RESEARCH

Distinct CSF biomarker-associated DNA methylation in Alzheimer's disease and cognitively normal subjects

Wei Zhang[1], Juan I. Young[2,3], Lissette Gomez[3], Michael A. Schmidt[3], David Lukacsovich[1], Achintya Varma[3], X. Steven Chen[1,4], Eden R. Martin[2,3] and Lily Wang[1,2,3,4*]

Abstract

Background Growing evidence has demonstrated that DNA methylation (DNAm) plays an important role in Alzheimer's disease (AD) and that DNAm differences can be detected in the blood of AD subjects. Most studies have correlated blood DNAm with the clinical diagnosis of AD in living individuals. However, as the pathophysiological process of AD can begin many years before the onset of clinical symptoms, there is often disagreement between neuropathology in the brain and clinical phenotypes. Therefore, blood DNAm associated with AD neuropathology, rather than with clinical data, would provide more relevant information on AD pathogenesis.

Methods We performed a comprehensive analysis to identify blood DNAm associated with cerebrospinal fluid (CSF) pathological biomarkers for AD. Our study included matched samples of whole blood DNA methylation, CSF $A\beta_{42}$, phosphorylated tau$_{181}$ (pTau$_{181}$), and total tau (tTau) biomarkers data, measured on the same subjects and at the same clinical visits from a total of 202 subjects (123 CN or cognitively normal, 79 AD) in the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort. To validate our findings, we also examined the association between premortem blood DNAm and postmortem brain neuropathology measured on a group of 69 subjects in the London dataset.

Results We identified a number of novel associations between blood DNAm and CSF biomarkers, demonstrating that changes in pathological processes in the CSF are reflected in the blood epigenome. Overall, the CSF biomarker-associated DNAm is relatively distinct in CN and AD subjects, highlighting the importance of analyzing omics data measured on cognitively normal subjects (which includes preclinical AD subjects) to identify diagnostic biomarkers, and considering disease stages in the development and testing of AD treatment strategies. Moreover, our analysis revealed biological processes associated with early brain impairment relevant to AD are marked by DNAm in the blood, and blood DNAm at several CpGs in the DMR on *HOXA5* gene are associated with pTau$_{181}$ in the CSF, as well as tau-pathology and DNAm in the brain, nominating DNAm at this locus as a promising candidate AD biomarker.

Conclusions Our study provides a valuable resource for future mechanistic and biomarker studies of DNAm in AD.

Keywords DNA methylation, CSF biomarkers, Alzheimer's disease

*Correspondence:
Lily Wang
lily.wangg@gmail.com
Full list of author information is available at the end of the article

Zhang *et al. Alzheimer's Research & Therapy*   (2023) 15:78

Page 2 of 19

Introduction

Late-onset Alzheimer's disease (LOAD), affecting about 1 in 9 people 65 years and older in the USA [1], has become a major public health problem and one of the most financially costly diseases [2]. Clinically, Alzheimer's disease (AD) is characterized by progressive deterioration of cognitive functions, eventually leading to a lack of ability to carry out even the simplest tasks, which places significant emotional, financial, and physical burdens on caregivers. Growing evidence has demonstrated that DNA methylation (DNAm), a widely studied epigenetic mechanism that modifies gene expression without changing the underlying DNA sequences, plays an important role in AD [3–5]. In particular, recent studies have identified and replicated a number of DNAm loci in the brain (e.g., *ANK1*, *RHBDF2*, and *HOXA*) that are robustly associated with AD neuropathology [6–10]. Encouragingly, it has become increasingly evident that DNAm differences can also be detected in the blood of AD subjects [11–16]. Most recently, our meta-analysis of two large clinical AD datasets revealed a number of DNAm loci in the blood significantly associated with AD diagnosis [17].

Given that it still is not practical to obtain methylation levels in brain tissues from living human subjects, most studies have correlated blood DNAm with AD diagnosis. However, as the pathophysiological process of AD can begin many years before the onset of clinical symptoms [18, 19], there is often disagreement between neuropathology and clinical phenotypes [20, 21]. Currently, there is still limited knowledge on the association of blood DNAm with changes in AD neuropathology.

CSF biomarkers are well-established AD endophenotypes, and their abnormality is predictive of the onset and progression of AD [22–27]. Encouragingly, premortem CSF biomarker values also correlate significantly with neuropathology scores measured on postmortem brain samples [28]. The hallmark of AD is the accumulation of aggregated amyloid and tau proteins in the brain. Under the AT(N) framework [29, 30], cerebrospinal fluid (CSF) levels of $A\beta_{42}$, phosphorylated tau at threonine 181 ($pTau_{181}$), and total tau corresponds to the accumulation of $A\beta$ plaque (A), fibrillary tau (T), and non-disease-specific neurodegeneration (N), respectively.

In this study, we performed a comprehensive analysis to identify blood DNA methylation associated with CSF biomarkers in the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort. In addition to a greater understanding of the regulatory changes associated with different pathological disease-associated processes in living individuals, compared to previous analyses that used clinical AD diagnosis as the endpoint, we also expected this analysis of CSF biomarkers, which are quantitative measurements, would help with improving statistical power. To prioritize the significant CSF biomarker-associated DNAm, we performed several integrative analyses that additionally included gene expression and genetics data, as well as a validation analysis which analyzed the London dataset with both premortem blood DNAm and postmortem brain neuropathology measured on a group of 69 subjects. Results from this study provide an improved understanding of the epigenetics underlying inter-individual variations in various pathological pathways involved in AD.

Methods

Study dataset

The ADNI is a longitudinal study that aims to define the progression of AD [31]. To create a dataset with independent samples, we only analyzed the last visit data of each subject from the longitudinal ADNI study. Our blood sample dataset included 202 DNA methylation samples (123 cognitively normal (CN) samples and 79 AD samples) with available CSF biomarkers information ($A\beta_{42}$, phosphorylated tau_{181}, and total tau) measured on the same subject at the same clinical visit in the ADNI study. To avoid the inclusion of early-onset AD subjects, only subjects older than 65 years of age were included. The study datasets can be accessed from the ADNI study website (adni.loni.usc.edu). Sample characteristics for the CN and AD groups were compared using Fisher's exact test for categorical variables and the Wilcoxon rank sum test for continuous variables.

Pre-processing of DNA methylation data

The DNA methylation samples were measured with the Illumina HumanMethylation EPIC beadchip, which includes more than 850,000 CpGs. Supplementary Table 1 shows the number of probes and samples removed at each step of quality control (QC). For the QC of probes, we first selected probes with a detection P-value < 0.01 in every sample. A small detection P-value (i.e., P-value < 0.01) indicates a significant difference between the signals in the probe and the background noise. Next, using the rmSNPandCH function from the DMRcate R package, we removed probes that are cross-reactive [32], located close to single nucleotide polymorphism (SNPs) (i.e., an SNP with minor allele frequency (MAF) ≥ 0.01 was present in the last five base pairs of the probe), or located on X or Y chromosomes. QC for samples included restricting our analysis to samples with good bisulfite conversion efficiency (i.e., $\geq 85\%$). In addition, principal component analysis (PCA) was used to remove the outlier samples. Specifically, PCA was performed using the 50,000 most variable CpGs, and we selected samples within ± 3 standard deviations from the mean of the first PC and second

PC. Finally, we excluded samples without matching clinical or CSF biomarkers information.

The quality-controlled methylation samples were then subjected to the QN.BMIQ normalization procedure [33], which included between-array quantile normalization (QN) followed by within-array β-mixture quantile normalization (BMIQ) [34]. For the QN step, we used the betaqn function in the wateRmelon R package (version 1.99.1) to remove systematic effects between samples. For the BMIQ procedure, which is also implemented in the wateRmelon R package, the distributions of beta values measured by type 1 and type 2 design probes were normalized within each Illumina array.

Immune cell type proportions, including B lymphocytes, natural killer cells, CD4+T lymphocytes, monocytes, and granulocyte, were estimated using the EpiDISH R package (version 2.12.0) [35]. Here, the granulocyte proportions were computed as the sum of neutrophils and eosinophils proportions since neutrophils and eosinophils are classified as granular leukocytes, as previously described [36, 37].

CSF biomarkers
We obtained information for CSF biomarkers (Aβ_{42}, pTau$_{181}$, and tTau), which were measured by Roche Elecsys immunoassay, from the "UPENNBIOMK9.CSV" file at the ADNI website (adni.loni.usc.edu). Standardized CSF biomarkers values were computed by log (base 2)-transformation followed by centering using the study means, as in previous analyses of CSF biomarkers [38, 39].

Identification of CSF biomarker-associated CpGs
To assess the associations between CSF biomarkers (Aβ_{42}, pTau$_{181}$, and tTau) and DNA methylation, we fitted the following linear regression model (Model 1) to CN and AD samples separately: standardized CSF biomarker ~ methylation.beta + age + methylation plate + sex + APOE4 + years of education + smoking history + immune cell-type proportions (B, NK, CD4T, Mono, Gran).

We also compared the effects of methylation-to-CSF biomarker associations in CN samples and AD samples, by fitting the following model (Model 2) to combined CN and AD samples: standardized CSF biomarker ~ methylation.beta + diagnosis + methylation.beta × diagnosis + age + methylation plate + sex + APOE4 + years of education + smoking history + immune cell-type proportions (B, NK, CD4T, Mono, Gran). Significant methylation.beta × diagnosis interaction effect corresponds to a significant difference in methylation-to-CSF biomarker associations in the CN samples and AD samples.

Inflation assessment and correction
We estimated genomic inflation factors (lambda values) using both the conventional approach [40] and the bacon method [41], which is specifically proposed for a more accurate assessment of inflation in EWAS. Supplementary Table 2 shows the estimated inflation and bias of the test statistics from Model 1 described above. Specifically, lambda values (λ) by the conventional approach ranged from 0.719 to 1.096, and lambdas based on the bacon approach (λ.bacon) ranged from 0.863 to 1.019. The estimated bias ranged from -0.097 to 0.117. Genomic correction using the bacon method [41], as implemented in the *bacon* R package, was then applied to obtain bacon-corrected effect sizes, standard errors, and *P*-values for each dataset to obtain a more accurate estimate of statistical significance. After bacon correction, the estimated bias ranged from -0.002 to 0.002, the estimated inflation factors ranged from $\lambda = 0.967$ to 1.042, and λ.bacon ranged from 0.974 to 1.000.

For each CSF biomarker, we considered CpGs with a false discovery rate (FDR) ≤ 0.05 as statistically significant. Given the modest number of samples with both DNA methylation and CSF biomarker measurements, we expected our analysis to be underpowered. Therefore, based on our experiences and previous studies in the analysis of EWAS measured in blood [37, 42, 43], we also prioritized CpGs with suggestive significance at the pre-specified significance threshold *P*-value $< 1 \times 10^{-5}$.

Differentially methylated regions (DMR) analysis
To identify the differentially methylated regions associated with CSF biomarkers, we used the comb-p software [44]. Briefly, comb-p takes single CpG *P*-values and locations of the CpG sites to scan the genome for regions enriched with a series of adjacent low *P*-values. In our analysis, we used the bacon-corrected *P*-values from Model 1 above as the input, and the parameter setting −seed 0.05 and −dist 750 (a *P*-value of 0.05 is required to start a region and extend the region if another *P*-value was within 750 base pairs), which was shown to have optimal statistical properties in our previous comprehensive assessment of the comb-p software [45]. As comb-p uses the Sidak method to correct *P*-values for multiple comparisons, we considered DMRs with Sidak-adjusted *P*-value < 0.05 as significant. To further reduce false positives, we imposed two additional criteria in our final selection of DMRs: (1) DMRs with nominal *P*-value $< 1 \times 10^{-5}$; (2) all CpGs within the DMR have a consistent direction of change in estimated effect sizes from Model 1 described above.

Zhang *et al. Alzheimer's Research & Therapy* (2023) 15:78

Page 4 of 19

Functional annotation of significant methylation associations

The significant methylation at individual CpGs and DMRs was annotated using both the Illumina (UCSC) gene annotation and Genomic Regions Enrichment of Annotations Tool (GREAT) software which associates genomic regions to target genes [46]. To assess the overlap between our significant CpGs and DMRs (CpG or DMR location ± 250 bp) with enhancers, we used enhancer–gene maps generated from 131 human cell types and tissues described in Nasser et al. (2021) [47] (https://www.engreitzlab.org/resources/). Specifically, we selected enhancer-gene pairs with "positive" predictions from the ABC model, which included only expressed target genes, did not include promoter elements, and had an ABC score higher than 0.015. In addition, we also required that the enhancer-gene pairs be identified in cell lines relevant to this study (https://github.com/Trans BioInfoLab/AD-meta-analysis-blood/blob/main/code/annotations/).

Pathway analysis

To identify biological pathways enriched with CSF biomarker-associated DNA methylation, we used the methylRRA function in the methylGSA R package [48] (version 1.14.0). The pathway analyses were performed separately for each of the three CSF biomarkers, and the most significant *P*-value among the 3 *P*-values (one for each CSF biomarker) was then selected as the final *P*-value for each pathway. In each analysis, we used the bacon-corrected *P*-values from Model 1 described above as the input for methylGSA. Briefly, methylGSA first computes a gene-wise ρ value by aggregating *P*-values from multiple CpGs mapped to each gene. Next, the different number of CpGs on each gene is adjusted by Bonferroni correction. Finally, a Gene Set Enrichment Analysis [49] (in pre-rank analysis mode) is performed to identify pathways enriched with significant CSF-associated DNAm. We analyzed pathways in the KEGG [50] and REACTOME [51] databases. Because of the relatively smaller number of gene sets being tested, a 25% FDR significance threshold, instead of the conventional 5% FDR, was suggested to be the default significance threshold for GSEA (https://software.broadinstitute.org/cancer/software/gsea/wiki/index.php/FAQ). Therefore, we considered pathways with FDR < 0.25 as statistically significant.

Integrative methylation-to-gene expression analysis

To evaluate the DNA methylation effect on the gene expression of nearby genes, we analyzed matched gene expression (Affymetrix Human Genome U 219 array) and DNA methylation (EPIC array) data from 263 independent subjects in the ADNI study (adni.loni.usc.edu). To reduce the effect of potential confounding, when testing methylation-to-gene expression associations, we first adjusted age at visit, sex, immune cell-type proportions (for B lymphocytes, natural killer cells, CD4 + T lymphocytes, monocytes, granulocytes), batch effects, number of APOE4 alleles, smoking history, and years of education in both DNA methylation and gene expression levels separately and extracted residuals from the linear models. Immune cell-type proportions were estimated using the R packages EpiDISH [35] and Xcell [52] (https://github.com/dviraran/xCell) for DNA methylation and gene expression data, respectively. A separate linear model was then used to test for the association between methylation residuals and gene expression residuals, separately for CN and AD samples. For the analysis of DMRs, we summarized each DMR by the median methylation value of all CpGs mapped within the DMR, and then fitted the linear models described above, by replacing the methylation value for the CpG with the median methylation value for the DMR.

Correlation and overlap with genetic susceptibility loci

We searched mQTLs using the GoDMC database [53], which was downloaded from http://mqtldb.godmc.org.uk/downloads. To select significant blood mQTLs in GoDMC, we used the same criteria as the original study [53], that is, considering a cis *P*-value smaller than 10^{-8} and a trans-*P*-value smaller than 10^{-14} as significant. The 24 LD blocks of genetic variants reaching genome-wide significance were obtained from Supplementary Table 8 of Kunkle et al. (2019) [54]. The CSF biomarker-associated genetic loci were obtained from Supplementary Tables 2–4 of Deming et al. (2017) [38].

Sensitivity analysis

Immune cell type proportions were estimated using the IDOL algorithm [55], as implemented in the estimate-CellCounts2 function in the R package FlowSorted.Blood.EPIC. We then fitted the same linear models described in "Identification of CSF biomarker-associated CpGs" above, except by replacing cell type proportions estimated by EpiDISH method with those estimated by IDOL algorithm.

Validation analysis using an independent dataset

The London dataset [7, 56], which consists of DNAm measured on premortem whole blood samples from 69 subjects, along with their postmortem neurofibrillary tangle burden as measured by AD Braak stage [57], as well as DNAm measured on the brain prefrontal cortex at autopsy, was downloaded from the GEO database (accession number: GSE29685). The blood and brain

Zhang *et al. Alzheimer's Research & Therapy* (2023) 15:78

Page 5 of 19

Table 1 Sample characteristics of the study dataset

	CN (N = 123)	AD (N = 79)	P- value
N. of females (%)	63 (51)	29 (37)	0.061
age in years, mean (sd)	76.18 (6.62)	77.17 (6.01)	0.217
APOE4 carrier, N (%)	30 (25)	56 (71)	<.001
smoking, N (%)	51 (41)	37 (47)	0.544
education in years, mean(sd)	16.63 (2.54)	16.11 (2.62)	0.168
CSF Aβ_{42}, median (range)	1332 (318 to 3555)	613 (257 to 3114)	<.001
CSF phosphorylated tau$_{181}$, median (range)	20 (8 to 68)	33 (10 to 90)	<.001
CSF total tau, median (range)	231 (112 to 631)	367 (121 to 870)	<.001
MMSE, mean (sd)	28.95 (1.34)	21.77 (4.14)	<.001

DNAm samples from the London dataset were pre-processed in the same way as described above. Given the relatively modest number of samples at some of the Braak stages, we modeled the Braak stage as a binary variable, with absent/low (Braak scores of 0, 1, 2) vs. intermediate/high (Braak scores 3–6) neurofibrillary tangle tau pathology, as previously described [28]. Specifically, to test the association between premortem blood DNAm and postmortem AD Braak stage, we fitted the model methylation M value ~ Braak stage (absence/low vs. intermediate/high) + sex + age at blood draw + batch. In the London dataset, none of the estimated blood cell-type proportions were significantly associated with the Braak stage (Supplementary Fig. 1), so they were unlikely to be confounding factors; therefore, we did not include them in the above linear model. To assess concordance between brain and blood DNAm at each CpG within the DMR located on the *HOXA5* gene, we computed Spearman correlations.

Results
Sample characteristics
To identify DNA methylation associated with CSF biomarkers, we studied matched whole blood DNA methylation, CSF Aβ_{42}, phosphorylated tau$_{181}$ (pTau$_{181}$), and total Tau (tTau) biomarkers data measured on the same subjects and at the same clinical visits in the ADNI study [31, 37]. Our study included samples from a total of 202 subjects (123 cognitively normal, 79 AD cases). Table 1 shows the demographic information of these subjects. There were no significant differences in

age, sex, smoking history, and educational attainment between the cognitively normal (CN) and AD subjects. Overall, the majority of the subjects are in their seventies (with an average age of 76.6), are highly educated (with an average of 16 years of education), and are fewer than half of the subjects smoked. Compared to CN subjects, the AD subjects have a higher proportion of APOE ε4 carriers (71% in AD vs. 25% in CN). Moreover, CSF Aβ_{42} levels were significantly lower in AD subjects, while CSF pTau$_{181}$ and tTau levels were significantly higher in AD subjects. Finally, Mini-Mental State Examination (MMSE) scores were significantly lower in AD subjects (an average of 22 points in AD vs. an average of 29 points in CN), indicating more cognitive dysfunction.

DNA methylation in the blood is significantly associated with CSF biomarkers at individual CpGs and genomic regions
To identify DNAm differences associated with CSF biomarkers at different stages of the disease, we analyzed CN and AD samples separately. Supplementary Table 3 presents a summary of the significant CpGs and DMRs. In CN samples, after adjusting covariate variables (age, sex, batch effects, years of education, number of *APOE4* alleles, smoking history, immune cell-type proportions), and correcting for genomic inflation in each dataset, we identified 1 CpG cg06171420, located in the vicinity of *PCBP3* gene, significantly associated with CSF levels of total tau (tTau) at 5% false discovery rate (FDR) (Supplementary
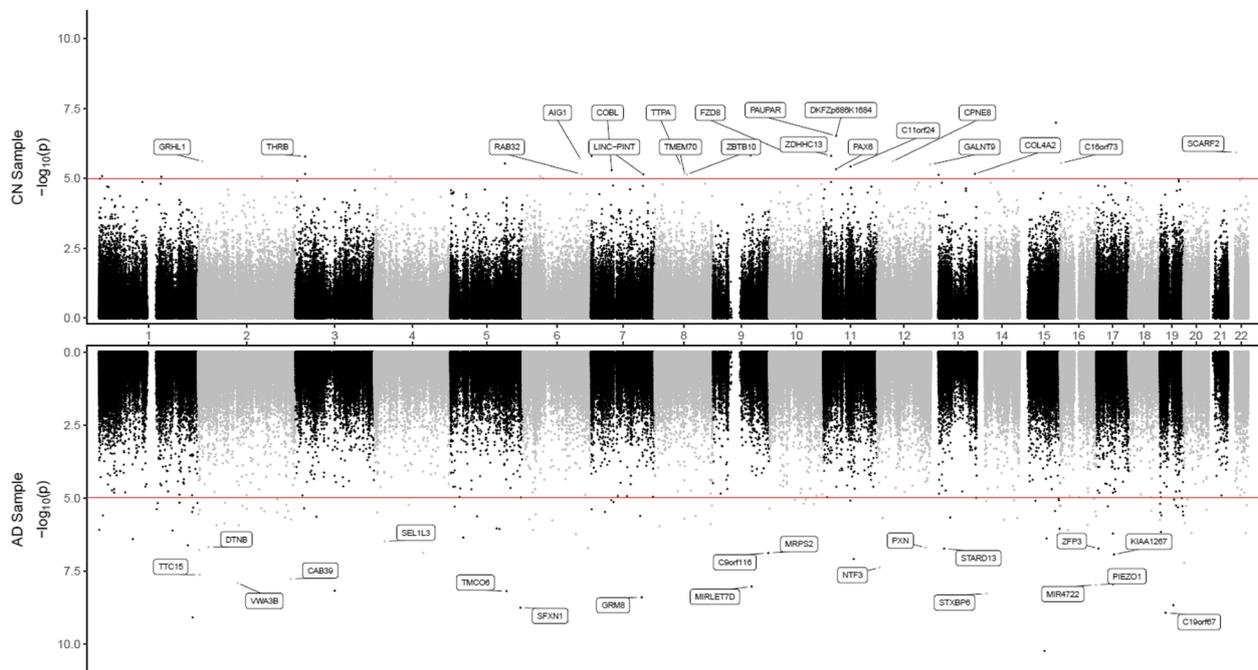
Fig. 1 Miami plot for CpGs significantly associated with CSF Aβ_{42} in the ADNI cohort. The *X*-axis shows chromosome numbers. The *Y*-axis shows –log$_{10}$ (*P*-value) of methylation-to-CSF Aβ_{42} association in cognitively normal (CN) subjects, or Alzheimer's disease (AD) subjects. The genes associated with the 20 most significant CpGs per subject group are highlighted. The red line indicates *P*-value < 10^{-5} significance threshold
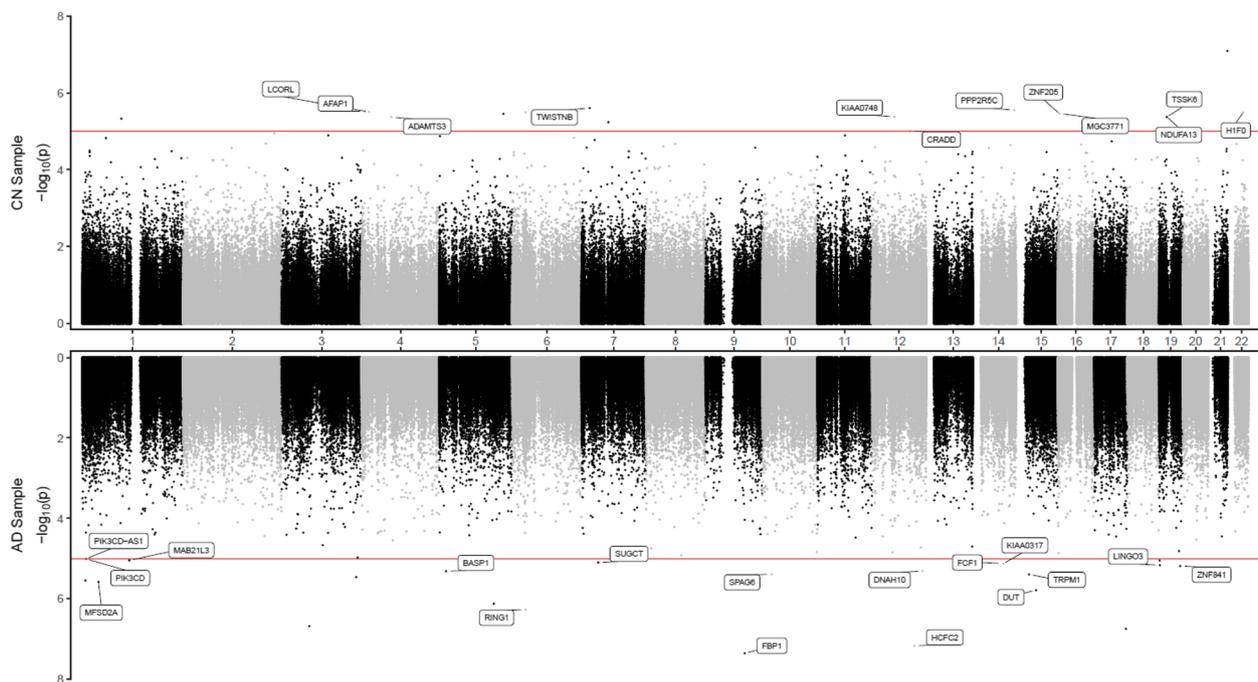
Fig. 2 Miami plot for CpGs significantly associated with CSF phosphorylated tau$_{181}$ (pTau$_{181}$) in the ADNI cohort. The *X*-axis shows chromosome numbers. The *Y*-axis shows –log$_{10}$ (*P*-value) of methylation- to-CSF pTau$_{181}$ association in cognitively normal (CN) subjects, or Alzheimer's disease (AD) subjects. The genes associated with the 20 most significant CpGs per subject group are highlighted. The red line indicates the *P*-value < 10^{-5} significance threshold

Table 2 Top 10 most significant CpGs associated with CSF Aβ$_{42}$ in cognitively normal (CN) and Alzheimer's disease (AD) subjects. Annotations include the location of the CpG based on hg19/GRCh37 genomic annotation (chr, position) and nearby genes based on GREAT (GREAT_annotation). Regression analysis results for CpG-to-CSF Aβ$_{42}$ association include effect estimate, standard error (se), and *P*-values after inflation correction using the bacon method (PMID: 28129774). Highlighted in red are gene promoter regions mapped to significant CpGs

cpg	chr	position	GREAT_annotation	estimate	se	pValue
significant in CN subjects						
cg01373819	chr15	91209329	BLM (-51229);CRTC3 (+136030)	-0.053	0.010	1.03E-07
cg10961330	chr11	31849190	RCN1 (-263260);PAX6 (-9682)	0.064	0.012	3.02E-07
cg07755173	chr10	35930525	FZD8 (-164)	-0.373	0.076	1.08E-06
cg04608146	chr22	20792323	SCARF2 (-178)	-0.119	0.024	1.20E-06
cg21481456	chr9	94351666	NFIL3 (-165523);ROR2 (+360778)	0.050	0.010	1.52E-06
cg14409029	chr11	19138243	ZDHHC13 (-433)	0.024	0.005	1.58E-06
cg04116766	chr7	1283255	UNCX (+10713);MICALL2 (+215883)	0.052	0.011	1.58E-06
cg27622506	chr3	24537160	THRB (-895)	-0.108	0.023	1.66E-06
cg17472152	chr6	143486011	AIG1 (+103999);ADAT2 (+285799)	-0.071	0.015	1.94E-06
cg26510017	chr12	39299546	CPNE8 (-114)	-0.143	0.030	2.46E-06
significant in AD subjects						
cg03492603	chr15	62513534	TLN2 (-340030);C2CD4B (-56053)	-0.188	0.029	5.99E-11
cg07637143	chr1	234833313	IRF2BP2 (-88043);TOMM20 (+458938)	-0.103	0.017	8.39E-10
cg09379609	chr19	14196637	PALM3 (-26667);SAMD1 (+4883)	-0.303	0.050	1.21E-09
cg24037493	chr5	174905263	SFXN1 (-135)	0.076	0.013	1.81E-09
cg16253115	chr19	33781998	SLC7A10 (-65243);CEBPA (+11472)	-0.082	0.014	2.19E-09
cg24700475	chr7	126879927	GRM8 (+13421)	-0.182	0.031	4.07E-09
cg24303533	chr14	25279500	GZMB (-176028);STXBP6 (+240003)	-0.094	0.016	5.47E-09
cg27658391	chr5	140023555	IK (-3800)	-0.079	0.014	6.59E-09
cg25163647	chr3	99113791	DCBLD2 (-493259);COL8A1 (-243528)	-0.070	0.012	6.87E-09
cg03275648	chr9	96940848	ZNF169 (-80745);PTPDC1 (+94103)	-0.080	0.014	9.53E-09

Table 4). At *P*-value $< 1 \times 10^{-5}$, we identified an additional 34, 15, and 11 CpGs significantly associated with CSF Aβ$_{42}$, pTau$_{181}$, and tTau levels, respectively (Figs. 1 and 2, Tables 2 and 3, Supplementary Table 4, 5, 6). Similarly, the analysis of AD samples revealed 125, 21, and 14 CpGs significantly associated with Aβ$_{42}$, pTau$_{181}$, and tTau at *P*-value $< 1 \times 10^{-5}$, respectively, among which 112, 4, and 3 CpGs also achieved 5% FDR. The greater number of DNAm with significant associations to Aβ$_{42}$ than tau (Supplementary Fig. 2) might be due to CSF Aβ$_{42}$ reduction occurring earlier in the disease process and thus is associated with more pervasive epigenetic effects.

Among these 198 significant CSF biomarker-associated CpGs in either CN or AD samples, the majority (61% or 120 CpGs) were negatively associated with increased levels of AD biomarkers; about two-thirds were located in distal regions of genes (65% or 129 CpGs); about half of the significant CpGs (51% or 100 CpGs) were located in CpG islands or shores, and only about a third of them were located in gene promoter regions (Supplementary Tables 4, 5 and 6).

At 5% Sidak adjusted *P*-value, comb-p software identified 81, 18, and 24 differentially methylated regions (DMRs) in CN samples, and 57, 15, and 13 DMRs in AD samples, which were significantly associated with Aβ$_{42}$, pTau$_{181}$, and tTau, respectively (Tables 4 and 5, Supplementary Tables 7-9). The number of CpGs in these DMRs ranged from 3 to 23. Among these 184 DMRs that were significant in either CN or AD samples analysis (Supplementary Table 3), about half (58%, 107 DMRs) were negatively associated with increased levels of AD biomarkers; about half of the DMRs (59%, 109 DMRs) were located in promoter regions; and the majority (80% or 147 DMRs) were located in CpG island or shores. Only a very small number of CpGs (16 CpGs), representing 8% of the total significant CpGs, overlapped with a small number of DMRs (14 DMRs) (Supplementary Fig. 3). Interestingly, among the significant CpGs and DMRs, 18% CpGs (36 CpGs) and 32% DMRs (59 DMRs) also overlapped enhancer regions

Table 3 Top 10 most significant CpGs associated with CSF phosphorylated tau$_{181}$ (pTau$_{181}$) in cognitively normal (CN) and Alzheimer's disease (AD) subjects. Annotations include the location of the CpG based on hg19/GRCh37 genomic annotation (chr, position) and nearby genes based on GREAT (GREAT_annotation). Regression analysis results for CpG-to-CSF pTau$_{181}$ association include effect estimate, standard error (se), and *P*-values after inflation correction using the bacon method (PMID: 28129774). Highlighted in red are gene promoter regions mapped to significant CpGs

		Annotations		Association with pTau$_{181}$		
cpg	chr	position	GREAT_annotation	estimate	se	pValue
significant in CN subjects						
cg06171420	chr21	47058388	SLC19A1 (-96004);PCBP3 (-5220)	-0.074	0.014	8.13E-08
cg20109393	chr7	19748466	TWISTNB (+244)	-0.543	0.115	2.49E-06
cg05020081	chr14	102263662	DYNC1H1 (-167203);PPP2R5C (+35528)	-0.332	0.071	2.82E-06
cg25912009	chr4	7844982	PSAPL1 (-408283);AFAP1 (+96671)	0.144	0.031	3.02E-06
cg05568762	chr4	18023126	LCORL (+357)	-0.340	0.073	3.13E-06
cg18001983	chr6	32913125	HLA-DMB (-4279)	-0.054	0.012	3.23E-06
cg01443426	chr22	38201594	GCAT (-2318);H1F0 (+481)	0.074	0.016	3.24E-06
cg27367045	chr16	3165677	ZNF213 (-19352);ZNF205 (+3117)	-0.080	0.017	3.43E-06
cg10015705	chr5	158713704	UBLCP1 (+23616);IL12B (+44191)	-0.063	0.014	3.53E-06
cg07880109	chr12	55379228	TESPA1 (-773)	-0.029	0.006	4.21E-06
significant in AD subjects						
cg14211930	chr9	97402452	FBP1 (+79)	0.086	0.016	4.47E-08
cg14999716	chr12	104477949	HCFC2 (+19715);NFYB (+54118)	-0.264	0.049	6.81E-08
cg03814390	chr17	77776716	CBX8 (-5802);CBX4 (+36512)	0.346	0.066	1.80E-07
cg01784297	chr3	68716133	FAM19A4 (+265628);FAM19A1 (+662775)	-0.474	0.091	2.10E-07
cg03037740	chr6	33176666	RING1 (+395)	-0.330	0.066	5.42E-07
cg10329462	chr5	135160936	CXCL14 (-245968);SLC25A48 (-9480)	-0.385	0.078	7.57E-07
cg02040583	chr15	48623118	DUT (-502)	-0.120	0.025	1.65E-06
cg15616998	chr1	40435396	CAP1 (-70509);MFSD2A (+14575)	-0.151	0.032	2.64E-06
cg20152585	chr1	8155696	SLC45A1 (-222190);ERRFI1 (-69329)	0.304	0.065	2.86E-06
cg27305835	chr3	184209995	EPHB3 (-69577);CHRD (+112135)	0.153	0.033	3.48E-06

(Supplementary Tables 4–9), which are regulatory DNA sequences that transcription factors bind to activate gene expressions [47, 58].

Blood DNAm associated with CSF biomarkers differed between diagnosis groups

Overall, we found the DNAm associated with CSF biomarkers were relatively distinct across diagnosis groups. Specifically, there was no overlap between the significant CpGs in AD samples and CN samples (Supplementary Fig. 2). Among the 184 significant DMRs that were significant in either CN or AD sample analysis (Supplementary Table 3), only 3 DMRs (chr15:69,744,390–69,744,763, chr6:30,130,819–30,131,284, and chr6:30,130,819–30,131,362), all of which are CSF Aβ$_{42}$ associated-DMRs, were significant in both CN and AD samples. Consistent with this result, there was only a modest and non-significant correlation between estimated effect sizes of CpG-to-CSF biomarker associations in CN samples vs. those in AD samples among significant CpGs (Spearman $\rho = 0.10$, 0.06, 0.18 for Aβ$_{42}$, pTau$_{181}$, and tTau-associated

CpGs, respectively) (Supplementary Fig. 4). Moreover, our interaction model (Model 2 in Methods), which analyzed the combined CN and AD samples, showed that for the majority of the significant CpGs in CN or AD sample analysis (70% or 139 out of a total of 198 CpGs) (Supplementary Tables 4–9), the DNAm × diagnosis interaction effect was significant, indicating significant different DNAm-to-CSF biomarker associations in the two groups.

Pathway analysis revealed DNA methylation associated with CSF biomarkers is enriched in a number of biological pathways in cognitively normal and AD subjects

To better understand biological pathways enriched with significant CSF biomarkers-associated DNA methylation, we next performed pathway analysis using the methylGSA software [48]. At 25% FDR (Methods), a total of 89 and 13 pathways were significant in CN and AD samples, respectively (Supplementary Table 10). Among them, 3 pathways (*calcium signaling pathway, regulation of actin cytoskeleton, neuroactive ligand-receptor interaction*) also reached

Table 4 Top 10 most significant DMRs associated with CSF $A\beta_{42}$ in cognitively normal (CN) and Alzheimer's disease (AD) subjects. For each DMR, annotations include the location of the DMR based on hg19/GRCh37 genomic annotation (chr, start, end) and nearby genes based on GREAT (GREAT_annotation). Direction indicates a positive or negative association between DNA methylation at a CpG located within the DMR and CSF biomarker. Highlighted in red are gene promoter regions mapped to significant DMRs

chr	start	end	n_probes	z_p	z_sidak_p	Direction	GREAT_annotation
(A) significant DMRs in CN subjects							
3	24536252	24537408	19	1.42E-14	8.98E-12	-------------------	THRB (-564)
6	42927959	42928547	23	2.06E-14	2.55E-11	----------------------	GNMT (-243)
6	30130819	30131362	10	5.16E-14	6.95E-11	++++++++++	TRIM10 (-2380);TRIM15 (+98)
7	142494148	142494596	9	5.91E-13	9.64E-10	+++++++++	EPHB6 (-58420);PRSS1 (+37042)
12	39299326	39299727	7	3.06E-12	5.57E-09	-------	CPNE8 (-94)
13	88323992	88324712	7	5.24E-10	5.32E-07	-------	SLITRK5 (-518)
3	158449734	158450252	7	1.34E-09	1.89E-06	-------	RARRES1 (+492)
2	139537745	139537846	5	2.88E-10	2.08E-06	-----	NXPH2 (+122)
4	30721353	30722153	6	2.41E-09	2.20E-06	------	PCDH7 (-1292)
18	25757202	25757711	7	9.86E-09	1.42E-05	-------	CDH2 (-47)
(B) significant DMRs in AD subjects							
2	3642400	3642968	9	1.89E-17	2.43E-14	+++++++++	COLEC11 (-6812);RPS7 (+19889)
17	41437877	41438403	3	4.76E-14	6.60E-11	+++	ARL4D (-38187);TMEM106A (+74246)
15	91473059	91473570	10	6.15E-13	8.79E-10	----------	HDDC3 (+2461);MAN2A2 (+25895)
16	1583391	1584119	10	5.49E-12	5.51E-09	++++++++++	TMEM204 (+181)
15	64673455	64673873	9	4.84E-12	8.46E-09	---------	KIAA0101 (+45)
17	79004850	79005441	7	7.41E-12	9.16E-09	-------	BAIAP2 (-3816)
21	36258423	36259624	11	1.32E-10	8.03E-08	-----------	RUNX1 (+162617);CLIC6 (+217336)
6	26224013	26224121	3	1.27E-11	8.61E-08	+++	HIST1H3E (-1316)
20	57607406	57607693	8	4.50E-11	1.15E-07	++++++++	ATP5E (-113)
12	120652604	120652845	4	4.57E-11	1.39E-07	----	RPLP0 (-13823);PXN (+50838)

Table 5 Top 10 most significant DMRs associated with CSF phosphorylated tau_{181} ($pTau_{181}$) in cognitively normal (CN) subjects and Alzheimer's disease (AD) subjects. For each DMR, annotations include the location of the DMR based on hg19/GRCh37 genomic annotation (chr, start, end), and nearby genes based on GREAT (GREAT_annotation). Direction indicates a positive or negative association between DNA methylation at a CpG located within the DMR and CSF biomarker. Highlighted in red are gene promoter regions mapped to significant DMRs

chr	start	end	n_probes	z_p	z_sidak_p	Direction	GREAT_annotation
(A) significant DMRs in CN subjects							
1	19110562	19110979	5	3.05E-11	5.34E-08	-----	TAS1R2 (+75405);PAX7 (+153271)
18	59997009	59997176	5	6.51E-11	2.85E-07	-----	ZCCHC2 (-193147);TNFRSF11A (+4545)
11	64703192	64703638	8	3.07E-10	5.03E-07	--------	GPHA2 (-55)
15	91473167	91473570	7	8.70E-08	1.58E-04	++++++	HDDC3 (+2407);MAN2A2 (+25949)
1	228362309	228362510	3	1.09E-07	3.98E-04	---	OBSCN (-33421);IBA57 (+8894)
1	110254709	110254897	5	4.57E-07	1.78E-03	+++++	GSTM5 (-74)
2	198650985	198651224	6	6.94E-07	2.12E-03	++++++	BOLL (+193)
21	44898090	44898207	4	5.11E-07	3.19E-03	++++	SIK1 (-51141);HSF2BP (+181225)
6	49681178	49681392	8	1.02E-06	3.49E-03	--------	CRISP2 (-11)
13	111522222	111522315	3	5.59E-07	4.38E-03	+++	ANKRD10 (+45147);ING1 (+154940)
(B) significant DMRs in AD subjects							
4	174203061	174203521	5	5.23E-12	8.31E-09	-----	HMGB2 (+52985);GALNT7 (+113387)
9	97402452	97402658	3	6.03E-10	2.14E-06	+++	FBP1 (-24)
17	7311742	7312082	5	4.20E-08	9.02E-05	+++++	TMEM256-PLSCR3 (-4496);TMEM256 (-4456);NLGN2 (+410)
19	55660514	55660859	6	7.39E-08	1.57E-04	------	TNNT1 (+35)
16	8961250	8961472	6	5.25E-08	1.73E-04	++++++	CARHSP1 (+897)
3	10149803	10149980	5	2.26E-07	9.33E-04	+++++	BRK1 (-7384);FANCD2 (+81794)
7	27183946	27184668	18	1.06E-06	1.07E-03	++++++++++++++++++	HOXA5 (-1020)
5	102898463	102898730	6	5.34E-07	1.46E-03	------	NUDT12 (-103)
1	204183116	204183582	7	4.74E-06	7.41E-03	+++++++	GOLT1A (-129)
10	134226361	134226548	3	2.06E-06	8.02E-03	+++	INPP5A (-124869);PWWP2B (+15783)

5% FDR in CN samples, and 2 pathways (*cardiac conduction* and *muscle contraction*) also reached 5% FDR in AD samples.

We next examined the overlap between significant pathways identified in CN samples and AD samples. Among the 95 pathways that reached 25% FDR in either CN or AD samples, only 7 pathways (7.4%) were significant in both groups (Supplementary Table 10). These seven pathways are *regulation of actin cytoskeleton, neuroactive ligand-receptor interaction, ubiquitin mediated, proteolysis, Wnt signaling pathway, MAPK signaling pathway, cardiac conduction*, and *muscle contraction*. We also found pathway enrichment of the significant CSF biomarker-associated CpGs to be independent in CN samples and AD samples (Supplementary Fig. 5). These pathway analysis results are consistent with those described above for individual CpGs, in which we observed little correlation between estimated effect sizes of CpG-to-CSF biomarkers associations in CN and in AD.

Correlation of DNA methylation at significant CSF biomarker-associated CpGs and DMRs with expressions of nearby genes

To prioritize significant DNAm with downstream functional effects, we next correlated DNA methylation levels of the significant DMRs or CpGs with the expression levels of genes found in their vicinity, using matched DNAm and gene expression samples generated from 263 independent subjects (84 AD cases and 179 CN) in the ADNI cohort. In CN subjects, after removing effects of covariate variables in both DNA methylation and gene expression levels separately (see the "Methods" section), at 5% FDR, we found DNAm at 2 CpGs, and 6 DMRs were significantly associated with target gene expression levels (Supplementary Table 11). Interestingly, aside from 1 CpG (cg14074117) located in the intergenic regions, all CpGs and DMRs were negatively associated with target gene expressions. Among them, 3 DMRs were located in gene promoter regions and negatively associated with expression levels of the target genes at *GSTM5*, *CAT*, and *CRISP2*. *GSTM5* belongs to the Glutathione S-Transferase family of genes, which encodes enzymes associated with oxidative stress in neurodegenerative diseases [59, 60]. Recently, *GSTM5* was observed to be significantly downregulated in the primary visual cortex brain tissues, an area mildly affected by tau pathology and corresponds to the "early" AD transcriptome [61]. This previous finding is consistent with our result that DNAm increases with $pTau_{181}$ and tTau levels and are negatively associated with the target gene. Similarly, the *CAT* gene encodes catalase, another key antioxidative enzyme that mitigates oxidative stress [62]. Defects in catalase have been

implicated in a number of neurological disorders, including AD [63].

On the other hand, in AD samples, we found DNAm at 5 CpGs and 5 DMRs were significantly associated with target gene expression levels. Half of these DNAm (4 CpGs and 1 DMR) had a negative correlation with target gene expression. Two DMRs, located in the promoter region of the *TNNT1* gene, were positively associated with the expression level of the *TNNT1* gene, which was shown to be a marker of central nervous system molecular stress associated with neuropsychiatric diseases [64]. Our results are consistent with previous observations that DNAm at some promoter regions is correlated with increased target gene expression [65–68]. While traditionally promoter methylation is thought to be associated with transcriptional silencing by blocking the binding of transcription factors (TFs), which are proteins that bind DNA to facilitate the transcription of DNA into RNA, recent studies suggest more complex patterns of protein–DNA interaction associated with the DNA methylome [69, 70]. In particular, several studies observed that the binding and activity of some TFs are enhanced by CpG methylation to activate gene expression [71–73]. In addition, the positive promoter DNAm to target gene association could also be due to a co-regulatory phenomenon in which both DNAm and target gene are altered by proteins associated with TFs [53, 69, 74, 75].

Correlation and overlap with genetic susceptibility loci

To identify methylation quantitative trait loci (mQTLs) for the significant DMRs and CpGs, we next performed look-up analyses using the GoDMC database [53] for mQTLs. In CN samples, among the 764 individual CpGs or CpGs located within DMRs that are significantly associated with the CSF biomarkers, 301 CpGs had mQTLs in *cis*, and 41 CpGs had mQTLs in *trans*. Similarly, among the 610 significant CpGs or CpGs located in the DMRs in AD samples, 281 and 55 CpGs had mQTLs in *cis* and in *trans*, respectively. Among them, 30,127 CpG–mQTL pairs, associated with 16 unique CpGs, were significant in both CN and AD sample analyses (Supplementary Table 12). These results suggested that approximately half of the CSF biomarker-associated CpGs are impacted by genetic variation, consistent with a recent large mQTL meta-analysis of blood samples, which estimated that genetic variants influence about 45% of CpGs on the Illumina array [53].

Next, to evaluate if the significant mQTLs in CN and AD overlapped with genetic risk loci implicated in AD, we compared the mQTLs with the 24 LD blocks of genetic variants reaching genome-wide significance in a recent meta-analysis of AD GWAS [54]. In CN samples, we found 1518 mQTLs, associated with DNA

methylation at 10 significant CpGs (all of which are located in DMRs), overlapped with the LD regions chr 6:32,395,036–32,636,434, and 19:1,050,130–1,075,979, which included genetic variants mapped to *HLA-DRA, HLA-DRB5, HLA-DRB1, HLA-DQA1, HLA-DQB1* on chromosome 6, and *ABCA7, ARHGAP45, HMHA1* on chromosome 19 (Supplementary Table 13). Similarly, in AD samples, we found 41 mQTLs, associated with DNA methylation at 9 significant CpGs (all of which are located in DMRs), overlapped with the LD regions chr 6:32,395,036–32,636,434 and chr 15:58,873,555–59,120,077, which included genetic variants mapped to *HLA-DRA, HLA-DRB5, HLA-DRB1, HLA-DQA1, HLA-DQB1* on chromosome 6, and *ADAM10, HSP90AB4P, LOC101928725, FAM63B* on chromosome 15 (Supplementary Table 14). Our comparison of the mQTLs with CSF biomarker-associated genetic loci [38] did not identify any overlapping variants. These results suggested that the majority of the CSF biomarker-associated CpGs, by and large, are not influenced by genetic variants at the GWAS loci for AD or AD biomarkers. Therefore, even though a substantial proportion of the CpGs are influenced by genetic variants, we found no evidence that genetic variations might be confounding variables in our DNAm to CSF biomarker associations because these genetic variations are not significantly associated with AD or AD biomarkers.

Finally, we also evaluated if our significant methylation loci overlapped with the genetic risk loci associated with AD diagnosis [54] or CSF AD biomarkers [38]. However, we found no overlap between the significant DNAm discovered in this study compared with AD diagnosis or CSF AD biomarker-associated genetic risk loci. This result is consistent with a previous study which also found no evidence of overlap between significant EWAS loci and GWAS loci in a meta-analysis of 11 blood-based EWAS of neurodegenerative disorders [36]. The lack of commonality between genetic and epigenetic loci in AD supports previous findings that DNA methylation and genetic variants play relatively independent roles in AD [4, 76].

Sensitivity analysis

We performed an additional analysis to evaluate the robustness of DNAm to CSF biomarker associations with regard to different methods for estimating cell type proportions. To this end, we estimated immune cell type proportions using an alternative method, the IDOL algorithm described in Salas et al. (2018) [55]. Our results show the cell type proportions estimated by the IDOL method and the EpiDISH method [35] we used in our primary analyses are highly concordant (Supplementary

Fig. 6). Next, we repeated our DNAm to CSF biomarkers association analyses by adjusting cell type proportions estimated by IDOL. Our results showed the blood DNAm to CSF biomarker associations obtained by adjusting IDOL cell type proportions are largely congruent with our primary analysis results. In particular, the $A\beta_{42}$-associated CpGs and $pTau_{181}$-associated CpGs remained highly significant, with *P*-values ranging from 1.10×10^{-10} to 1.81×10^{-4} (Supplementary Table 15), and 1.39×10^{-8} to 2.92×10^{-3} (Supplementary Table 16), respectively, indicating our results are robust to different algorithms for estimating cell type proportions.

Validation analysis using an independent dataset

To validate our findings, we also studied DNAm associated with brain pathology in an independent dataset. To this end, we analyzed DNAm measured on premortem blood samples from 69 subjects, along with their postmortem neurofibrillary tangle burden in the brain prefrontal cortex determined at autopsy, as measured by AD Braak stage [57] in the London dataset [7, 56]. At a nominal *P*-value less than 0.05, a number of CSF biomarker-associated CpGs and DMRs that we identified in the ADNI dataset are also significantly associated with the Braak stage in the London dataset (Supplementary Tables 17, 18). These DNAm are located at the *ERO1LB, MBTPS1, HOXA5, TRIM15, TYW3, MME, HMSD, CHAD, SEMA3C* genes, and the intergenic regions. Note that because CSF $A\beta_{42}$ decreases and brain tau-pathology increases in AD subjects, we selected CpGs or DMRs with opposite directions in blood DNAm-to-CSF $A\beta_{42}$ and blood DNAm-to-Braak stage associations.

After correcting for multiple comparisons, at Sidak adjusted *P*-value less than 0.05, we observed blood DNAm at two DMRs, located on the *HOXA5* and *CHAD* genes, were significantly associated with AD Braak stage in the London dataset, and overlapped with CSF $pTau_{181}$ or $A\beta_{42}$ associated DMRs in the ADNI dataset. Of particular interest is the strong replication association signal located in the promoter region of the *HOXA5* gene. In ADNI (discovery) dataset, blood DNAm at DMR chr7:27,183,946–27,184,668 is significantly associated with CSF $pTau_{181}$ (*P*-value $= 1.06 \times 10^{-6}$, Sidak-adjusted *P*-value $= 1.07 \times 10^{-3}$); in London (replication) dataset, blood DNAm at this locus (at DMR chr7: 27,183,133–27,184,451) is also significantly associated with Braak stage in the brain (*P*-value $= 7.27 \times 10^{-20}$, Sidak-adjusted *P*-value $= 2.49 \times 10^{-17}$) (Supplementary Table 18). Previously, Smith et al. (2018) also observed significant hypermethylation across the *HOXA* gene cluster in the brain significantly associated with AD Braak stage in the Mt. Sinai, London, and ROSMAP brain datasets [8]. Intriguingly, we also observed significant correlations between

brain and blood DNAm at 7 CpGs located within the DMR (Supplementary Fig. 7), as well as a significant association between the DMR with target gene expression (Supplementary Fig. 8). Together, these results suggested the DMR at *HOXA5* is a promising biomarker robustly associated with tau-pathology in both brain and the blood.

Discussion

In this study, we analyzed samples from the CN and AD subjects separately, as we reasoned that the CSF biomarker-associated DNAm discovered in CN samples would most likely be associated with AD risk; in contrast, after the onset of disease, the CSF biomarker-associated DNAm in AD samples would most likely be associated with both AD risk as well as changes caused by AD pathologies that accumulate in the brain. Supporting this premise, we found that the significant DNAm identified in AD and CN samples were largely distinct (Supplementary Fig. 2). There was also little correlation between DNAm-to-AD biomarker associations in the two groups of subjects, both at the levels of CpGs (Supplementary Fig. 4) and pathways (Supplementary Fig. 5). These results suggest that the epigenetics associated with different pathological processes in cognitively normal subjects (some of which might later proceed to develop AD) and AD patients vary, supporting the recommendation of considering the patients' disease stage in developing treatment strategies [77, 78].

Our comprehensive analyses identified a number of DNAm differences significantly associated with CSF biomarkers $A\beta_{42}$, $pTau_{181}$, and tTau, many of which were associated with genes previously implicated in AD pathogenesis. Specifically, in the analysis of CN subjects, we identified 1 CpG (cg06171420) mapped to around 5 kb upstream of the *PCBP3* gene, significantly associated with tTau at 5% FDR (Supplementary Table 4, Supplementary Fig. 9). The *PCBP3* gene encodes the RNA-binding protein hnRNPE3 (poly(rC) binding protein 3), which regulates alternative splicing of the tau gene [79, 80]. In Down syndrome, AD, and other neurodegenerative diseases, an abnormal ratio of tau protein isoforms often results in aggregated tau, a major component of neurofibrillary tangles. In the region-based analysis, the most significant CSF $A\beta_{42}$-associated DMR is located in the promoter of the *THRB* gene (Supplementary Fig. 10), which encodes a receptor for the thyroid hormone, previously observed to be dysregulated in AD subjects [81–83].

In AD subjects, we identified significantly more DNA methylation associated with the CSF biomarkers; a total of 112, 4, and 3 CpGs reached 5% FDR in their association with $A\beta_{42}$, $pTau_{181}$, and tTau, respectively. Among

the top 10 most significant CpGs associated with $A\beta_{42}$ (Table 2), cg24037493 maps to the promoter of the *SFXN1* gene and is significantly associated with CSF $A\beta_{42}$ in AD subjects (Supplementary Fig. 11). *SFXN1* encodes the mitochondrial serine transporter, which helps to maintain mitochondrial iron homeostasis [84]. It has been observed that iron levels accumulate in the brains of AD subjects and correlate significantly with cognitive decline [85–87]. Similarly, among the top 10 most significant $pTau_{181}$ and tTau-associated CpGs (Table 2), cg03037740 maps to the promoter of the *RING1* gene and is significantly associated with CSF $pTau_{181}$ (Supplementary Fig. 12). *RING1* encodes a protein that interacts with the polycomb protein BMI1, which plays a critical role in AD pathogenesis. Remarkably, it has been demonstrated that reduced expression of BMI1 protein alone is sufficient to induce both amyloid and tau pathologies in both cellular and animal models [88, 89]. The most significant promoter DMR associated with $A\beta_{42}$ is located at the *TMEM204* gene (Supplementary Fig. 13), which encodes a transmembrane protein that functions as a cell surface marker for infiltrating microglia in the CNS during neuroinflammation [90]. Similarly, the most significant promoter DMR associated with $pTau_{181}$ is located at the *FBP1* gene (Supplementary Fig. 14), which encodes an enzyme that regulates glucose and energy metabolism. It has been observed the expression levels of *FBP1* are reduced in the brains of patients at risk for AD [91, 92], consistent with our observed hypermethylation at the promoter of the *FBP1* gene in samples with increased levels of $pTau_{181}$. Taken together, these results demonstrated that our analysis nominated biologically meaningful DNA methylation loci in the blood associated with AD and, importantly, that changes in the different pathological processes in the CSF, both before and after the clinical diagnosis of AD, are reflected in the epigenome.

In AD samples, the most significant pathways that reached 5% FDR are *cardiac conduction* (P-value $=2.76 \times 10^{-4}$, FDR $=2.54 \times 10^{-2}$) and *muscle conduction* (P-value $=1.42 \times 10^{-4}$, FDR $=2.54 \times 10^{-2}$), which also achieved 25% FDR in CN samples (P-value $=3.58 \times 10^{-4}$, FDR $=6.58 \times 10^{-2}$; P-value $=5.63 \times 10^{-4}$, FDR $=7.85 \times 10^{-2}$). In recent years, the interaction between the heart and brain has increasingly been recognized [93]. Cardiovascular disease, even subclinical cardiac damage, has been shown to be a significant risk factor for dementia [94–97].

In CN samples, interestingly, among the most significant pathways enriched with significant CpGs is the KEGG pathway "Alzheimer's disease", which was curated based on recent AD literature and included genes that confer AD risks, such as *APOE*, *PSENEN*, *MAPT*, *CALM3*, *MME*, and others. Also, in CN samples, the

most significant pathway is the *calcium signaling pathway* (P-value $= 2.39 \times 10^{-4}$,   FDR $= 9.09 \times 10^{-3}$),   consistent with the calcium hypothesis of AD, which posits that dysregulated neuronal calcium homeostasis induces impaired synaptic plasticity and defective neurotransmission, promotes accumulation of Aβ and tau proteins, and subsequently leads to neuronal apoptosis in the brain [98, 99]. Moreover, increased levels of free intracellular calcium have also been observed in normal aging, the strongest risk factor for AD [100, 101]. The second most significant pathway is the *regulation of actin cytoskeleton* (P-value $= 1.61 \times 10^{-3}$,   FDR $= 2.51 \times 10^{-2}$),   consistent with the observation that synapse degeneration is a key early feature of AD pathogenesis [102, 103], and stability of the actin cytoskeleton is crucial for maintaining functional integrity of the dendritic spines at sites for neurotransmission in the brain [104]. These results suggest that some of the brain impairment during the early stages of the disease (i.e., preclinical) is also reflected in the blood epigenome.

Although the majority of the CSF biomarker-associated DNAm differed in CN and AD samples, our analyses also identified a small number of DMRs that were significantly associated with CSF biomarkers in both groups (Supplementary Fig. 2), which could serve as candidate biomarkers in future studies of AD progression. Specifically, three DMRs, all of which were associated with Aβ$_{42}$, reached Sidak adjusted P-value < 0.05 in both CN and AD sample analyses. The first DMR chr15:69,744,390–69,744,763 is located at the promoter of the *RPLP1* gene, which encodes a subunit protein of the ribosome. A defective ribosomal function is associated with decreased capacity for protein synthesis, reduced number of synapses, and has been observed as an early feature of AD preceding neuronal loss [105, 106]. Another noteworthy result is two overlapping DMRs significantly associated with CSF Aβ$_{42}$, at chr6:30,130,819–30,131,284 in AD samples and chr6:30,130,819–30,131,362 in CN samples, both are located in the promoter of the *TRIM15* gene, which encodes a member of the TRIM protein family involved in the ubiquitin system responsible for degrading misfolded protein aggregates and plays important roles in neurodegenerative diseases [107, 108].

To validate our findings, we studied premortem blood DNAm associated with postmortem Braak stage measured on prefrontal cortex samples in an independent dataset, previously described as the London dataset [7]. Encouragingly, we found a number of CSF-biomarker-associated blood DNAm also correlated significantly with the Braak stage, which corresponds to neurofibrillary tangle tau pathology burden in the brain (Supplementary Tables 17, 18). In the London dataset, we observed a strong blood DNAm to Braak stage association signal

located at a DMR in the promoter region of the *HOXA5* gene. Interestingly, this locus also showed a significant association to CSF pTau$_{181}$ in the ADNI dataset (Supplementary Table 18, Supplementary Fig. 15). Moreover, we also observed a significant correlation between brain DNAm and blood DNAm at a subset of 7 CpGs within the DMR (Supplementary Fig. 7), as well as a significant association between the DMR and downstream target gene expression (Supplementary Fig. 8). Consistent with previous studies, which discovered the extensive hypermethylation in the brain at the *HOXA* gene clusters significantly associated with tau neuropathology [7], our study provided strong evidence that these hypermethylated CpGs can also be observed in the blood epigenome, and are significantly associated with pTau$_{181}$ levels in the CSF (Supplementary Table 18). Taken together, these results nominate hypermethylation at the *HOXA5* locus in the blood as a plausible biomarker for tau pathology.

On the other hand, given brain and blood cells originate from different developmental cell lineages, previous studies also suggested that DNA methylation profiles are, by and large, distinct between brain and blood [7, 17, 109]. Consistent with these previous results, our comparison of the significant blood DNAm from this study with significant brain DNAm associated with AD pathology in two large recent meta-analyses of postmortem brain tissues [9, 110] shows only a few overlapping DNAm (3 CpGs and 8 DMRs), mapped to *PRSSL1, LINGO3, SPRED2, HOXA2, NR2F1, CPT1B, HOXA5, ZFPM1* genes, and intergenic regions, were significant with both blood DNAm-to-CSF Aβ$_{42}$/pTau$_{181}$ association and brain DNAm-to-brain Aβ/tau association (Supplementary Tables 4–9). Also, there is not any overlap between significant blood DNAm associated with the CSF AD biomarkers and significant blood DNAm associated with clinical AD from our previous meta-analyses of two large clinical AD datasets [17, 111]. This is not surprising, given the disconnection between brain pathology and clinical diagnosis in AD; it has been observed that a substantial proportion of cognitively normal subjects also have AD pathology in the brain [2021].

This study has several limitations. First, we analyzed the methylation levels measured on whole blood, which contains a complex mixture of cell types. To reduce confounding effects due to different cell types, we included estimated cell-type proportions as covariate variables in all our analyses. Future studies that utilize single-cell technology for gene expression and DNAm could improve power and shed more light on the particular cell types affected at the DNAm loci discovered in this study. Second, to study DNAm associated with CSF biomarkers in subjects at different stages of the disease (i.e., preclinical or clinical), we separately analyzed samples from cognitively normal and AD subjects, which reduced the

sample sizes of the analysis datasets considerably. Given the modest sample size, we pre-defined a more liberal significance threshold (i.e., P-value $< 10^{-5}$) based on previous analyses of blood DNA methylation data [17, 37, 43], to select a small number of loci that were then further prioritized using additional integrative analyses. Future studies with larger sample sizes are needed to identify and replicate DNAm loci at more stringent significance thresholds. Third, we did not consider MCI subjects in this study because there is considerable heterogeneity among MCI subjects, with subjects converting to AD at different trajectories [112]. As ADNI is currently conducting additional phases of the study, future analyses with a larger sample size will make it possible to detect DNA methylation to CSF AD biomarker associations in different subgroups of MCI subjects. Fourth, although women make up about two-thirds of AD patients in the general US population [1], our study cohort (which had both CSF biomarkers and blood DNAm available in ADNI) had a disproportionately lower proportion of females in the AD group (37% females in AD group vs. 51% females in CN group) (Table 1). Therefore, our study cohort may not represent a random sample from the general population. In all our analyses, we adjusted the variable sex in addition to other covariate variables, so the DNAm-to-CSF biomarkers associations we identified are independent of sex. Large and diverse community-based cohort studies that validate our findings are needed. Fifth, as recent autopsy studies revealed that about a quarter of CN subjects also shows AD neuropathology in the brain [20, 21], the CSF biomarker-associated methylation we observed in CN subjects could potentially be markers of an early feature in AD that precedes clinical diagnosis. Future studies that develop DNAm-based prediction models for diagnosing AD and compare their performance with state-of-the-art plasma biomarkers of AD are needed. Finally, the associations we identified do not necessarily reflect causal relationships. Future studies are needed to establish the causality of the nominated DNA methylation markers.

Conclusions

In this study, we leveraged AD biomarkers as quantitative outcomes to identify DNAm associated with various AD pathology. Our study found a number of novel associations between blood DNAm and CSF $A\beta_{42}$, phosphorylated tau_{181}, and total tau, which are proxy biomarkers of AD pathophysiology, demonstrating that changes in various pathological processes in the CSF are reflected in the blood epigenome. Overall, the CSF biomarker-associated DNA methylome is relatively distinct in CN and AD subjects, highlighting the importance of analyzing omics data measured on cognitively normal subjects (which includes preclinical AD subjects) to identify diagnostic biomarkers, and considering disease stages in the development and testing of AD treatment strategies. Our analysis of blood samples of cognitively normal subjects pointed to a number of potential therapeutic targets relevant to the treatment of AD, such as calcium channel blockers associated with *calcium signaling pathway* [98], and spine stabilizing therapy associated with *regulation of actin cytoskeleton* [104]. Moreover, we found blood DNAm at several CpGs in the DMR on the *HOXA5* gene are not only associated with CSF $pTau_{181}$, but also tau-pathology in the brain, as well as brain DNAm at the same locus in an independent dataset, nominating DNAm at this locus as a promising candidate AD biomarker. In summary, our study provides a valuable resource for future mechanistic and biomarker studies in AD.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13195-023-01216-7.

Additional file 1: Supplementary Table 1. Quality control (QC) information on pre-processing of DNA methylation samples and probes. **Supplementary Table 2.** Estimates of inflation and bias in the analysis CpG to CSF biomarker associations in Alzheimer's disease (AD) and cognitively normal (CN) samples. The bacon approach (PMID: 28129774) was implemented using the R package bacon. Conventional approach for inflation estimate is based on the method described in Devlin and Roeder (PMID: 11315092). **Supplementary Table 3.** Summary of number of significant CpGs and DMRs associated with CSF AD biomarkers. **Supplementary Table 4.** Significant CpGs associated with CSF total tau in cognitive normal (CN) and Alzheimer's disease (AD) subjects. Highlighted in red are significant DNAm to CSF biomarker associations with P-value $< 10^{-5}$ or FDR < 0.05, or disease by DNAm interaction with P-value < 0.05. Also highlighted in red are gene promoter regions mapped with significant CpGs. The significant DNAm were compared to analysis results of brain samples in Zhang et al. (2020) (PMID: 33257653) and Shireby et al. (2022) (PMID: 36153390). **Supplementary Table 5.** Significant CpGs associated with CSF $A\beta_{42}$ in cognitive normal (CN) and Alzheimer's disease (AD) subjects. Highlighted in red are significant associations with P-value $< 10^{-5}$, FDR < 0.05, or disease by DNAm interaction with P-value < 0.05. Highlights in yellow indicates overlap with significant DNAm in previous literature. Also highlighted in red are gene promoter regions mapped with significant CpGs. The significant DNAm were compared to analysis results of brain samples in Zhang et al. (2020) (PMID: 33257653) and Shireby et al. (2022) (PMID: 36153390). **Supplementary Table 6.** Significant CpGs associated with CSF phosphorylated tau_{181} in cognitive normal (CN) and Alzheimer's disease (AD) subjects. Values highlighted in red are significant associations with P-value $< 10^{-5}$, FDR < 0.05, or disease by DNAm interaction with P-value < 0.05. Also highlighted in red are genes with significant CpGs in the promoter regions. Highlights in yellow indicates overlap with significant DNAm in previous literature. The significant DNAm were compared to analysis results of brain samples in Zhang et al. (2020) (PMID: 33257653) and Shireby et al. (2022) (PMID: 36153390). **Supplementary Table 7.** Significant DMRs at 5% Sidak adjusted P-value (z_sidak_p) associated with CSF $A\beta_{42}$ in cognitively normal (CN) subjects and Alzheimer's disease (AD) subjects. For each DMR, annotations include location of the DMR based on hg19/GRCh37 genomic annotation (chr, start, end), nearby genes based on GREAT (GREAT_annotation), and Illumina gene annotations (UCSC_RefGene_Name), location with respect to CpG islands (Relation_to_Island), and overlap with enhancers described in Nasser et al. (2021) study (PMID: 33828297). Direction indicates positive or negative association between DNA methylation at a CpG located within the DMR

and CSF biomarker. Highlights in yellow indicates overlap with significant DNAm in previous literature. Highlights in red indicate gene promoter regions mapped with significant DMRs. The significant DNAm were compared to analysis results of brain samples in Zhang et al. (2020) (PMID: 33257653) and Shireby et al. (2022) (PMID: 36153390). **Supplementary Table 8.** Significant DMRs at 5% Sidak adjusted *P*-value (z_sidak_p) associated with CSF pTau in cognitively normal (CN) subjects and Alzheimer's disease (AD) subjects. For each DMR, annotations include location of the DMR based on hg19/GRCh37 genomic annotation (chr, start, end), nearby genes based on GREAT (GREAT_annotation) and Illumina gene annotations (UCSC_RefGene_Name), location with respect to CpG islands (Relation_to_Island), and overlap with enhancers described in Nasser et al. (2021) study (PMID: 33828297). Direction indicates positive or negative association between DNA methylation at a CpG located within the DMR and CSF biomarker. Highlights in yellow indicates overlap with significant DNA methylation loci in previous literature. The significant DNAm were compared to analysis results of brain samples in Zhang et al. (2020) (PMID: 33257653) and Shireby et al. (2022) (PMID: 36153390). **Supplementary Table 9.** Significant DMRs at 5% Sidak adjusted *P*-value (z_sidak_p) associated with CSF total Tau in cogntively normal (CN) subjects and Alzheimer's disease (AD) subjects. For each DMR, annotations include location of the DMR based on hg19/GRCh37 genomic annotation (chr, start, end), nearby genes based on GREAT (GREAT_annotation) and Illumina gene annotations (UCSC_RefGene_Name), location with respect to CpG islands (Relation_to_Island), and overlap with enhancers described in Nasser et al. (2021) study (PMID: 33828297). Direction indicates positive or negative association between DNA methylation at a CpG located within the DMR and CSF biomarker. Highlights in yellow indicates overlap with significant DNA methylation loci in previous literature. The significant DNAm were compared to analysis results of brain samples in Zhang et al. (2020) (PMID: 33257653) and Shireby et al. (2022) (PMID: 36153390). **Supplmentary Table 10.** Results of pathway analysis using methylGSA (PMID: 30346483). In cognitively normal (CN) subjects, at 25% FDR, a total of 13 KEGG pathways and 76 Reactome pathways were enriched with CSF biomarker-associated CpGs, among which 3 KEGG pathways also reached 5% FDR. In AD samples, at 25% FDR, 10 KEGG pathways and 3 Reactome pathways were enriched with CSF biomarker-associated CpGs, among which 2 Reactome pathways also reached 5% FDR. Shown in red are FDR values less than 0.25, values of FDR < 0.05 are additionally highlighted in bold. Description of pathways that reached FDR < 0.25 in both CN and AD samples are also highlighted in red. **Supplementary Table 11.** Significant associations between CSF biomarker-associated CpGs and DMRs with target genes in blood samples of cognitively normal (CN) subjects and Alzheimer's disease (AD) subjects. We analyzed CpGs located in the promoter regions and distal regions separately. More specifically, for CpGs located in the promoter region (within ± 2 kb around the transcription start sites (TSS)), we tested the association between CpG methylation with expression levels of the target genes; for CpGs in the distal regions (> 2 kb from TSS), we tested associations between CpG methylation with expression levels of ten nearest genes upstream and downstream from the CpG. **Supplementary 12.** A total of 30127 CpG - mQTL pairs, associated with 16 unique CpGs, were significant in both cognitively normal (CN) and Alzheimer's disease (AD) sample analyses. The blood mQTLs were obtained from the GoDMC database. Definitions for columns under "Blood mQTLs" can be obtained from README file at http://mqtldb.godmc.org.uk/downloads. **Supplementary Table 13.** In cognitively normal (CN) samples, a total of 1518 mQTLs overlapped with the 24 GWAS nominated LD blocks in Kunkle et al. (2019) (PMID: 30820047). The mQTLs in blood were obtained from the GoDMC database. Annotations for CpGs include location of the CpG based on hg19/GRCh37 genomic annotation (Chr, Position), Illumina gene annotation (UCSC_RefGene_Name), the type of associated genomic feature (UCSC_RefGene_Group), and location with respect to CpG islands (Relation_to_Island). **Supplementary Table 14.** In Alzheimer's disease (AD) samples, a total of 41 mQTLs overlapped with the 24 GWAS nominated LD blocks in Kunkle et al. (2019) (PMID: 30820047). The mQTLs in blood were obtained from the GoDMC database. Annotations for CpGs include location of the CpG based on hg19/GRCh37 genomic annotation (Chr, Position), Illumina

gene annotation (UCSC_RefGene_Name), the type of associated genomic feature (UCSC_RefGene_Group), and location with respect to CpG islands (Relation_to_Island). **Supplementary Table 15.** Sensitivity analysis for model that adjust cell type proportions estimated by the IDOL algorithm (PMID: 29843789), as implemented by estimateCellCounts2 function in R package FlowSorted.Blood.EPIC. All $A\beta_{42}$-associated CpGs remained highly significant, with *P*-values ranging from 1.10×10^{-10} to 1.81×10^{-4}. **Supplementary Table 16.** Sensitivity analysis for model that adjust cell type proportions estimated by the IDOL algorithm (PMID: 29843789), as implemented by estimateCellCounts2 function in R package FlowSorted.Blood.EPIC. All $pTau_{181}$-associated CpGs remained highly significant, with *P*-values ranging from 1.39×10^{-8} to 2.92×10^{-3}. **Supplementary Table 17.** CpGs with significant associations to both CSF AD biomarkers (in ADNI dataset) and Braak stage (in London dataset). Highlighted in red are associations that reached $P < 10^{-5}$ in ADNI dataset analysis and $P < 0.05$ in London dataset. **Supplementary Table 18.** CSF biomarker-associated DMRs with significant associations to both CSF AD biomarkers (in ADNI dataset) and brain pathology (Braak stage in the independent London dataset).

Additional file 2: Supplementary Figure 1. The estimated blood cell-type proportions from pre-mortem DNA methylation samples are not significantly associated with Braak scores measured on postmortem brain samples in the London dataset. **Supplementary Figure 2.** Overlap between significant individual CpGs (A) and significant DMRs (B) in the analysis of samples from AD subjects and cognitively normal subjects. Abbreviations: AD = Alzheimer's disease, CN = cognitivelynormal. **Supplementary Figure 3.** Overlap between significant individual CpGs (*P* < 10^{-5}) and CpGs in significant DMRs (Sidak adjusted *P* < 0.05) associated with CSF biomarkers in (A) Alzheimer's disease subjects and (B) cognitively normal subjects. **Supplementary Figure 4.** The correlations of estimated effect sizes for CpG-to-CSF biomarker associations in cognitively normal (CN) subjects vs. those in Alzheimer's disease (AD) subjects are modest and non-significant. **Supplementary Figure 5.** The *P*-values for pathway enrichment of CSF biomarker-associated DNA methylation in cognitively normal (CN) samples are independent of those in Alzheimer's disease (AD) samples. Shown are 95 pathways (Supplementary Table 9) that reached FDR < 0.25 in either CN sample analysis or AD sample analysis. Abbreviations: R = Spearman correlation, p = *P*-value. **Supplementary Figure 6.** Immune cell type proportions estimated using the EpiDISH (PMID: 28193155) and the IDOL (labeled as "estimateCell") (PMID: 29843789) algorithms are highly concordant. **Supplementary Figure 7.** A subset of CpGs located in DMR chr7: 27183946 – 27184668 at *HOXA5* geneshowed both (A) Significant association with Braak stage in London dataset (information extracted fromSupplementary Table 18). (B) Significant correlation between brain DNAm and blood DNAm. **Supplementary Figure 8.** DNA methylation (DNAm) at the DMR chr7:27183946 - 27184668 is significantly correlated with HOXA5 gene expression. The ADNI dataset with matched gene expression and DNAm measured on 263 subjects was used for this analysis. Residuals were obtained by adjusting DNAm and gene expression separately by age, sex, immune cell-type proportions, batch effect, number of APOE4 alleles, smoking and years of education. Methylation level for the DMR was estimated by median methylation level over all CpGs mapped within the DMR. **Supplementary Figure 9.** DNA methylation at the CpG cg06171420 is significantly associated with CSF total tau (TAU) in cognitively normal samples in the ADNI dataset. To remove confounding effects from covariate variables, residuals were obtained by fitting model log CSF total tau ~ age + methylation plate + sex + APOE4 + years of education + smoking history + immune cell-type proportions. **Supplementary Figure 10.** $A\beta_{42}$-associated DMR on chromosome 3. The highlighted CpG next to purple line / purple dot is the most significant CpG in the DMR. **Supplementary Figure 11.** DNA methylation at the CpG cg24037493 is significantly associated with CSF $A\beta_{42}$ in Alzheimer's disease subjects in the ADNI dataset. To remove confounding effects from covariate variables, residuals were obtained by fitting model log CSF Aβ42 ~ age + methylation plate + sex + APOE4 + years of education + smoking history + immune cell-type proportions. *An outlier sample with beta value of 0.5, log abeta residual of 0.04 was omitted to improve resolution of the figure for all other data points. The fitted line represents estimated

Zhang *et al. Alzheimer's Research & Therapy* (2023) 15:78

Page 16 of 19

linear model applied to all data points, including the omitted outlier. Robust linear model, which is often used to model data with outlier, showed similar P-value for this CpG cg24037493 (P-value $= 8.79 \times 10\text{-}7$) as the P-value in linear regression (P-value $= 1.81 \times 10^{-9}$). **Supplementary Figure 12.** DNA methylation at the CpG cg03037740 is significantly associated with CSF pTau$_{181}$ in Alzheimer's disease subjects in the ADNI dataset. To remove confounding effects from covariate variables, residuals were obtained by fitting model log CSF pTau$_{181}$ ~ age $+$ methylation plate $+$ sex $+$ APOE4 $+$ years of education $+$ smoking history $+$ immune cell-type proportions. **Supplementary Figure 13.** The Aβ_{42}-associated DMR on chromosome 16. The highlighted CpG next to purple line / purple dot is the most significant CpG in the DMR. **Supplementary Figure 14.** The pTau181-associated DMR on chromosome 9. The highlighted CpG next to purple line / purple dot is the most significant CpG in the DMR. **Supplementary Figure 15.** The pTau$_{181}$-associated DMR at chr7: 27183946 - 27184668 in the ADNI dataset. The highlighted CpG next to purple line / purple dot is the most significant CpG in the DMR.

Availability of data and materials

The ADNI datasets can be accessed from http://adni.loni.usc.edu. The scripts for the analysis performed in this study can be accessed at https://github.com/TransBioInfoLab/AD-ATN-biomarkers-and-DNAm.

Declarations

Ethics approval and consent to participate

The study procedures were approved by the institutional review boards of all participating centers (https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf), and written informed consent was obtained from all participants or their authorized representatives. The study was conducted in accordance with the Declaration of Helsinki and all study participants provided written informed consent for data collection. All work complied with ethical regulations for working with human participants. Ethics approval was obtained from the institutional review boards of each institution involved: Oregon Health and Science University; University of Southern California; University of California—San Diego; University of Michigan; Mayo Clinic, Rochester; Baylor College of Medicine; Columbia University Medical Center; Washington University, St. Louis; University of Alabama at Birmingham; Mount Sinai School of Medicine; Rush University Medical Center; Wien Center; Johns Hopkins University; New York University; Duke University Medical Center; University of Pennsylvania; University of Kentucky; University of Pittsburgh; University of Rochester Medical Center; University of California, Irvine; University of Texas Southwestern Medical School; Emory University; University of Kansas, Medical Center; University of California, Los Angeles; Mayo Clinic, Jacksonville; Indiana University; Yale University School of Medicine; McGill University, Montreal-Jewish General Hospital; Sunnybrook Health Sciences, Ontario; U.B.C. Clinic for AD & Related Disorders; Cognitive Neurology—St. Joseph's, Ontario; Cleveland Clinic Lou Ruvo Center for Brain Health; Northwestern University; Premiere Research Inst (Palm Beach Neurology); Georgetown University Medical Center; Brigham and Women's Hospital; Stanford University; Banner Sun Health Research Institute; Boston University; Howard University; Case Western Reserve University; University of California, Davis—Sacramento; Neurological Care of CNY; Parkwood Hospital; University of Wisconsin; University of California, Irvine—BIC; Banner Alzheimer's Institute; Dent Neurologic Institute; Ohio State University; Albany Medical College; Hartford Hospital, Olin Neuropsychiatry Research Center; Dartmouth-Hitchcock Medical Center; Wake Forest University Health Sciences; Rhode Island Hospital; Butler Hospital; UC San Francisco; Medical University South Carolina; St. Joseph's Health Care Nathan Kline Institute; University of Iowa College of Medicine; Cornell University and University of South Florida: USF Health Byrd Alzheimer's Institute. The investigators within the ADNI contributed to the design and implementation of the ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found online (http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Division of Biostatistics, Department of Public Health Sciences, University of Miami Miller School of Medicine, 1120 NW 14Th Street, Miami, FL 33136, USA. [2]Dr. John T Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL 33136, USA. [3]John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL 33136, USA. [4]Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, FL 33136, USA.

References

1. Rajan KB, et al. Population estimate of people with clinical Alzheimer's disease and mild cognitive impairment in the United States (2020–2060). Alzheimer's Dement. 2021;17(12):1966–75.
2. Hurd MD, Martorell P, Delavande A, Mullen KJ, Langa KM. Monetary costs of dementia in the United States. N Engl J Med. 2013;368:1326–34.
3. Lord J, Cruchaga C. The epigenetic landscape of Alzheimer's disease. Nat Neurosci. 2014;17:1138–40.
4. Klein HU, Bennett DA, De Jager PL. The epigenome in Alzheimer's disease: current state and approaches for a new path to gene discovery and understanding disease mechanism. Acta Neuropathol. 2016;132:503–14.

Zhang *et al. Alzheimer's Research & Therapy*   (2023) 15:78

Page 17 of 19

5. Stoccoro A, Coppede F. Role of epigenetics in Alzheimer's disease pathogenesis. Neurodegener Dis Manag. 2018;8:181–93.

6. De Jager PL, et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. Nat Neurosci. 2014;17:1156–63.

7. Lunnon K, et al. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. Nat Neurosci. 2014;17:1164–70.

8. Smith RG, et al. Elevated DNA methylation across a 48-kb region spanning the HOXA gene cluster is associated with Alzheimer's disease neuropathology. Alzheimers Dement. 2018;14:1580–8.

9. Zhang L, et al. Epigenome-wide meta-analysis of DNA methylation differences in prefrontal cortex implicates the immune processes in Alzheimer's disease. Nat Commun. 2020;11:6114.

10. Zhang L, et al. Sex-specific DNA methylation differences in Alzheimer's disease pathology. Acta Neuropathol Commun. 2021;9:77.

11. Fransquet PD, et al. DNA methylation analysis of candidate genes associated with dementia in peripheral blood. Epigenomics. 2020;12:2109–23.

12. Fransquet P.D, et al. Blood DNA methylation signatures to detect dementia prior to overt clinical symptoms. Alzheimers Dement (Amst). 2020;12:e12056.

13. Kobayashi N, et al. Increased blood COASY DNA methylation levels a potential biomarker for early pathology of Alzheimer's disease. Sci Rep. 2020;10:12217.

14. Roubroeks JAY, et al. An epigenome-wide association study of Alzheimer's disease blood highlights robust DNA hypermethylation in the HOXB6 gene. Neurobiol Aging. 2020;95:26–45.

15. Madrid A, et al. DNA hypomethylation in blood links B3GALT4 and ZADH2 to Alzheimer's disease. J Alzheimers Dis. 2018;66:927–34.

16. Mitsumori R, et al. Lower DNA methylation levels in CpG island shores of CR1, CLU, and PICALM in the blood of Japanese Alzheimer's disease patients. PLoS ONE. 2020;15: e0239196.

17. Silva, TC, et al. Cross-tissue analysis of blood and brain epigenome-wide association studies in Alzheimer's disease. Nat Commun. 2022;13:4852.

18. Jack CR Jr, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol. 2010;9:119–28.

19. Sperling RA, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7:280–92.

20. Jansen WJ, Janssen O, Tijms BM, et al. Prevalence Estimates of Amyloid Abnormality Across the Alzheimer Disease Clinical Spectrum. JAMA Neurol. 2022;79(3): 228–43.

21. Ossenkoppele R, et al. Prevalence of amyloid PET positivity in dementia syndromes: a meta-analysis. JAMA. 2015;313:1939–49.

22. Llano D.A, Bundela S, Mudar R.A, Devanarayan V, Alzheimer's Disease Neuroimaging, I. A multivariate predictive modeling approach reveals a novel CSF peptide signature for both Alzheimer's Disease state classification and for predicting future disease progression. PLoS One. 2017;12:e0182098.

23. Blennow K, et al. Predicting clinical decline and conversion to Alzheimer's disease or dementia using novel Elecsys Abeta(1–42), pTau and tTau CSF immunoassays. Sci Rep. 2019;9:19024.

24. Clark CM, et al. Cerebrospinal fluid tau and beta-amyloid: how well do these biomarkers reflect autopsy-confirmed dementia diagnoses? Arch Neurol. 2003;60:1696–702.

25. Shaw LM, et al. Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. Ann Neurol. 2009;65:403–13.

26. Strozyk D, Blennow K, White LR, Launer LJ. CSF Abeta 42 levels correlate with amyloid-neuropathology in a population-based autopsy study. Neurology. 2003;60:652–6.

27. Tapiola T, et al. Cerebrospinal fluid beta-amyloid 42 and tau proteins as biomarkers of Alzheimer-type pathologic changes in the brain. Arch Neurol. 2009;66:382–9.

28. Grothe MJ, et al. Associations of Fully Automated CSF and Novel Plasma Biomarkers With Alzheimer Disease Neuropathology at Autopsy. Neurology. 2021;97(12):e1229–42.

29. Jack CR Jr, et al. NIA-AA Research Framework: toward a biological definition of Alzheimer's disease. Alzheimers Dement. 2018;14:535–62.

30. Jack CR Jr, et al. A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. Neurology. 2016;87:539–47.

31. Veitch DP, et al. Understanding disease progression and improving Alzheimer's disease clinical trials: recent highlights from the Alzheimer's Disease Neuroimaging Initiative. Alzheimers Dement. 2019;15:106–52.

32. Chen YA, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. 2013;8:203–9.

33. Wang T, et al. A systematic study of normalization methods for Infinium 450K methylation data using whole-genome bisulfite sequencing data. Epigenetics. 2015;10:662–9.

34. Teschendorff AE, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013;29:189–96.

35. Teschendorff AE, Breeze CE, Zheng SC, Beck S. A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. BMC Bioinformatics. 2017;18:105.

36. Nabais MF, et al. Meta-analysis of genome-wide DNA methylation identifies shared associations across neurodegenerative disorders. Genome Biol. 2021;22:90.

37. Vasanthakumar A, et al. Harnessing peripheral DNA methylation differences in the Alzheimer's Disease Neuroimaging Initiative (ADNI) to reveal novel biomarkers of disease. Clin Epigenetics. 2020;12:84.

38. Deming Y, et al. Genome-wide association study identifies four novel loci associated with Alzheimer's endophenotypes and disease modifiers. Acta Neuropathol. 2017;133:839–56.

39. Deming Y, et al. Sex-specific genetic predictors of Alzheimer's disease biomarkers. Acta Neuropathol. 2018;136:857–72.

40. Delvin B, Roeder K. Genomic control for association studies. Biometrics. 1999;55:997–1004.

41. van Iterson, M., van Zwet, E.W., Consortium, B. & Heijmans, B.T. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. Genome Biol. 2017;18:19.

42. Silva TC, et al. Cross-tissue meta-analysis of blood and brain epigenome-wide association studies in Alzheimer's disease. Nature Communications: In Press; 2022.

43. Starnawska A, et al. Epigenome-wide association study of cognitive functioning in middle-aged monozygotic twins. Front Aging Neurosci. 2017;9:413.

44. Pedersen BS, Schwartz DA, Yang IV, Kechris KJ. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. Bioinformatics. 2012;28:2986–8.

45. Mallik S, et al. An evaluation of supervised methods for identifying differentially methylated regions in Illumina methylation arrays. Brief Bioinform. 2019;20:2224–35.

46. McLean CY, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010;28:495–501.

47. Nasser J, et al. Genome-wide enhancer maps link risk variants to disease genes. Nature. 2021;593:238–43.

48. Ren X, Kuan PF. methylGSA: a Bioconductor package and Shiny app for DNA methylation data length bias adjustment in gene set testing. Bioinformatics. 2019;35:1958–9.

49. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545–50.

50. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 2012;40:D109–14.

51. Gillespie M, et al. The reactome pathway knowledgebase 2022. Nucleic Acids Res. 2022;50:D687–92.

52. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. 2017;18:220.

53. Min JL, et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. Nat Genet. 2021;53:1311–21.

54. Kunkle BW, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. Nat Genet. 2019;51:414–30.

55. Salas LA, et al. An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. Genome Biol. 2018;19:64.

56. Hannon E, Lunnon K, Schalkwyk L, Mill J. Interindividual methylomic variation across blood, cortex, and cerebellum: implications for epigenetic studies of neurological and neuropsychiatric phenotypes. Epigenetics. 2015;10:1024–32.

57. Braak H, Braak E. Staging of Alzheimer's disease-related neurofibrillary changes. Neurobiol Aging. 1995;16:271–8 discussion 278–84.

58. Mukherjee S, Erickson H, Bastia D. Enhancer-origin interaction in plasmid R6K involves a DNA loop mediated by initiator protein. Cell. 1988;52:375–83.

59. Mazzetti A.P, Fiorile M.C, Primavera A, Lo Bello M. Glutathione transferases and neurodegenerative diseases. Neurochem Int. 2015;82:10–8.

60. Kumar A, et al. Role of Glutathione-S-transferases in neurological problems. Expert Opin Ther Pat. 2017;27:299–309.

61. Guennewig B, et al. Defining early changes in Alzheimer's disease from RNA sequencing of brain regions differentially affected by pathology. Sci Rep. 2021;11:4865.

62. Nandi A, Yan LJ, Jana CK, Das N. Role of catalase in oxidative stress- and age-associated degenerative diseases. Oxid Med Cell Longev. 2019;2019:9613090.

63. Habib LK, Lee MT, Yang J. Inhibitors of catalase-amyloid interactions protect cells from beta-amyloid-induced oxidative stress and toxicity. J Biol Chem. 2010;285:38933–43.

64. Lowe X, Wyrobek A. Characterization of the early CNS stress biomarkers and profiles associated with neuropsychiatric diseases. Curr Genomics. 2012;13:489–97.

65. Spainhour JC, Lim HS, Yi SV, Qiu P. Correlation patterns between DNA methylation and gene expression in The Cancer Genome Atlas. Cancer Inform. 2019;18:1176935119828776.

66. Bahar Halpern K, Vana T, Walker M.D. Paradoxical role of DNA methylation in activation of FoxA2 gene expression during endoderm development. J Biol Chem. 2014;289:23882–92.

67. Hantusch B, Kalt R, Krieger S, Puri C, Kerjaschki D. Sp1/Sp3 and DNA-methylation contribute to basal transcriptional activation of human podoplanin in MG63 versus Saos-2 osteoblastic cells. BMC Mol Biol. 2007;8:20.

68. Niesen MI, et al. Activation of a methylated promoter mediated by a sequence-specific DNA-binding protein. RFX J Biol Chem. 2005;280:38914–22.

69. Zhu H, Wang G, Qian J. Transcription factors as readers and effectors of DNA methylation. Nat Rev Genet. 2016;17:551–65.

70. Heberle E, Bardet AF. Sensitivity of transcription factors to DNA methylation. Essays Biochem. 2019;63:727–41.

71. Yin Y, et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science. 2017;356(6337):eaaj2239.

72. Hu S, et al. DNA methylation presents distinct binding sites for human transcription factors. Elife. 2013;2: e00726.

73. Silva TC, Young JI, Martin ER, Chen XS, Wang L. MethReg: estimating the regulatory potential of DNA methylation in gene transcription. Nucleic Acids Res. 2022;50:e51–e51.

74. Bonder MJ, et al. Disease variants alter transcription factor levels and methylation of their binding sites. Nat Genet. 2017;49:131–8.

75. Banovich NE, et al. Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. PLoS Genet. 2014;10: e1004663.

76. Chibnik LB, et al. Alzheimer's loci: epigenetic associations and interaction with genetic factors. Ann Clin Transl Neurol. 2015;2:636–47.

77. Sperling R.A, Jack C.R. Jr, Aisen P.S. Testing the right target and right drug at the right stage. Sci Transl Med. 2011;3:111cm33.

78. Mortberg MA, Vallabh SM, Minikel EV. Disease stages and therapeutic hypotheses in two decades of neurodegenerative disease clinical trials. Sci Rep. 2022;12:17708.

79. Wang Y, Gao L, Tse SW, Andreadis A. Heterogeneous nuclear ribonucleoprotein E3 modestly activates splicing of tau exon 10 via its proximal downstream intron, a hotspot for frontotemporal dementia mutations. Gene. 2010;451:23–31.

80. Fisher E, Feng J. RNA splicing regulators play critical roles in neurogenesis. Wiley Interdiscip Rev RNA. 2022;13(6):e1728.

81. Bensenor IM, Lotufo PA, Menezes PR, Scazufca M. Subclinical hyperthyroidism and dementia: the Sao Paulo Ageing & Health Study (SPAH). BMC Public Health. 2010;10:298.

82. Kalmijn S, et al. Subclinical hyperthyroidism and the risk of dementia. The Rotterdam study Clin Endocrinol (Oxf). 2000;53:733–7.

83. van Osch LA, Hogervorst E, Combrinck M, Smith AD. Low thyroid-stimulating hormone as an independent risk factor for Alzheimer disease. Neurology. 2004;62:1967–71.

84. Tifoun N, et al. A High-Throughput Search for SFXN1 Physical Partners Led to the Identification of ATAD3, HSD10 and TIM50. Biology (Basel). 2022;11(9):1298.

85. Du L, et al. Increased Iron Deposition on Brain Quantitative Susceptibility Mapping Correlates with Decreased Cognitive Function in Alzheimer's Disease. ACS Chem Neurosci. 2018;9:1849–57.

86. Sun Y, et al. Characterizing Brain Iron Deposition in Patients with Subcortical Vascular Mild Cognitive Impairment Using Quantitative Susceptibility Mapping: A Potential Biomarker. Front Aging Neurosci. 2017;9:81.

87. Qin Z, et al. Quantitative Susceptibility Mapping of Brain Iron Relating to Cognitive Impairment in Hypertension. J Magn Reson Imaging. 2022;56:508–15.

88. Hogan R, et al. The Role of BMI1 in Late-Onset Sporadic Alzheimer's Disease. Genes (Basel). 2020;11(7):825.

89. Flamier A, et al. Modeling Late-Onset Sporadic Alzheimer's Disease through BMI1 Deficiency. Cell Rep. 2018;23:2653–66.

90. DePaula-Silva AB, et al. Differential transcriptional profiles identify microglial- and macrophage-specific gene markers expressed during virus-induced neuroinflammation. J Neuroinflammation. 2019;16:152.

91. Duran-Aniotz C, Hetz C. Glucose Metabolism: A Sweet Relief of Alzheimer's Disease. Curr Biol. 2016;26:R806–9.

92. Gonzalez A, Calfio C, Churruca M, Maccioni RB. Glucose metabolism and AD: evidence for a potential diabetes type 3. Alzheimers Res Ther. 2022;14:56.

93. Nolte CH, Endres M. The heart of the matter: a link between troponin and dementia? Eur Heart J. 2014;35:1779–81.

94. Kovacic JC, Castellano JM, Fuster V. The links between complex coronary disease, cerebrovascular disease, and degenerative brain disease. Ann N Y Acad Sci. 2012;1254:99–105.

95. Emanuele E, Martinelli V, Abbiati V, Ricevuti G. Linking atherosclerosis to Alzheimer's disease: focus on biomarkers. Front Biosci (Elite Ed). 2012;4:700–10.

96. Sahathevan R, Brodtmann A, Donnan GA. Dementia, stroke, and vascular risk factors; a review. Int J Stroke. 2012;7:61–73.

97. Schneider AL, et al. High-sensitivity cardiac troponin T and cognitive function and dementia risk: the atherosclerosis risk in communities study. Eur Heart J. 2014;35:1817–24.

98. Tong BC, Wu AJ, Li M, Cheung KH. Calcium signaling in Alzheimer's disease & therapies. Biochim Biophys Acta Mol Cell Res. 2018;1865:1745–60.

99. Chami M. Calcium Signalling in Alzheimer's Disease: From Pathophysiological Regulation to Therapeutic Approaches. Cells. 2021;10(1):140.

100. Raza M, et al. Aging is associated with elevated intracellular calcium levels and altered calcium homeostatic mechanisms in hippocampal neurons. Neurosci Lett. 2007;418:77–81.

101. Oh MM, Oliveira FA, Waters J, Disterhoft JF. Altered calcium metabolism in aging CA1 hippocampal pyramidal neurons. J Neurosci. 2013;33:7905–11.

102. Selkoe DJ. Alzheimer's disease is a synaptic failure. Science. 2002;298:789–91.

103. Davies CA, Mann DM, Sumpter PQ, Yates PO. A quantitative morphometric analysis of the neuronal and synaptic content of the frontal and temporal cortex in patients with Alzheimer's disease. J Neurol Sci. 1987;78:151–64.

104. Penzes P, Vanleeuwen JE. Impaired regulation of synaptic actin cytoskeleton in Alzheimer's disease. Brain Res Rev. 2011;67:184–92.

105. Ding Q, Markesbery WR, Chen Q, Li F, Keller JN. Ribosome dysfunction is an early event in Alzheimer's disease. J Neurosci. 2005;25:9171–5.

106. Hernandez-Ortega K, Garcia-Esparcia P, Gil L, Lucas JJ, Ferrer I. Altered machinery of protein synthesis in Alzheimer's: from the nucleolus to the ribosome. Brain Pathol. 2016;26:593–605.

107. Watanabe M, Hatakeyama S. TRIM proteins and diseases. J Biochem. 2017;161:135–44.

108. Zhu Y, Afolabi LO, Wan X, Shim JS, Chen L. TRIM family proteins: roles in proteostasis and neurodegenerative diseases. Open Biol. 2022;12: 220098.

Zhang *et al. Alzheimer's Research & Therapy*　　(2023) 15:78

Page 19 of 19

109. Yu L, et al. Methylation profiles in peripheral blood CD4+ lymphocytes versus brain: the relation to Alzheimer's disease pathology. Alzheimers Dement. 2016;12:942–51.
110. Shireby G, et al. DNA methylation signatures of Alzheimer's disease neuropathology in the cortex are primarily driven by variation in non-neuronal cell-types. Nat Commun. 2022;13:5620.
111. Silva TC, et al. Distinct sex-specific DNA methylation differences in Alzheimer's disease. Alzheimers Res Ther. 2022;14:133.
112. Nettiksimmons J, DeCarli C, Landau S, Beckett L, Alzheimer's Disease Neuroimaging, I. Biological heterogeneity in ADNI amnestic mild cognitive impairment. Alzheimers Dement. 2014;10:511-521 e1.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.