. & Miller, K. M. Making it or breaking it: DNA methylation and genome integrity. *Essays Biochem.* **64**, 687–703 (2020).
46. Han, L., Witmer, P. D., Casey, E., Valle, D. & Sukumar, S. DNA methylation regulates MicroRNA expression. *Cancer Biol. Ther.* **6**, 1284–1288 (2007).
47. Villela, D. et al. Differential DNA methylation of MicroRNA genes in temporal cortex from Alzheimer's Disease individuals. *Neural Plast.* **2016**, 2584940 (2016).
48. Glaich, O. et al. DNA methylation directs microRNA biogenesis in mammalian cells. *Nat. Commun.* **10**, 5657 (2019).
49. Van den Hove, D. L. et al. Epigenetically regulated microRNAs in Alzheimer's disease. *Neurobiol. Aging* **35**, 731–745 (2014).
50. Zhang, M. et al. Ten-eleven translocation 1 mediated-DNA hydroxymethylation is required for myelination and remyelination in the mouse brain. *Nat. Commun.* **12**, 5091 (2021).
51. Britton, J. S. et al. Agora: An open-access platform for the exploration of nascent targets for Alzheimer's disease therapeutics. *Alzheimer's Dement.* **19**, e079328 (2023).
52. Shirwany, N. A., Payette, D., Xie, J. & Guo, Q. The amyloid beta ion channel hypothesis of Alzheimer's disease. *Neuropsychiatr. Dis. Treat.* **3**, 597–612 (2007).
53. He, B. et al. Tissue-specific 5-hydroxymethylcytosine landscape of the human genome. *Nat. Commun.* **12**, 4249 (2021).
54. Yu, L. et al. Association of brain DNA methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 with pathological diagnosis of Alzheimer disease. *JAMA Neurol.* **72**, 15–24 (2015).
55. Yu, M. et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**, 1368–1380 (2012).
56. Bennett, D. A. et al. Religious orders study and rush memory and aging project. *J. Alzheimer's Dis.* **64**, S161–s189 (2018).

57. Schneider, J. A., Arvanitakis, Z., Leurgans, S. E. & Bennett, D. A. The neuropathology of probable Alzheimer disease and mild cognitive impairment. *Ann. Neurol.* **66**, 200–208 (2009).
58. Wilson, R. S., Arnold, S. E., Schneider, J. A., Tang, Y. & Bennett, D. A. The relationship between cerebral Alzheimer's disease pathology and odour identification in old age. *J. Neurol. Neurosurg. Psychiatry* **78**, 30–35 (2007).
59. Bennett, D. A. et al. Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology* **66**, 1837–1844 (2006).
60. Bennett, D. A. et al. Natural history of mild cognitive impairment in older persons. *Neurology* **59**, 198–205 (2002).
61. Bennett, D. A. et al. Decision rules guiding the clinical diagnosis of Alzheimer's disease in two community-based cohort studies compared to standard practice in a clinic-based cohort study. *Neuroepidemiology* **27**, 169–176 (2006).
62. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
63. Risso, D., Ngai, J., Speed, T. P. & Dudoit, S. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* **32**, 896–902 (2014).
64. Kondili, M. et al. UROPA: a tool for Universal RObust Peak Annotation. *Sci. Rep.* **7**, 2593 (2017).
65. Ping, L. et al. Global quantitative analysis of the human brain proteome and phosphoproteome in Alzheimer's disease. *Sci. Data* **7**, 315 (2020).
66. Yu, L. et al. Cortical proteins associated with cognitive resilience in community-dwelling older persons. *JAMA Psychiatry* **77**, 1172–1180 (2020).
67. Mostafavi, S. et al. A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat. Neurosci.* **21**, 811–819 (2018).
68. Mathys, H. et al. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019).
69. Jew, B. et al. Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nat. Commun.* **11**, 1971 (2020).
70. Engebretsen, S. & Bohlin, J. Statistical predictions with glmnet. *Clin. Epigenet.* **11**, 123 (2019).
71. Dunham, I. et al. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
72. Harrow, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012).
73. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
74. Szklarczyk, D. et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–d646 (2023).
75. Giambartolomei, C. et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLOS Genet.* **10**, e1004383 (2014).
76. Zhao, J., Gu, T., Gao, C., Miao, G. & Palma-Gudiel, H. Brain 5-hydroxymethylcytosine alterations are associated with Alzheimer's disease neuropathology (Version v1). https://doi.org/10.5281/zenodo.14804357 (2025).

Acknowledgements

We thank the ROSMAP participants for their participation in these studies, and the staff of the Rush Alzheimer's Disease Center. This work was supported by the National Institutes of Health (NIH) grants: RF1AG052476 (J.Z.), R01DK091369 (J.Z.), R01MH097018 (J.Z.), R01AG064786 (J.Z.), R01DK107532 (J.Z.), P30AG10161(D.A.B.), P30AG72975 (D.A.B.), R01AG15819 (D.A.B.), R01AG17917 (D.A.B.), R01AG16042 (D.A.B.), U01AG46152 (D.A.B.), U01AG61356 (D.A.B.), and R01AG36836 (D.A.B.).

Author contributions

Y.L., J.L., and R.L. conducted the laboratory experiments. T.G., C.G., G.M., and H.P. conducted the data analysis and contributed to drafting the manuscript. J.Z. conceived the idea, directed the project, and prepared the manuscript with input from all the other authors. J.Z., P.J., and D.A.B. supervised the project and obtained funding. All other authors (L.Y., J.Y., Y.W., B.Y., H.W., J.A.S., N.S., F.G., P.L.D.J.) provided critical feedback and contributed to the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-58159-w.

Correspondence and requests for materials should be addressed to Jinying Zhao, Peng Jin or David A. Bennett.

Peer review information *Nature Communications* thanks Reid Alisch, Adam Smith, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.