nature communications

Article

https://doi.org/10.1038/s41467-026-68864-9

Integrative epigenomic landscape of Alzheimer's Disease brains reveals oligodendrocyte molecular perturbations associated with tau

Received: 19 February 2025

Accepted: 16 January 2026

Published online: 03 March 2026

Check for updates

Stephanie R. Oatman[1], Joseph S. Reddy[2], Amin Atashgaran[1], Xue Wang[2], Yuhao Min[1], Zachary Quicksall[2], Floor Vanelderen[1], Minerva M. Carrasquillo[1], Chia-Chen Liu[1], Yu Yamazaki[1], Thuy T. Nguyen[1], Michael Heckman[2], Na Zhao[1], Michael DeTure[1], Melissa E. Murray[1], Guojun Bu[1], Takahisa Kanekiyo[1], Dennis W. Dickson[1], Mariet Allen[1] & Nilüfer Ertekin-Taner[1,3]

Alzheimer's disease (AD) brains have variable neuropathologic and biochemical changes. Capturing epigenetic factors associated with this variability can reveal novel biological insights into AD pathophysiology. Here, we conduct an epigenome-wide association study of DNA methylation in 472 AD brains with neuropathologic and biochemical brain protein levels core to AD pathogenesis. Using a novel regional methylation (rCpGm) approach, we identify 5478 significant associations, 99.7% of which associate with tau biochemical measures, and 93 concordant associations in external datasets. Transcriptome-methylome integration reveals enrichment in oligodendrocyte genes, including known AD risk gene *BIN1*, myelination genes *MYRF, MBP* and *MAG* previously implicated in AD, and novel genes like *LDB3*. Further characterization of these perturbations in independent AD and primary tauopathy datasets highlights consistent tau-related associations. In summary, we uncover the integrative epigenomic landscape of AD, demonstrate tau-related oligodendrocyte gene perturbations as a common potential pathomechanism across tauopathies and share findings via our Multiomic Atlas.

Amyloid beta (Aβ) plaques and tau neurofibrillary tangles (NFT) are core neuropathological hallmarks of Alzheimer's Disease (AD)[1,2]. The characteristic distribution and levels of these amyloid and tau hallmarks in the brain are often measured by Thal phase[3] and Braak stage[4], respectively. These core measures show high inter- and intra-individual brain variability within AD[5–8] suggesting the presence of distinct mechanisms controlling the levels and biochemical states of their underlying proteins. This is supported by our previous findings that key proteins involved in the pathophysiology of AD including Aβ, tau, and apolipoprotein E (APOE) each have unique genetic architectures that associate with their specific biochemical states and levels in the AD brain[9]. Characterization of these endophenotype specific molecular mechanisms in the AD brain are important as each protein and state has potentially unique functions and consequences[10]. Insoluble states of Aβ42 and Aβ40 are major components enriched in Aβ plaques in the brain parenchyma and cerebral amyloid angiopathy

[1]Department of Neuroscience, Mayo Clinic, Jacksonville, FL, USA. [2]Department of Quantitative Health Sciences, Mayo Clinic, Jacksonville, FL, USA. [3]Department of Neurology, Mayo Clinic, Jacksonville, FL, USA. e-mail: taner.nilufer@mayo.edu

(CAA) deposits, respectively[6,11]. CAA is a co-pathology often seen in AD characterized by amyloid deposits in the cerebral vasculature which can impair blood flow and may contribute to clinical deficits beyond AD[12–14]. Insoluble states of tau, particularly hyperphosphorylated tau (pTau), are core components of NFT which are often seen after brain levels of soluble total tau rise and become hyperphosphorylated[15]. Recent studies have also shown that membrane bound states of Aβ and tau correlate with clinical severity and that tau increases its propensity to fibrillize when contacting a plasma membrane[16–21]. Moreover, the APOE protein which is important in transporting and processing lipids in the brain, is co-deposited in insoluble Aβ plaques and its gene harbors the largest known genetic risk factor for AD, the *APOEε4* tagging variant[22–24].

Although there has been great progress identifying risk for AD including genetic and environmental/lifestyle factors[25,26], the underlying molecular mechanisms that these risk factors act through to impact specific AD endophenotypes remains to be fully characterized. It is likely that these risk factors affect both shared and distinct mechanisms that underpin the hallmarks of AD and investigating these factors will be important in discovering the molecular underpinnings of AD. Epigenetic mechanisms are known to mediate genetic as well as environmental and lifestyle factor effects on gene expression. Previous epigenetic studies of AD found broad dysregulation of epigenetic mechanisms corresponding to transcriptomic dysregulation[27,28]. A key epigenetic regulator of gene expression is DNA methylation (DNAm). Previous studies have found differential DNAm sites near known AD risk genes including *BIN1, APOE, APP, ADAM10, MAPT, SPI1, TREM2* and others[29–37]. DNAm changes have also been linked to AD risk, onset and progression[30–32,38,39]. These results suggest that epigenetic regulation, particularly that of DNAm, may be important in modulating expression of known AD risk genes and thereby influencing core AD endophenotypes. Nonetheless, large scale, integrative studies have yet to investigate how dysregulation of DNAm impacts specific brain endophenotypes beyond global neuropathology levels. Identification of AD-endophenotype specific molecular perturbations is required to deconstruct the underlying pathophysiology of AD.

We hypothesize that, within the AD brain, neuropathology, levels and biochemical states of core AD-related proteins are influenced by variations in DNAm. To test this, we performed epigenome-wide association studies (EWAS) using an innovate approach to analyze regional levels of DNAm (rCpGm) from the temporal cortex (TCX) and cerebellum (CER) with AD-related neuropathologic measures (Braak, Thal, and CAA) and brain biochemical levels of five proteins (APOE, Aβ40, Aβ42, tau, and pTau) (Fig. 1). Through our deep endophenotype EWAS and integrative omics approach leveraging a large cohort of neuropathologically confirmed AD cases, we identified widespread associations of DNAm with levels of distinct biochemical states of tau connected to significant dysregulation of nearby oligodendrocyte and myelin related gene expression. Interrogation of these tau related, epigenetically dysregulated oligodendrocyte genes and loci revealed

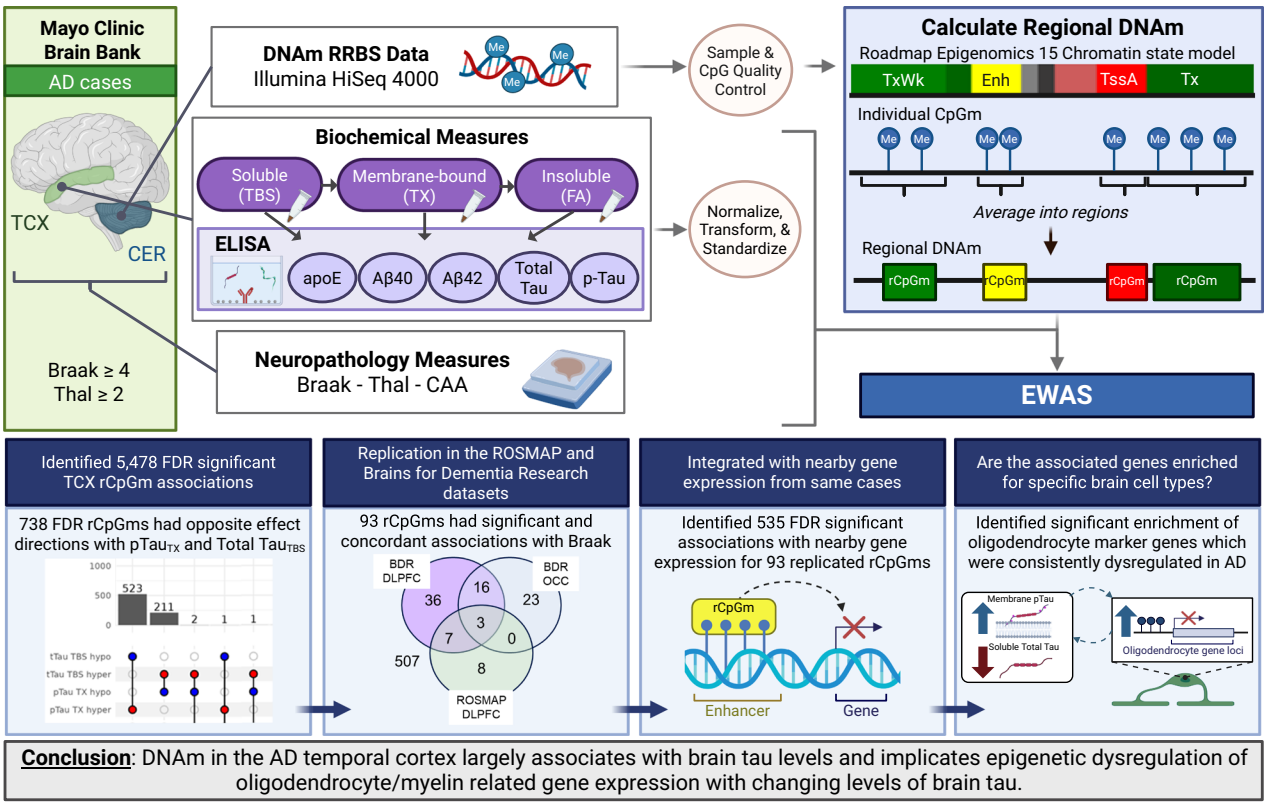


Fig. 1 | Graphical abstract. Graphical description of the study design. Neuropathologically diagnosed AD cases were identified from the Mayo Clinic Brain Bank. Samples had neuropathology measures, biochemical measures from the temporal cortex, and DNA methylation data from the temporal cortex and a subset of samples from the cerebellum. The DNA methylation data was grouped into regions based on windows defined by the Roadmap Epigenomics 15 Chromatin state model temporal lobe predictions (Kundaje et al.[48]). These rCpGms were tested for association with the biochemical and neuropathology measures. Top EWAS hits were then investigated identifying important perturbed biological pathways related to tau and oligodendrocyte gene expression. Aβ Amyloid-β, AD Alzheimer's Disease, CAA Cerebral Amyloid Angiopathy, CER Cerebellum, CpGm DNA methylation at CpG sites, DLPFC Dorsolateral prefrontal cortex, DNAm DNA methylation, Enh Enhancer, EWAS Epigenome wide association study, FA Formic Acid, FDR False discovery rate, Me Methylation, OCC Occipital cortex, rCpGm Regional DNA methylation grouped by chromatin state, RRBS Reduced Representation Bisulfite Sequencing, SNP Single Nucleotide Polymorphism, TBS Tris Buffered Saline, TCX Temporal cortex, TssA Transcription Start Site Active, TX Triton-X, Tx Strong Transcription, TxWk Weak Transcription, Adapted from Oatman et al.[9], *Molecular Neurodegeneration*; Created in Biorender.com.

concordant tau related associations across multiple independent datasets of AD and other tauopathies. Our findings suggest a key relationship between brain levels and biochemical states of tau and expression of oligodendrocyte genes regulated through DNAm perturbations. This may be a common pathomechanism across AD and other tauopathies. We broadly share our data and findings facilitated through our interactive web application, Multiomic Atlas of AD Brain Endophenotypes.

Results

We analyzed levels of DNA methylation at CpG sites using reduced representation bisulfite sequencing (RRBS) from two brain regions, TCX and CER, in a dataset of neuropathologically confirmed AD donors with multiple AD-related endophenotype measurements. These endophenotypes included neuropathological variables (Braak stage, Thal phase, and CAA scores) and TCX biochemical measures of five AD-related proteins[10] (APOE, Aβ40, Aβ42, total tau, and phospho-Tau (pTau) (Thr231)) from three tissue fractions (soluble (TBS), membrane-bound (TX), and insoluble (FA)) (Fig. S1). Our study dataset consisted of 472 individuals (53% female) with CpGm measured from the TCX and a subset of 200 from the CER (Supplementary Data 1). After stringent quality control and data reduction steps, a total of 455 TCX and 191 CER samples with 1,958,174 TCX and 1,858,631 CER CpGm sites were available for downstream analytics (Supplementary Data 2, Fig. S2).

Grouping CpGm by predicted chromatin state is a biologically relevant method to investigate regional levels of DNA methylation

A commonly used EWAS approach to identify AD-related CpGm loci is to associate individual CpGm sites with a phenotype of interest. However, this individual CpGm based EWAS does not account for methylation variations at neighboring CpGm sites which are highly correlated and likely work together to have a functional effect[40–43]. Currently available methods address this by grouping nearby significant individual CpGm sites into regions, however, these approaches are based on post-hoc statistical groupings which lack biological basis[44–47]. In this study, we aimed to investigate CpGm on a regional level defined by the biological function of the genomic region to identify functionally relevant CpGm loci. To do this, we developed a novel method that averages CpGm sites into regions (rCpGm) based on biologically based genomic windows defined by the Roadmap Epigenomics Consortium 15-chromatin state model generated from the human temporal lobe[48], a region pertinent to our human brain dataset (Fig. 1).

First, to determine whether our novel method yields biologically consistent results, we investigated the average methylation levels of rCpGm sites grouped by brain chromatin states[48]. We found that the average methylation levels for each chromatin state were highly consistent with the predicted biology of these states (Fig. 2). Quiescent, repressed, and heterochromatin annotations had higher methylation, typically associated with lower gene expression levels[49], whereas active transcription start sites (TSS), flanking TSS, and bivalent/poised TSS regions had low methylation levels. Chromatin regions that are expected to have more variability across cell types, like enhancer regions, have around 50% methylation in our bulk brain tissue rCpGm data. Our chromatin state specific rCpGm averages were also highly consistent with published whole-genome bisulfite sequencing (WGBS) DNA methylation data[48] demonstrating the utility of our method for defining biologically relevant CpGm regions.

Next, we compared the average rCpGm levels between the TCX and CER brain regions for each chromatin state annotation. TCX and CER have mostly similar levels of rCpGm for each chromatin state with the largest differences in enhancer and transcription states which may reflect the distinct cell-specific epigenetic regulation of these two brain regions (Fig. 2). Investigating the 190 donors with both TCX and CER rCpGm measures, we found that average global rCpGm levels for each sample were similar between these brain regions (range: 0.27–0.34) and highly correlated in the same individual ($r = 0.88 - 0.94$, $p < 2.2e\text{-}16$), although TCX had consistently higher global methylation levels than CER (Fig. 2, Fig. S2).

We next explored the applicability of our novel method in two independent brain DNAm datasets measured via array technologies which are more often used in large-scale AD EWAS. These two datasets are the ROSMAP[30] ($n = 686$) and Brains for Dementia Research (BDR)[32] cohorts with CpGm measured from the dorsolateral prefrontal cortex (DLPFC) in both datasets and also the occipital cortex (OCC) in BDR ($n_{DLPFC} = 581$, $n_{OCC} = 576$) (Supplementary Data 3). Quality control (QC) processing of the DNAm levels from these datasets through our array-specific pipeline resulted in 429,723 CpGm sites in ROSMAP and 747,577 CpGm sites in BDR. Using our novel rCpGm method, we identified 136,201 rCpGms in ROSMAP and 203,435 rCpGms in BDR (Fig. S3). These independent datasets are similar in the levels and patterns of their average rCpGm for each chromatin state annotation across these datasets (Fig. S3) and in comparison to our dataset (Fig. 2), despite differences in donors, brain regions, collection sites and methylation measurement methods highlighting the replicability of our novel analytic approach. These average rCpGm patterns are also consistent even when examining only those rCpGms common to all datasets (Fig. S3). Notably, the array-based rCpGms have a narrower range of values (Fig. S3) in comparison to our RRBS-based measurements, particularly on the hypomethylated end which may potentially be due to lower precision and dynamic range for the former methodology[50,51]. Even within array-based datasets, we observe differences across chromatin states where 450 K arrays have collectively higher rCpGm averages than the EPICv1 array. These differences indicate potential methodology-specific impacts on absolute methylation value measurements, which are unlikely to influence our results since we analyze methylation levels within each dataset. We do not recommend comparisons of absolute rCpGm values across datasets measured with independent technologies. Further, rCpGm levels for each chromatin annotation are more similar between DLPFC and OCC regions for the BDR cohort than for the DLPFC region across BDR and ROSMAP cohorts (Fig. S3). This suggests that inter-individual differences may be greater than intra-individual brain region differences in methylation.

Collectively, these results underscore the robust and replicable rCpGm measurements obtained with our analytic approach which demonstrate the broad utility and biological relevance of our novel method.

DNA methylation associates primarily with tau related brain endophenotypes

To determine DNA methylation (DNAm) associations with AD-related endophenotypes we took two EWAS approaches, namely the more common method of individual CpGm associations and regional methylation associations using our novel method that averages CpGms into regions based on the 15-chromatin state model[48]. We also compared results from our novel regional CpGm association method (rCpGm) to the published post-hoc regional aggregation *comb-p* method[45].

Neuropathology endophenotypes (Braak stage, Thal phase, CAA score) were available for all study donors (i.e., both TCX and CER samples), whereas brain biochemical endophenotypes (protein levels of APOE, Aβ40, Aβ42, Total tau, and pTau measured in soluble (TBS), membrane-bound (TX), and insoluble (FA) brain tissue fractions) were measured only from TCX tissue (Fig. S1). We performed EWAS of individual CpGm levels from both TCX and CER with neuropathology and those from TCX also with biochemical endophenotypes adjusting for age at death, sex, and the first 3 genetic principal components. We

https://doi.org/10.1038/s41467-026-68864-9

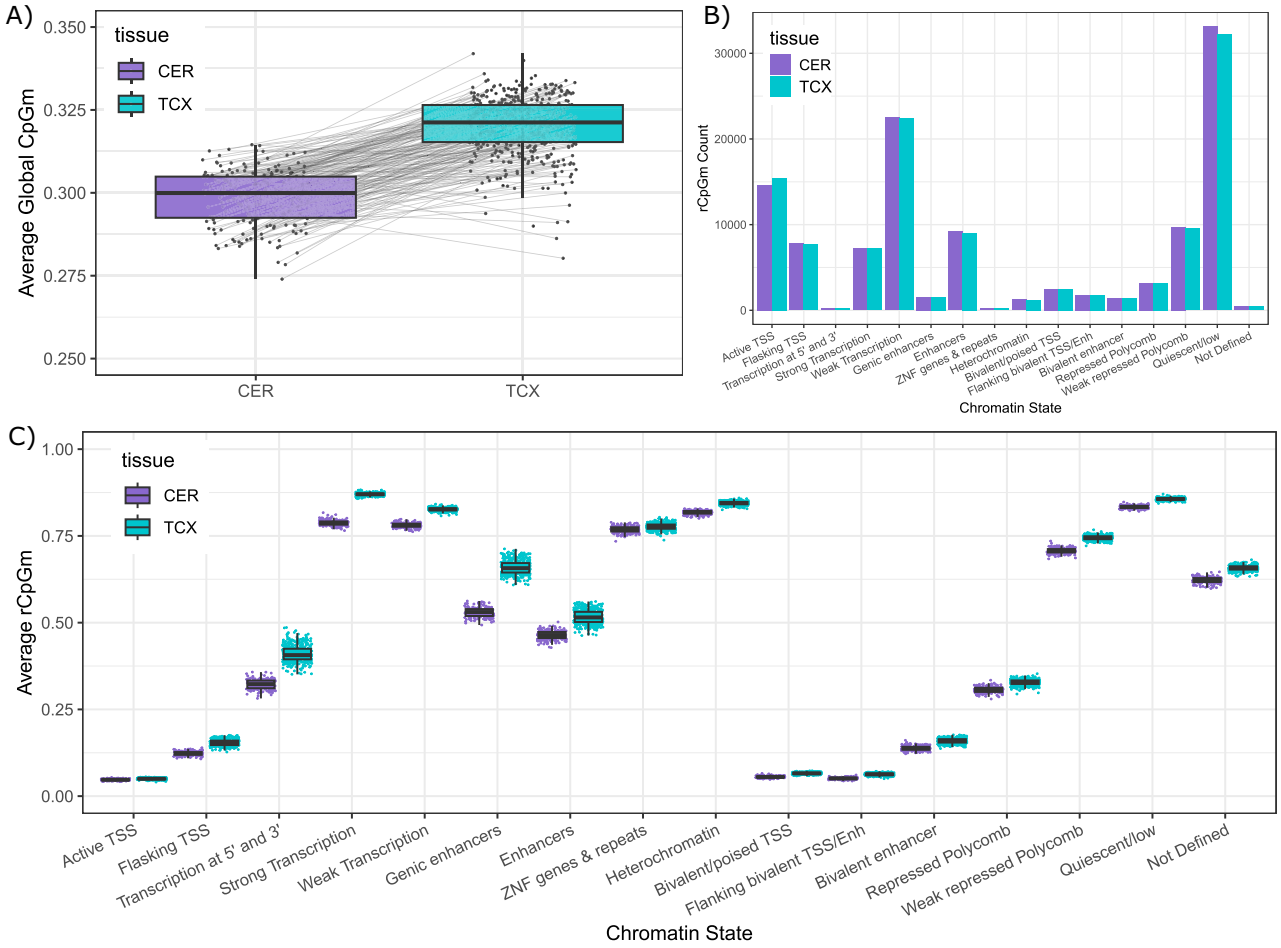


Fig. 2 | Characterization of global DNAm and identified rCpGms in our dataset. **A** Average global CpG methylation calculated for each sample ($N = 191$ CER and $N = 455$ TCX) before data reduction QC steps. Samples with both CER and TCX CpGm measures ($N = 190$) are connected by a gray line. **B** Number of rCpGm sites in the CER and TCX that were annotated to each predicted chromatin state. **C** Average rCpGm for each sample ($N = 191$ CER and $N = 455$ TCX) split by chromatin state and tissue type. Chromatin states annotated as "Not Defined" are the regions that did not lift over from hg19 to hg38 in the Epigenomics Roadmap model and so do not have a defined annotation.; Using two sided *t*-tests, we found the TCX and CER average rCpGm across each annotation were also significantly different from each other ($p < 1E\text{-}05$) as well as the average global CpGm between the CER and TCX. Boxplots show the first quartile, median, and third quartiles with the whiskers indicating ±1.5 times the inter-quartile range from the third and first quartile, respectively. CER Cerebellum; CpGm DNA methylation at CpG sites, DNAm DNA methylation, Enh Enhancer, rCpGm CpG methylation grouped by predicted chromatin state, TCX Temporal Cortex, TSS Transcription Start Sites.

identified 2489 FDR significant TCX CpGm associations with 10 endophenotypes and 50 FDR significant CER CpGm associations only for Thal (Fig. S4, Supplementary Data 4, Supplementary Data 5). The majority (97%) of the FDR significant TCX CpGm associations were with tau biochemical measures ($n = 2415$). The most frequently associated endophenotype was Total tau_{TBS} with 2268 FDR significant TCX CpGms, 21 of which were also associated with $pTau_{TX}$. The remaining FDR significant CpGms were uniquely associated with a single endophenotype. Top TCX CpGm associations for each endophenotype were located in or near the genes *ACSF3* (Total tau_{FA}), *SLC16A1-AS1* (Total tau_{TBS}), *YES1* ($AB40_{TX}$), *TRAPPC14* (Thal phase), *HR* ($pTau_{TX}$), *RGS10* ($APOE_{TX}$), *PLEKHJ1* ($AB42_{TX}$), *CIB1/GDPGP1* ($APOE_{FA}$), *ARPC1B* ($AB42_{FA}$), and *UMODL1* ($APOE_{TB}$)$_S$. Interestingly, CpGms near *ACSF3, SLC16A1, HR, RGS10, PLEKHJ1, ARPC1B* and *UMODL1* have previously been identified in meta-analyses with Braak Stage, although for different CpGm positions[32,52]. *ACSF3, ARPC1B, CIB1, HR, PLEKHJ1, RSG10, UMODL1* also had nominal CpGm associations with CERAD score and Thal phase[32]. Our top 5 CER CpGm associations with Thal phase were near genes *ANKRD33, LINC02314, RAB30, ASIC1*, and *UBE2Q2*. A previous meta-analyses of cortical tissue found CpGms near *RAB30* and *UBE2Q2* to be nominally associated with Braak stage and Thal phase, respectively[32].

Using our novel regional methylation analysis method, we performed EWAS with 116,014 TCX and 116,811 CER rCpGms with the same endophenotypes. We identified 5478 FDR significant TCX rCpGm associations for 5 endophenotypes (Total tau_{TBS}, $pTau_{TX}$, $pTau_{FA}$, $APOE_{TX}$, $A\beta42_{FA}$) (Fig. 3, Supplementary Data 6). There were no FDR significant TCX or CER rCpGm associations with any of the neuropathology endophenotypes (Supplementary Data 6, Supplementary Data 7). Concordant with the individual CpGm findings, the majority of the significant TCX rCpGm associations were with Total tau_{TBS} ($n = 4424$) followed by $pTau_{TX}$ ($n = 1029$). There were 738 rCpGms significantly associated with brain levels of both Total tau_{TBS} and $pTau_{TX}$. All 738 rCpGms had opposite directions of association with Total tau_{TBS} and $pTau_{TX}$, which is consistent with the negative correlation of these two endophenotypes ($\rho = -0.21$, $p < 0.001$)[10]. This suggests that Total tau_{TBS} and $pTau_{TX}$ may have shared but opposing molecular mechanisms that impact their levels in the brain. Each FDR significant rCpGm was on average composed of 23 individual CpGm sites (range: 1–895 CpGms) indicating that most of these associations

https://doi.org/10.1038/s41467-026-68864-9

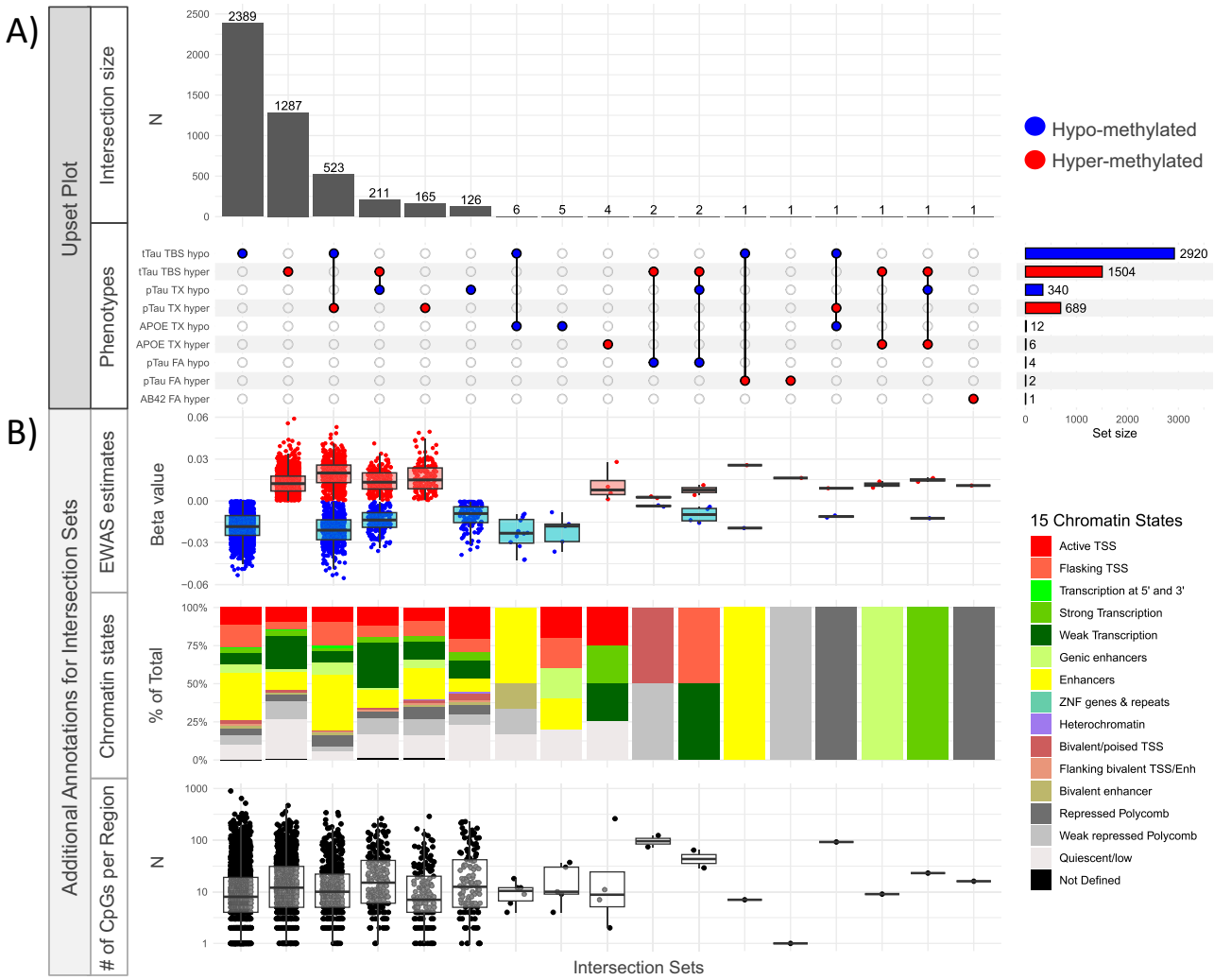


Fig. 3 | FDR significant rCpGm associations in the temporal cortex across AD endophenotypes. A Upset plot showing all FDR significant rCpGm associations across the AD endophenotypes tested in the temporal cortex samples ($N = 455$). **B** Boxplot and bar graphs showing additional annotations for each intersection set of the FDR significant associations from (**A**). EWAS estimates show the regression beta value for each FDR significant association colored by direction of effect. Chromatin states depict the Epigenetic Roadmap 15-chromatin state model annotation breakdown for each intersection set. "Number of CpGs per region" depicts the number of individual CpGms that make up the rCpGm tested. Red indicates hyper-methylation or a positive direction of association and blue indicates hypo-methylation or a negative direction of association with the tested endophenotype. Boxplots show the first quartile, median, and third quartiles with the whiskers indicating ±1.5 times the inter-quartile range from the third and first quartile, respectively. Aβ Amyloid-β, EWAS Epigenome-wide association study, FA Insoluble formic acid fraction, FDR false discovery rate, N Number, pTau phospho-tau, TBS soluble tris buffered saline fraction, TSS Transcription start site, tTau total tau, TX membrane associated Triton-X fraction.

are not driven by a single CpGm but are a composite methylation measure of the region. The majority (69%) of the FDR significant TCX rCpGm associations reside in more active chromatin regions including active TSS (TssA, $N = 578$), Flanking TssA ($N = 652$), transcription at gene 5′ and 3′ ends ($N = 47$), Strong transcription ($N = 175$), weak transcription ($N = 707$), genic enhancer ($N = 256$) and enhancer ($N = 1363$) annotations (Fig. 3) suggesting that these significant rCpGms may be functionally important in gene expression regulation.

To ensure that these FDR significant rCpGm associations are robust to any biological and technical variations, we repeated these analyses after further adjustments for Braak stage and Thal phase, neuronal cell type proportion estimates, oligodendrocyte cell type proportion estimates, and sequencing batch. Comparisons between the primary and further adjusted models by Pearson correlation revealed high, significant positive correlations of both the coefficient estimates ($r = 0.998$-0.9997) and p values ($r = 0.66$–0.989) (Fig. S5). We also investigated if the normal distribution assumptions were violated or if these associations were driven by outliers via a non-parametric Spearman rank correlation test. We found 99.3% (5442/5478) of the FDR significant associations also had a nominally significant ($p \le 0.05$) association in the Spearman rank test (Fig. S5). These highly consistent test statistics between our primary and adjusted models and their persistent significance in a non-parametric test demonstrate the robustness of these associations to potential biological and technical confounders.

To compare our rCpGm approach to the previously published *comb-p* method[45], we applied *comb-p* which groups proximal CpGms into regions based a sliding window approach and association significance. We identified 2362 significant associations (Sidak $p \le 0.05$) in 1833 unique regions by *comb-p* (Fig. S6, Supplementary Data 8). Unlike our rCpGm approach, *comb-p* does not take into account direction of effect when determining significant regions. Further, since *comb-p* defines regions based on arbitrary windows and statistical significance, different regions are identified for each endophenotype EWAS. This hinders a direct comparison of *comb-p* regions across endophenotypes, a feature which is possible by our rCpGm approach. To resolve

the *comb-p* shortcomings on effect direction and regional variability, we defined positive and negative *comb-p* associations by the effect direction for the majority of the individual CpGs making up that region. We considered regions across different endophenotype EWAS to be common if there was any genomic overlap between these regions. *Comb-p* based regional associations largely followed similar trends as the individual CpGm and rCpGm EWAS where Total tau_{TBS} had the highest number of significant associations (70.2%) followed by $pTau_{TX}$ (21.6%) and $APOE_{TX}$ (6.4%). Additionally, the largest overlap of significant associations between endophenotypes were for Total tau_{TBS} and $pTau_{TX}$, all with opposing directions of association, consistent with rCpGm results. There were 1628 regional methylation associations common to *comb-p* and rCpGm, however rCpGm identified 3763 additional associations whereas *comb-p* uniquely identified 677 associations.

Although both the individual CpGm followed by *comb-p* regional grouping and our novel rCpGm analyses found the most associations with brain tau biochemical endophenotypes, there were twice as many FDR significant loci using the latter approach. Of the 2489 FDR significant individual TCX CpGms, 68% were included in at least one of the FDR significant rCpGms. Notably, 1671 individually significant CpGms were part of the 1000 rCpGms associated with Total tau_{TBS}, and 26 such CpGms were within 23 rCpGm associated with $pTau_{TX}$. Taken together, our novel rCpGm approach identifies both those loci captured by the more traditional single CpGms followed by *comb-p* grouping but importantly also many others that are missed by the latter. This suggests that clustering CpGms based on chromatin states is both a more powerful EWAS approach and also provides biological context for these associations. Consequently, we focused our subsequent analyses on the rCpGm results and approach. Importantly, regardless of the CpGm analytic approach, most DNAm across the genome primarily associate with tau protein levels in the AD brain.

Replication of DNAm associations with tau neuropathology and other AD-related endophenotypes in independent datasets

We next sought replication of our findings in the ROSMAP[30] ($n = 686$) and Brains for Dementia Research (BDR)[32] ($n_{DLPFC} = 581$, $n_{OCC} = 576$) cohorts that have DLPFC DNAm measured with the Illumina 450 K Human Methylation and both DLPFC and OCC DNAm measured with the EPICv1 850 K arrays, respectively (Supplementary Data 3). Although to our knowledge there are no DNAm datasets that also have the precise brain biochemical measures of AD-related proteins available in our dataset, ROSMAP and BDR are two of the largest cohorts with both brain DNAm and AD-related endophenotypes.

We already demonstrated the broad similarities of rCpGm levels and patterns for each chromatin state annotation across these datasets and ours (Fig. S3, Fig. 2), despite the fact that the individual CpGm sites measured by their DNAm platforms are not identical. To determine whether overlapping rCpGms that are obtained from different datasets correlate with each other, we performed pairwise correlations (Fig. S7). Using DNAm levels averaged across all samples within each dataset for all rCpGms overlapping with those from the other datasets, we conducted Spearman correlations. There are strong positive correlations for DNAm levels of overlapping rCpGm sites for each pairwise dataset analysis. As expected, these correlations are strongest for BDR DLPFC and BDR OCC datasets (Spearman Coefficient = 0.996) obtained from the same brain donors, followed by BDR DLPFC and ROSMAP DLPFC that have measurements from the same brain region both using similar array technologies (Spearman Coefficient = 0.914) and BDR OCC and ROSMAP DLFPC (Spearman Coefficient=0.918). Nonetheless, overlapping rCpGms from our RRBS-based TCX also had high correlations with those from the other array-based datasets (Spearman Coefficients: 0.74–0.828) (Fig. S7). Thus, while not identical to our DNAm measurements, rCpGms from these other AD brain cohorts could serve as replication datasets for our findings.

For replications, we focused on the 738 TCX rCpGms that are significantly associated with brain levels of both Total tau_{TBS} and $pTau_{TX}$ (Fig. 3). These 738 rCpGms are high confidence both due to their FDR-significant and biologically congruent directions of associations with two brain biochemical AD endophenotypes. Of these 738 rCpGms, 600 were present in at least one of the replication datasets. Although BDR and ROSMAP did not have brain biochemical AD-related protein measurements, they both have Braak stages, which is a measure of tau pathology. To consider an rCpGm as having a biologically congruent association in the replication datasets, the direction of effect would be the same for Braak stage and $pTau_{TX}$ and opposite for Total tau_{TBS}. This is based on prior work[9] and positive correlations of Braak stage with $pTau_{TX}$ (Spearman correlation $\rho = 0.289$, pval = 3.72E-10), but negative with Total tau_{TBS} (Spearman correlation $\rho = -0.30$, pval = 6.29E-11) in our dataset (Fig. S8).

Using these criteria, 93 out of 600 tested rCpGms had nominally significant ($p \le 0.05$) and biologically congruent associations with Braak stage in BDR and/or ROSMAP (Fig. 4, Supplementary Data 9, Supplementary Data 10). Of these 93, 3 replicated in all 3 datasets (tier 1), 23 in 2 of the datasets (tier 2), and 67 in 1 of the replication datasets (tier 3). We also performed a meta-analysis of Braak stage including the RRBS-TCX, BDR-DLPFC, and ROSMAP-DLPFC datasets and found that 47 of the 93 replicated rCpGms were significant (fixed $p \le 0.05$) including all of tier 1 rCpGms, 16/23 tier 2 rCpGms, and 28/67 tier 3 rCpGms (Fig. 4, Supplementary Data 11). We further tested the association of these rCpGms with Thal phase and AD diagnosis in the replication datasets and also found biological congruence of associations across these additional AD measures (Fig. 4, Supplementary Data 9, Supplementary Data 10). In other words, rCpGms with FDR-significant negative associations with Total tau_{TBS} and positive associations with $pTau_{TX}$ in our dataset, had positive associations with Braak, Thal and AD diagnosis in one or more of the replication datasets and vice versa.

When we annotated these 93 rCpGms for proximal (±1 Mb) variants identified in AD genome-wide studies[53], we found that a number of them reside near known AD risk loci including *JAZF1, SIGLEC11, SCIMP, KLF16, BIN1, SPI1, ABI3, ANKH, WDR12, and SIGLEC11*[53] (Supplementary Data 12). Collectively, these results support independent replication of our 93 high confidence TCX rCpGms for their association with brain tau-related and other AD endophenotypes, which may also have implications for AD risk.

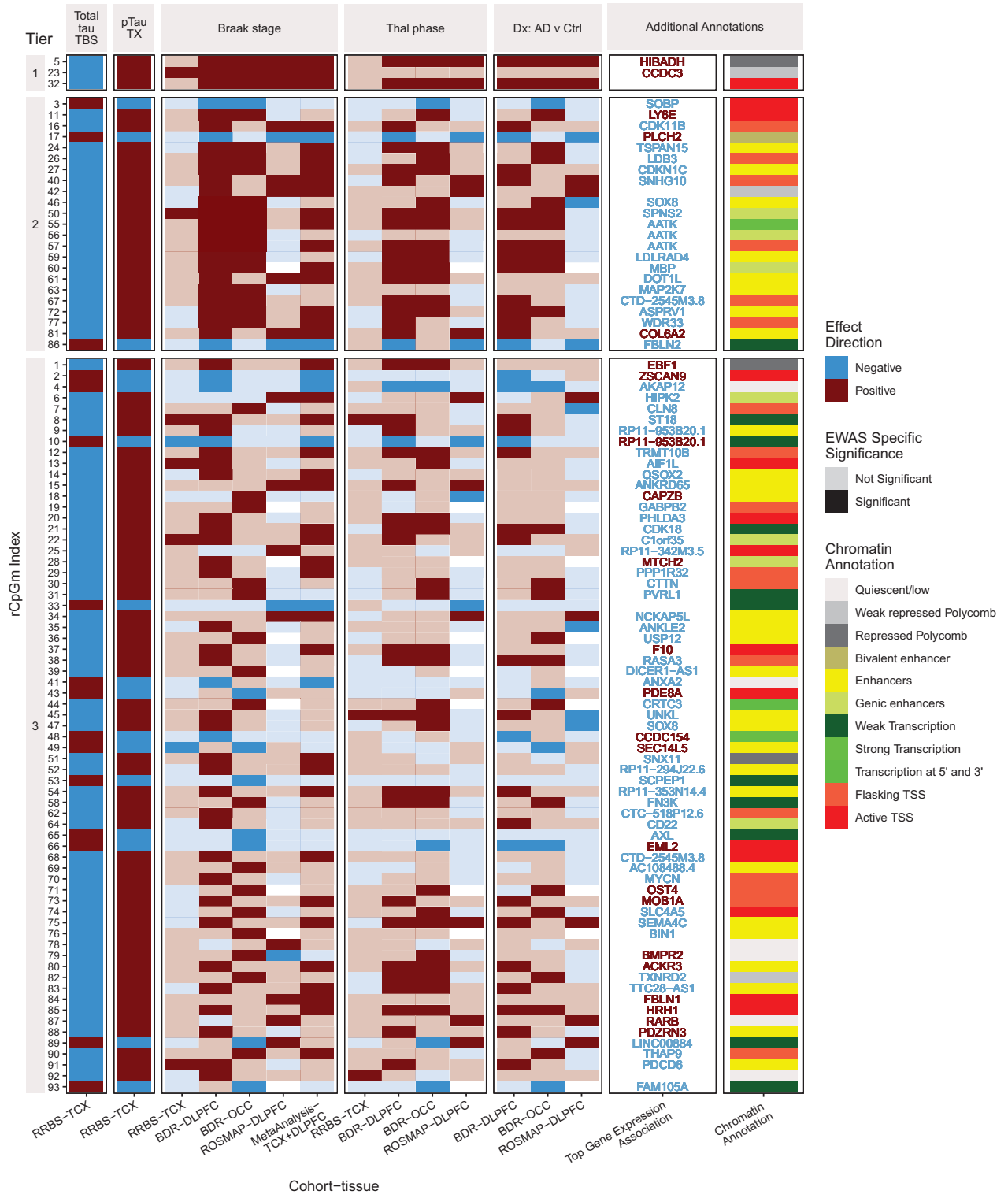
Tau related rCpGms associate with expression of nearby oligodendrocyte and myelin related genes

Most of the FDR significant AD endophenotype associated rCpGms in our dataset are located in active chromatin regions (Figs. 3, 4). To further characterize the associations between these rCpGms and brain gene expression levels, we leveraged bulk CER and TCX RNAseq data from the same individuals. After RNAseq QC, we paired rCpGms to genes ±500 kb from the start and end of each rCpGm yielding 1,612,499 TCX and 1,834,887 CER *cis*-rCpGm-gene pairs. We then tested the association between *cis*-rCpGm-gene expression pairs using linear regression.

We identified 12,973 TCX and 1413 CER FDR significant *cis*-rCpGm-gene pairs (FDR $p \le 0.05$) (Supplementary Data 13, Supplementary Data 14). Many of the significant *cis*-rCpGm-gene pair associations had rCpGms annotated to expression regulation terms including active TSS, flanking active TSS, and enhancer sites (51% in TCX and 32% in CER). Genes previously reported to have AD-related methylation changes nearby, including *RHBDF2*[30,31], *IRF8*[31], *CCND1*[31], and *SPI1*[32] had both FDR significant rCpGm associations with AD endophenotypes and *cis*-rCpGm-gene expression associations in our dataset (Supplementary Data 13, Supplementary Data 14).

To further annotate the 93 high-confidence tau and other AD-endophenotype associated rCpGms, we tested their corresponding

https://doi.org/10.1038/s41467-026-68864-9



1464 *cis*-rCpGm-gene expression pairs (1225 unique genes). Many of these 93 rCpGms reside in more active chromatin regions such as Active TSS, Flasking TSS, or Enhancers, suggesting their functional consequences in gene expression (Fig. 4). Indeed, when tested for their effects on brain gene expression, these rCpGms had FDR significant associations with nearby genes measured from the same donors (Supplementary Data 12–14). There were 535 significant *cis*-rCpGm-gene expression associations (459 unique genes) which included previously implicated AD genes like *BIN1*[23,53,54], *ANXA2*[55,56], CDK18[55,57,58], *AKAP13*[59], *RORA*[60,61], *KIFC1*[62], *and KLK6*[63,64](Supplementary Data 12).

To determine if this set of 459 genes, which correspond to the 93 tau-related high confidence rCpGms, implicates common pathways or cell types, we performed gene set enrichment analysis. The top three nominally enriched gene ontology terms for these genes relate to positive regulation of myelination (GO:0031643, p= 3.4E-5), gliogenesis (GO:0042063, p = 1.4E-4), and microtubule nucleation (GO:0007020, p = 1.5E-4) (Supplementary Data 15). Consistent with these GO terms, this gene set was also enriched for oligodendrocyte marker genes. Using the top 1000 BRETIGEA[65] cell type markers genes, there were 45 oligodendrocyte genes amongst these 459 (p = 9.8E-05),

Fig. 4 | Replication of FDR significant brain tau-associated rCpGms in independent datasets. The y-axis depicts the 93 rCpGms that are FDR significant (FDR $p < 0.05$) in the Mayo Clinic RRBS-TCX total tau_{TBS} and $pTau_{TX}$ EWAS and have opposite directions of associations with these tau endophenotypes. These rCpGms are nominally significant ($p < 0.05$) and have congruent direction of associations with Braak stage in ROSMAP and/or BDR through linear regression tests. Tier 1 rCpGms ($N = 3$) are associated with Braak stage in all three replication datasets (BDR-DLPFC, BDR-OCC, and ROSMAP-DLPFC); Tier 2 ($N = 23$) are associated with Braak stage in two of the replication datasets; Tier 3 ($N = 67$) are associated with Braak stage in one of the replication datasets. In addition, a meta-analysis of Braak stage was conducted including the RRBS-TCX, BDR-DLPFC, and ROSMAP-DLPFC datasets. Meta-analysis fixed-effect estimates and p values are reported in Supplementary Data 11. Also included are additional linear regression associations of the rCpGms with Thal phase and AD diagnosis (Dx). rCpGm associations with Braak stage, Thal phase, and diagnosis are considered significant if the p value is <0.05. Non-colored associations in ROSMAP indicate that this region was not measured on the array. Each rCpGm is additionally annotated with the top FDR significant gene expression association (GE - rCpGm) colored by direction of effect and by their chromatin state annotation from the Roadmap Epigenomics 15 chromatin state model. See Supplementary Data 12 to connect rCpGm Index numbers with rCpGm positional data. Red indicates a positive direction of effect, blue indicates a negative direction of effect, and the color transparency indicates if the association passed nominal significance threshold. BDR Brains for dementia research, CpGm DNA methylation at CpG sites, DLPFC Dorsolateral prefrontal cortex, Dx Diagnosis, EWAS Epigenome-wide association study, FDR false discovery rate, GE Gene expression, N Number, OCC Occipital cortex, pTau phospho-tau, rCpGm Regional DNA methylation grouped by chromatin state, RRBS Reduced representation bisulfite sequencing, TBS soluble tris buffered saline fraction, TCX Temporal cortex, TX membrane associated Triton-X fraction.

11 of which were also in the top 100 BRETIGEA[65] markers ($p = 4.0E-05$). These top 11 oligodendrocyte marker genes with rCpGm associations are *MAG, KLK6, ST18, MYRF, LDB3, MBP, CD22, TGFA, RTKN, SOX8*, and *CLMN*. We next investigated these 11 genes and an additional gene *CKD18* further in both our and external datasets. Although part of the 1000, and not the 100 top oligodendrocyte marker genes[65], CKD18 was previously implicated in tau phosphorylation including at the T231 site and tau aggregation[58,66], found contained within NFT[58], differentially expressed in TCX tissue of individuals with AD compared to controls[58], associated with AD DNAm changes[67], and helps prevent DNA damage and promotes genomic stability[68], thereby leading us to include this gene in additional analyses.

All 12 oligodendrocyte genes have negative associations of their brain expression and paired rCpGm methylation levels (Fig. 5). There were 16 rCpGm-gene expression pairs for these 12 oligodendrocyte genes and 15 unique rCpGms. All 15 rCpGms had FDR significant negative associations with Total tau_{TBS} and positive associations with $pTau_{TX}$ (Fig. 5). These rCpGm - tau endophenotype associations remained significant even after adjusting for oligodendrocyte cell proportion estimates (Supplementary Data 16), indicating that they are not driven by changes in proportions of oligodendrocytes in bulk brain tissue. These rCpGms had additional nominally significant associations with other AD-related endophenotypes, the most consistent of which were negative associations with $APOE_{TX}$ levels and positive associations with $pTau_{FA}$ (Fig. S9). Almost all of the individual CpGms that comprise these oligodendrocyte rCpGms have the same direction of association with Total tau_{TBS} or $pTau_{TX}$ as their corresponding rCpGm, although few individually reach FDR significance (Fig. S10). This again underscores the superior power of our novel and biologically relevant rCpGm in comparison to individual CpGm analysis in detecting DNAm associations.

We further evaluated the top oligodendrocyte genes with tau-related rCpGm associations using our dataset and independent studies. Within the same brain donors, TCX expression levels of these oligodendrocyte genes also had inverse directions of associations with TCX levels of Total tau_{TBS} and $pTau_{TX}$. Of these 12 top oligodendrocyte genes, 5 nominally associated with both Total tau_{TBS} and $pTau_{TX}$ levels (*CDK18, LDB3, MBP, CD22, CLMN*) and 2 with $pTau_{TX}$ (*KLK6, ST18*) in the TCX (Fig. 5A, Supplementary Data 17). All but one of these associations (*TGFA*- $pTau_{TX}$) had directions of effect that are biologically congruent with both rCpGm-tau and rCpGm-gene associations. Notably, *TGFA*, unlike the other marker genes, is not only expressed in oligodendrocyte cells but is also expressed in oligodendrocyte precursor cells which may underlie this discrepancy. Upon adjustment for oligodendrocyte cell proportion estimates, *CDK18, MBP*, and *CD22* remained nominally associated with both Total tau_{TBS} and $pTau_{TX}$ levels with consistent effect directions (Supplementary Data 17).

Further investigation of the 15 rCpGms in Illumina 450k DNAm data derived from NeuN+ and NeuN- sorted occipital cortex nuclei[36], revealed 14 rCpGms with positive associations with AD in the glia enriched NeuN- population while only 6 rCpGms had positive associations with AD in the neuronal enriched NeuN+ population (Fig. S11). The regression estimate magnitudes of the glia enriched NeuN- population were also larger than the NeuN+ population suggesting that the associations identified in the bulk DNAm data are likely originating from DNAm changes in the glial cells rather than neuronal cells[32].

Using three independent brain bulk RNAseq datasets of AD and control brains, we next investigated the association of these top oligodendrocyte genes with Braak stage in the Mayo TCX[69], Mount Sinai BM22[70], and ROSMAP[71] datasets from the AMP-AD study[72]. Expression of *LDB3* was associated with Braak stage in the Mayo TCX (beta = −0.19; FDR $p = 2.54E-2$) and ROSMAP (beta = −0.058; FDR $p = 2.52E-2$) datasets with a biologically congruent negative direction of effect (Supplementary Data 18). Using AMP-AD data, *MYRF* and *CDK18* have also been nominated as potential therapeutic targets for AD[73], providing additional support for a role of these oligodendrocyte genes in tau-related outcomes.

To determine whether these brain rCpGm-oligodendrocyte gene-tau associations are unique to AD vs. conserved across other tauopathies, we investigated available data from methylation or gene expression studies of progressive supranuclear palsy (PSP) and Pick's disease. Further, we analyzed oligodendrocyte cell clusters from published brain single nucleus datasets of tauopathies to establish these associations in the cell type of interest. There was one brain DNAm study conducted from the prefrontal cortex of PSP and control donors[74] to which we applied our QC, rCpGm method and conducted an EWAS of PSP risk and Braak stage (90 PSP and 66 controls) (Supplementary Data 3 and 19, Fig. 5B). We determined biologically congruent associations with Braak stage across the 12 top oligodendrocyte rCpGms that also had consistent PSP risk associations for all but three of the genes (*TGFA, ST18, CDK18*). This finding supports a model where rCpGm-oligodendrocyte gene-tau associations are not unique to AD but may be a shared disease mechanism across tauopathies.

We next analyzed all available single nucleus RNAseq (snRNAseq) data from sizeable brain datasets of AD and other tauopathies[75–81] to replicate the top oligodendrocyte gene expression associations with disease risk and tau endophenotypes in an oligodendrocyte cell-type specific manner. These independent snRNAseq datasets and brain regions were from Is et al.[75] (24 TCX, AD and controls), Mathys et al. (427 prefrontal cortex, AD and controls)[76], Gamache et al. (24 TCX, AD and controls)[77], Morabito et al. (18 prefrontal cortex, AD and controls)[78], Anderson et al. (15 dorsolateral prefrontal cortex, AD and controls)[79], Rexach et al. (41 insular cortex, AD, PSP, Pick's disease and controls)[80], and Min et al. (34 TCX, PSP and controls)[81]. In addition to these snRNAseq data from a total of 583 independent brain samples from multiple regions and diagnoses, Min et al.[81] also had bulk RNAseq

https://doi.org/10.1038/s41467-026-68864-9

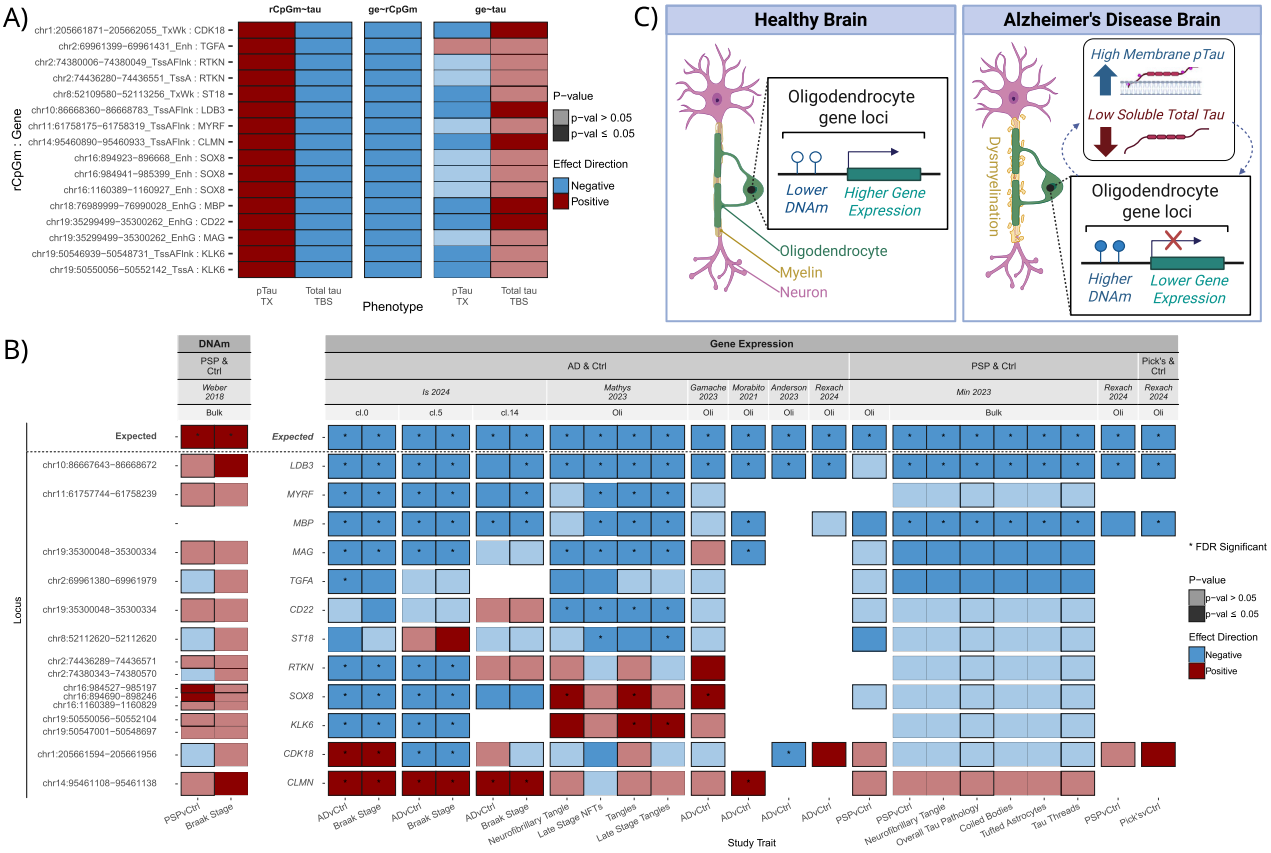


Fig. 5 | Top rCpGm and oligodendrocyte gene loci consistently associate with tau-related phenotypes across studies. A Linear regression association results of the top oligodendrocyte specific rCpGm-gene pairs related to TCX tau levels in our dataset across three models: (1) rCpGm ~ tau phenotype, (2) Gene expression ~ rCpGm, (3) Gene expression ~ tau-phenotype. **B** Study specific association results of top tau-related rCpGm or oligodendrocyte gene expression levels to tau-related endophenotypes (bottom x-axis) from recently published studies including Weber 2018 ($N = 90$ PSP, $N = 66$ Control), Is 2024 ($N = 12$ AD, $N = 12$ Control), Mathys 2023 ($N = 238$ AD, $N = 189$ non-AD), Gamache 2023 ($N = 12$ AD, $N = 12$ control), Morabito 2021 ($N = 11$ AD, $N = 7$ control), Anderson 2023 ($N = 7$ AD, $N = 8$ Control), Rexach 2024 ($N = 10$ AD, $N = 11$ PSP, $N = 10$ Pick's, $N = 10$ Control), and Min 2023 (snRNAseq: $N = 18$ PSP, $N = 16$ Control; Bulk: $N = 281$ PSP, $N = 127$ Control). Top X-axis is faceted first by whether DNAm or gene expression was tested, second the disease type examined, third the study, fourth whether bulk data, pseudobulk oligodendrocyte clusters from single-cell/nuclei data or specific oligodendrocyte assigned clusters were analyzed. The first row "Expected" depicts expected association direction based on results from our dataset. Locus annotations on the y-axis include the rCpGms defined by CpGms present in the Weber et al.[74] study that overlap with our TCX rCpGms of interest in (**A**). White areas indicate a result for that association was not reported from the study. **C** Hypothesized relationship between TCX tau levels, DNAm, and oligodendrocyte gene expression in the AD brain. **C** created in BioRender.com. Red = positive associations, blue = negative associations, * = FDR significant, nominal p values shown. abs absolute value, AD Alzheimer's Disease, cl. cluster, Ctrl Control, DNAm DNA methylation, FDR False discovery rate, ge Gene expression, Oli Pseudobulk Oligodendrocytes, PSP Progressive Supranuclear Palsy, pTau phospho-tau, p-val p value, rCpGm Regional DNA methylation grouped by chromatin state, TBS soluble tris buffered saline fraction, tTau total tau, TX membrane-associated Triton-X fraction.

data from 408 PSP and control TCX samples. We tested the top brain rCpGm associated oligodendrocyte genes for their oligodendrocyte-specific expression associations with risk for AD[75–80], PSP[80,81], Pick's disease[80], tau-related (Fig. 5B) and other available endophenotypes (Fig. S12) in these datasets. Despite differences in the cohorts, brain regions and diagnoses, there was highly consistent association for oligodendrocyte-specific expression for many of these top 12 genes with these outcomes in a direction that would be expected based on the rCpGm-oligodendrocyte gene-tau associations we identified. Most of these genes had lower expression levels in AD and other tauopathies and were negatively associated with tau-endophenotypes such as Braak stage and NFT. Most consistent associations were observed for *LDB3, MYRF, MBP, MAG, TGFA*, and *CD22* loci where the implicated rCpGms positively associate with $pTau_{TX}$, negatively with Total tau_{TBS} and brain gene expression in our data (Fig. 5A, Supplementary Data 16–18). These rCpGms associate positively with PSP risk and Braak in Weber et al.[74] DNAm data (Fig. 5B, Supplementary Data 19). Further, oligodendrocyte cluster specific expression of these genes associate negatively with tauopathy risk and tau traits[75–81] (Fig. 5B, Fig. S12).

To determine whether genetic variants could account for the tau-related DNAm changes observed for these top oligodendrocyte gene related rCpGms, we performed a *cis*-methylation quantitative trait locus (mQTL) analysis between these rCpGms and imputed genetic variant dosages within a distance of ±500 kb. There were 160 FDR significant SNP-rCpGm associations (Supplementary Data 20). The rCpGms implicated by the mQTL associated with *ST18* (14 SNP-rCpGm pairs), *MBP* (122), *CLMN* (2), *MAG* & *CD22* (2), *KLK6* (1), and *RTKN* (19) gene expression. Of these 160 unique SNPs, 13 near *MBP* nominally ($p \le 0.05$) associated with Total tau_{TBS} levels, the top 2 of which were also trending towards association ($p < 0.052$) with $pTau_{TX}$ levels (rs13381407, rs12457073: $R^2 = 1$, $D' = 1$)[9] (Supplementary Data 20), but none associated with AD risk[53]. These results suggest that genetic variation is unlikely to account for the rCpGm-gene-tau relationships at these loci, with the possible exception of the *MBP* locus.

Collectively, these findings support a model for tauopathies such that with increasing brain levels of tau neuropathology, $pTau_{TX}$, and decreasing Total tau_{TBS}, DNA hyper-methylation occurs in oligodendrocyte/myelin gene regions leading to reduced expression of these genes and potentially downstream oligodendrocyte dysfunction and dysmyelination (Fig. 5C). However, it is equally plausible that other untested genetic or environmental factors lead to DNA hyper-methylation, subsequently leading to downregulated myelin genes and oligodendrocyte dysfunction, culminating in brain accumulation of toxic tau forms[82].

We have compiled our findings and data in our interactive web application, Multiomic Atlas of AD Brain Endophenotypes. Our web application has the results from the following analyses: *GWAS:* Association between genetic variant dosages and TCX brain biochemical measures. *rCpGm EWAS*: Association between regional CpGm (rCpGm) levels and neuropathology or TCX brain biochemical measures. *Individual CpGm EWAS:* Association between individual CpGm levels and neuropathology or TCX brain biochemical measures. *mQTL:* Association between regional CpGm (rCpGm) levels and nearby genetic variant dosages (+/- 500 kb). *rCpGm to GE:* Association between regional CpGm (rCpGm) levels and nearby gene expression measures, where rCpGms were tested for association with genes +/- 500 kb from start and end of region. These data can be searched by gene names or chromosomal locations, and results are presented as both tables and interactive plots.

Discussion

AD is a complex and heterogeneous disease with variability not only in its clinical presentation but also neuropathological features and brain biochemical levels of AD-related proteins[5–8] like Aβ, tau, and apolipoprotein E (APOE). Capturing the genomic basis of this variability can reveal novel molecular mechanisms of AD and identify causal factors that underlie disease heterogeneity. These are essential to devise precision therapies for this neurodegenerative disease much like is being done in other complex and heterogeneous diseases, like cancer[83]. We[84] and others[85] have utilized deep AD neuropathology endophenotypes in GWAS to uncover genetic variants associated with variability of these neuropathologic measures. Previous brain methylation studies in AD focused on associations with clinical features of this disease[30–32,38,39], with fewer studies focusing on AD neuropathology[32,86]. We previously identified genetic variants that may influence variability in brain biochemical measures of AD-related proteins[9], demonstrating their significant potential as deep endophenotypes in AD genomic studies. However, to our knowledge, there are no studies that previously explored the combined genetic, epigenetic and transcriptomic landscape of both AD neuropathology and brain biochemical measures.

This study employed a deep endophenotyping approach to identify important molecular mechanisms of DNAm that influence levels of neuropathology and biochemical states of core AD related proteins in AD brain tissue. By leveraging the Roadmap Epigenomics 15-chromatin state model to group individual CpGm sites into rCpGms, we established a novel methodology to investigate grouped CpGms in a biologically relevant manner increasing our ability to identify robust associations with these endophenotypes. Through our innovative methodology to investigate regional CpGm, EWAS, and integrative analyses, we found that although all the tested endophenotypes are core to AD pathophysiology, each have distinct CpGm architectures underlying their variability in the AD brain. This establishes the role of epigenetic mechanisms in influencing heterogeneity of these AD brain endophenotypes (Fig. 1).

We found highly consistent and biologically congruent patterns of average rCpGm levels across three independent datasets from Mayo Clinic (this study), ROSMAP[30] and BDR[32], collectively comprising 2515 samples representing four different brain regions (TCX, CER, DLPFC, OCC) (Fig. 2, Fig. S3). The consistent similarities in the rCpGm levels and patterns, despite variabilities in cohort, brain region, DNAm measurement technologies and methylation coverage in these cohorts underscores the strong biological underpinning of our novel rCpGm method and supports its broad utility in analyzing regional levels of DNA methylation across measurement platforms, tissue types, and disease states.

Using our rCpGm grouping approach, we identified 5478 FDR significant TCX rCpGms where the majority (99.7%) associated with brain tau levels, i.e., Total tau_{TBS} and/or $pTau_{TX}$ (Fig. 3). We found 738 rCpGm associations with shared but opposite directions of effect between Total tau_{TBS} and $pTau_{TX}$ suggesting that there are shared DNAm related mechanisms that inversely impact levels of soluble total tau and membrane-bound phosphorylated tau (pTau) in the AD brain. These findings may have broader implications for therapeutic development as targeting these shared pathways may influence levels of both brain Total tau_{TBS} and $pTau_{TX}$. Conversely, similar to previous AD CpGm studies, we found no FDR significant associations of CER rCpGm with neuropathology[31,39,87], although there were some individual CER CpGm associations, but to a much lesser extent than those for TCX CpGm. This may suggest that CER, which has less AD neuropathology, may consequently have less DNAm perturbations. Together, these results demonstrate the effectiveness of using a deep endophenotyping approach combined with integrative multi-omics to identify important and phenotype specific molecular mechanisms that likely contribute to the pathophysiology of AD.

The finding that the largest number of significant brain rCpGm associations occur with brain tau related measures suggests that epigenetic dysregulation may occur more within the context of tau rather than amyloid in the AD brain. This parallels a previous epigenetic study which suggested that large-scale changes in chromatin states marked by H3K9ac, a typical open chromatin marker, occur as a result of tau rather than Aβ burden[28]. Interestingly, these tau related epigenetic associations juxtapose our previous GWAS findings in this same dataset where we identified that genetic variants associate with brain amyloid rather than tau proteins[9]. In sum, these findings may suggest a model where distinct AD-related proteins in the brain may be associated with different genomic perturbations. Specifically, genetic variants primarily drive brain biochemical measures of Aβ, whereas brain methylation changes associate with brain tau levels. These results demonstrate how integrating multi-omics measures with distinct deep AD phenotypes can begin to unravel the molecular basis of different facets of AD, although the precise genomic mechanisms underlying this distinction remain to be uncovered.

The regional DNAm associations we found with brain tau endophenotypes in our Mayo Clinic cohort were replicated in the two large, independent datasets of AD and control brains from the ROSMAP[30] and BDR[32] cohorts (Fig. 4). As these replication cohorts do not have brain tau biochemical measurements, we used Braak stage as a tau neuropathology endophenotype. Remarkably, we found that of the 600 tau-related rCpGms in our cohort which were also available from these replication datasets, 93 showed significant and biologically congruent associations in BDR and/or ROSMAP of which half were also significant in a Braak stage meta-analysis. Ten of these replicated rCpGms are located near (±1MB) known AD risk GWAS loci[53] suggesting an epigenetic mechanism in conferring AD risk at these loci. Characterization of the 93 replicated rCpGms by integrating matched brain gene expression measures from the same brain donors revealed a likely functional role of these rCpGms on expression of nearby genes. Many of these genes also have previous connections with tau, such as *BIN1* which is known to interact with tau proteins in extracellular vesicles[88,89] including tau Thr231[90] which is the pTau species measured in our study, *ANXA2*[55,56], as well as *CDK18*[57,58] and *AKAP13*[59] which can phosphorylate tau. Nonetheless, future functional experiments will

need to verify the casual direction of effect between these rCpGms and their related genes.

Further interrogation of the genes associated with the tau-related rCpGms revealed enrichment in oligodendrocyte marker and myelination-related genes and GO terms. Importantly, higher brain levels of $pTau_{TX}$, which is considered to be a toxic tau species[9,91] is associated with hypermethylation at the rCpGms near the oligodendrocyte genes and biologically consistent lower brain gene expression levels for many of them. Downregulation of oligodendrocyte gene expression in AD brains and other tauopathies have previously been reported by our group[81,92,93] and others[94]. Our integrated brain methylation and gene expression results reported here suggest that the transcriptional downregulation of oligodendrocyte and myelin related genes in tauopathies may be mediated by pTau related CpG hypermethylation near these genes. However, other untested factors may also be driving CpG hypermethylation, resulting in oligodendrocyte gene expression reductions and dysmyelination, which culminates in toxic tau propagation[82].

We further characterized and validated the rCpGm and gene expression associations for these oligodendrocyte genes using additional, independent data generated by both our group[75,81] and others[74,76–80]. Considering our Mayo Clinic discovery, ROSMAP[30] and BDR[32] replication cohorts in this study together with these additional validation cohorts, these top oligodendrocyte genes were assessed in 2671 rCpGm, 1080 bulk RNA and 583 snRNAseq brain sample measurements representing 8 different brain regions (Fig. 5). Of the oligodendrocyte genes thus characterized, *LDB3, MBP*, *MAG*, and *MYRF* had the most consistent brain rCpGm, bulk or single nucleus gene expression associations with disease risk or tau endophenotypes across AD and other tauopathy cohorts.

LDB3 encodes LIM Domain Binding 3 (LDB3) which is a protein that interacts with the cytoskeleton and plays a role in targeting and clustering membrane proteins often studied in muscle tissues, but it is also expressed in the central nervous system[95,96] and oligodendrocytes in the brain[97] (proteinatlas.org/ENSG00000122367-LDB3)[98]. Although genetic variants in *LDB3* are often implicated in myopathies[95,99,100], there are suggestive associations with AD risk in Caribbean Hispanics[101,102] and differential expression in AD[77,103]. There is also evidence that *Ldb3* decreases in a rat model of demyelinating epilepsy[104] and expression of *LDB3* is downregulated in Huntington's disease human brain tissue and mouse models[105]. In our study, brain *LDB3* gene expression and its rCpGm are associated with brain tau endophenotypes, AD and other tauopathies. Like *LDB3*, tau is a cytoskeletal protein that interacts with and likely stabilizes microtubules[91,106]. This raises the possibility that these two proteins may interact through cytoskeletal activity and collectively contribute to maintenance of oligodendrocyte and neuronal activity in health or their demise in the presence of toxic tau species (Fig. 5C). In addition, *MBP, MAG*, and *MYRF* encode important factors for myelination. Myelin basic protein (MBP) and myelin associated glycoprotein (MAG) are major components of myelin[107], and myelin regulatory factor (MYRF) is a transcription factor regulating oligodendrocyte differentiation and myelin-related gene expression[108]. Previous cell culture models showed that knockdown of tau mRNA led to a decrease in protein levels of MBP, disruption of microtubule formation, and disturbance of myelin sheath formation[109]. This finding is aligned with our results suggesting that Total tau may have a protective role in maintaining oligodendrocyte epigenetic and transcriptional homeostasis, whereas pTau has the opposite, toxic effect (Figs. 3–5).

A relationship between myelination and tau is supported by many previous studies by our group and others[109–115] including downregulation of myelination related gene expression networks in AD as well as the primary tauopathy of PSP. There are white matter abnormalities and myelin loss in human AD patients[116–118] and mouse models of tauopathy[119–121]. Further, higher levels of tau-PET correlate with lower myelin levels[122]. Importantly, oligodendrocytes can propagate tau pathology in the absence of neuronal transmission, suggesting a potential role of the myelin sheath in pathologic tau propagation[123]. Primary tauopathies, like PSP, can present with pathological tau inclusions in oligodendrocytes[124]. Previous studies of mice expressing human mutant tau (P301L) exclusively in oligodendrocytes found evidence that oligodendrocytic tau inclusions preceded myelin and neuronal axon disruption and that they may in part cause neurodegeneration[125].

Our epigenome-wide association study of deep brain AD protein biochemical levels, coupled with our novel rCpGm approach, and integrated transcriptome and epigenome data of donors from >4000 collective brain samples demonstrate strong, replicable and biologically consistent associations between brain tau and hypermethylation near oligodendrocyte genes possibly leading to their lower expression, dysmyelination and neurodegeneration in AD and other tauopathies (Fig. 5C). However, functional experimental studies are needed to decipher the actual sequence of events in this cascade. While future studies are needed to precisely delineate the functional mechanisms behind these results, previous studies also provide evidence for these conclusions. Expression levels of *BIN1*, an AD risk gene[23,53,54], associate with AD and there is loss of *BIN1* expression and demyelination in multiple sclerosis lesions[126]. *BIN1* also emerged as a gene with brain tau, rCpGm and expression associations in this study. Others have shown hypermethylation of individual CpGms near *MBP* in AD brain[37,52,86,127] and plasma[128] as well as in multiple sclerosis lesions[129], and suggested that Braak associated CpGms in bulk brain tissue likely reflect changes in glia cells including oligodendrocyte enriched populations[32]. Moreover, the importance of DNAm in oligodendrocyte differentiation and function has been appreciated where disruption of normal DNAm mechanisms can cause oligodendrocyte and myelin dysfunction[130,131]. These observations support the connection between DNAm and tau in oligodendrocyte function and may open up potentially important opportunities for therapeutic interventions for neurodegenerative diseases[132] targeting correction of disrupted myelination as well as DNAm mechanisms.

Our study has multiple strengths including our discovery[9] and replication[30,32] cohorts of AD and control donors with DNAm measures from four brain regions of 2515 samples which we analyzed using our novel rCpGm analytic approach. Mayo Clinic discovery cohort has deep endophenotypes of AD neuropathology and precise biochemical measures of core AD proteins. In this cohort, we generated DNAm measures via RRBS, which allowed us to investigate close to two million high confidence CpGm sites, more than doubling the number of sites that can be investigated by the frequently used array platforms. We developed a novel, biologically relevant methodology to investigate DNAm on a regional, rather than individual CpGm level. We determined that our regional grouping method can be applied to DNAm data generated with different technologies, yielding highly consistent rCpGm levels across independent datasets and tissue types demonstrating broad utility of our innovative analytic approach. Additionally, investigating DNAm grouped by chromatin regions[48], rather than individual CpGm or regions grouped by arbitrary sliding windows and significance as with *comb-p*[45], improves identification of biologically relevant results. We integrated additional omics measures from the same donors including brain gene expression and genotype data and identified 93 rCpGms corresponding to 535 genes that have replicable and biologically congruent associations with brain tau levels. These associations are enriched for myelination pathways and genes, which we further validated using additional brain methylation, bulk and single nucleus expression data from AD and other tauopathy cohorts totaling >4000 samples. Our results uncovered opposite roles for brain Total tau (likely beneficial) and pTau (detrimental) and support two hypothetical models, one where brain tauopathy influences rCpGm resulting in hypermethylation and reduced expression of

oligodendrocyte genes, and another where reduced expression of oligodendrocyte genes from rCpGm hyper-methylation disrupts tau homeostasis, both of which could culminate in dysmyelination. Importantly, we built an interactive application tool Multiomic Atlas of AD Brain Endophenotypes through which we share our findings and data.

Despite these strengths, there are some limitations to our study. As with many DNAm measurement approaches, RRBS cannot differentiate between 5mC and 5hmC methylation, with the latter found at higher levels in central nervous system tissue types compared to other tissue types[133,134]. While this should not affect our highly replicable findings, we may have missed some specific methylation associations of either 5mC or 5hmC, as they can have opposite functional consequences[135–139] including in AD[29,140–143]. Additionally, we extracted CER DNA using a manual rather than the automated approach used for TCX DNA, to achieve similar quality DNA from both tissue regions. Although our pilot studies (see "Methods") demonstrate no significant contributions to CpGm variation by DNA extraction method, we cannot entirely rule out this potential for all CpGm sites. Our study is conducted on bulk brain tissue, and individual cell types can have different CpGm compositions[32,144]. Although the associations for the top hits we identified remained even after adjusting for neuronal or oligodendrocyte proportions and showed consistent and larger magnitude associations in glial enriched nuclei, the marker-based deconvolution methods we used to estimate cell type proportions have limitations, particularly in tissue impacted by pathology. As a result, future single cell type specific DNAm studies, particularly those from glial cell types[32], will be important in parsing out potential cell type specific rCpGm effects or the role of cellular heterogeneity in these associations. In addition, the results of this study are based on cross-sectional data at the time of autopsy and future longitudinal blood or CSF DNAm studies and functional experimentation investigating our top hits will be important in understanding the temporal relationship between DNAm, gene expression changes and tau pathology. Moreover, our investigation of DNAm regulating gene expression was restricted to nearby genes, so it is possible we may have missed important longer range regulatory networks which future studies may be able to identify by incorporation of chromatin conformation data. Similarly, future large scale mQTL studies would be beneficial to further understand underlying mechanisms of genetic regulation as well as aid in methods of causal inference. Lastly, given the lack of brain multi-omics data from diverse populations, the cohorts in this study are from non-Hispanic white brain donors. This is a major gap in the field, which we anticipate will be addressed with foundational brain multi-omics data recently emerging from multi-ethnic brain donors[145].

In conclusion, the landscape of regional DNAm integrated with gene expression reveal a distinct epigenomic architecture underlying variability of brain tau biochemical species and neuropathology in AD. Our robust and replicable findings suggest that tau-related epigenetic dysregulation of gene expression impacts oligodendrocytes and may be a critical mechanism of dysmyelination in AD and other tauopathies. Our study identifies novel genes implicated in AD, such as *LDB3*, and many others that are replicated across multiple independent datasets. This study demonstrates the potential of applying rigorous and innovative multi-omics analysis to deep endophenotypes to unravel the complex heterogeneity of AD, which also has implications for other neurodegenerative diseases. The plethora of data and knowledge derived from this study is shared through our interactive application tool Multiomic Atlas of AD Brain Endophenotypes to aid further discoveries by the research community.

Methods

This study was approved by the appropriate Mayo Clinic Institutional Review Board.

Brain samples

Post-mortem superior temporal gyrus cortex (TCX) and cerebellum (CER) tissue samples were collected by the Department of Neuroscience, Mayo Clinic Brain Bank (Jacksonville, Florida, USA). These samples are a part of the Mayo Clinic AD-CAA (MC-CAA) study, previously described[9,84] and also detailed on the AD-knowledge portal (https://adknowledgeportal.synapse.org, see Data Sharing section). All samples had a confirmed AD neuropathological diagnosis with a Braak stage ≥4 and Thal phase ≥ 2, were scored for CAA pathology, had an age at death greater than 55 years, and due to availability, were all non-Hispanic white (Supplementary Data 1). Donors with co-pathologies were not excluded to allow for inclusion of the full spectrum of neuropathologies in AD. Braak stage, Thal phase, and CAA scores were measured using previously established protocols[3,4,10,146,147]. Intermediate Braak stages were grouped with the next lowest stage as follows, stage 3.5 is 3, 4.5 is 4, and 5.5 is 5 as detailed previously[10,84] (Supplementary Data 1, Fig. S1).

Biochemical measures

Biochemical measures from a subset of superior temporal cortex brain samples included the MC-CAA study described previously[9,10] were utilized for this study. Biochemical measures include five AD-related proteins (APOE, Aβ40, Aβ42, tau, and phospho-tau (Thr231)) from three tissue fractions (soluble, membrane-bound, insoluble). Briefly, supernatant fractions were collected after three sequential buffer treatments of tissue homogenate and resulting pellets: first with tris-buffered saline buffer (TBS), second with detergent (TBS/1% Triton X) buffer (TX), and lastly with formic acid buffer (FA). Protein quantification was performed via ELISA for each fraction and normalized against total protein quantities. All biochemical measures were transformed by either the natural log or square root to approximate a normal distribution including the Aβ40/42 ratio. Transformed measures were then standardized with a mean of zero and standard deviation of one using the z-score formula to allow for comparisons of effect sizes between biochemical measure analyses. Of the 469 MC-CAA samples with biochemical measures, 452 also had DNA methylation data (Fig. S1).

Genotype data

Genome-wide genotypes (GWG) were generated in previous studies and are described in detail elsewhere[9,84]. Briefly, DNA was isolated for GWG from brain tissue using the AutoGen245T instrument according to manufacturer's instructions including an RNase A (Qiagen) digestion. GWG were collected using the Infinum Omni2.5 Exome 8 v1.3 genotyping array, formatted in PLINK (v1.9)[148,149] and underwent rigorous quality control (QC). Variants passing QC were imputed to the haplotype reference consortium (HRC) keeping only variants with an imputation quality $R^2 \geq 0.7$ and minor allele frequency (MAF) ≥ 2%. Variant base pair (BP) positions were lifted from hg19 to hg38 using the *liftOver* package and corresponding chain files provided by UCSC Genome Browser. After mapping coordinates in hg19 to hg38, positional information in the plink files were updated using the '--update-map' flag. Of the $N = 6{,}724{,}981$ imputed and quality controlled genetic variants aligned to the hg19 reference genome, $N = 6{,}721{,}677$ successfully lifted over to the hg38 reference and could be analyzed.

DNA methylation

DNA from 40-70 mg of TCX tissue was isolated using the AutoGen245T instrument according to manufacturer's instructions including an RNase A (Qiagen) digestion prior to loading. The CER samples failed DNA isolation using the AutoGen instrument, so DNA was isolated from 10 to 15 mg of CER tissue using the same AutoGen Reagent kit and instrument steps manually. Pilot experiments were performed comparing TCX DNA isolated in triplicate from 3 samples using the AutoGen instrument vs manual methods as well as 2 CER samples in

triplicate manually to determine if extraction method significantly impacted overall DNA methylation (DNAm) levels (Fig. S13). Following DNA extraction of all samples, Reduced Representation Bisulfite Sequencing (RRBS) was performed on 25 ng of genomic DNA for each sample by the Mayo Clinic Genome Analysis Core (GAC) in Rochester, MN following manufacturer's recommendations for the NuGen Ovation RRBS Methyl-Seq System 1–16 (Tecan Genomics). The EZ DNA Methylation Kit (Zymo Research) was used for bisulfite conversion. RRBS libraries were sequenced on the HiSeq 4000 (Illumina) with 8 samples per lane and 50 bp paired end reads. The Streamlined Analysis and Annotation Pipeline for Reduced Representational Bisulfite Sequencing (SAAP-RRBS)[150] was used to align reads to the hg38 reference genome and to get CpGm ratio information. Briefly, FASTQs were trimmed to remove diversity and adaptor sequences using cutadapt, and any reads with less than 15 bp were discarded. Trimmed FASTQs were then aligned against the hg38 reference genome using BSMAP[151]. Samtools[152] was used to get mpileup and custom PERL scripts were used to determine CpG and non-CpG methylation levels along with bisulfite conversion ratios. All samples had a conversion ratio >98% and we kept only CpG sites with at least 10x coverage. One triplicate TCX sample (*TCX_Autogen_1c*) had low read coverage and so was not included in the downstream comparisons (Fig. S13a). In a Principal Component Analysis (PCA) of the top 50% most variable CpG sites with 10x coverage ($N = 535{,}165$ CpGs), we found that sample replicates and brain tissue types were tightly clustered independent of extraction method (Fig. S13b). We also found CpGm levels were highly correlated within sample replicates regardless of extraction method and to a much greater degree than between tissue types of the same samples (Fig. S13c). These results indicated that extraction method (instrument vs manual) does not significantly contribute to differences in DNA methylation levels. Consequently, we performed automated DNA extraction for TCX samples and manual for CER samples.

All study samples were randomized for brain tissue (TCX, CER), sex, age at death, *APOE*ε4 dose and CAA score for sample prep and sequencing. Library preparation and sequencing was performed in the same manner as that for the pilot study. One TCX sample failed sequencing resulting in $N = 200$ CER and $N = 471$ TCX samples with available DNAm data. Raw sequencing reads were analyzed using the SAAP-RRBS pipeline described above. Rigorous QC of the CpG sites was then performed independently for each brain tissue type by (1) removing samples with a bisulfite conversion ratio <98%, (2) removing sites with less than 10x total read coverage and those that mapped to the mitochondrial chromosome, (3) removing samples that failed genome-wide association study QC performed in a previous study[84], (4) removing sites with a total read coverage in the 99.9th percentile likely indicative of PCR bias, (5) removing samples that failed sex check by calculating the average X and Y chromosome read coverage normalized to the average read coverage of chromosome 1 defining failed samples as those that are beyond 4 standard deviations of the mean of their respective female and male clusters, (6) removing CpG sites that mapped to the sex chromosomes, (7) masking individual sites for each sample separately that are C to T nucleotide transitions on the forward strand or G–A transitions on the reverse strand as determined from the imputed genotype data[84], (8) removing outlier samples determined from a principle component analysis (PCA) of the top 50% most variable CpGm sites based on standard deviation using the R function *prcomp*(.scale = TRUE, center = TRUE) with outliers defined as those 4 times the standard deviation from the mean of principal components (PC) 1-3, (9) removing sites not present in over 50% of the samples, and (10) removing samples with less than 50% of the CpGs that passed QC (Supplementary Data 2).

For analyses investigating regional CpGm (rCpGm), we leveraged the ChromHMM core 15-state chromatin model developed by the Roadmap Epigenomics Consortium[48] for the adult brain Inferior Temporal Lobe (E072). The hg38lift predicted chromatin bed file was downloaded from https://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/coreMarks/jointModel/final/ on June 29, 2022. Although there were regions in this file that were not annotated to a specific chromatin state as a result of the lift over from hg19 to hg38, these individual regions were kept and annotated as "16_NA" or "Not Defined" for downstream analyses. We grouped all CpGm sites that passed QC into regions based on the 15-state chromatin model annotations and averaged each group to determine rCpGm levels (Fig. 1). rCpGm start and stop sites were defined as the first and last CpGm present in each rCpGm group. As a predicted chromatin state model for the CER did not exist at the time of this study and the fact that we want to analyze rCpGms across tissue types, we used the Inferior Temporal Lobe model to define rCpGms for all brain tissue types. In this study, we had two tiers of DNA methylation measures, one at an individual CpGm site level (CpGm) and one at a regional level according to our novel method defined by the chromatin state model (rCpGm). For each tier, we then performed data reduction steps to remove non-variable sites/regions that are not statistically informative and that would unnecessarily contribute to increasing the multiple testing correction burden. These data reduction steps include, (11) removing CpGm/rCpGm with a methylation ratio range less than 1% as these are likely not biologically informative, and (12) removing sites/regions where 95% or more of the samples are fully methylated (100%) or fully unmethylated (0%) as these provide unreliable statistical estimates (Supplementary Data 2).

Gene expression measures

As part of the MC-CAA study, bulk RNAseq data was also generated for 477 TCX and a subset of 200 CER samples. *Sample Preparation and Sequencing*: Total RNA was extracted for each tissue sample using Trizol® reagent and cleaned using Qiagen RNeasy columns with DNase treatment. RNA integrity (RIN) was measured using an Agilent Technologies 2100 Bioanalyzer and only samples with a RIN > 5.5 were sent to the Mayo Clinic Genome Analysis Core (GAC) for library preparation and sequencing. The TruSeq RNA Library Prep Kit v2 (Illumina, San Diego, CA) was used for library preparation (non-stranded). Library concentration and size distribution were determined on an Agilent Bioanalyzer DNA 1000 chip. Samples were randomized across sequencing flowcells with respect to average CAA score, sex, age at death, and *APOE*ε4 genotype. The Illumina HiSeq 4000 was used for sequencing with 100 bp paired-end reads, multiplexing six samples per flowcell lane. *Quality Control*: The Mayo Clinic MAP-Rseq pipeline v.3.0.2[153] was used for read alignment to hg38 implementing STARv2.5.2b[154], and featureCounts from the Subread package v1.5.1 for counting[155]. Raw read counts were log2-transformed and normalized using conditional quantile normalization (CQN) via the Bioconductor package; accounting for sequencing depth, gene length, and GC content[156]. FastQC was used for QC of raw sequence reads, and RSeQC[157] was used for QC of mapped reads. We excluded samples with <50 million reads mapped to genes, <50% of reads mapped to genes, and outliers (1.5 inter-quartile range) based on distribution of % junction reads. Sex check was performed by comparing expected expression of Y chromosome genes to recorded sex. Samples that were outliers (>4 standard deviations from the mean) on PCs 1 and 2 in a PCA (*prcomp*, CPM, R statistical software) and multiple other QC variables (RSeQC) were also excluded. We also excluded samples that failed previous genetic based QC steps including relatedness and heterozygosity checks as described previously[84]. There were 456 TCX and 187 CER samples that passed QC and of these, 449 TCX and 184 CER samples had matching RRBS data. We included in our downstream analysis only autosomal genes expressed above background which we defined based on the distribution of CQN values filtering out genes with a CQN value less than 1 for both the CER and TCX datasets.

AMP-AD datasets

The Mayo RNAseq study[69], The Mount Sinai Brain Bank (MSBB) study[70], and The Religious Orders Study and Memory and Aging Project (ROSMAP) Study[71] were obtained from the AD-knowledge portal (https://adknowledgeportal.synapse.org). The RNA-seq data previously underwent consensus reprocessing by Accelerating Medicines Partnership-AD (AMP-AD, RNAseq Harmonization Study)[72]. Additional QC and diagnosis harmonization of these datasets based on neuropathological measures are described in detail elsewhere[84].

Statistical analyses

Power calculation. Power calculations were performed with the *pwr.r.test* function from the *pwr* package v1.3-0 in R (v4.0.2). For a sample size of 472 in a two-sided test, we have 80% power to detect a correlation coefficient of 0.239 and 0.129 at an alpha of 1E-05 and 0.05, respectively. For a sample size of 200 in a two-sided test, we have 80% power to detect a correlation coefficient of 0.359 and 0.197 at an alpha of 1E-05 and 0.05, respectively.

Phenotypes. Spearman rank correlation between all biochemical measures and CAA have been reported previously[10].

Cell type proportion estimates. Neuronal and oligodendrocyte cell type proportions were estimated from the bulk tissue RNAseq data using the Digital Sorting Algorithm (DSA)[158] via the *DSA* R package (v1.0). The top 100 genes from BRETIGEA[65] were used as the cell type specific marker genes for these estimations. Oligodendrocyte precursor cell (OPC) types were not investigated as we previously did not observe robust correlations among the defined OPC marker genes as described in Wang et al.[93].

Epigenome wide association studies (EWAS). Primary association models were run using the linear regression *lm*() function in R v4.1.2 adjusting for age at death, sex, and the first three genetic PCs. The CpGm variable (CpGm or rCpGm) was included in each association model as the dependent variable. There were three samples with a Thal phase measure of 2 (2 TCX and 1 with both TCX and CER) that passed DNAm QC. Given the low number of this measure in a single group and the possibility that small variations in this group may considerably and unreliably impact our statistical estimates, we set these Thal phase 2 measures to *NA* in analyses that included Thal phase as a variable. False Discovery Rate (FDR) corrected p values were calculated with the R *p.adjust* function using the Benjamini & Hochberg method.

***Comb-p* regional aggregation**. To compare our rCpGm method with the published regional aggregation method *comb-p*[45], test statistics from each individual CpGm EWAS were independently used an inputs to the python based *comb-p* software following the default "pipeline" steps[45]. *Comb-p* identifies and groups spatially correlated p values through a sliding window approach. Input parameters included regions with ≥ 3 CpGs, a seed of <1E-04, a sliding window size of 500 bp, and significant regions defined as those with a Sidak corrected $p \leq 0.05$. While *comb-p* does not take into account the CpG regression estimates, significant regions were defined as positively or negatively associated with a phenotype by the effect directions of the majority (>50%) of CpGs included in the defined region. When comparing across EWAS, overlapping regions were identified and concatenated using the *GenomicRanges::reduce* function.

Association with gene expression measures. In the MC-CAA dataset, there were $N = 449$ TCX and $N = 184$ CER samples that had both RRBS and bulk RNA seq data that could be investigated. *Cis*-rCpGm-Gene pairs were identified by connecting rCpGm with all genes ±500 kb away from the start and end of each rCpGm for the TCX and CER separately. Residuals of rCpGm were extracted by fitting the linear regression model: rCpGm ~ AgeAtDeath + Sex + genetic PC 1–3. Bulk gene expression (GE) residuals were extracted by fitting the mixed effects linear regression model: GE ~ AgeAtDeath + Sex + RIN + (1|batch). The association between rCpGm and bulk gene expression residuals were tested by fitting the linear regression model: Gene_Expression_residuals ~ rCpGm_residuals for each *cis*-rCpGm-Gene pair independently for each tissue type.

To identify gene expression associations with the AD endophenotypes, we tested the association of normalized gene expression data with the endophenotype of interest. For continuous phenotypes (biochemical measures and CAA), we ran the following linear regression model: cqn(gene expression) ~ ADphenoytpe + Age at Death + sex + RIN + (1|flowcell) using the *lmer* function in the *lme4* package[159]. For ordinal variables (Braak stage and Thal phase), we first extracted residuals from the gene expression values using the following model: cqn(gene expression) ~ Age at Death + sex + RIN + (1|flowcell), and then associated the gene expression residuals with each phenotype in an ordinal regression model using the *clm* function in the *ordinal* R package. As with the DNAm data, because there were only a small number of samples with a Thal phase of 2, these were set to *NA* in analyses that included Thal as a variable. FDR p values were calculated with the R *p.adjust* package using the Benjamini & Hochberg method.

The AMP-AD datasets were used to investigate the association between CQN normalized gene expression levels of implicated genes with Braak stage in non-Hispanic white, neuropathologically defined AD cases and controls using linear regression in R with the *lm()* function. Mayo TCX models adjusted for age at death, sex, RNA integrity number (RIN), and sequencing flow cell rounding down intermediate Braak stage numbers to the next lowest stage. MSBB BM22 models adjusted for age at death, sex, RIN, and batch. ROSMAP models adjusted for age at death, sex, RIN, sequencing batch, and study (ROS/MAP). Individuals with an age at death above 90 were set to 90, in compliance with HIPAA rules. Oligodendrocyte proportions were further adjusted for in each model by including the *OLIG2* (ENSG00000205927) normalized expression levels as a covariate.

mQTL analysis. *Cis*-methylation Quantitative Trait Locus (*cis*-mQTL) analysis was performed using the R package *MatrixEQTL*[160] for all genetic variant dosage and rCpGms pairs ±500 kb apart for the TCX and CER separately adjusting for age at death, sex, and the first 3 genetic PCs. In total, we investigated 289,078,059 TCX rCpGm-SNP pairs and 291,400,204 CER rCpGm-SNP pairs. False Discovery Rate (FDR) corrected p values were calculated with the R *p.adjust* package using the Benjamini & Hochberg method.

Sensitivity analyses. We performed post-hoc sensitivity analyses to test the robustness of the top associations to effects of batch, neuropathological changes, neuronal and oligodendrocyte cell type proportions. To determine if top associations were influenced by batch, we included batch in a linear regression model as a random effect variable using the *lmer()* function of the *lme4* R package[159]: rCpGm ~ Endophenotype + Age at death + Sex + Genetic PCs 1-3 + (1|FlowCell). To determine if top biochemical associations were influenced by neuropathological changes, we included Braak stage and Thal phase as fixed-effect covariates in a linear regression model: rCpGm ~ Endophenotype + Age at death + Sex + Genetic PCs 1-3 + Braak stage + Thal phase. To determine if top associations were influenced by neuronal cell type proportions, we included DSA estimated neuronal cell type proportions as a fixed-effect covariate in the linear regression model: rCpGm ~ Endophenotype + Age at death + Sex + Genetic PCs 1-3 + Neuronal Cell Proportion. To determine if top associations were influenced by oligodendrocyte cell type proportions, we included DSA estimated oligodendrocyte cell type proportions as a fixed-effect covariate in the linear regression model: rCpGm ~ Endophenotype + Age at death + Sex + Genetic PCs 1-3 + Oligodendrocyte Cell

Proportion. Lastly, to determine if distribution of a phenotype influenced the top associations, we performed Spearman rank correlation analysis in R using the *cor.test()* function of the *stats* package.

Pathway analysis. Pathway enrichment was tested using the Gene Ontology Biological Processes database in the *anRichment* R package with *p* values computed via a one-sided hypergeometric test.

Annotations. CpGm sites were annotated according to their regions (Island, shore, shelf, inter-genic) and nearby genes (±50 kb) using the R package *annotatr()* v1.21.0[161] querying AnnotationHub_3.2.2 in R (v4.1.2) or with built-in hg38 genes, respectively.

DNAm replication datasets. Data from The Religious Orders Study and Memory and Aging Project (ROSMAP) Study was downloaded from the AD knowledge portal for 707 Non-Hispanic White, dorsolateral prefrontal cortex (DLPFC) samples with DNAm data measured on the Illumina HumanMethylation450 BeadChip array[30] including IDAT and metadata files on May 17th, 2023 (Supplementary Data 3). Additional and updated metadata for these ROSMAP samples was also obtained from the Rush Alzheimer's Disease Center (RADC) on May 22nd, 2023. Data from the Brains for Dementia Research (BDR) Dataset was downloaded for 630 individuals with DLPFC and occipital cortex (OCC) DNAm measured on the Illumina EPICv1 DNA methylation array[32] including IDAT files from the Gene Expression Omnibus (GEO, accession number GSE197305) on February 23rd, 2023 and metadata from the UK Brain Bank Network (UKBBN) database after project approval (https://brainbanknetwork.ac.uk/) on May 26th, 2023 (Supplementary Data 3). PSP data from Weber et al.[74] was downloaded from GEO (GSE75704) on April 12th, 2023 for 94 PSP and 72 control individuals with prefrontal lobe DNAm measured using the Illumina HumanMethylation450 BeadChip. Sorted nuclei data from Gasparoni et al.[36] was downloaded from GEO (GSE66351) on January 31st, 2023 for 31 occipital cortex samples using the Illumina HumanMethylation450 BeadChip. The ROSMAP, BDR, Weber PSP, and Gasparoni sorted datasets were consensus reprocessed through a unified QC pipeline for each brain tissue type separately using R (v4.1.2). For consistency with the Mayo Clinic discovery cohort, individuals with an indicated ethnicity other than non-Hispanic white or white were removed, individuals with unknown ethnicity were kept. Raw IDAT files were imported as a RGChannelSet object using *read.metharray.exp* in *minfi*[162]. Sample and probe QC was performed using the following steps: (1) excluding probes that mapped to non-autosomal chromosomes; (2) excluding samples with a low (<80%) bisulfite conversion rate estimated by *bscon* in *wateRmelon*[163]; (3) using detection *p* values obtained by *detectionP* in *minfi* to exclude probes where <99% of the samples had a detection $p > 0.05$; (4) removing samples where <99% of the probes had a detection $p > 0.05$; (5) removing samples that failed sex check by clustering samples on the first PCs of previously identified robust sex-related probes on the X and Y chromosomes as implemented by *estimateSex*[164] in *wateRmelon*; (6) removing probes that have previously been identified to have cross-hybridizing potential or polymorphic targets (with European MAF > 0.01)[165], as well as the SNP and control probes; (7) removing outlier samples defined by the *outlyx* function in *wateRmelon*. Subsequently, to minimize intra- and inter-array variation, as well as probe design (I/II) bias, quantile normalization was performed using *dasen*[163] in *wateRmelon*, and outlier check was performed again on the normalized data. Individual sex and age at death were compared when available from different sources and when discrepancies were identified, individuals were removed from the analysis. PCA was performed with the normalized β-values using *prcomp*(*.scale* = TRUE, *center* = TRUE) on the top 50% most variable CpGm sites based on standard deviation. Non-CpGm sites were removed and CpGm sites were lifted to the hg38 genome using the *liftOver* package and corresponding chain files provided by UCSC Genome Browser. rCpGms were calculated and data reduction steps were performed as described for the RRBS data. Intermediate Braak stages were grouped with the next lowest stage. AD cases were neuropathologically defined as individuals with a Braak stage ≥4 and a Thal phase ≥ 2. Control samples were defined as individuals with a Braak stage <4 and a Thal phase <2 (Supplementary Data 3). Any donor who was not neuropathologically defined as an AD case or control was designated as "Other". For the Gasparoni sorted dataset, ADs were defined as having a Braak stage ≥4 and defined as AD by the original study while controls were defined as those with a Braak stage ≤ 3 and defined as a control by the original study[36]. For the Weber PSP dataset, PSP and control sample definitions were used as described[36,74]. For ROSMAP and BDR, EWAS were performed in each dataset and tissue type separately for Braak stage, Thal phase, and AD diagnosis looking at AD vs Controls using the *lmer* function in the *lmerTest* R package. For the ROSMAP linear mixed models, we adjusted for age at death, sex, study (ROS/MAP), and batch as fixed effects as well as slide as a random effect variable. For the BDR linear mixed models, we adjusted for age at death, sex, and CpGm PC's 1-3 as fixed effects as well as slide as a random effect variable. We included CpGm PC's 1-3 in the BDR models because batch, although not available, has previously been found to be a source of variation in this dataset and captured by PC's[32]. PC's 1-3 were chosen because these explained more than 5% of variation (PC1-23%, PC2-13%, PC3-6%) cumulating to a total of 41%. For the Gasparoni sorted dataset, EWAS were run using *lmer* and linear mixed models for AD diagnosis adjusting for age at death and sex as fixed effects as well as slide as a random effect. For Weber PSP, EWAS were run using *lmer* and linear mixed models for PSP diagnosis and Braak stage adjusting for age at death, sex, and CpGm PC's 1-3 as fixed effects as well as slide as a random effect variable. Similarly, because we did not have information on batch, we included PC's 1-3 explaining 25%, 12%, and 5% of the variation, respectively. Next, due to consistently high level of genomic inflation, we used the *bacon* R package[166] to adjust for inflation in each of the EWAS models (Supplementary Data 21).

Meta-analysis. Meta-analysis of Braak stage was performed in R with the *meta::metagen* function including rCpGm EWAS estimate and standard error statistics from the RRBS-TCX, BDR- DLPFC, and ROSMAP-DLPFC datasets. Test statistics from both fixed and random effect models are reported (Supplementary Data 11).

Previously published single nuclei/cell RNAseq studies. Oligodendrocyte specific gene expression association results from previously published single nuclei/cell RNAseq studies depicted in Fig. 5B and Fig. S12 were downloaded in October, 2024 from the following publications: Is 2024[75], Mathys 2023[76], Gamache 2023[77], Morabito 2021[78], Anderson 2023[79], Rexach 2024[80], Min 2023[81].

Cell type marker gene enrichment. One sided hypergeometric tests using the *phyper(lower.tail = FALSE)* function from the R *stats* package were used to test for enrichment of the top 100 and top 1000 cell type marker genes identified by BRETIGEA[65] in our dataset. We used the total number of genes expressed above background tested in our gene expression analysis ($N = 18{,}383$ genes) as the total universe number.

Reporting summary

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

Data availability

The data generated in this study have been deposited to the AD Knowledge Portal and can be accessed via the manuscript dataset page (https://www.synapse.org/Synapse:syn64679864). The AD Knowledge Portal is a platform for accessing data, analyses and tools generated by the Accelerating Medicines Partnership (AMP-AD) Target Discovery

Program and other National Institute on Aging (NIA)-supported programs to enable open- science practices and accelerate translational learning. The data, analyses and tools are shared early in the research cycle without a publication embargo on secondary use. Data is available for general research use according to the following requirements for data access and data attribution (https://adknowledgeportal.synapse.org/DataAccess/Instructions). We have also made summary data generated in this study available in the supplementary data files as well as easily accessible via our interactive Shiny web application Multiomic Atlas of AD Brain Endophenotypes (https://rtools.mayo.edu/MC-CAA/).

Code availability

Code used in this manuscript for data analyses can be accessed via the AD Knowledge Portal: https://www.synapse.org/Synapse:syn64465911.

References

1. Montine, T. J. et al. National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease: a practical approach. *Acta Neuropathol.* **123**, 1–11 (2012).
2. Hyman, B. T. et al. National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimers Dement* **8**, 1–13 (2012).
3. Thal, D. R., Rüb, U., Orantes, M. & Braak, H. Phases of A beta-deposition in the human brain and its relevance for the development of AD. *Neurology* **58**, 1791–1800 (2002).
4. Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol.* **82**, 239–259 (1991).
5. Mehta, R. I. & Schneider, J. A. What is 'Alzheimer's disease'? The neuropathological heterogeneity of clinically defined Alzheimer's dementia. *Curr. Opin. Neurol.* **34**, 237–245 (2021).
6. Lau, H. H. C., Ingelsson, M. & Watts, J. C. The existence of Abeta strains and their potential for driving phenotypic heterogeneity in Alzheimer's disease. *Acta Neuropathol.* https://doi.org/10.1007/s00401-020-02201-2 (2020).
7. Lam, B., Masellis, M., Freedman, M., Stuss, D. T. & Black, S. E. Clinical, imaging, and pathological heterogeneity of the Alzheimer's disease syndrome. *Alzheimers Res Ther.* **5**, 1–1 (2013).
8. Ferreira, D., Nordberg, A. & Westman, E. Biological subtypes of Alzheimer disease: a systematic review and meta-analysis. *Neurology* **94**, 436–448 (2020).
9. Oatman, S. R. et al. Genome-wide association study of brain biochemical phenotypes reveals distinct genetic architecture of Alzheimer's disease related proteins. *Mol. Neurodegener.* **18**, 2 (2023).
10. Liu, C.-C. et al. Tau and apolipoprotein E modulate cerebrovascular tight junction integrity independent of cerebral amyloid angiopathy in Alzheimer's disease. *Alzheimer's. Dement.* **16**, 1372–1383 (2020).
11. Serrano-Pozo, A., Frosch, M. P., Masliah, E. & Hyman, B. T. Neuropathological alterations in Alzheimer disease. *Cold Spring Harb. Perspect. Med.* **1**, a006189 (2011).
12. DeTure, M. A. & Dickson, D. W. The neuropathological diagnosis of Alzheimer's disease. *Mol. Neurodegener.* **14**, 32 (2019).
13. Vidoni, E. D. et al. Cerebral β-amyloid angiopathy is associated with earlier dementia onset in Alzheimer's disease. *Neurodegener. Dis.* **16**, 218–224 (2016).
14. Charidimou, A. et al. Emerging concepts in sporadic cerebral amyloid angiopathy. *Brain* **140**, 1829–1850 (2017).
15. Iqbal, K., Liu, F., Gong, C. X. & Grundke-Iqbal, I. Tau in Alzheimer disease and related tauopathies. *Curr. Alzheimer Res.* **7**, 656–664 (2010).
16. Steinerman, J. R. et al. Distinct pools of beta-amyloid in Alzheimer disease-affected brain: a clinicopathologic study. *Arch. Neurol.* **65**, 906–912 (2008).
17. Roberts, B. R. et al. Biochemically-defined pools of amyloid-β in sporadic Alzheimer's disease: correlation with amyloid PET. *Brain* **140**, 1486–1498 (2017).
18. Ait-Bouziad, N. et al. Discovery and characterization of stable and toxic Tau/phospholipid oligomeric complexes. *Nat. Commun.* **8**, 1678 (2017).
19. Ekinci, F. J. & Shea, T. B. Phosphorylation of tau alters its association with the plasma membrane. *Cell Mol. Neurobiol.* **20**, 497–508 (2000).
20. Jones, E. M. et al. Interaction of tau protein with model lipid membranes induces tau structural compaction and membrane disruption. *Biochemistry* **51**, 2539–2550 (2012).
21. Koss, D. J. et al. Soluble pre-fibrillar tau and β-amyloid species emerge in early human Alzheimer's disease and track disease progression and cognitive decline. *Acta Neuropathol.* **132**, 875–895 (2016).
22. Kanekiyo, T., Xu, H. & Bu, G. ApoE and Aβ in Alzheimer's disease: accidental encounters or partners? *Neuron* **81**, 740–754 (2014).
23. Lambert, J. C. et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
24. Wightman, D. P. et al. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat. Genet.* **53**, 1276–1282 (2021).
25. Livingston, G. et al. Dementia prevention, intervention, and care: 2020 report of the *Lancet* Commission. *Lancet* **396**, 413–446 (2020).
26. Andrews, S. J. et al. The complex genetic architecture of Alzheimer's disease: novel insights and future directions. *eBioMedicine* **90**, https://doi.org/10.1016/j.ebiom.2023.104511 (2023).
27. Gil, L., Niño, S. A., Guerrero, C. & Jiménez-Capdeville, M. E. Phospho-tau and chromatin landscapes in early and late Alzheimer's disease. *Int. J. Mol. Sci.* **22**, 10283 (2021).
28. Klein, H.-U. et al. Epigenome-wide study uncovers large-scale changes in histone acetylation driven by tau pathology in aging and Alzheimer's human brains. *Nat. Neurosci.* **22**, 37–46 (2019).
29. Fetahu, I. S. et al. Epigenetic signatures of methylated DNA cytosine in Alzheimer's disease. *Sci. Adv.* **5**, eaaw2880 (2019).
30. De Jager, P. L. et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat. Neurosci.* **17**, 1156–1163 (2014).
31. Smith, R. G. et al. A meta-analysis of epigenome-wide association studies in Alzheimer's disease highlights novel differentially methylated loci across cortex. *Nat. Commun.* **12**, 3517 (2021).
32. 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.* **13**, 5620 (2022).
33. Piras, I. S. et al. Integrated DNA methylation/RNA profiling in middle temporal gyrus of Alzheimer's disease. *Cell. Mol. Neurobiol.* **43**, 2289–2307 (2023).
34. Iwata, A. et al. Altered CpG methylation in sporadic Alzheimer's disease is associated with APP and MAPT dysregulation. *Hum. Mol. Genet.* **23**, 648–656 (2013).
35. Smith, A. R. et al. Increased DNA methylation near TREM2 is consistently seen in the superior temporal gyrus in Alzheimer's disease brain. *Neurobiol. Aging* **47**, 35–40 (2016).
36. Gasparoni, G. et al. DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex. *Epigenetics Chromatin* **11**, 41 (2018).
37. Hernández, H. G. et al. Alzheimer's disease DNA methylome of pyramidal layers in frontal cortex: laser-assisted microdissection study. *Epigenomics* **10**, 1365–1382 (2018).
38. Lunnon, K. et al. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat. Neurosci.* **17**, 1164–1170 (2014).

39. Semick, S. A. et al. Integrated DNA methylation and gene expression profiling across multiple brain regions implicate novel genes in Alzheimer's disease. *Acta Neuropathol.* **137**, 557–569 (2019).
40. Moore, L. D., Le, T. & Fan, G. DNA methylation and its basic function. *Neuropsychopharmacology* **38**, 23–38 (2013).
41. Lövkvist, C., Dodd, I. B., Sneppen, K. & Haerter, J. O. DNA methylation in human epigenomes depends on local topology of CpG sites. *Nucleic Acids Res.* **44**, 5123–5132 (2016).
42. Eckhardt, F. et al. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.* **38**, 1378–1385 (2006).
43. Bell, J. T. et al. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.* **12**, R10 (2011).
44. Klein, H.-U. & Hebestreit, K. An evaluation of methods to test predefined genomic regions for differential methylation in bisulfite sequencing data. *Brief. Bioinform.* **17**, 796–807 (2015).
45. Pedersen, B. S., Schwartz, D. A., Yang, I. V. & Kechris, K. J. Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics* **28**, 2986–2988 (2012).
46. Piao, Y., Xu, W., Park, K. H., Ryu, K. H. & Xiang, R. Comprehensive Evaluation of Differential Methylation Analysis Methods for Bisulfite Sequencing Data. *Int. J. Environ. Res. Public Health* **18**, https://doi.org/10.3390/ijerph18157975 (2021).
47. Campagna, M. P. et al. Epigenome-wide association studies: current knowledge, strategies and recommendations. *Clin. Epigenetics* **13**, 214 (2021).
48. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
49. Dor, Y. & Cedar, H. Principles of DNA methylation and their implications for biology and medicine. *Lancet* **392**, 777–786 (2018).
50. Mansell, G. et al. Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. *BMC Genom.* **20**, 366 (2019).
51. Dugué, P.-A. et al. Reliability of DNA methylation measures from dried blood spots and mononuclear cells using the HumanMethylation450k BeadArray. *Sci. Rep.* **6**, 30317 (2016).
52. 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.* **11**, 6114–6114 (2020).
53. Bellenguez, C. et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436 (2022).
54. Kunkle, B. W. et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
55. Gauthier-Kemper, A. et al. The frontotemporal dementia mutation R406W blocks tau's interaction with the membrane in an annexin A2–dependent manner. *J. Cell Biol.* **192**, 647–661 (2011).
56. Gauthier-Kemper, A. et al. Annexins A2 and A6 interact with the extreme N terminus of tau and thereby contribute to tau's axonal localization. *J. Biol. Chem.* **293**, 8065–8076 (2018).
57. Pan, Y. et al. Cyclin-dependent kinase 18 promotes oligodendrocyte precursor cell differentiation through activating the extracellular signal-regulated kinase signaling pathway. *Neurosci. Bull.* **35**, 802–814 (2019).
58. Herskovits, A. Z. & Davies, P. The regulation of tau phosphorylation by PCTAIRE 3: implications for the pathogenesis of Alzheimer's disease. *Neurobiol. Dis.* **23**, 398–408 (2006).
59. Azorsa, D. O. et al. High-content siRNA screening of the kinome identifies kinases involved in Alzheimer's disease-related tau hyperphosphorylation. *BMC Genom.* **11**, 25 (2010).
60. Acquaah-Mensah, G. K., Agu, N., Khan, T. & Gardner, A. A regulatory role for the insulin- and BDNF-linked RORA in the hippocampus: implications for Alzheimer's disease. *J. Alzheimer's. Dis.* **44**, 827–838 (2015).
61. Dharshini, S. A. P., Taguchi, Y. H. & Gromiha, M. M. Investigating the energy crisis in Alzheimer disease using transcriptome study. *Sci. Rep.* **9**, 18509 (2019).
62. Rajabli, F. et al. Admixture mapping identifies novel Alzheimer's disease risk regions in African Americans. *Alzheimer's. Dement.* **19**, 2538–2548 (2023).
63. Ashby, E. L., Kehoe, P. G. & Love, S. Kallikrein-related peptidase 6 in Alzheimer's disease and vascular dementia. *Brain Res.* **1363**, 1–10 (2010).
64. Goldhardt, O. et al. Kallikrein-related peptidases 6 and 10 are elevated in cerebrospinal fluid of patients with Alzheimer's disease and associated with CSF-TAU and FDG-PET. *Transl. Neurodegener.* **8**, 25 (2019).
65. McKenzie, A. T. et al. Brain cell type specific gene expression and co-expression network architectures. *Sci. Rep.* **8**, 8868 (2018).
66. Garwain, O., Yerramilli, V. S., Romero, K. & Scarlata, S. The Gαq/phospholipase Cβ signaling system represses tau aggregation. *Cell. Signal.* **71**, 109620 (2020).
67. Xu, D. et al. Mapping enhancer and chromatin accessibility landscapes charts the regulatory network of Alzheimer's disease. *Comput. Biol. Med.* **168**, 107802 (2024).
68. Barone, G. et al. Human CDK18 promotes replication stress signaling and genome stability. *Nucleic Acids Res.* **44**, 8772–8785 (2016).
69. Allen, M. et al. Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Sci. Data* **3**, 160089 (2016).
70. Wang, M. et al. The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Sci. Data* **5**, 180185 (2018).
71. De Jager, P. L. et al. A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci. Data* **5**, 180142 (2018).
72. Wan, Y. W. et al. Meta-analysis of the Alzheimer's disease human brain transcriptome and functional dissection in mouse models. *Cell Rep.* **32**, 107908 (2020).
73. Greenwood, A. K. et al. Agora: an open platform for exploration of Alzheimer's disease evidence. *Alzheimer's. Dement.* **16**, e046129 (2020).
74. Weber, A. et al. Epigenome-wide DNA methylation profiling in Progressive Supranuclear Palsy reveals major changes at DLX1. *Nat. Commun.* **9**, 2929 (2018).
75. İş, Ö et al. Gliovascular transcriptional perturbations in Alzheimer's disease reveal molecular mechanisms of blood brain barrier dysfunction. *Nat. Commun.* **15**, 4758 (2024).
76. Mathys, H. et al. Single-cell atlas reveals correlates of high cognitive function, dementia, and resilience to Alzheimer's disease pathology. *Cell* **186**, 4365–4385.e4327 (2023).
77. Gamache, J. et al. Integrative single-nucleus multi-omics analysis prioritizes candidate cis and trans regulatory networks and their target genes in Alzheimer's disease brains. *Cell Biosci.* **13**, 185 (2023).
78. Morabito, S. et al. Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat. Genet.* **53**, 1143–1155 (2021).
79. Anderson, A. G. et al. Single nucleus multiomics identifies ZEB1 and MAFB as candidate regulators of Alzheimer's disease-specific cis-regulatory elements. *Cell Genom.* **3**, 100263 (2023).
80. Rexach, J. E. et al. Cross-disorder and disease-specific pathways in dementia revealed by single-cell genomics. *Cell* **187**, 5753–5774.e5728 (2024).
81. Min, Y. et al. Cross species systems biology discovers glial DDR2, STOM, and KANK2 as therapeutic targets in progressive supranuclear palsy. *Nat. Commun.* **14**, 6801 (2023).

82. Kedia, S. & Simons, M. Oligodendrocytes in Alzheimer's disease pathophysiology. *Nat. Neurosci.* **28**, 446–456 (2025).
83. National Academies of Sciences, Engineering, and Medicine; Health and Medicine Division; Behavioral and Social Sciences and Education; Board on Health Sciences Policy; Board on Behavioral, Cognitive, and Sensory Sciences; Committee on Research Priorities for Preventing and Treating Alzheimer's Disease and Related Dementias; Yost OC, Downey A, Powell T, editors. Preventing and Treating Dementia: Research Priorities to Accelerate Progress. (Washington (DC), National Academies Press (US), 2024).
84. Reddy, J. S. et al. Genome-wide analysis identifies a novel LINC-PINT splice variant associated with vascular amyloid pathology in Alzheimer's disease. *Acta Neuropathol. Commun.* **9**, 93 (2021).
85. Shade, L. M. P. et al. GWAS of multiple neuropathology endophenotypes identifies new risk loci and provides insights into the genetic risk of dementia. *Nat. Genet.* **56**, 2407–2421 (2024).
86. Zhang, L. et al. Sex-specific DNA methylation differences in Alzheimer's disease pathology. *Acta Neuropathol. Commun.* **9**, 77 (2021).
87. Pellegrini, C. et al. A meta-analysis of brain DNA methylation across sex, age, and Alzheimer's disease points for accelerated epigenetic aging in neurodegeneration. *Front. Aging Neurosci.* **13**, https://doi.org/10.3389/fnagi.2021.639428 (2021).
88. Crotti, A. et al. BIN1 favors the spreading of tau via extracellular vesicles. *Sci. Rep.* **9**, 9477 (2019).
89. Calafate, S., Flavin, W., Verstreken, P. & Moechars, D. Loss of Bin1 promotes the propagation of tau pathology. *Cell Rep.* **17**, 931–940 (2016).
90. Sottejeau, Y. et al. Tau phosphorylation regulates the interaction between BIN1's SH3 domain and Tau's proline-rich domain. *Acta Neuropathol. Commun.* **3**, 58 (2015).
91. Wang, Y. & Mandelkow, E. Tau in physiology and pathology. *Nat. Rev. Neurosci.* **17**, 5–21 (2016).
92. Allen, M. et al. Conserved brain myelination networks are altered in Alzheimer's and other neurodegenerative diseases. *Alzheimers Dement.* **14**, 352–366 (2018).
93. Wang, X. et al. Deciphering cellular transcriptional alterations in Alzheimer's disease brains. *Mol. Neurodegener.* **15**, 38–38 (2020).
94. McKenzie, A. T. et al. Multiscale network modeling of oligodendrocytes reveals molecular components of myelin dysregulation in Alzheimer's disease. *Mol. Neurodegener.* **12**, 82 (2017).
95. Pathak, P. et al. Myopathy associated LDB3 mutation causes Z-disc disassembly and protein aggregation through PKCα and TSC2-mTOR downregulation. *Commun. Biol.* **4**, 355 (2021).
96. Blech-Hermoni, Y. et al. Expression of LIM domain-binding 3 (LDB3), a striated muscle Z-band alternatively spliced PDZ-motif protein in the nervous system. *Sci. Rep.* **13**, 270 (2023).
97. Zhang, Y. et al. Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron* **89**, 37–53 (2016).
98. Siletti, K. et al. Transcriptomic diversity of cell types across the adult human brain. *Science* **382**, eadd7046 (2023).
99. Koopmann, T. T. et al. Biallelic loss of LDB3 leads to a lethal pediatric dilated cardiomyopathy. *Eur. J. Hum. Genet.* **31**, 97–104 (2023).
100. Lopez-Ayala, J. M. et al. A mutation in the Z-line Cypher/ZASP protein is associated with arrhythmogenic right ventricular cardiomyopathy. *Clin. Genet.* **88**, 172–176 (2015).
101. Bis, J. C. et al. Whole exome sequencing study identifies novel rare and common Alzheimer's-Associated variants involved in immune response and transcriptional regulation. *Mol. Psychiatry* **25**, 1859–1875 (2020).
102. Zhang, X. et al. A rare missense variant of CASP7 is associated with familial late-onset Alzheimer's disease. *Alzheimer's. Dement.* **15**, 441–452 (2019).
103. Sun, L. L., Yang, S. L., Sun, H., Li, W. D. & Duan, S. R. Molecular differences in Alzheimer's disease between male and female patients determined by integrative network analysis. *J. Cell Mol. Med.* **23**, 47–58 (2019).
104. Wang, P. et al. Predicting signaling pathways regulating demyelination in a rat model of lithium-pilocarpine-induced acute epilepsy: a proteomics study. *Int. J. Biol. Macromol.* **193**, 1457–1470 (2021).
105. Lim, R. G. et al. Huntington disease oligodendrocyte maturation deficits revealed by single-nucleus RNAseq are rescued by thiamine-biotin supplementation. *Nat. Commun.* **13**, 7791 (2022).
106. Kadavath, H. et al. Tau stabilizes microtubules by binding at the interface between tubulin heterodimers. *Proc. Natl. Acad. Sci. USA* **112**, 7501–7506 (2015).
107. Osso, L. A. & Hughes, E. G. Dynamics of mature myelin. *Nat. Neurosci.* **27**, 1449–1461 (2024).
108. Xin, W. & Chan, J. R. Myelin plasticity: sculpting circuits in learning and memory. *Nat. Rev. Neurosci.* **21**, 682–694 (2020).
109. Seiberlich, V. et al. Downregulation of the microtubule associated protein Tau impairs process outgrowth and myelin basic protein mRNA transport in oligodendrocytes. *Glia* **63**, 1621–1635 (2015).
110. LoPresti, P. Tau in oligodendrocytes takes neurons in sickness and in health. *Int. J. Mol. Sci.* **19**, https://doi.org/10.3390/ijms19082408 (2018).
111. Richter-Landsberg, C. Protein aggregate formation in oligodendrocytes: tau and the cytoskeleton at the intersection of neuroprotection and neurodegeneration. *Biol. Chem.* **397**, 185–194 (2016).
112. Klein, C. et al. Process outgrowth of oligodendrocytes is promoted by interaction of fyn kinase with the cytoskeletal protein tau. *J. Neurosci.* **22**, 698–707 (2002).
113. Ferrer, I. et al. Involvement of oligodendrocytes in tau seeding and spreading in tauopathies. *Front. Aging Neurosci.* **11**, https://doi.org/10.3389/fnagi.2019.00112 (2019).
114. Koutsodendris, N. et al. Neuronal APOE4 removal protects against tau-mediated gliosis, neurodegeneration and myelin deficits. *Nat. Aging* **3**, 275–296 (2023).
115. Braak, H. & Braak, E. Development of Alzheimer-related neurofibrillary changes in the neocortex inversely recapitulates cortical myelogenesis. *Acta Neuropathol.* **92**, 197–201 (1996).
116. Nasrabady, S. E., Rizvi, B., Goldman, J. E. & Brickman, A. M. White matter changes in Alzheimer's disease: a focus on myelin and oligodendrocytes. *Acta Neuropathol. Commun.* **6**, 22 (2018).
117. Salat, D. H. et al. Regional white matter volume differences in nondemented aging and Alzheimer's disease. *Neuroimage* **44**, 1247–1258 (2009).
118. Ihara, M. et al. Quantification of myelin loss in frontal lobe white matter in vascular dementia, Alzheimer's disease, and dementia with Lewy bodies. *Acta Neuropathol.* **119**, 579–589 (2010).
119. Sahara, N. et al. Age-related decline in white matter integrity in a mouse model of tauopathy: an in vivo diffusion tensor magnetic resonance imaging study. *Neurobiol. Aging* **35**, 1364–1374 (2014).
120. Desai, M. K. et al. Triple-transgenic Alzheimer's disease mice exhibit region-specific abnormalities in brain myelination patterns prior to appearance of amyloid and tau pathology. *Glia* **57**, 54–65 (2009).
121. Chen, J.-F. et al. Enhancing myelin renewal reverses cognitive dysfunction in a murine model of Alzheimer's disease. *Neuron* **109**, 2292–2307.e2295 (2021).
122. Rubinski, A. et al. Higher levels of myelin are associated with higher resistance against tau pathology in Alzheimer's disease. *Alzheimers Res. Ther.* **14**, 139 (2022).
123. Narasimhan, S. et al. Human tau pathology transmits glial tau aggregates in the absence of neuronal tau. *J. Exp. Med.* **217**, https://doi.org/10.1084/jem.20190783 (2019).

124. Chung, D.-E. C., Roemer, S., Petrucelli, L. & Dickson, D. W. Cellular and pathological heterogeneity of primary tauopathies. *Mol. Neurodegener.* **16**, 57 (2021).
125. Higuchi, M. et al. Axonal degeneration induced by targeted expression of mutant human tau in oligodendrocytes of transgenic mice that model glial tauopathies. *J. Neurosci.* **25**, 9434–9443 (2005).
126. De Rossi, P. et al. Predominant expression of Alzheimer's disease-associated BIN1 in mature oligodendrocytes and localization to white matter tracts. *Mol. Neurodegener.* **11**, 59 (2016).
127. Shen, Y. et al. Integrated analyses of 5 mC, 5hmC methylation and gene expression reveal pathology-associated AKT3 gene and potential biomarkers for Alzheimer's disease. *J. Psychiatr. Res.* **178**, 367–377 (2024).
128. Guemri, J. et al. Methylated ccfDNA from plasma biomarkers of Alzheimer's disease using targeted bisulfite sequencing. *Epigenomics* **14**, 451–468 (2022).
129. Tiane, A. et al. From methylation to myelination: epigenomic and transcriptomic profiling of chronic inactive demyelinated multiple sclerosis lesions. *Acta Neuropathol.* **146**, 283–299 (2023).
130. Moyon, S. et al. Functional characterization of DNA methylation in the oligodendrocyte lineage. *Cell Rep.* **15**, 748–760 (2016).
131. Arthur-Farraj, P. & Moyon, S. DNA methylation in Schwann cells and in oligodendrocytes. *Glia* **68**, 1568–1583 (2020).
132. Fodder, K., de Silva, R., Warner, T. T. & Bettencourt, C. The contribution of DNA methylation to the (dys)function of oligodendroglia in neurodegeneration. *Acta Neuropathol. Commun.* **11**, 106 (2023).
133. Nestor, C. E. et al. Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Res.* **22**, 467–477 (2012).
134. Globisch, D. et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **5**, e15367 (2010).
135. Szulwach, K. E. et al. Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet.* **7**, e1002154 (2011).
136. Choi, I., Kim, R., Lim, H.-W., Kaestner, K. H. & Won, K.-J. 5-hydroxymethylcytosine represses the activity of enhancers in embryonic stem cells: a new epigenetic signature for gene regulation. *BMC Genom.* **15**, 670 (2014).
137. Hahn, M. A. et al. Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. *Cell Rep.* **3**, 291–300 (2013).
138. Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S. & Heintz, N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151**, 1417–1430 (2012).
139. He, B. et al. Tissue-specific 5-hydroxymethylcytosine landscape of the human genome. *Nat. Commun.* **12**, 4249 (2021).
140. Chouliaras, L. et al. Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiol. Aging* **34**, 2091–2099 (2013).
141. Bradley-Whitman, M. A. & Lovell, M. A. Epigenetic changes in the progression of Alzheimer's disease. *Mech. Ageing Dev.* **134**, 486–495 (2013).
142. Coppieters, N. et al. Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. *Neurobiol. Aging* **35**, 1334–1344 (2014).
143. Zhao, J. et al. A genome-wide profiling of brain DNA hydroxymethylation in Alzheimer's disease. *Alzheimer's. Dement.* **13**, 674–688 (2017).
144. Loyfer, N. et al. A DNA methylation atlas of normal human cell types. *Nature* **613**, 355–364 (2023).
145. Reddy, J. S. et al. Bridging the gap: multi-omics profiling of brain tissue in Alzheimer's disease and older controls in multi-ethnic populations. *Alzheimers Dement.* **20**, 7174–7192 (2024).
146. Murray, M. E. et al. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. *Lancet Neurol.* **10**, 785–796 (2011).
147. Murray, M. E. et al. Clinicopathologic and 11C-Pittsburgh compound B implications of Thal amyloid phase across the Alzheimer's disease spectrum. *Brain* **138**, 1370–1381 (2015).
148. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, https://doi.org/10.1186/s13742-015-0047-8 (2015).
149. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
150. Sun, Z. et al. SAAP-RRBS: streamlined analysis and annotation pipeline for reduced representation bisulfite sequencing. *Bioinformatics* **28**, 2180–2181 (2012).
151. Xi, Y. & Li, W. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinform.* **10**, 232 (2009).
152. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
153. Kalari, K. R. et al. MAP-RSeq: Mayo Analysis Pipeline for RNA sequencing. *BMC Bioinform.* **15**, 224 (2014).
154. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
155. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
156. Hansen, K. D., Irizarry, R. A. & Wu, Z. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* **13**, 204–216 (2012).
157. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
158. Zhong, Y., Wan, Y.-W., Pang, K., Chow, L. M. L. & Liu, Z. Digital sorting of complex tissues for cell type-specific gene expression profiles. *BMC Bioinform.* **14**, 89 (2013).
159. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
160. Shabalin, A. A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353–1358 (2012).
161. Cavalcante, R. G. & Sartor, M. A. annotatr: genomic regions in context. *Bioinformatics* **33**, 2381–2383 (2017).
162. Aryee, M. J. et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
163. Pidsley, R. et al. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genom.* **14**, 293 (2013).
164. Wang, Y. et al. DNA methylation-based sex classifier to predict sex and identify sex chromosome aneuploidy. *BMC Genom.* **22**, 484 (2021).
165. McCartney, D. L. et al. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genom. Data* **9**, 22–24 (2016).
166. van Iterson, M., van Zwet, E. W. & Heijmans, B. T. & the, B. C. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol.* **18**, 19 (2017).
167. Bennett, D. A. et al. Religious orders study and rush memory and aging project. *J. Alzheimers Dis.* **64**, S161–S189 (2018).

Acknowledgements

We thank the patients and families for their participation and tissue donation, without whom these studies would not have been possible. We thank Özkan İş, PhD, Jianna Tan, and Jeremiah Bergman for their technical assistance. We thank the Mayo Clinic Genome Analysis Core (GAC), Co-Directors, Julie M. Cunningham, PhD and Eric Wieben, PhD, and supervisor Julie Lau, for their collaboration in collection of omics

data. We also thank our colleague Saurabh Baheti at the Mayo Clinic Bioinformatic Core (BIC) for their collaboration in data quality control. *AD Knowledge Portal*: The results published here are in whole or in part based on data obtained from the AD Knowledge Portal (https://adknowledgeportal.org). *Mayo Clinic AD-CAA*: The Mayo Clinic AD-CAA study was led by Dr. Guojun Bu and Dr. Nilüfer Ertekin-Taner at the, Mayo Clinic, Jacksonville, FL as part of the multi-PI RF1AG051504 (MPIs Bu and Ertekin-Taner) using samples from the Mayo Clinic Brain Bank. Data collection was supported through funding by NIA grants P50AG016574, R37AG027924, Cure Alzheimer's Fund, and support from Mayo Foundation. *AMP-AD datasets*: (https://doi.org/10.7303/syn2580853). *Mayo Clinic*: The Mayo RNAseq study data was led by Dr. Nilüfer Ertekin-Taner, Mayo Clinic, Jacksonville, FL as part of the multi-PI U01 AG046139 (MPIs Golde, Ertekin-Taner, Younkin, Price). Samples were provided from the following sources: The Mayo Clinic Brain Bank and Banner Sun Health Research Institute. Data collection was supported through funding by NIA grants P50 AG016574, R01 AG032990, U01 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216, R01 AG003949, NINDS grant R01 NS080820, CurePSP Foundation, and support from Mayo Foundation. Study data includes samples collected through the Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona. The Brain and Body Donation Program is supported by the National Institute of Neurological Disorders and Stroke (U24 NS072026 National Brain and Tissue Resource for Parkinsons Disease and Related Disorders), the National Institute on Aging (P30 AG19610 Arizona Alzheimers Disease Core Center), the Arizona Department of Health Services (contract 211002, Arizona Alzheimers Research Center), the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901 and 1001 to the Arizona Parkinson's Disease Consortium) and the Michael J. Fox Foundation for Parkinsons Research. *MSBB*: These data were generated from post-mortem brain tissue collected through the Mount Sinai VA Medical Center Brain Bank and were provided by Dr. Eric Schadt from Mount Sinai School of Medicine. *ROSMAP*: Study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by NIA grants P30AG10161 (ROS), R01AG15819 (ROSMAP; genomics and RNAseq), R01AG17917 (MAP), R01AG30146, R01AG36042 (5hC methylation, ATACseq), RC2AG036547 (H3K9Ac), R01AG36836 (RNAseq), R01AG48015 (monocyte RNAseq) RF1AG57473 (single nucleus RNAseq), U01AG32984 (genomic and whole exome sequencing), U01AG46152 (ROSMAP AMP-AD, targeted proteomics), U01AG46161(TMT proteomics), U01AG61356 (whole genome sequencing, targeted proteomics, ROSMAP AMP-AD), the Illinois Department of Public Health (ROSMAP), and the Translational Genomics Research Institute (genomic). Additional phenotypic data can be requested at www.radc.rush.edu. *AGORA*: The results published here are in whole or in part based on data obtained from Agora, a platform initially developed by the NIA-funded AMP-AD consortium that shares evidence in support of AD target discovery (agora.adknowledgeportal.org/). *RADC*: We thank the study participants and staff of the Rush Alzheimer's Disease Center. ROSMAP[167] is supported by P30AG10161, P30AG72975, R01AG15819, R01AG17917, U01AG46152, and U01AG61356. ROSMAP resources can be requested at https://www.radc.rush.edu and www.synpase.org. All ROSMAP participants are enrolled without known dementia and agreed to detailed clinical evaluation and brain donation at death. Both ROS and MAP studies were approved by an Institutional Review Board of Rush University Medical Center. Each participant signed an informed consent, Anatomic Gift Act, and an RADC Repository consent allowing their data and biospecimens to be repurposed. *UK Brain Bank Network: Brains for Dementia Research*: The provision of data used in this study was supported by the UK Brain Banks Network with funding from the UK Medical Research Council and BDR (Brains for Dementia Research). This work was supported by National Institute on Aging [RF1 AG051504 to N.E.T. and G.B., U01 AG046139 R01 AG061796, and U19 AG074879N.E.T.]. N.E.T. is also supported by the Alzheimer's Association Zenith Fellows Award (ZEN-22-969810).

Author contributions

S.R.O. and N.E.T. wrote the manuscript; S.R.O., N.E.T., and M.A. designed the study; S.R.O., Z.Q., A.A., J.S.R., X.W., M.M.C., and F.V. performed data analyses; MH consulted on statistical methods; S.R.O. built the web application; Y.M. assisted with web application development; C.C.L., Y.Y., N.Z., T.K., and G.B. collected biochemical measures; D.W.D., M.D., and M.E.M. provided neuropathological data and tissue samples; T.T.N. isolated DNA from tissue samples; N.E.T. oversaw the study and provided direction, funding, and resources. All authors reviewed and contributed to the manuscript. The authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-68864-9.

Correspondence and requests for materials should be addressed to Nilüfer Ertekin-Taner.

Peer review information *Nature Communications* thanks Rebecca Smith, Vivek Swarup, who co-reviewed with Neelakshi Soni, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.