nature neuroscience

ARTICLES

https://doi.org/10.1038/s41593-018-0291-1

Epigenome-wide study uncovers large-scale changes in histone acetylation driven by tau pathology in aging and Alzheimer's human brains

Hans-Ulrich Klein[1,2], Cristin McCabe[2], Elizabeta Gjoneska[2,3], Sarah E. Sullivan[4], Belinda J. Kaskow[4], Anna Tang[2], Robert V. Smith[4], Jishu Xu[2], Andreas R. Pfenning[5], Bradley E. Bernstein[6], Alexander Meissner[7], Julie A. Schneider[8], Sara Mostafavi[9,10,11], Li-Huei Tsai[2,3], Tracy L. Young-Pearse[2,4], David A. Bennett[8,12] and Philip L. De Jager[1,2,12]*

Accumulation of tau and amyloid-β are two pathologic hallmarks of Alzheimer's disease. We conducted an epigenome-wide association study using the histone 3 lysine 9 acetylation (H3K9ac) mark in 669 aged human prefrontal cortices; in contrast with amyloid-β, tau protein burden had a broad effect on the epigenome, affecting 5,990 of 26,384 H3K9ac domains. Tau-related alterations aggregated in large genomic segments reflecting spatial chromatin organization, and the magnitude of these effects correlated with the segment's nuclear lamina association. Functional relevance of these chromatin changes was demonstrated by (1) consistent transcriptional changes in three independent datasets and (2) similar findings in two mouse models of Alzheimer's disease. Finally, we found that tau overexpression in induced pluripotent stem cell-derived neurons altered chromatin structure and that these effects could be blocked by a small molecule predicted to reverse the tau effect. Thus, we report broad tau-driven chromatin rearrangements in the aging human brain that may be reversible with heat-shock protein 90 (Hsp90) inhibitors.

Alzheimer's disease (AD) is a chronic neurodegenerative disease characterized pathologically by the accumulation of amyloid-β plaques and tau tangles, which leads to neuronal cell death, cognitive impairment and, ultimately, a diagnosis of dementia. Although loci harboring genetic risk factors have been identified in genome-wide studies[1], a large portion of late-onset AD dementia risk remains unexplained, thus indicating the need for complementary approaches such as exploring the epigenome of the aging brain. Epigenomic alterations can be caused by genetic and non-genetic risk factors such as life experiences and environmental exposures, but they can also occur as a consequence of AD pathologies[2]. Thus, studying the AD epigenome may also be helpful for understanding molecular events resulting from the toxicity of AD pathologies. Evidence for epigenomic perturbations in AD was found in smaller human studies[3]. For example, a recent study characterized the redistribution of histone 4 lysine 16 acetylation (H4K16ac) in the temporal cortex of 12 subjects with AD compared with controls[4], but, so far, large AD-related epigenome-wide association studies of the human cortex have been limited to DNA methylation[5,6]. These studies demonstrated reproducible DNA methylation changes in subjects with AD, but did not explicitly distinguish between amyloid-β- and tau-related alterations. Notably, tau pathology has recently been associated with epigenetic changes in model systems. In *Drosophila*, tau overexpression has been shown to relax heterochromatin[7,8]. There is also evidence for a physiological function of nuclear tau in maintaining and regulating heterochromatin, which may be lost by pathologically phosphorylated tau[9]. However, whether these mechanisms cause major chromatin alterations in the human AD brain, translate to transcriptional alterations, and are restricted to heterochromatic regions remains unknown.

Here, we studied the acetylation of the ninth lysine of histone 3 (H3K9ac), which marks transcriptionally active open chromatin genome-wide in the dorsolateral prefrontal cortex (DLPFC) of 669 subjects. Our data support the hypothesis that tau but not amyloid-β causes widespread chromatin remodeling. We mapped the location of these alterations genome-wide, characterized the architecture of these large areas of coordinated chromatin remodeling, replicated them in transcriptomic data, demonstrated that tau is sufficient to cause such chromatin rearrangements prior to tangle formation, and identified a compound that may attenuate this chromatin perturbation.

Results

Tau but not amyloid-β pathology has a broad effect on histone acetylation in the human brain. We studied H3K9ac in the DLPFC of 669 participants enrolled in either the Religious Order Study (ROS) or the Rush Memory and Aging Project (MAP), two longitudinal studies of aging and dementia[10,11]. Participants did not

[1]Center for Translational & Computational Neuroimmunology, Department of Neurology, Columbia University Medical Center, New York, NY, USA. [2]Broad Institute, Cambridge, MA, USA. [3]Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, MA, USA. [4]Department of Neurology, Ann Romney Center for Neurologic Diseases, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [5]Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA, USA. [6]Department of Pathology, Massachusetts General Hospital, Boston, MA, USA. [7]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA. [8]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA. [9]Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada. [10]Department of Statistics, University of British Columbia, Vancouver, British Columbia, Canada. [11]Centre for Molecular Medicine and Therapeutics, Vancouver, British Columbia, Canada. [12]These authors contributed equally: David A. Bennett, Philip L. De Jager. *e-mail: pld2115@cumc.columbia.edu

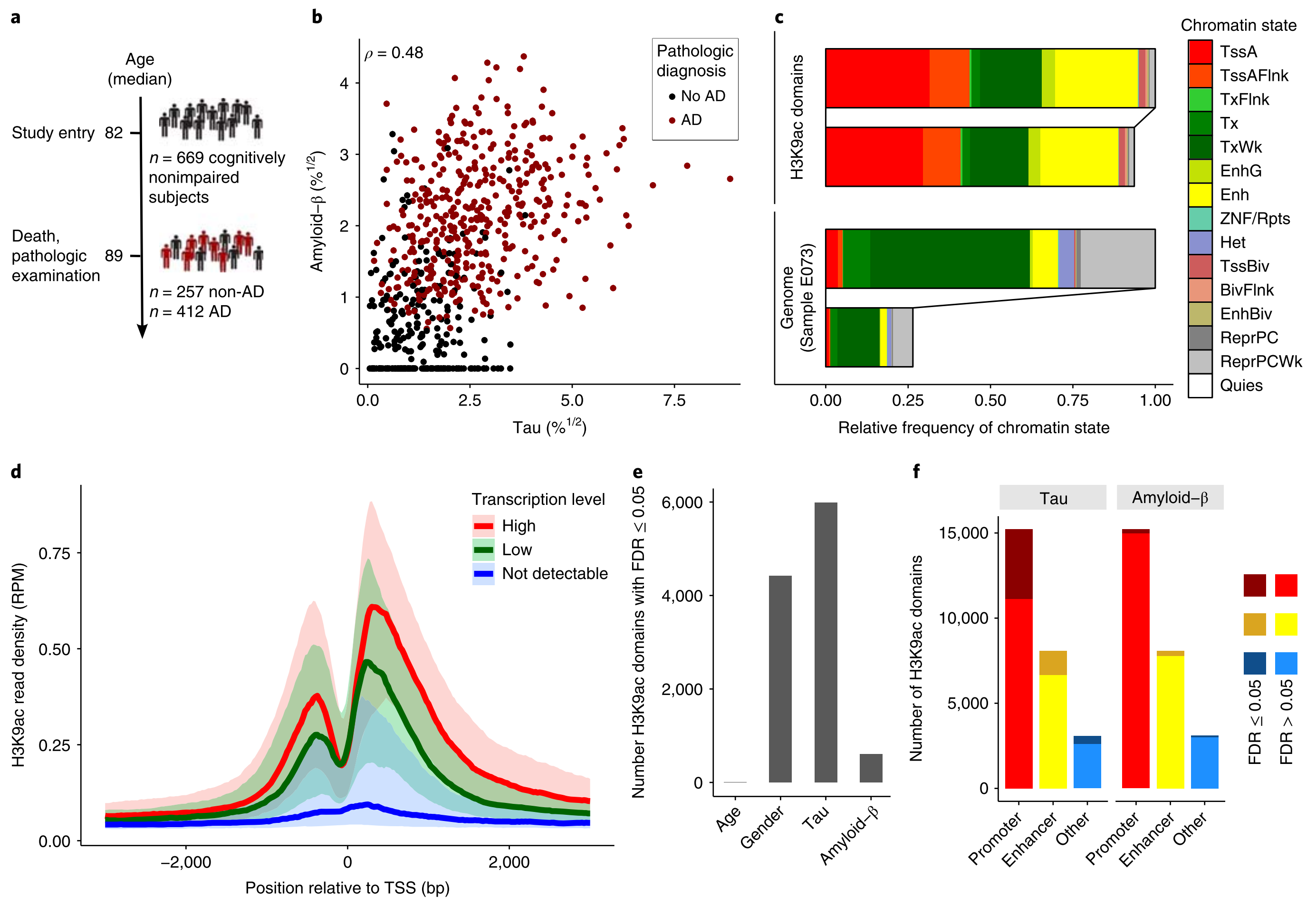


Fig. 1 | The active histone mark H3K9ac is broadly associated with tau pathology in the human cortex. **a**, H3K9ac was studied in the prefrontal cortex of 669 subjects. **b**, Pearson correlation between tau and amyloid-β levels was 0.48 in this cohort. Amyloid-β peptide levels were below the detection threshold for 103 samples. **c**, Horizontal bars depict the relative frequency of each chromatin state within the H3K9ac domains (97 Mbp) and the whole genome (3.1 Gbp). Chromatin states were obtained from a DLPFC sample with minimal neuropathology (E073) included in the Roadmap Epigenomics Project: 15-state model (TssA, active TSS TssA; TssAFlnk, flanking active TSS; TxFlnk, transcription at gene 5′ and 3′; Tx, strong transcription; TxWk, weak transcription; EnhG, genic enhancer; Enh, enhancer; ZNF/Rpts, ZNF genes and repeats; Het, heterochromatin; TssBiv, bivalent or poised TSS; BivFlnk, flanking bivalent TSS and Enh; EnhBiv, bivalent enhancer; ReprPC, repressed polycomb; ReprPCWk, weak repressed polycomb; Quies, quiescent or low). For better clarity, the respective upper bars depict the relative frequencies without the quiescent state. **d**, Median H3K9ac ChIP-seq read coverage is shown on the *y* axis for highly transcribed, lowly transcribed and not detectable RefSeq transcripts within ±3 kb around the TSS. Transcripts were grouped by median transcription values (high: $x > 22$ reads per kilobase of transcript, per million mapped reads (RPKM) (0.95 quantile), low: 1 RPKM $< x <$ 1.5 RPKM (0.6–0.65 quantiles), not detectable: $x = 0$). The median of the H3K9ac read density across all samples was calculated for each transcript. Then, the median of all transcripts within each of the three groups was calculated, plotted as a solid line. The 25% and 75% quantiles for each group are indicated by the transparent bands. **e**, Bars depict the number of H3K9ac domains that were significantly associated at an FDR of 0.05 with age, gender, tau levels and amyloid-β levels. **f**, Bars depict the number of H3K9ac domains stratified by promoter, enhancer and other domains. Darker shading indicates the number of domains whose H3K9ac levels were significantly associated with tau or amyloid-β.

have dementia upon study entry (Fig. 1a). At autopsy, neuropathologic examination was performed, and quantitative measurements of the density of phosphorylated tau tangles and the burden of amyloid-β were obtained (Table 1). A wide spectrum of tau tangles and amyloid-β burdens was observed in our subjects, and there is a moderate correlation of $\rho = 0.48$ between these two AD pathologies (Fig. 1b). Chromatin immunoprecipitation sequencing (ChIP-seq) was performed in isolated gray matter from frozen DLPFC samples to generate genome-wide H3K9ac profiles. A median of 55 million 36 base pair (bp) single-end reads were sequenced per sample (Supplementary Table 1).

We found 26,384 H3K9ac peaks or 'domains' that showed a distinct signal above the genome-wide background in at least 100 of the 669 subjects (Supplementary Table 2). As with DNA methylation[5], there is limited inter-individual variability; the average correlation of domains between two individuals was 0.98, despite the vast differences in life experiences between the older individuals. Nearly half (41%) of the genomic regions covered by H3K9ac domains in our data were annotated as containing an active transcription start site (TSS), and more than one-quarter (27%) were annotated as being in an enhancer site (Fig. 1c), based on the chromatin state annotation generated for this cortical region by the Roadmap Epigenomics Project (sample E073; a MAP subject with minimal neuropathology)[12]. We thus binned H3K9ac peaks into 'promoter' ($n = 15{,}225$), 'enhancer' ($n = 8{,}071$) and 'other' ($n = 3{,}088$) domains (Supplementary Fig. 1a–d) for subsequent analyses. Using RNA-seq data from the same region in a subset of subjects ($n = 500$), we verified the expected positive correlation between H3K9ac and transcriptional activity (Fig. 1d).

To distinguish between tau- and amyloid-β-related epigenomic changes, we modeled both pathologies simultaneously as explanatory

Table 1 | Summary characteristics of the ROS/MAP dataset

	H3K9ac ($n=669$; ChIP-seq)	Transcription ($n=500$; RNA-seq)	DNA methylation ($n=729$; BeadChip)
Age (years)	88.3 (± 6.5)	88.5 (± 6.6)	88.0 (± 6.6)
Male gender	233 (35%)	189 (38%)	267 (37%)
Tau protein load (%$^{-1/2}$)	2.2 (± 1.4)	2.1 (± 1.3)	2.1 (± 1.4)
Amyloid-β load (%$^{-1/2}$)	1.5 (± 1.1)	1.5 (± 1.1)	1.5 (± 1.1)
Pathologic AD diagnosis	412 (62%)	290 (58%)	441 (60%)
Post mortem interval (h)	7.5 (± 5.9)	7.1 (± 4.9)	7.4 (± 5.8)

Mean values and standard deviation in parentheses are shown for quantitative variables. Number of cases and percentages in parentheses are shown for categorical variables. H3K9ac and transcription data were available for 452 samples, H3K9ac and DNA methylation data were available for 645 samples, and transcription and DNA methylation data were available for 496 samples. All three data types were available for 449 samples.

variables in a regression model with the H3K9ac levels as the outcome. At a false discovery rate (FDR) of 0.05, 23% of H3K9ac domains showed an association with tau, whereas only 2% were associated with amyloid-β (Fig. 1e and Supplementary Table 3). Only a few domains ($n=88$) had acetylation levels that were significantly associated with both pathologies. This striking and unexpected difference in the impact of tau and amyloid-β may be related to the fact that tau pathology initially accumulates intracellularly and may directly affect neuronal chromatin organization, whereas amyloid-β is secreted and affects cells extracellularly. For tau, the largest proportion of associated domains was observed among promoter regions (Fig. 1f), but the estimated effect sizes were similar for domains found in promoters, enhancers and other regions (Supplementary Fig. 1e,f), indicating that the mechanism driving these associations is not specific for a type of domain. The greater average read depth at promoters probably explains the larger number of significant domains in these regions. Interestingly, known AD risk loci[1] were not enriched in tau-associated H3K9ac domains: out of 39 H3K9ac domains within 50 kb of 19 known AD loci, nine were associated with tau, corresponding to the genome-wide relative frequency of 23% (Supplementary Table 4).

Next, we investigated whether tau-related changes in H3K9ac led to functional consequences by analyzing our RNA-seq data. We repeated the simultaneous evaluation of amyloid and tau pathology for 18,257 out of 24,594 active transcripts that could be mapped to an H3K9ac domain. We observed a positive but weak correlation of 0.14 between the coefficients for tau (0.12 for amyloid) from the transcription and the H3K9ac data. In general, epigenomic and transcriptomic data showed the same direction of effect (Supplementary Fig. 1g,h and Supplementary Table 5), with notable exceptions that indicate the presence of other regulatory mechanisms. The modest extent of the correlation is probably attributable to the fact that H3K9ac is not sufficient by itself to define a transcriptionally active region (Supplementary Fig. 1i).

Spatial pattern of tau-related alterations in H3K9ac reflects higher-order chromatin organization. Because of the large number of tau-associated H3K9ac domains, we evaluated the spatial distribution of tau-associated domains using a plot that presents the physical location, significance, and directionality of each tested domain's association with tau (Fig. 2a and Supplementary Fig. 2a). Throughout the genome, we found large-scale genomic segments whose H3K9ac domains were coordinately enriched for associations with tau. Unlike genome-wide single nucleotide polymorphism (SNP) studies in which we see small genomic regions associated with disease because linkage disequilibrium is observed only over ~18 kilobase pairs (kbp) in most regions of the human genome[13], we see a clustering of disease-related epigenomic associations in certain segments of the genome that cover several megabase pairs (Mbp). These results suggest that a physical aspect of the chromosome may be implicated in disease. To define the boundaries of these large-scale segments, we applied a segmentation algorithm[14] to our data, and this method divided the genome into 178 segments, within which domains display a similar tau association (Fig. 2a). The median size of these segments was 5.2 Mbp, and they contain a median of 79.5 H3K9ac domains per segment (Supplementary Table 6). In contrast to tau, domains associated with amyloid-β pathology showed no similar spatial pattern (Supplementary Fig. 2b), thus suggesting a fundamental difference in the relation of these pathologies to the human cortex epigenome.

At the megabase-pair scale, chromatin is spatially organized into type A and type B compartments[15]. Type A compartments are primarily characterized by open chromatin, and type B compartments are characterized by closed chromatin. To study whether the spatial chromatin architecture underlies the observed spatial pattern of H3K9ac alterations, we derived a compartment map from previously published Hi-C data of the cortical and subcortical plate of a human fetal brain[16]. The genome was divided into 636 compartments, and H3K9ac domains were mapped to compartments (locations of H3K9ac domains and type A and B compartments for chromosome 1 are shown in Figure 2a). Genome-wide, the majority of H3K9ac domains (70%) were located in type A compartments, but we also observed 223 type B compartments out of 507 compartments that contained at least five H3K9ac domains. For these 507 compartments, we calculated the mean tau effect on H3K9ac and found a significantly larger tau effect in type A compartments compared with that in type B compartments (Fig. 2b), indicating that the tau effect on H3K9ac is associated with the spatial chromatin organization.

A critical element of chromatin structure is the nuclear lamina because it provides anchor points that couple chromatin to the lamina. Nuclear lamina is associated with repressive chromatin and type B compartments. Recent work in *Drosophila* suggests that tau can disrupt lamina function and thereby induce heterochromatin relaxation[7]. To elucidate the role of the lamina in our observations that center on euchromatin, we calculated the lamina association of our tau-defined segments as the proportion of lamina-associated domains (LADs) within a segment using previously published genome-wide DamID lamin B1 data from the human HT1080 fibrosarcoma cell line[17]. Although some LADs are cell-type specific, most LADs are strongly conserved, so that LADs derived from human fibroblasts are still useful for annotating whole brain tissue[17,18]. We plotted the segments' lamina association versus the average tau effect of all H3K9ac domains within a segment (Fig. 2c). The negative Pearson correlation ($\rho=-0.86$) indicates that the effect of tau pathology on single H3K9ac domains varies depending on the lamina association of the surrounding genomic region. The 178 genomic segments within which tau associations are correlated explain 29% of the variance of the tau effects observed across all 26,384 H3K9ac domains. Thus, a large portion of the tau effect can be explained by chromatin organization.

We next evaluated our RNA-seq data and found that each segment's mean tau effect obtained from the RNA-seq data was correlated ($\rho=0.83$) with the mean tau effects from the H3K9ac data (Fig. 2d), reflecting a shift in average gene transcription that is related to the burden of tau pathology and is concordant with the chromatin alterations of the segments. However, in line with the weak correlation of alterations in H3K9ac and RNA levels at

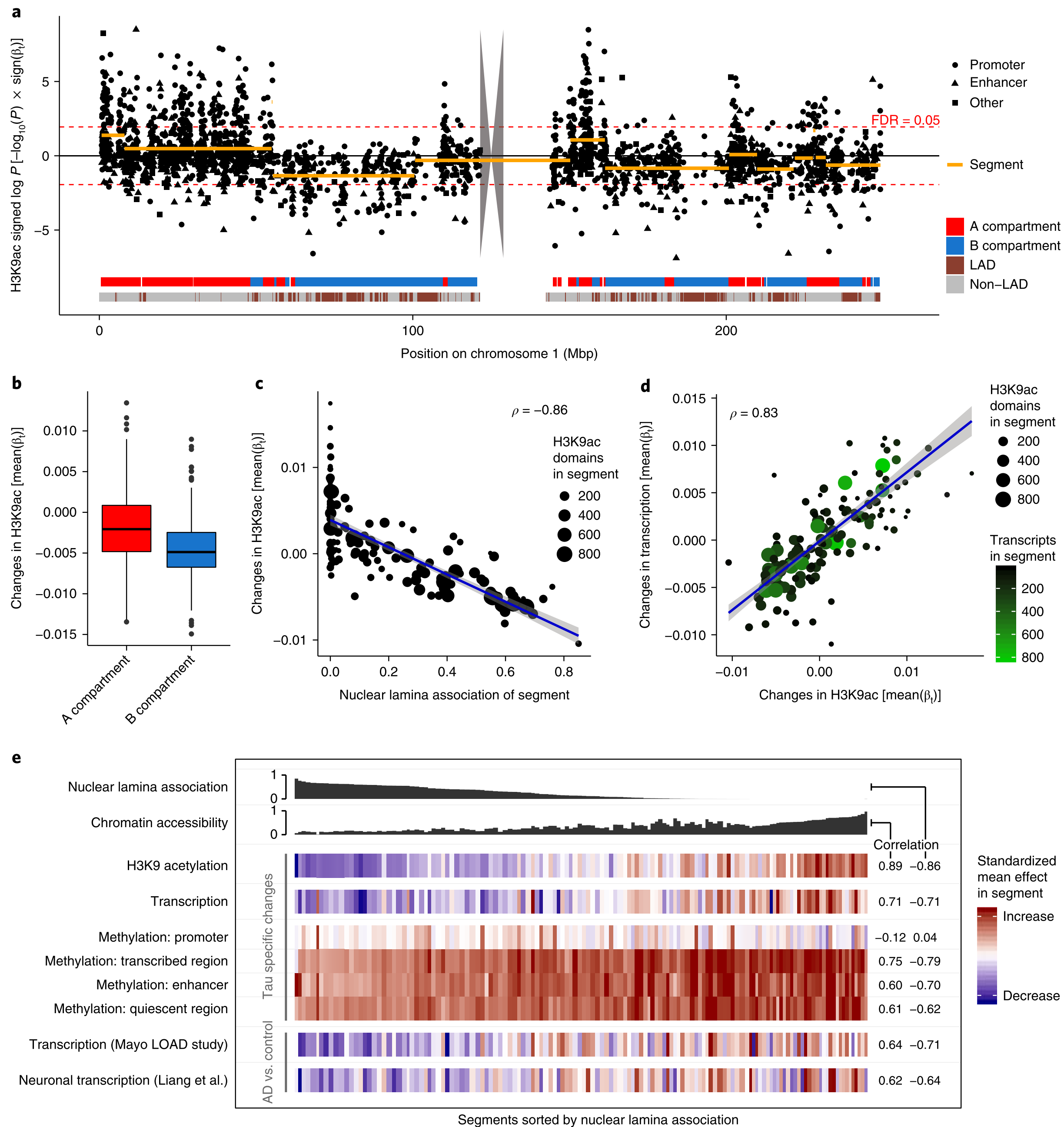


Fig. 2 | Tau-related chromatin alterations show spatial patterns. a, Manhattan plot depicts log-transformed unadjusted *P* values for the tau coefficients of all H3K9ac domains along chromosome 1 (two-sided Wald tests; $n = 669$ subjects). Transformed *P* values were plotted with the sign of the respective coefficient to distinguish between positive and negative associations. Red dashed lines indicate FDR threshold of 0.05 when adjusting for all 26,384 domains. Broad genomic segments covering H3K9ac domains showing associations with tau similar in direction and strength are depicted as orange lines. The gray polygon represents the chromosome's centromere. Two ribbons at the bottom show chromatin-structure-related annotation. The top ribbon indicates type A and B compartments. The bottom ribbon indicates LADs. **b**, Boxplot shows larger mean tau-related changes in H3K9ac levels in A compartments ($n = 284$) than in B compartments ($n = 223$) (Mann–Whitney–Wilcoxon test, $P = 8 \times 10^{-13}$). The center line of the boxplot shows the median, the outer edges correspond to the lower and upper quartiles, and the whiskers extend to the most extreme observed value within 1.5 × IQR from the edges. Compartments with no or very few H3K9ac domains (<5) were discarded. **c**,**d**, Scatter plots show the association between segments' mean tau-related changes in H3K9ac levels and nuclear lamina association (**c**) and mean tau-related changes in mRNA levels (**d**). Each dot represents a segment ($n = 138$; sex chromosomes were excluded). Weighted linear regression line (blue line), 95% confidence interval of the mean change (shaded region), and weighted Pearson correlation are depicted in the plots. Segments were weighted by the number of H3K9ac domains. **e**, Columns of the heat map represent segments sorted by lamina association first and then by chromatin accessibility to break ties, because multiple segments were completely free of LADs. Chromatin accessibility was calculated using DLPFC-specific annotation from the Roadmap Epigenomics Project (Methods). Column widths reflect segment sizes. In rows 1–6, color indicates the mean tau effect size observed in H3K9ac ($n = 669$ subjects), RNA transcription ($n = 500$ subjects), and DNA methylation ($n = 729$ subjects) at active TSSs, weakly transcribed regions, enhancers and quiescent regions. In the bottom two rows, color indicates the mean difference between AD and control samples observed in published transcription data from temporal cortex tissue[21] (row 7, $n = 292$ subjects) and dissected neurons[24] (row 8, $n = 34$ subjects). Weighted Pearson correlations between the mean tau effect in the segments and histone accessibility (left column) and nuclear lamina association (right column) are shown on the right. Segments were weighted by the number of H3K9ac domains in the segments.

the single-transcript level, only 2% of the variance of tau-related effects in RNA-seq data could be explained by the segments defined with the chromatin data. This finding suggests that, while large-scale epigenomic changes explain some of the effect of tau on the transcriptome, other regulatory mechanisms must also be influenced by this neuropathologic process.

We then speculated that other epigenetic marks may show similar tau-related patterns. In evaluating our DNA methylation data from the same brain region[5], we found consistent evidence for a perturbation of the epigenomic architecture. Specifically, we binned CpGs into four groups using the tissue-specific reference chromatin state map from the Roadmap Epigenomics Project[12] (sample E073): active TSSs, enhancers, weakly transcribed regions and quiescent regions. These four chromatin states cover most of the CpGs in our dataset, and we repeated the segment-based analysis for each group of CpGs. A negative correlation between tau effects on DNA methylation and nuclear lamina association was observed in enhancers, quiescent regions and weakly transcribed regions (as shown in Fig. 2e), which is in agreement with the paradigm that high methylation of gene bodies is a feature of active genes[19]. The variance of tau effects in DNA methylation data that could be explained by the H3K9ac-derived segments ranged 6% for enhancer regions, 4% for transcribed regions, and 1% for quiescent regions (Supplementary Table 7). By contrast, no distinct correlation was observed at active TSSs, in line with previous work reporting that promoter regions are spared from altered methylation in other diseases[20].

To validate our findings in an independent set of samples, we assessed a published transcriptomic dataset from the temporal cortex of subjects from a late-onset AD (LOAD) case–control study[21]. Because quantitative amyloid-β and tau burdens were not available, we compared samples from individuals with an AD diagnosis ($n = 202$) with control subjects ($n = 90$; individuals who do not meet diagnostic criteria for AD or progressive supranuclear palsy diagnosis) and calculated the mean AD-related change in transcription over the genes within each of the 178 genomic segments defined using our H3K9ac data. Although we could not distinguish the tau from the amyloid-β effects in these data, we did confirm the presence of a negative correlation ($\rho = -0.71$) between the segments' nuclear lamina association and AD-related transcriptional changes (Fig. 2e). Similar to our RNA-seq data, 4% of the variance of AD effects in the transcriptomic data of the LOAD study could be explained by the segments. Nevertheless, as in our RNA-seq data, the mean tau-related changes in the H3K9ac data were positively correlated ($\rho = 0.65$) with the mean AD-related changes in the transcription data from this LOAD study, thus validating our observation.

Because the data analyzed so far were generated from bulk tissue, it is critical to assess whether the observed alterations may be driven by changing cell type compositions. We estimated the proportion of neurons, astrocytes, oligodendrocytes, myeloid cells and endothelial cells from our RNA-seq data using a set of five marker genes per cell type[22] (Supplementary Table 8). The proportion of neurons was modestly decreased ($P = 0.043$, two-sided t test, $n = 452$ subjects) in subjects with AD, indicating that cell type composition might be a confounder. However, the coefficients for tau derived from a model adjusted for cell type proportions fitted on a subset of $n = 452$ subjects that had both RNA-seq and H3K9ac data were highly correlated ($\rho > 0.99$) with the tau coefficients from the unadjusted model used in our primary analysis (Supplementary Fig. 3). Similarly, P values did not change noticeably after adjustment for cell type proportions. While we cannot exclude confounding by changing cell type compositions completely[23], these results indicate that varying cell type compositions are probably not the main driver of our observations, and we decided to further investigate which cell type or types may be affected by tau pathology.

Altered transcription due to tau-related chromatin remodeling occurs in neurons. A priori, neurons are the most likely candidate cell type, because intracellular phosphorylated tau is known to accumulate and form tangles in neurons. To explore this hypothesis, we repurposed a gene transcription dataset generated from laser-capture microdissected neurons from the superior frontal gyrus of individuals with AD ($n = 23$) and control subjects ($n = 11$)[24]. Because tau and amyloid-β loads were not reported in these subjects, we estimated the AD effect instead of the tau effect for each gene. The AD effects were then averaged within each of the 178 genomic segments defined in our H3K9ac data, and the correlation with the segment's nuclear lamina association was calculated. Again, we observed a negative correlation of $\rho = -0.64$ (Fig. 2e), and the correlation of these neuronal data with the average tau effect in our H3K9ac data was $\rho = 0.67$. Similar to the tissue transcriptomic data, 2% of the AD effect variance was explained by the segments. Interestingly, the original report of these neuronal data specifies that neurons lacking neurofibrillary tangles were selected for laser capture, indicating that the observed AD-related transcriptional changes are occurring early in pathogenesis, prior to the accumulation of tangles[24]. Because the transcriptional changes consistent with our epigenomic changes are found in purified neurons, they are likely to be cell autonomous for neurons; however, we cannot rule out that other cell types may also be affected.

H3K9ac data from tau mouse models suggest similar structural changes in the murine brain epigenome that involve the nuclear lamina. To explore whether the epigenomic changes that we found in human brain are recapitulated in mouse models known to accumulate tau, we generated hippocampal H3K9ac ChIP-seq profiles from two different mouse models at (1) an early time point and (2) a late stage of neurodegeneration (Supplementary Table 9). Specifically, we studied 6- and 11-month old mutant tau mice (*MAPT* P301S), which start to accumulate phosphorylated tau in neurons by 6 months[25]. Wild-type mice of the same age were used as controls. The second mouse model was the CK-p25 model, which is characterized by increased amyloid-β levels early after p25 induction followed by increased tau phosphorylation and neuronal loss at later stages[26,27]. Three-month-old CK-p25 mice were studied 2 weeks and 6 weeks after p25 induction and were compared to CK littermate controls. We found 44,165 H3K9ac domains across the epigenome of both mouse models. Tau mice showed less significantly altered H3K9ac domains compared with the CK-p25 mice, especially at the early stage of 6 months (Fig. 3a). Chromosome-wide plots did not show the explicit spatial pattern observed in the human cortex, possibly because of the low sample size of 3 mice per time point. We then classified the mouse H3K9ac domains into two bins: 'close to lamina' (if domain was <50 kb from a LAD) or 'distant to lamina' based on published DamID lamin B1 data from mouse embryonic fibroblasts[17]. As anticipated, we observed smaller differences between AD mice and control mice in H3K9ac domains that were located in the proximity of LADs compared with those that were distant (Fig. 3b,c). This observation was consistent for both mouse models and more distinct at the later time points, during which the pathology accumulated more extensively, confirming our observations from the human cortex, where the effect of tau was smaller in segments enriched with LADs.

Induced neurons overexpressing *MAPT* demonstrate chromatin structure alterations. Given the evidence from neuronal expression data that the observed changes in chromatin modifications are likely to occur in neurons, we assessed whether overexpression of the 4R isoform of *MAPT*, which encodes the tau protein, in forebrain neurons derived from human induced pluripotent stem cells (iNs) could recapitulate the epigenomic changes that we found in our cortical H3K9ac data. In iNs bearing a familial AD mutation,

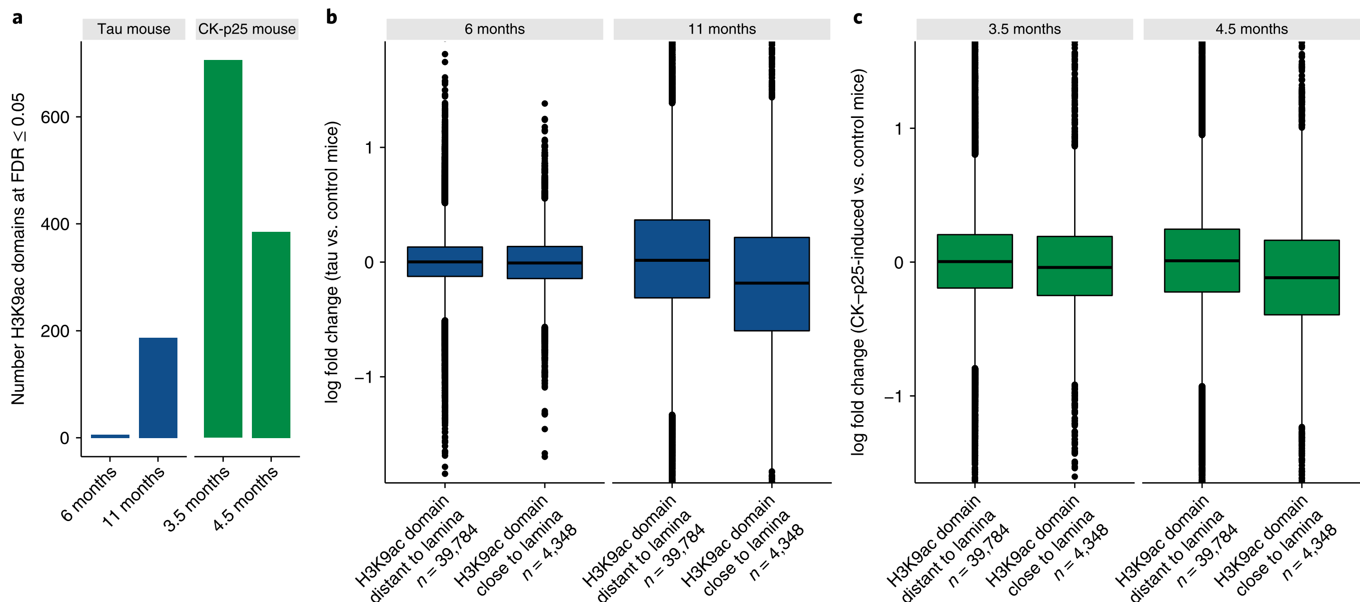


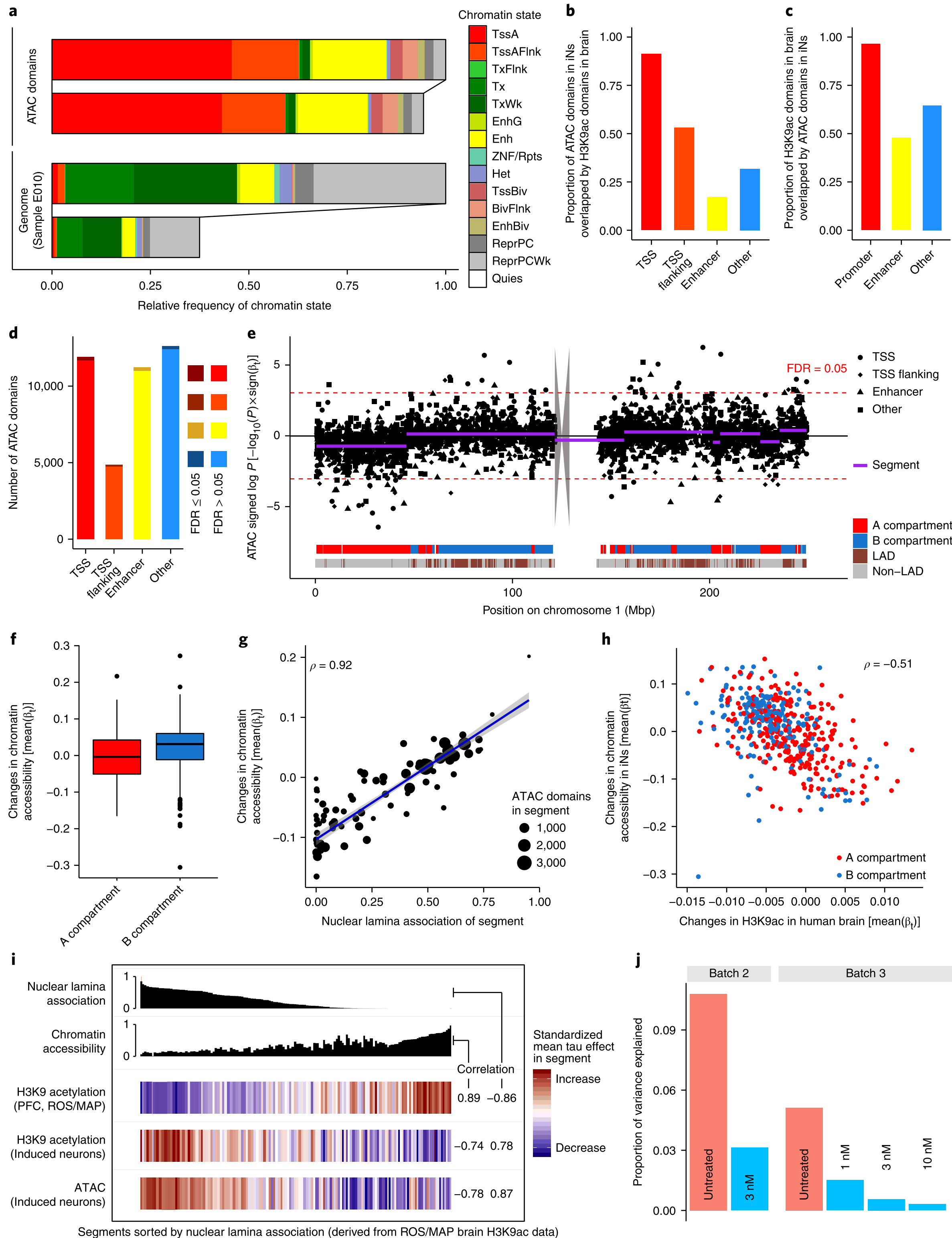
Fig. 3 | H3K9ac alterations in AD mouse models reflect spatial pattern observed in the human cortex. a, Bar plot shows the number of domains with significantly different H3K9ac levels observed in the tau mouse model (blue bars) and the CK-p25 mouse model (green bars) compared with respective control mice for different time points. CK-p25 mice were 3 months old when p25 was induced. **b**,**c**, Boxplots depict differences in H3K9ac levels between tau mice (**b**) or CK-p25 mice (**c**) and respective control mice separately for H3K9ac domains distant to and close to LADs. The center line shows the median, the outer edges correspond to the lower and upper quartiles, and the whiskers extend to the most extreme observed value within 1.5 × IQR from the edges. The median observed log fold changes at lamina-free and lamina-associated H3K9ac domains differed by 0.20 ($P \leq 1 \times 10^{-16}$, Mann–Whitney–Wilcoxon test) at 11 months (0.01 at 6 months) for the tau model, and by 0.13 ($P \leq 1 \times 10^{-16}$, Mann–Whitney–Wilcoxon test) at 4.5 months (0.04 at 3.5 months) for the CK-p25 model.

this model system induces AD-related features such as the intracellular accumulation of phosphorylated tau[28]. To characterize the epigenomic changes in this in vitro model system, we used the Assay for Transposase-Accessible Chromatin with sequencing (ATAC-seq) to map genome accessibility[29]. In contrast with anti-H3K9ac ChIP-seq, ATAC-seq generates reads at all open regulatory elements

Fig. 4 | Tau induces chromatin alterations in iPSC-derived neurons. a, Horizontal bars depict the relative frequency of each chromatin state within the ATAC domains (31 Mbp) and the whole genome (3.1 Gbp). Chromatin states were obtained from H9-derived cultured neurons (E010) included in the Roadmap Epigenomics Project. For better clarity, the respective upper bars depict the relative frequencies without the quiescent state. **b**, Bars depict proportion of ATAC domains in neurons that were overlapped by an H3K9ac domain in the DLPFC. **c**, Bars depict proportion of H3K9ac domains in the DLPFC that were overlapped by an ATAC domain in neurons. **d**, Bars depict the number of ATAC domains stratified by TSS, TSS flanking, enhancer and other domains. Darker shading indicates the number of domains whose chromatin accessibility differed significantly between *MAPT*-overexpressing neurons and controls at an FDR of 0.05. **e**, Manhattan plot depicts the log-transformed unadjusted P values for differences between *MAPT*-overexpressing neurons and controls of all ATAC domains along chromosome 1 (two-sided t tests, $n = 18$ experiments). Transformed P values were plotted with the sign of the test statistic to distinguish between increased and decreased chromatin accessibility. Red dashed lines indicate an FDR threshold of 0.05 when adjusting for all 40,637 domains. Broad genomic segments covering ATAC domains showing similar changes in chromatin accessibility are depicted as purple lines. The gray polygon represents the chromosome's centromere. Two ribbons at the bottom show chromatin structure-related annotation. The top ribbon indicates A and B compartments. The bottom ribbon indicates LADs. **f**, Boxplot shows different mean tau-related changes in chromatin accessibility in A compartments ($n = 306$) and B compartments ($n = 283$) ($P = 4 \times 10^{-7}$, Mann–Whitney–Wilcoxon test). The center line of the boxplot shows the median, the outer edges correspond to the lower and upper quartiles, and the whiskers extend to the most extreme observed value within 1.5 × IQR from the edges. Compartments with no or very few ATAC domains (<5) were discarded. **g**, Scatter plot shows association between segments' nuclear lamina association and mean change in chromatin accessibility. Each dot represents a segment ($n = 89$; sex chromosomes were excluded). Weighted linear regression line (blue line), 95% confidence interval (shaded region), and weighted Pearson correlation are depicted in the plot. Segments were weighted by the number of ATAC domains. **h**, Each dot represents either an A ($n = 284$) or B ($n = 221$) compartment derived from fetal human brain Hi-C data. For each compartment, the mean tau-related change in H3K9ac levels in the human DLPFC on the x axis is plotted vs. the mean change in chromatin accessibility between *MAPT*-overexpressing neurons and controls on the y axis (Pearson correlation $\rho = -0.51$). Compartments with fewer than five H3K9ac domains or fewer than five ATAC domains were discarded. **i**, Columns of the heat map represent the 138 genomic segments (sex chromosomes were excluded) derived from the DLPFC H3K9ac data and were first sorted by lamina association and then by chromatin accessibility. The color in the first row indicates the mean tau effect size observed in the DLPFC H3K9ac data ($n = 669$ subjects) as shown in Fig. 2e. The second and third rows depict the differences in H3K9ac ($n = 3$ experiments) and chromatin accessibility ($n = 18$ experiments) observed between *MAPT* overexpressing neurons and control neurons. Weighted Pearson correlations between the mean tau effect in the segments and histone accessibility (left column) and nuclear lamina association (right column) is shown on the right side. Segments were weighted by the number of DLPFC H3K9ac domains in the segments. **j**, Bars depict the variance of differences between *MAPT* overexpressing and respective control iNs that can be explained by segments (Supplementary Table 13). Blue bars indicate that the *MAPT*-overexpressing iNs were treated with 17-DMAG.

where the Tn5 transposase can access exposed DNA. We therefore expected to observe ATAC-seq peaks in the vicinity of nucleosomes carrying the H3K9ac mark and additional peaks at enhancers and insulators that might not be marked by H3K9ac.

All iNs were derived from the same induced pluripotent stem cell (iPSC) line, and we observed 40,637 ATAC domains with a distinct peak in at least four of our 18 samples. Each sample consisted of two or three pooled individual wells of transfected iNs. Nine samples represented independent transductions of *MAPT*-overexpressing iNs; the remaining nine samples were control iNs transduced with a GFP-overexpressing construct. Samples were generated in three experimental batches of independent differentiations. ATAC domains were annotated with chromatin states of human embryonic stem cell (hESC)-derived neurons (sample E010) from the



Roadmap Epigenomics Project[12] (Fig. 4a) and clustered into four groups: TSS, TSS-flanking region, enhancer and other domains (Supplementary Fig. 4a–c). As expected, most of the TSS domains (91%) overlapped one of the H3K9ac domains from our whole-brain tissues, but only 17% of the enhancer domains were included in our set of cortical H3K9ac enhancer domains (Fig. 4b), indicating that H3K9ac is not an explicit enhancer marker. Assessing these domains in the other direction, the majority of cortical H3K9ac domains were overlapped by at least one ATAC domain (Fig. 4c). Thus, while there are important differences between the two sets of profiles, there is good overlap, particularly in TSSs, where the bulk of the tau effect is observed in our H3K9ac tissue-level data.

In total, 747 of 40,637 ATAC domains showed a difference in chromatin accessibility at an FDR of 0.05 when comparing the *MAPT*-overexpressing iNs with control iNs infected with a non-relevant vector (Fig. 4d). The plots in Fig. 4e and Supplementary Fig. 2c depict large genomic segments of predominantly concordant tau-related changes; segments were defined using the same segmentation algorithm that we used for the cortical H3K9ac data. In iNs, we found 99 segments that explained 11% of the variance of tau effects observed for single ATAC domains (Supplementary Table 10). As seen in the H3K9ac data, the tau effect differed significantly between type A and type B compartments (Fig. 4f), and the average tau effect of the segments correlated ($\rho = 0.92$) with the proportion of LADs in the segment (Fig. 4g). These results were found consistently in each of the three batches of data (Supplementary Fig. 4d–j). Interestingly, in contrast with the H3K9ac changes observed in human and in mouse brain tissue, segments containing many LADs showed a large positive average tau effect, indicating that chromatin relaxation preferentially occurred in lamina-associated segments in this in vitro system. Concordantly, when looking at the human fetal brain A/B compartments, the tau effect observed in human brain H3K9ac data was negatively correlated with the tau effect in iNs (Fig. 4h). To verify that the model system presents a tau effect that is the inverse of what we see in the human cortex, we repeated the iN experiment with anti-H3K9ac ChIP-seq and estimated the effect size of tau for each of the 34,988 detected H3K9ac domains. Then, using the 178 genomic segments derived from the cortical H3K9ac data, we calculated the segments' average tau effect in (i) iN H3K9ac data and (ii) iN ATAC data to compare it with the brain tissue H3K9ac data (Fig. 4i). The H3K9ac and ATAC data from iNs consistently indicate that tau is associated with chromatin opening in genomic segments with a high fraction of lamina association. Thus, segments of the genome are significantly affected by tau pathology in vivo and in vitro, but the direction of effect is reversed in relation to tau pathology, which accumulates over decades in the human brain and over 14 days in our *MAPT*-overexpressing iN model system.

The Hsp90 inhibitor 17-DMAG is a candidate drug for attenuating tau-related alterations of chromatin structure. To find putative drugs with the potential to protect neurons from tau toxicity, we scanned the Connectivity Map database[30] for compounds whose gene expression signature was negatively correlated with our H3K9ac signature for tau pathology derived from ROS/MAP cortical profiles. H3K9ac domains were mapped to genes based on TSSs inside the domains. Domains without a TSS were discarded. Out of 1,309 compounds in the database, the N-terminal Hsp90 inhibitor 17-(dimethylaminoethylamino)-17-demethoxygeldanamycin (17-DMAG, alvespimycin) achieved the smallest *P* value according to the database's search algorithm and the fifth smallest correlation coefficient (Supplementary Table 11). Hsp90 and other heat shock proteins are involved in protein folding, and Hsp90 has been previously evaluated in the context of AD[31,32].

To assess the effect of 17-DMAG, we treated three independent cultures of *MAPT*-overexpressing iNs with 17-DMAG at a concentration of 3 nM for 24 h before harvesting (on day 31), based on published studies[33]. The differences between the treated *MAPT*-overexpressing iNs and the respective controls were calculated for each of the 40,637 ATAC domains. The variance of these tau-induced differences was then decomposed into between-segment and within-segment variance using the 99 tau-associated chromosomal segments derived for the iNs using ATAC-seq data. If 17-DMAG attenuates the tau effect on chromatin organization, the variance explained by the segments (that is, the between-segment variance) should decrease for the 17-DMAG-treated iNs. In our experiment, we observed such a protective effect of 17-DMAG (Fig. 4j). We then confirmed these results in a second experiment and a dose–response curve: triplicates of *MAPT*-overexpressing iN cell cultures with 17-DMAG at concentrations of 1 nM, 3 nM or 10 nM. As controls, we cultured *MAPT*-overexpressing iNs and control iNs in DMSO solution. While our experiments indicate that 17-DMAG may protect neurons from chromatin alterations related to tau (Fig. 4j and Supplementary Tables 12, 13), further studies with different lines of iNs are necessary to replicate the effect of 17-DMAG and to explore the interplay between 17-DMAG and tau pathology.

Discussion

Our epigenome-wide association study revealed that tau pathology is associated with broad changes in the brain's epigenome. We observed large genomic segments of several megabases within which H3K9ac domains showed similar tau-associated gains or losses of histone acetylation. At the megabase-pair scale, the genome is organized into two major types of compartments, A and B, defined by patterns observed in Hi-C interaction maps[15]. Type A compartments are characterized by euchromatin, which is transcriptionally active, whereas the inactive type B compartments exhibit high chromatin density. These two major types of compartments are essential for chromatin organization[34] and cell identity[35], and they impose a cis- and trans-correlation structure on epigenetic marks[36]. We demonstrated that the effect of tau on H3K9ac differed between type A and type B compartments and that the segments observed (in Fig. 2a) reflect higher-order chromatin structure. Thus, tau appears to be preferentially affecting certain elements of nuclear architecture and chromosomal organization.

The nuclear lamina is a key element in the spatial organization of chromatin and is associated with inactive B compartments. Previous work in *Drosophila* has suggested that the lamin nucleoskeleton is disrupted in tauopathies, which causes heterochromatin relaxation and mediates neuronal cell death[7,8]. In line with this hypothesis, we observed a strong correlation between the effect of tau in a genomic segment and the nuclear lamina association of that segment. However, we measured the euchromatic mark H3K9ac, and it remains unclear how this mark relates to heterochromatin relaxation. This intriguing association between the effect of tau on H3K9ac and key elements of spatial chromatin organization needs to be explored further using techniques such as those that map chromatin contact points.

The association of tau pathology with alterations in chromatin structure is not specific to the H3K9ac chromatin mark and has functional consequences: we replicated the presence of the segments with coordinated alterations in relation to tau pathology in both DNA methylation and transcriptomic data from the same persons. Further, we find the same pattern of results in publically available transcriptomic data from a different cohort, and a second dataset from laser-captured neurons further refines the observation by implicating a specific cell population. Even though the pattern was less distinct in the various transcriptomic datasets, demonstrating the presence of these transcriptional perturbations in relation to tau is important in understanding the consequences of these alterations. Not all epigenetic changes directly translate to altered transcription. Additionally, if chromatin alterations induce the transcription of epigenetically silenced genes, transcription may occur at a low level

in a tissue sample, and we may thus be underestimating the extent of transcriptional alterations in tissue-level RNA-seq data.

We employed AD mouse models and a *MAPT*-overexpressing iN in vitro model to further explore our findings and to address some of the limitations of our human cortex study. Looking at the results from our iN experiments, we demonstrated that *MAPT* overexpression is sufficient to induce chromatin reorganization in neurons. Further, this event occurs prior to tangle formation and neuronal cell death in our model system, as well as in the repurposed RNA dataset from laser-captured neurons[24]. Perturbation of epigenomic architecture may therefore be an early event in tau pathology. This is an important point that we cannot address with our cross-sectional human H3K9ac brain data. In the longitudinal CK-p25 mouse model data, we detected altered H3K9ac levels depending on the proximity of the H3K9ac domain to the nuclear lamina two weeks after p25 induction. For the same mouse model at the same time point, a study reported altered localization of lamin in hippocampal neurons, indicating a dispersion of the nuclear lamina membrane that could affect chromatin organization[37]. This effect was stronger over time and preceded apoptosis[37]. Overall, we hypothesize that tau-induced epigenomic changes are early events that occur downstream of pathologic tau accumulation but before neurofibrillary tangles develop.

Interestingly, we observed large positive tau effects on H3K9ac in lamina-free regions of the human cortex and mouse brain, but the direction of the association was reversed for iNs. This is an intriguing result and suggests a complex interaction between tau and chromatin structure. iNs grow in vitro in the absence of the three-dimensional context of the cerebral cortex and do not interact with non-neuronal cells. The inverse directionality that we see at an early time point after *MAPT* overexpression could be an initial response of iNs to a rapid accumulation of tau (rather than accumulation over decades in the human brain), in line with recent work suggesting that nuclear tau also has a physiological function in assembling and maintaining heterochromatin[9]. Importantly, the segmental pattern and the association of the tau effect with the proportion of LADs in the segments is seen in both the human cortex and iNs, suggesting the involvement of the same mechanism. More work is required to delineate the molecular events happening in vivo and in vitro, and to study whether the mechanism is specific for tau or can occur with other stimuli. Further, we used publicly available data from different cell types to annotate our results; therefore, the extent to which spatial organization changes occur with tau pathology needs to be interrogated directly. iNs overexpressing tau may be a useful model system with which to address these questions. This is illustrated by our testing of an Hsp90 inhibitor that is predicted to block the effects of tau on the basis of our brain data: 17-DMAG attenuates the chromatin alterations induced by tau in iNs.

Overall, we have identified large-scale changes throughout the epigenome of the human AD brain and shown that tau-induced alterations of chromatin structure are much more profound than the changes that are attributable to amyloid pathology. The effect of these alterations propagate into the transcriptome. Our data supports the hypothesis that the nuclear lamina is a key element in mediating tau toxicity in vivo and in vitro, and we demonstrated that iNs are a useful model system to further study tau-associated chromatin changes and to screen for drugs such as 17-DMAG that might prevent or reverse these chromatin alterations.

Accession Codes. Synapse ID: syn4896408; GEO accession numbers: GSE97560 and GSE97409.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41593-018-0291-1.

Received: 29 March 2018; Accepted: 13 November 2018;
Published online: 17 December 2018

References

1. 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).
2. Klein, H. U., Bennett, D. A. & De Jager, P. L. The epigenome in Alzheimer's disease: current state and approaches for a new path to gene discovery and understanding disease mechanism. *Acta Neuropathol.* **132**, 503–514 (2016).
3. Lardenoije, R. et al. The epigenetics of aging and neurodegeneration. *Prog. Neurobiol.* **131**, 21–64 (2015).
4. Nativio, R. et al. Dysregulation of the epigenetic landscape of normal aging in Alzheimer's disease. *Nat. Neurosci.* **21**, 497–505 (2018).
5. 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).
6. Lunnon, K. et al. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat. Neurosci.* **17**, 1164–1170 (2014).
7. Frost, B., Bardai, F. H. & Feany, M. B. Lamin dysfunction mediates neurodegeneration in Tauopathies. *Curr. Biol.* **26**, 129–136 (2016).
8. Frost, B., Hemberg, M., Lewis, J. & Feany, M. B. Tau promotes neurodegeneration through global chromatin relaxation. *Nat. Neurosci.* **17**, 357–366 (2014).
9. Mansuroglu, Z. et al. Loss of Tau protein affects the structure, transcription and repair of neuronal pericentromeric heterochromatin. *Sci. Rep.* **6**, 33047 (2016).
10. Bennett, D. A., Schneider, J. A., Arvanitakis, Z. & Wilson, R. S. Overview and findings from the religious orders study. *Curr. Alzheimer. Res.* **9**, 628–645 (2012).
11. Bennett, D. A. et al. Overview and findings from the rush Memory and Aging Project. *Curr. Alzheimer. Res.* **9**, 646–663 (2012).
12. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
13. Gabriel, S. B. et al. The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).
14. Olshen, A. B., Venkatraman, E. S., Lucito, R. & Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–572 (2004).
15. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
16. Won, H. et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016).
17. Meuleman, W. et al. Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res.* **23**, 270–280 (2013).
18. Kind, J. et al. Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163**, 134–147 (2015).
19. Klein, H. U. & De Jager, P. L. Uncovering the role of the methylome in dementia and neurodegeneration. *Trends Mol. Med.* **22**, 687–700 (2016).
20. Schoofs, T. et al. DNA methylation changes are a late event in acute promyelocytic leukemia and coincide with loss of transcription factor binding. *Blood* **121**, 178–187 (2013).
21. Zou, F. et al. Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genet.* **8**, e1002707 (2012).
22. Zhong, Y., Wan, Y. W., Pang, K., Chow, L. M. & Liu, Z. Digital sorting of complex tissues for cell type-specific gene expression profiles. *BMC Bioinformatics* **14**, 89 (2013).
23. Jaffe, A. E. & Irizarry, R. A. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome. Biol.* **15**, R31 (2014).
24. Liang, W. S. et al. Alzheimer's disease is associated with reduced expression of energy metabolism genes in posterior cingulate neurons. *Proc. Natl Acad. Sci. USA.* **105**, 4441–4446 (2008).
25. Yoshiyama, Y. et al. Synapse loss and microglial activation precede tangles in a P301S tauopathy mouse model. *Neuron* **53**, 337–351 (2007).
26. Cruz, J. C., Tseng, H. C., Goldman, J. A., Shih, H. & Tsai, L. H. Aberrant Cdk5 activation by p25 triggers pathological events leading to neurodegeneration and neurofibrillary tangles. *Neuron* **40**, 471–483 (2003).
27. Gjoneska, E. et al. Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature* **518**, 365–369 (2015).
28. Muratore, C. R. et al. Cell-type dependent Alzheimer's disease phenotypes: probing the biology of selective neuronal vulnerability. *Stem Cell Rep.* **9**, 1868–1884 (2017).
29. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).

30. Lamb, J. et al. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929–1935 (2006).
31. Dickey, C. A. et al. The high-affinity HSP90-CHIP complex recognizes and selectively degrades phosphorylated tau client proteins. *J. Clin. Invest.* **117**, 648–658 (2007).
32. Luo, W. et al. Roles of heat-shock protein 90 in maintaining and facilitating the neurodegenerative phenotype in tauopathies. *Proc. Natl Acad. Sci. USA.* **104**, 9511–9516 (2007).
33. Gao, L. et al. Discovery of the neuroprotective effects of alvespimycin by computational prioritization of potential anti-Parkinson agents. *FEBS J.* **281**, 1110–1122 (2014).
34. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
35. Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
36. Fortin, J. P. & Hansen, K. D. Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. *Genome Biol.* **16**, 180 (2015).
37. Chang, K. H. et al. Nuclear envelope dispersion triggered by deregulated Cdk5 precedes neuronal death. *Mol. Biol. Cell* **22**, 1452–1462 (2011).

Acknowledgements

This work has been supported by NIH grants U01 AG046152, R01 AG036836, RF1 AG015819, R01 AG017917, RC2 AG036547. L-H. T. has been supported by NIH/NINDS/NIA (R01 NS078839) and the Robert A. and Renee R. Belfer Family Foundation. The Mayo Clinic Alzheimer's Disease Genetic Studies were led by N. Ertekin-Taner and S. G. Younkin, Mayo Clinic, Jacksonville, FL, using samples from the Mayo Clinic Study of Aging, the Mayo Clinic Alzheimer's Disease Research Center and the Mayo Clinic Brain Bank. 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.

Author Contributions

P.L.D., D.A.B., T.L.Y-P., A.M. and B.E.B. conceived the study. C.M., E.G., A.T., S.E.S., B.J.K., A.T. and R.V.S. conducted experiments. J.A.S. and D.A.B. and contributed post mortem brain tissues. E.G. and L.-H.T. contributed mouse models. S.E.S. and T.L.Y-P. contributed neuronal models. H-U.K., J.X. A.R.P. and P.L.D. analyzed data. H-U.K., E.G., S.M., T.L.Y-P., D.A.B. and P.L.D. interpreted data and designed follow-up experiments. H.-U.K. and P.L.D. wrote the manuscript with contributions from all co-authors.

Competing interests

B.E.B. owns equity in Fulcrum Therapeutics, 1CellBio Inc., Nohla Therapeutics and HiFiBio Inc., and is an advisor for Fulcrum Therapeutics, HiFiBio Inc. and Cell Signaling Technologies. The other authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41593-018-0291-1.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to P.L.D.J.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2018

Methods

ROS/MAP cohort and pathologic characterization. All participants were enrolled in the Religious Orders Study (ROS)[10] or the Memory and Aging Project (MAP)[11]. Studies were approved by the Institutional Review Boards of Rush University Medical Center and Partners Healthcare. Participants were not demented at the time of enrollment and agreed to donate their brain upon death. Amyloid-β and tau tangles were assessed postmortem as previously described[38,39]. Briefly, brains were cut into 1-cm-thick coronal slabs and immersion fixed in 4% paraformaldehyde. Tissue blocks from 8 brain regions (hippocampus -CA1/ subiculum-, angular gyrus, entorhinal, superior frontal, dorsolateral prefrontal, inferior temporal, anterior cingulate, and calcarine cortices) were embedded in paraffin and sectioned at 20 µm. Paraffin-embedded sections were immunostained for amyloid-β using 1 of 3 monoclonal anti-human antibodies: 4G8 (1:9000; Covance Labs, Madison, WI), 6F/3D (1:50; Dako North America Inc., Carpinteria, CA), and 10D5 (1:600; Elan Pharmaceuticals, San Francisco, CA). Paired helical filament (PHF) tau tangles were labeled with an antibody specific to phosphorylated tau (AT8, Thermoscientific, Waltham, MA, USA). A computerized sampling procedure combined with image analysis software was used to calculate the percentage area occupied with amyloid-β and the density of PHFtau tangles. Composite scores were computed for overall amyloid-β burden and PHFtau tangle density by averaging the scores obtained from the eight brain regions. Amyloid-β and tau tangle scores were square root transformed for better statistical properties. Brain samples were randomly selected and sent to ChIP-seq processing in an arbitrary order. Individuals generating ChIP-seq data were blinded to the outcome measures. No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to those reported in previous studies[5,40].

Human brain H3K9ac ChIP-seq. We identified the Millipore anti-H3K9ac antibody (catalog # 06–942, lot: 31636) as a robust antibody validated for ChIP-seq (validation data from Millipore available at http://www.emdmillipore.com/US/en/product/Anti-acetyl-Histone-H3-Lys9-Antibody,MM_NF-06-942). 50 mg of gray matter was dissected on ice from biopsies of the DLPFC of the ROS/MAP cohorts. The tissue was minced and cross-linked with 1% formaldehyde at room temperature for 15 min and quenched with 0.125 M glycine. The tissue was then homogenized in cell lysis buffer using a tissue lyser and a 5 mm stainless steel bead. Then, the nuclei were lysed in cell lysis buffer, and chromatin was sheared by sonication. Samples were incubated overnight at 4 °C with 2.5 µl of the H3K9ac antibody with a final volume of 3 ml using the ChIP Dilution Buffer. Chromatin bound to the antibody was purified with protein A Sepharose beads. The final DNA was extracted and used for Illumina library construction following usual methods of end repair, adapter ligation and gel size selection. Samples were pooled and sequenced on the Illumina HiSeq (36 bp single-end reads).

Human brain H3K9ac ChIP-seq data pre-processing, peak detection and annotation. Single-end reads were aligned by the BWA algorithm against the human reference genome GRCh37, and reads with a mapping quality of 0 were removed[41]. Peaks were detected for each sample individually by MACS2 using the broad peak option, a stringent q-value cutoff of 0.001, and pooled genomic DNA of seven samples as a negative control library[42]. A combination of different ChIP-seq quality measures were employed to remove low-quality samples[43], ie, samples that did not reach (i) $\geq 15 \times 10^6$ unique reads, (ii) non-redundant fraction ≥ 0.3, (iii) cross-correlation ≥ 0.03, (iv) fraction of reads in peaks ≥ 0.05 and (v) $\geq 6{,}000$ peaks were removed. After quality control, 669 out of 712 samples remained (Supplementary Table 1). We defined our H3K9ac domains by calculating all genomic regions that were detected as a peak in at least 100 (15%) of our 669 samples. Regions neighbored within 100 bp were merged, and very small regions of less than 100 bp were removed, resulting in 26,384 H3K9ac domains. H3K9ac levels were quantified by counting the number of fragments overlapping the H3K9ac domains in each sample. Fragments were estimated by extending the short sequence reads toward the 3′ end by the optimal shift obtained by maximizing the cross-correlation for each sample. The mean optimal shift was 271 bp. Only non-duplicated uniquely mapped reads were used to quantify H3K9ac levels. H3K9ac domains were annotated using chromatin states (core 15-state model) from an unaffected DLPFC sample included in the Roadmap Epigenomics project[12] (E073). Proportion of chromatin states were calculated for each H3K9ac domain. Then, k-means clustering was applied to partition the H3K9ac domains according to their chromatin states into $k = 5$ clusters that explained 61% of the variance of the domain annotation. Chromatin states of the five cluster centroids (Supplementary Fig. 1b) revealed that two clusters represented H3K9ac domains at promoters and another two clusters represented H3K9ac domains at enhancers. Promoter and enhancer clusters were merged, resulting in three classes of H3K9ac domains: promoter, enhancer and other.

Mouse models. All mouse work was approved by the Committee on Animal Care of the Division of Comparative Medicine at MIT and complied with the relevant ethical regulations. Adult (3 months old) female double-transgenic CK-p25 mice[26] (crosses of p25 Tg mice, see https://www.jax.org/strain/005706, and Camk2a-tTA mice, see https://www.jax.org/strain/003010) and their respective control littermates were used for the experiments. Brain tissue was collected at either 2 or 6 weeks after p25 induction. For the tau (P301S) mouse model[25] (see https://www.jax.org/strain/008169), adult female 6 months old and male 11 months old transgenic mice with age- and gender-matched wild-type mice as controls were used. Hippocampal tissue was extracted for ChIP-seq and flash-frozen in liquid nitrogen. No animals were excluded from the study. Supplementary Table 9 summarizes the number of animals available at each time point.

Mouse brain H3K9ac ChIP-seq and data pre-processing. H3K9ac ChIP of mouse hippocampal tissue was performed as described for the human brain tissue using the Abcam antibody ab#4441 (lot numbers GR196840-1 and GR207546-1) at a concentration of 1.5 µl (=1.5 µg) of antibody per 1 ml of chromatin sample. Validation data is available from the manufacturer at https://www.abcam.com/histone-h3-acetyl-k9-antibody-chip-grade-ab4441.html. ChIP-seq data were pre-processed similarly to the human DLPFC ChIP-seq data: BWA was used to map reads against the reference assembly NCBI m37. The same quality measures were calculated (Supplementary Table 14) as those for the human data, but we did not require samples to have $\geq 15 \times 10^6$ unique reads, because CKp25 samples were sequenced at a lower coverage. MACS2 was used for peak detection, and H3K9ac domains were defined as regions that were covered by a peak in at least 4 (17%) out of the 23 samples ($n = 12$ CKp25; $n = 11$ tau) resulting in 44,165 H3K9ac domains.

Induction of human neuronal cells overexpressing *MAPT*. The iPSC line used in this study (YZ1) was originally generated from the IMR-90 cell line (ATCC), and characterization of this line was described previously[44]. Due to a karyotype abnormality in a small subset of cells, monoclonal isolates were obtained and confirmed to be karyotypically normal and pathogen free prior to initiating this study. The iPSC line used was confirmed to be of the correct identity prior to and at the conclusion of the study using short tandem repeat (STR) profiling (Genetica Cell Line Testing). Neurons were generated from the direct conversion of induced pluripotent stem cells by transduction with Neurogenin 2, as previously described[45]. Neurons were plated on Matrigel-coated 96-well plates on DIV4 and maintained in media consisting of 485 ml neurobasal medium (Gibco), 5 ml Glutamax, 7.5 ml 20% dextrose, 2.5 ml MEM NEAA with 1:50 B27, BDNF, CNTF, GDNF and doxycycline. After 17 days in vitro, cells were transfected with a 1:1 dilution of lentivirus-packaged open reading frames (ORF) expressing human *MAPT* (titer 6.2×10^6), resulting in overexpression of tau protein, or a GFP-expressing construct (pRosetta). After 18-h incubation, virus-containing media was removed and replaced with fresh media to incubate until day 31 in vitro, when harvested for ATAC-seq. RNA-sequencing revealed a tenfold increase of *MAPT* mRNA in cells transfected with *MAPT* cDNA compared with controls. In total, nine *MAPT*-overexpressing and nine control iNs were generated in three batches (differentiation rounds) of triplicate wells.

17-DMAG treatment of induced neurons. The second and third batch of experiments contained cell cultures of induced neurons treated with 1, 3, or 10 nM 17-DMAG for 24 h prior to collection. Concentrations were chosen based on previous work[33]. To account for vehicle-specific effects, *MAPT*-overexpressing cells and control cells were cultured in concentration-matched DMSO solution for 24 h prior to collection. Experimental design of the iN experiments is shown in Supplementary Table 12.

Assay for transposase-accessible chromatin using sequencing (ATAC-seq). ATAC-seq was performed as previously described[29] with three modifications: (i) to reduce mitochondrial DNA contamination, we substituted 0.1% Ipegal with 0.01% Tween in the nuclear lysis buffer; (ii) additional tagmentation buffer (TD; 2×) was prepared as follows: 20 mM Tris(hydroxymethyl)aminomethane and 10 mM $MgCl_2$, adjusted to pH 7.6 with 100% acetic acid before addition of 20% (vol/vol) dimethylformamide[46]; (iii) we performed the PCR amplification in a 25 µl total volume rather than 50 µl to save half of our tagmented DNA for backup. After 31 days in vitro, iPSC neurons were trypsinized and counted to provide 50,000 live cells for ATAC-seq. iPSC-derived neurons were lysed with 0.01% Tween, 10 mM Tris-Cl, pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, and nuclei were pelleted, washed and transposed upon addition of 2.5 µl Nextera Tn5 Transposase in 1× TD buffer (Nextera Kit, Illumina) and incubation at 37 °C for 30 min. Immediately after transposition, DNA fragments were purified using a Qiagen MinElute PCR Purification Kit (Qiagen) following the manufacturer's instructions. Transposed DNA was amplified with dual indexed Nextera PCR primers for 5 cycles prior to removing a 5 µl aliquot from each sample to test further amplification requirements by qPCR. Rounding up the cycle threshold from each sample determines how many additional cycles each sample needs to be amplified. For these samples, we needed to re-array and continue amplifying the remaining 20 µl for 5–8 cycles. Final libraries were cleaned twice with 1.5× Ampure XP SPRI beads and eluted in 12 µl of water. We quantified with Qubit HS DNA assay and assessed for quality using a Bioanalyzer High-Sensitivity DNA analysis Kit. 20 ng of each sample were pooled and diluted to 4 nM using an estimated library length of 300 bp for all and submitted to the Broad Institute's Genomics Platform for Illumina HiSeq2500 25 bp paired-end sequencing.

ATAC-seq data pre-processing, peak detection and annotation. Paired-end reads from ATAC-seq were aligned by the BWA algorithm against the human reference genome GRCh37[41]. MACS2 was applied to each sample using fragment sizes obtained from the paired alignments and a q-value cutoff of 0.001[42]. Our ChIP-seq quality control pipeline was adapted for ATAC-seq. Instead of cross-correlation, we calculated the median insert size and verified that the distribution of insert sizes showed a periodicity equal to that of the helical pitch of DNA (Supplementary Table 15). No experiments were excluded. ATAC domains were defined as genomic regions covered by a peak in at least four (22%) of our 18 samples ($n=9$ *MAPT* overexpressing; $n=9$ controls). Domains less than 50 bp away from each other were merged. In total, we obtained 40,637 ATAC domains. ATAC domains were annotated with chromatin states obtained from hESC-derived neurons included in the Roadmap Epigenomics project (sample E010, core 15-state model)[12]. k-means clustering was applied to partition the ATAC domains according to their chromatin states into $k=4$ clusters that explained 65% of the variance of the domain annotation. Chromatin states of the four cluster centroids (Supplementary Fig. 4a) revealed that the first three clusters represent ATAC domains at TSSs, TSS-flanking regions and enhancers. The fourth cluster represents ATAC domains with various other chromatin states.

H3K9ac ChIP-seq of induced neurons. H3K9ac ChIP-seq data from iNs were generated and processed as described for the DLPFC tissue samples. No experiments were excluded based on quality measures (Supplementary Table 15). Pooled input DNA from all three samples ($n=2$ *MAPT* overexpressing, $n=1$ control) were used as a negative control library during peak detection. Peaks that were detected in at least two of the three samples were defined as neuronal H3K9ac domains ($n=34{,}988$) and used for subsequent analyses.

Statistical analysis. The numbers of reads Y within each of the $m=26{,}384$ H3K9ac domains and $n=669$ individuals were calculated and assumed to follow a negative binomial distribution $Y \sim NB(\mu,\theta)$ with mean μ and dispersion parameter θ[17]. A log-linear regression model

$$\log(E(Y)) = \log(n) + \beta_0 + x_t\beta_t + x_a\beta_a + z\gamma \quad (1)$$

was fitted for each H3K9ac domain separately using the R package MASS. The offset log*(n)* accounts for different total numbers of reads n (summed over all domains) per sample. Coefficients β_t and β_a model the effect of the observed square-root-transformed tau and amyloid-β levels on the H3K9ac levels. Coefficients were tested by Wald tests and the false discovery rate (FDR) was calculated to account for multiple testing. Vector γ models the biological covariates age and gender as well as the technical covariates post mortem interval, batch and cross-correlation.

Normalized ROS/MAP transcription data from the DLPFC were downloaded from Synapse (Synapse: syn3388564). RNA-seq data generation and quantification with RSEM have been described previously[40,48]. Pathologic variables were available for $n=500$ individuals. Only transcripts with RPKM values ≥2 in at least 25% of the samples were considered active in the human brain and kept in the dataset ($m=24{,}594$ transcripts). A linear regression model was fitted for each transcript with the $\log_2$-transformed RPKM values as outcome variable and tau and amyloid-β as explanatory variables. Models were adjusted for age, gender, RIN score (RNA integrity number), post mortem interval, study index (ROS or MAP) and logarithmized total number of aligned reads. Coefficients were tested by t tests, and the FDR was calculated to account for multiple testing.

Mouse H3K9ac data were analyzed using the voom method implemented in the R package limma[49]. For each mouse model and time point, we compared $n=3$ AD mice ($n=2$ for the tau mouse model at $t=6$ months) to $n=3$ respective control mice. Voom has been designed for small sample sizes and provides reliable domain-specific estimates of the variance by estimating the mean-variance trend using information from all H3K9ac domains. The voom method is incorporated in a weighted linear regression framework with $\log_2$-transformed reads per million ($\log_2$ RPM) as outcome. A model was fitted for each of the $m=44{,}165$ domains, and the cross-correlation was added as technical covariate in addition the AD/control indicator variable. Significance of the AD/control variable was assessed by moderated t tests, and the FDR was calculated to adjust for multiple testing[50].

Neuronal ATAC-seq data of *MAPT*-overexpressing iNs ($n=9$ experiments) and control iNs ($n=9$ experiments) from three batches were analyzed using voom[49]. For each of the $m=40{,}637$ ATAC domains, $\log_2$-transformed RPM values were regressed on the tau indicator variable (*MAPT* overexpressing or control) adjusted for batch and logarithmized total number of reads in all domains of the sample. Significance of the tau variable was assessed by moderated t tests, and the FDR was calculated to adjust for multiple testing[50]. Hierarchical clustering (Euclidian distance, average linkage) shown on Supplementary Fig. 4d was applied to logarithmized RPM values after subtracting the estimated effect of the total number of reads. For the comparison of batches (Supplementary Fig. 4e–j), the same regression models but without batch variable were applied to each batch ($n=3$ *MAPT*-overexpressing iNs, $n=3$ control iNs) separately. The effect of 17-DMAG treatment was analyzed separately for each batch/concentration combination (batch 2: 3 nM, batch 3: 1, 3 and 10 nM) using linear regression models within the voom framework. A binary variable encoding treatment or control condition ($n=3$ *MAPT*-overexpressing iNs + 17-DMAG, $n=3$ control iNs) and the logarithmized total number of reads were used as explanatory variables.

Accounting for altered cell type proportions. The Digital Sorting Algorithm (DSA) was applied to estimate the proportions of neurons, astrocytes, oligodendrocytes, myeloid cells and endothelial cells in the cortical tissue from RNA-seq data of $n=452$ individuals that had both H3K9ac and RNA-seq data[22]. A set of five marker genes was derived for each cell type (Supplementary Table 8) from published RNA-seq data of purified cells[51]. We calculated the mean RPKM value for each gene in each cell type and required a marker gene for a specific cell type to have a mean RPKM value ≥25 in that cell type and a mean RPKM value ≤3 in the other four cell types. Next, we selected the top five marker genes per cell type with the smallest coefficients of variation in our tissue-level RNA-seq data to avoid selecting genes with highly variable transcription, which are likely to be affected by various cellular processes resulting in imprecise estimations. For cell type proportion estimations, we used gene-level transcription values derived from the RNA-seq data available at Synapse (Synapse: syn3505732) instead of the isoform-level values used in the main analysis. Estimated cell type proportions are given in Supplementary Table 1. The proportion of neurons was decreased ($P=0.043$), and proportions of oligodendrocytes ($P=0.003$) and myeloid cells ($P=0.016$) were increased in subjects with AD (two-sided t tests, $n=452$ individuals, unadjusted P values). The main analysis of the H3K9ac data for the subset of 452 samples with RNA-seq data was repeated using (i) no adjustment, (ii) the proportion of one of the five cell types as additional covariate and (iii) the proportions of all five cell types as additional covariates in the regression model used for the main analysis given by equation (1). Pearson correlation between the tau coefficients of the unadjusted model and the tau coefficients of any adjusted model was >0.99 (Supplementary Fig. 3a–f). In total, 2,514 H3K9ac domains were significantly associated with tau (FDR ≤ 0.05) in the unadjusted model with the reduced sample size of $n=452$. None of the significant H3K9ac domains lost significance (FDR > 0.1) in any of the adjusted models. When adjusting for all cell types, an additional 11 H3K9ac domains were significantly (FDR ≤ 0.05) associated with tau that were not associated (FDR > 0.1) in the unadjusted model. Although cell type proportion affected the tissue-level H3K9ac levels, these results indicate that changing cell type proportions did not severely confound the tau effect, and we therefore decided to not adjust our primary analysis for cell type proportions so that we can utilize the full sample size of $n=669$ individuals.

Determining genomic segments of concordant tau-related changes. Genomic segments for the DLPFC H3K9ac data (Fig. 2a and Supplementary Table 6) were defined on the P values for the tau coefficient β_t from model (1). P values were log-transformed, and the signs of the respective tau coefficients β_t were assigned to the transformed P values. The circular binary segmentation algorithm implemented in the R package DNAcopy was then applied to the transformed P values with the corresponding genomic positions of the H3K9ac domains[14]. We identified 178 genomic segments using the default value of $\alpha=0.01$ for acceptance of new change-points. The same approach was applied to the transformed P values from the neuronal ATAC-seq data (comparing *MAPT*-overexpressing with control iNs) to identify 99 genomic segments in iNs (Fig. 4e and Supplementary Table 10).

Calculating mean tau effects of genomic segments. The mean tau effect of a segment was defined as the weighted mean of the tau coefficients β_t of all H3K9ac (or ATAC) domains located within the segment. The tau coefficients were weighted by their inverse standard deviation. Similarly, for transcription data, the TSS was used to map genes to segments. Coefficients and their standard deviations for DLPFC H3K9ac, DLPFC RNA-seq and iN ATAC-seq data were obtained from the models described in the section "Statistical analysis". In addition, Figs. 2e and 4i depict segments' mean tau effects from DLPFC methylation data, Mayo LOAD study transcription data, neuronal transcription data, and iN H3K9ac data. Mean effects were calculated in the same manner using coefficients and standard deviations obtained from the datasets and models described below.

Normalized ROS/MAP DNA methylation data from the DLPFC were downloaded from Synapse (Synapse: syn3157275). Data generation and processing have been described previously[5]. Pathologic variables were available for $n=729$ individuals. Relative methylation levels ("beta values") were assumed to follow a beta distribution[52]. A beta regression model with probit link function was fitted for each CpG, with the methylation level as outcome using the R package betareg[53]. Tau, amyloid-β, age, gender, post mortem interval, batch and bisulfite conversion rate were used as explanatory variables. Most CpGs were located in the chromatin states (sample E073, 15-state model) quiescent (24%), active transcriptional start site (21%), weakly transcribed (15%) and enhancer (7%). Segments' mean tau effects in DNA methylation data were calculated for these four chromatin states separately, because DNA methylation levels vary between different functional regions[54].

Normalized gene expression data from the Mayo LOAD study were downloaded from Synapse (Synapse: syn3157225)[21,55]. Individuals with progressive supranuclear palsy (PSP) were excluded, resulting in $n=202$ samples with AD and $n=90$ non-AD/PSP samples. Array probes with a low call rate (<50%) were excluded, reducing the number of probes from 24,526 to 17,910. For each probe, we fitted a linear regression model with the log-transformed expression value as

the outcome. Because quantitative measurements of AD pathologies were not available, a binary variable encoding the AD diagnosis was used as an explanatory variable to model the effect of AD on transcription. Models were adjusted for age, gender, PCR plate and RIN score.

Raw neuronal expression data were downloaded from Gene Expression Omnibus (GEO GSE5281)[24]. The RMA method implemented in the R package affy was applied to normalize data and summarize intensity values at the probeset level[56]. Only samples collected from the superior frontal gyrus were used ($n=23$ AD, $n=11$ non-AD samples). The R package limma was used to fit a linear regression model for each of the 54,675 probesets, with the log-transformed transcription level as outcome[50]. AD diagnosis and gender were used as explanatory variables.

H3K9ac ChIP-seq data from iNs ($n=2$ *MAPT* OE iNs, $n=1$ control iNs) were processed as described for the DLPFC tissue samples. The number of reads within each of the $m=34{,}988$ detected H3K9ac domains were calculated for each sample. Then, the voom method was applied to test for differences between *MAPT* overexpression and control neurons[49]. Cross-correlation was calculated and added to the models to adjust for technical variability.

A/B compartment and lamina-associated domain annotation. Human fetal brain A/B compartments were derived from a published Hi-C dataset (GEO GSE77565), as described in the original publication[16]. Briefly, Pearson correlation matrices were calculated from the intrachromosomal observed/expected 40 kbp binned interaction matrices. For each chromosome, the sign of the first principal component obtained from the correlation matrix was used to categorize bins as compartment type A or B. Mean tau effects in compartments (Figs. 2b and 4h) were calculated as described for genomic segments.

The 178 genomic segments derived from DLPFC H3K9ac data were annotated with nuclear lamina association using published DamID data for Lamin B1 in human fibroblasts[17] (GEO GSE22428). Probes of the DamID array were classified into nuclear lamina bound or unbound in the original work. We remapped the probes against the genome version GRCh37 and defined the lamina association of a segment as the ratio of nuclear lamina bound probes to all probes in the segment. To sort segments in Figs. 2e and 4i, chromatin accessibility was calculated from chromatin state annotation of sample E073 (DLPFC) from the Roadmap Epigenomics Project[12]. Chromatin accessibility was defined as the fraction of the segment that was annotated with any chromatin state other than heterochromatin, repressed polycomb, weak repressed polycomb or quiescent/low. These four chromatin states exhibited the lowest DNA accessibility measured by DNase-seq[12]. As expected, chromatin accessibility was negatively correlated with lamina association in the DLPFC (Spearman correlation $\rho=-0.86$, $n=178$ segments); that is, segments of more open chromatin were less likely to include a LAD. Nuclear lamina association for the 99 genomic segments derived from iN ATAC-seq data was calculated in the same manner using the same DamID data for lamin B1 from human fibroblasts. Published lamin B1 data from mouse embryonic fibroblasts were used to annotate our mouse H3K9ac data[17]. Genomic segments and their lamina association are given in Supplementary Tables 6 ($n=178$ DLPFC H3K9ac segments) and 10 ($n=99$ iN ATAC-seq segments). Small segments with <20 H3K9ac domains or <20 active transcripts (<20 ATAC domains) and segments located at sex chromosomes were removed when calculating the correlation between lamina association and mean tau effects shown in Fig. 2c–e (Fig. 4g,i), reducing the number of DLPFC H3K9ac and iN ATAC-seq segments to 138 and 89, respectively.

Analysis of the overall effect of *MAPT* overexpression. We overexpressed *MAPT* in iNs to replicate the spatial pattern of tau-related H3K9ac alterations in the human brain data (Fig. 2a). We found a similar pattern of tau-related alterations in the Manhattan plot derived from the ATAC-seq data of the iNs (Fig. 4e), but the *P* values in the Manhattan plot were not independent, because all tests were conducted on the same set of samples, and a few samples could drive the pattern. To address this question, we defined a segmentation score that quantifies whether the ATAC domain values of a single sample follow the spatial pattern that we observed in the human brain H3K9ac data. To calculate the score, we first regressed the log rpm values of each ATAC domain on the technical variables batch and logarithmized total number of reads. This is the same model as that used for the main analysis, but we excluded the variable that indicates whether the sample overexpresses *MAPT*. The residuals r_{ij} for sample i and domain j from these domain-wise regression models were extracted and still contain the *MAPT* overexpression effect. Next, the linear model $r_{ij} = a_i + b_i\,S_{\iota(j)} + e_{ij}$ was fitted for each sample i. S_k denotes the mean tau effect observed for segment k in the human brain H3K9ac data, and ι is a function mapping ATAC domain j to the segment k where the domain is located. Factor b_i is the segmentation score for sample i. The rationale is to model a sample's residuals using the segments and the segments' mean tau effects from the H3K9ac data. Segmentation scores for the nine *MAPT*-overexpressing iNs and the nine control iNs are shown in Supplementary Fig. 4k. Because residuals are centered at 0, the score is also centered at 0. The score was significantly different ($P=0.02$; Wilcox–Mann–Whitney test) between the two groups. In line with the inverse tau effect in the Manhattan plot (Fig. 4e), the score was smaller for *MAPT*-overexpressing iNs.

Search for drug candidates in the Connectivity Map database. Only H3K9ac domains spanning a TSS of an active transcript (≥100 samples with RPKM ≥ 2 in the RNA-seq data) were considered. Gene annotations of the TSS were used to map H3K9ac domains to genes in the Connectivity Map database[30]. The search algorithm provided by the Connectivity Map 2.0 requires a set of up- and downregulated genes as input. As suggested in the Connectivity Map documentation, we selected the top 250 positively tau-associated genes and the top 250 negatively tau-associated genes that had the smallest *P* values in our H3K9ac data and calculated the enrichment score and *P* values on the web server (https://www.broadinstitute.org/cmap). In addition, we calculated the Spearman rank correlation coefficient between the vector of standardized coefficients $\beta_t/\sigma_{\beta t}$ for tau obtained from our H3K9ac data and the differential expression statistic from the Connectivity Map for each compound ($n=1{,}309$) in the database. A negative correlation indicates that genes showing increased H3K9ac levels with tau were repressed by the respective compound.

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

Code availability. Analysis code is provided in the Supplementary Software.

Data availability

Human H3K9ac ChIP-seq data has been deposited at Synapse (Synapse: syn4896408). Mouse H3K9ac ChIP-seq data (GEO GSE97560) and ATAC-seq data from iNs (GEO GSE97409) have been deposited at Gene Expression Omnibus. Phenotype variables for the ROS/MAP studies can be requested from the RADC Research Sharing Hub (https://www.radc.rush.edu).

References

38. Bennett, D. A., Schneider, J. A., Tang, Y., Arnold, S. E. & Wilson, R. S. The effect of social networks on the relation between Alzheimer's disease pathology and level of cognitive function in old people: a longitudinal cohort study. *Lancet Neurol.* **5**, 406–412 (2006).
39. Bennett, D. A., Schneider, J. A., Wilson, R. S., Bienias, J. L. & Arnold, S. E. Neurofibrillary tangles mediate the association of amyloid load with clinical Alzheimer disease and level of cognitive function. *Arch. Neurol.* **61**, 378–384 (2004).
40. Mostafavi, S. et al. A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat. Neurosci.* **21**, 811–819 (2018).
41. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
42. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome. Biol.* **9**, R137 (2008).
43. Landt, S. G. et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
44. Zeng, H. et al. Specification of region-specific neurons including forebrain glutamatergic neurons from human induced pluripotent stem cells. *PLoS One* **5**, e11853 (2010).
45. Zhang, Y. et al. Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78**, 785–798 (2013).
46. Wang, Q. et al. Tagmentation-based whole-genome bisulfite sequencing. *Nat. Protoc.* **8**, 2022–2032 (2013).
47. Lun, A. T. & Smyth, G. K. De novo detection of differentially bound regions for ChIP-seq data using peaks and windows: controlling error rates correctly. *Nucleic Acids Res.* **42**, e95 (2014).
48. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
49. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
50. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, e3 (2004).
51. 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).
52. Hebestreit, K., Dugas, M. & Klein, H. U. Detection of significantly differentially methylated regions in targeted bisulfite sequencing data. *Bioinformatics* **29**, 1647–1653 (2013).
53. Cribari-Neto, F. & Zeileis, A. Beta regression in R. *J. Stat. Softw.* **34**, 1–24 (2010).
54. Schübeler, D. Function and information content of DNA methylation. *Nature* **517**, 321–326 (2015).
55. Allen, M. et al. Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. *Sci. Data* **3**, 160089 (2016).
56. Irizarry, R. A. et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).

nature research

Corresponding author(s): Philip L. De Jager

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
☒	☐	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Raw sequence data was generated by Illumina's HiSeq Control Software (HCS) and Real-Time Analysis (RTA) software. The StereoInvestigator software and ImageJ were used to quantify amyloid-beta and tau pathology burden respectively.

Data analysis

Analysis of ChIP-seq and ATAC-seq data:
Sequence alignment: BWA 0.7.4
BAM file processing: picard tools 1.737
Peak calling: MACS version 2
ChIP/ATAC-seq QC: R-packages GenomicRanges 1.24.3 + epigenomix 1.12.0
Negative binomial regression: R-package MASS 7.3-47
CBS segmentation algorithm: R-package DNAcopy 1.46.0
Connectivity map: https://portals.broadinstitute.org/cmap/ (build 2)
Linear regression with precision weights (voom): R-package limma 3.28.21

For human transcription and methylation data:
Linear regression: R-package stats 3.3.0
Beta regression: R-package betareg 3.1.0

Figures were generated using R and ggplot2 2.2.1.
All software is freely available.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Human H3K9ac ChIP-seq data has been deposited at Synapse (Synape: syn4896408). Mouse H3K9ac ChIP-seq data (GEO: GSE97560) and ATAC-seq data from induced neurons (GEO: GSE97409) have been deposited at Gene Expression Omnibus. Phenotype variables for the ROS/MAP studies can be requested from the RADC Research Sharing Hub (https://www.radc.rush.edu).

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

No sample-size calculation was performed, because the effect size of tau on genome-wide H3K9ac was unknown during study design. The number of n=669 samples (after QC) was determined by the availability of high-quality brain specimens of ROS/MAP study participants and is comparable to or larger than sample sizes of similar published studies.

Data exclusions

In total, 43 out of 712 ChIP-seq samples were excluded due to insufficient quality. We adapted the guidelines from the ENCODE project based on H3K9ac brain data from the Roadmap Epigenomics project to pre-establish quality control measures. We required H3K9ac ChIP-seq samples to have (i) ≥ 15x10^6 unique reads, (ii) non-redundant fraction ≥ 0.3, (iii) cross-correlation ≥ 0.03, (iv) fraction of reads in peaks ≥ 0.05 and (v) ≥ 6000 detected peaks. The QC process was independent of the samples' pathological characteristics.

Replication

We used different cohorts and data types to replicate/validate our findings in the human H3K9ac brain data:
1) Validation in human brain transcriptome of the same ROS/MAP cohort
2) Validation in human brain transcriptome of a different cohort (Mayo)
3) Validation in H3K9ac data from two different AD mouse models
4) Validation in ATAC-seq data from MAPT overexpressing induced neurons
5) Validation in H3K9Ac data from MAPT overexpressing induced neurons
Overall, these data consistently indicate an effect of tau on the epigenome. No attempts at replication failed and were excluded from the study.

Randomization

No allocation into groups was performed. Participants of the ROS and MAP studies were not cognitively impaired at enrollment. After death, pathological assessment was performed to measure tau and amyloid-beta burden. Subsequently, brain samples were send to H3K9ac ChIP-seq processing in an arbitrary order.

Blinding

Individuals generating the brain-derived ChIP-seq data were blinded to the outcome measures.

Reporting for specific materials, systems and methods

Materials & experimental systems		
n/a	Involved in the study	
☐	☒	Unique biological materials
☐	☒	Antibodies
☐	☒	Eukaryotic cell lines
☒	☐	Palaeontology
☐	☒	Animals and other organisms
☐	☒	Human research participants

Methods		
n/a	Involved in the study	
☐	☒	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials

Additional human samples are available from the individuals in our study through the RUSH University Alzheimer's Disease Center upon reasonable request. The CK-p25 mouse model is a cross between the strains "C57BL/6-Tg(tetO-CDK5R1/GFP)337Lht/J" and "B6;CBA-Tg(Camk2a-tTA)1Mmay/J" which can be obtained from The Jackson Laboratory (Stock No: 005706 and 003010). The tau mouse model (strain "B6;C3-Tg(Prnp-MAPT*P301S)PS19Vle/J") can also be ordered from The Jackson Laboratory (Stock No: 008169).

Antibodies

Antibodies used

For the human brain ChIP-seq, Millipore anti-H3K9ac (catalog # 06-942, lot: 31636) was used at a concentration of 2.5 µl of antibody / 3 ml of chromatin sample.
For the mouse brain ChIP-seq, Abcam anti-H3K9ac (catalog # ab4441, lots: GR196840-1 and GR207546-1) was used at a concentration of 1.5 µl of antibody / 1 ml of chromatin sample.

Validation

We assessed four different anti-H3K9ac antibodies, including two being sold as “ChIP grade” products. We prepared 6 chromatin samples from the frontal cortex and performed a ChIP with each of the antibodies. We the interrogated the efficacy of the ChIP using a nanostring panel developed by the Broad Institute’s Roadmap Epigenomics team that consists of 500 probes distributed throughout the genome in different chromatin states. Of the four antibodies tested, the Millipore anti-H3K9ac mAb captured the most DNA fragments for probes located in open chromatin, providing the best enrichment signal to background noise of random sequences found elsewhere in the genome. Validation data is availble from the manufacturer at http://www.emdmillipore.com/US/en/product/Anti-acetyl-Histone-H3-Lys9-Antibody,MM_NF-06-942

Validation data for the Abcam anti-H3K9ac Ab used for the mouse experiments is available at https://www.abcam.com/histone-h3-acetyl-k9-antibody-chip-grade-ab4441.html

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

iPSC line YZ1, derived from IMR90, from UCONN stem cell core

Authentication

We used STR profiling to confirm identity.

Mycoplasma contamination

Cells were tested negative for mycoplasma contamination.

Commonly misidentified lines (See ICLAC register)

The iPSC line used is not in the database of commonly misidentified cell lines.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

Female double-transgenic CK-p25 mice (C57BL/6-Tg(tetO-CDK5R1/GFP)337Lht/J) were 3 months old. Brain tissue was collected 2 or 6 weeks after p25 induction. Respective littermates were used as controls. For the Tau (P301S) mouse model (B6;C3-Tg(Prnp-MAPT*P301S)PS19Vle/J), adult female 6 months old and male 11 months old transgenic mice with age- and gendermatched wild-type mice as controls were used.

Wild animals

Our study did not involve wild animals.

Field-collected samples | Our study did not involve samples collected from the field.

Human research participants

Policy information about studies involving human research participants

Population characteristics

The mean age of death of the 669 subjects with H3K9ac profiles was 88.3 years. 233 of the 669 subjects were male, and 412 subjects were diagnosed with AD at pathological assessment. All analyses were adjusted for age and gender. The ROS and MAP studies are described in detail at the RADC Research Resource Sharing Hub: http://www.radc.rush.edu

Recruitment

Subjects participated in one of two very similar studies (ROS and MAP):
Religious Orders Study (ROS) (Enrollment began in 1994)
Design: Longitudinal, epidemiologic clinical-pathologic cohort study of aging and Alzheimer's disease (AD)
Participants: Older Catholic nuns, priests, and brothers from more than 40 groups across US
- Without known dementia at enrollment
- Agreeing to annual clinical evaluation and cognitive testing, baseline blood draw (subset with annual blood draw)
- Agreeing to brain donation after death (some in Chicago area also agree to spinal cord, nerve, and muscle donation)

Memory and Aging Project (MAP) (Enrollment began in 1997)
Design: Longitudinal, epidemiologic clinical-pathologic cohort study of common chronic conditions of aging with emphasis on decline in cognitive and motor function and risk of AD
Participants: Older adults recruited from retirement communities and subsidized senior housing facilities throughout Chicagoland and northeastern Illinois
- Without known dementia at enrollment
- Agreeing to annual clinical evaluation, cognitive testing, and blood draw
- Agreeing to organ donation (spinal cord, muscle and nerve, and brain)

ROS and MAP are studies of homogeneous communities and participants have already reached an old age without cognitive impairment at the time of enrollment. We cannot exclude self-selection biases, but since we studied the effect of tau pathology on the histone acetylome, we think that it is unlikely that a potential selection bias could affect our results.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

Human H3K9ac ChIP-seq data has been deposited at Synapse (Synapse: syn4896408):
https://www.synapse.org/#!Synapse:syn4896408
Mouse H3K9ac ChIP-seq data has been deposited at Gene Expression Omnibus (GEO: GSE97560):
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE97560
ATAC-seq data from iNs has been deposited at Gene Expression Omnibus (GEO: GSE97409):
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE97409

Files in database submission

Human brain H3k9ac ChIP-seq data:
- bam files with aligned sequence reads for all samples
- H3K9acDomains.csv (Location of H3K9ac domains)
- ReadCounts.csv (Matrix with read counts per sample and H3K9ac domain)
- QualityControl.csv (Quality measures for all samples)

Mouse brain H3K9ac ChIP-seq data:
- Raw sequencing data for all mouse samples
- GSE97560_domains.bed.gz (Location of H3K9ac domains)
- GSE97560_readCounts.csv.gz (Matrix with read counts per sample and H3K9ac domain)

ATAC-seq data from induced neurons:
- Raw sequencing data from all iN experiments
- GSE97409_domains.bed.gz (Location of ATAC domains)
- GSE97409_fragmentCounts.csv.gz (Matrix with read counts per sample and ATAC domain)

Genome browser session (e.g. UCSC)

no longer applicable

Methodology

Replicates

Study consists of ChIP-seq profiles from 669 subjects with various tau burden. Mouse model ChIP-seq data consists of 3 AD mice and 3 control mice for each time point. ATAC-seq data was generated in three different batches, each batch consisted of 3 MAPT overexpressing iN and 3 control iN cell cultures (n=18 experiments in total). All ATAC-seq samples were derived from the same iPSC line, but cultured and transduced independently.

Sequencing depth

A median of 55mio single-end short reads (36bp) were generated for human brain H3K9ac-ChIP-seq samples. Short paired-

Sequencing depth

end reads were generated for ATAC-seq. The exact numbers of reads generated and mapped for each sample are given in Supplementary Tables 1, 14 and 15.

Antibodies

Human brain: Millipore anti-H3K9Ac (catalog # 06-942, lot: 31636)
Mouse brain: Abcam anti-H3K9Ac (catalog # ab4441, lots: GR196840-1 and GR207546-1)

Peak calling parameters

The BWA program was used for read mapping with the single-end option "samse" for H3K9ac ChIP-seq data and the paired-end option "sampe" for ATAC-seq data. MACS 2.0 was applied to detect peaks in each sample individually. For the H3K9ac data, we set the options "broad", "keep-dup auto" and "qvalue 0.001", and used a pooled bam file of genomic DNA as control. For the ATAC-seq data, no control file was used, option "broad" was not set, but option "bampe" was set to make use of the paired-end information.

Data quality

We adapted the guidelines from the ENCODE project based on H3K9ac brain data from the Roadmap Epigenomics project to pre-establish quality control measures. We required human brain H3K9ac ChIP-seq samples to have (i) $\geq$ 15x10^6 unique reads, (ii) non-redundant fraction $\geq$ 0.3, (iii) cross-correlation $\geq$ 0.03, (iv) fraction of reads in peaks $\geq$ 0.05 and (v) $\geq$ 6000 detected peaks. The QC process was independent of the samples' pathological characteristics. QC metrics for all samples are given in Supplementary Tables 1, 14 and 15.

Software

The R software was used to analyze ChIP-seq and ATAC-seq data. R-packages GenomicRanges and epigenomix were used to count the number of reads for each sample in each domain. Read counts were analyzed using negative binomial regression models implemented in the R-package MASS. R code is provided as Supplementary Software linked to this article.