A histone acetylome-wide association study of Alzheimer's disease identifies disease-associated H3K27ac differences in the entorhinal cortex

Sarah J. Marzi[1,2], Szi Kay Leung[3^], Teodora Ribarska[4^], Eilis Hannon[3], Adam R. Smith[3], Ehsan Pishva[3,5], Jeremie Poschmann[3,6], Karen Moore[3], Claire Troakes[1], Safa Al-Sarraj[1], Stephan Beck[7], Stuart Newman[8], Katie Lunnon[3], Leonard C. Schalkwyk[8+], Jonathan Mill[3+*]

[1] Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK.
[2] The Blizard Institute, Queen Mary University of London, London, UK.
[3] University of Exeter Medical School, University of Exeter, Exeter, UK.
[4] Oslo University Hospital, Oslo, Norway.
[5] Department of Psychiatry and Neuropsychology, Maastricht University Medical Centre, Maastricht, The Netherlands
[6] Centre de Recherche en Transplantation et Immunologie, Inserm, Université de Nantes, Nantes, France.
[7] UCL Cancer Institute, University College London, London, UK.
[8] University of Essex, Colchester, UK
[^/+] Equal contributions

[*] **Correspondence: Professor Jonathan Mill, University of Exeter Medical School, RILD Building (Level 4), Royal Devon & Exeter Hospital, Barrack Rd, Exeter. EX2 5DW. UK.**

Abstract

We quantified genome-wide patterns of lysine H3K27 acetylation (H3K27ac) in entorhinal cortex samples from Alzheimer's disease (AD) cases and matched controls using chromatin immunoprecipitation and highly parallel sequencing (ChIP-seq). We observed widespread acetylomic variation associated with AD neuropathology, identifying 4,162 differential peaks (FDR < 0.05) between AD cases and controls. Differentially acetylated peaks were enriched in disease-related biological pathways and included regions annotated to genes involved in the progression of Aβ and tau pathology (e.g. *APP*, *PSEN1*, *PSEN2*, and *MAPT*), as well as regions containing variants associated with sporadic late-onset AD. Partitioned heritability analysis highlighted a highly-significant enrichment of AD risk variants in entorhinal cortex H3K27ac peak regions. AD-associated variable H3K27ac was associated with transcriptional variation at proximal genes including *CR1*, *GPR22*, *KMO*, *PIM3*, *PSEN1* and *RGCC*. In addition to identifying molecular pathways associated with AD neuropathology, we present a framework for genome-wide studies of histone modifications in complex disease.

Alzheimer's disease (AD) is a chronic neurodegenerative disorder characterized by cognitive decline and memory loss that contributes substantially to the global burden of disease, affecting in excess of 26 million people worldwide[1]. The symptoms of AD are associated with progressive neuropathology in the neocortex, with regions surrounding the entorhinal cortex being particularly affected early in the disease[2]. These neuropathological hallmarks of AD include the extracellular deposition of neurotoxic amyloid-β (Aβ) in the form of amyloid plaques and an accumulation of intracellular neurofibrillary tangles composed of hyperphosphorylated tau[3]. Despite progress in understanding risk factors contributing to AD progression, the mechanisms involved in disease progression are not fully understood and long-term treatments, reversing the cellular disease process in the cortex, are elusive.

There has been considerable success in identifying genetic risk factors for AD[4]. While autosomal dominant mutations in three genes (*APP*, *PSEN1*, and *PSEN2*) can explain early-onset (< 65 years) familial AD, these account for only 1-5% of the total disease burden[5]. Most cases of AD are late-onset (> 65 years), non-Mendelian and highly sporadic, with susceptibility attributed to the action of highly prevalent genetic variants of low penetrance. In addition to the well-established risk associated with the *APOE* locus[6] there has been notable success in identifying novel AD-associated variants capitalising on the power of genome-wide association studies (GWAS) in large sample cohorts; a recent large GWAS meta-analysis of AD, incorporating > 74,000 samples, identified 19 genome-wide significant risk loci for sporadic AD[7]. Despite these advances, little is known about the functional mechanisms by which risk variants mediate disease susceptibility.

Increased understanding about the functional complexity of the genome has led to growing recognition about the likely role of non-sequence-based regulatory variation in health and disease. Building on the hypothesis that epigenomic dysregulation is important in the etiology and progression of AD neuropathology[8], we and others recently performed the first genome-scale cross-tissue analyses of DNA methylation in AD identifying robust DNA methylation differences associated with AD neuropathology across multiple independent human post-mortem brain cohorts[9,10]. To date, however, no study has systematically examined other types of regulatory genomic modifications in AD. In this study, we focus on lysine H3K27 acetylation (H3K27ac), a robust mark of active enhancers and promoters that is strongly correlated with gene expression and transcription factor binding[11]. Interestingly, histone deacetylase (HDAC) inhibitors have been shown to ameliorate symptoms of cognitive decline and synaptic dysfunction in mouse models of AD[12] and are promising targets for novel human AD treatments[13]. Despite this, investigations into global levels of histone acetylation in AD have thus far been inconclusive[14-16] and no study has taken a

genome-wide approach. In fact, few studies have systematically profiled H3K27ac across large numbers of samples in the context of complex disease, and optimal methods for these analyses are still being developed[17].

In this study, we used chromatin immunoprecipitation combined with highly-parallel sequencing (ChIP-seq) to quantify levels of H3K27ac across the genome in post-mortem entorhinal cortex samples from AD patients and matched controls, incorporating cell-type-specific DNA methylation and gene expression marks to control for cellular heterogeneity. We identify regulatory genomic signatures associated with AD, including variable H3K27ac across discrete regions annotated to genomic loci mechanistically implicated in the onset of both tau and amyloid pathology, associating many of these differences with variation in gene expression. This is the first study of variable H3K27ac yet undertaken for AD; in addition to identifying molecular pathways associated with AD neuropathology, we present a framework for genome-wide studies of this modification in complex disease.

Results

Genome-wide profiling of inter-individual variation in H3K27ac in the entorhinal cortex

We generated high-quality H3K27ac ChIP-seq data using post-mortem entorhinal cortex tissue dissected from 47 elderly individuals (mean age = 77.43, SD = 9.66, range = 58-97) comprising both AD cases (n = 24, mean Braak stage = 6.00, SD = 0.00) and age-matched low pathology controls (n = 23, mean Braak stage = 1.30, SD = 1.11) **(Supplementary Table 1**). Raw H3K27ac ChIP-seq data is available for download from the Gene Expression Omnibus (GEO) (accession number GSE102538). Genome-wide SNP data was used to confirm that each of the samples included in our analysis was of Western European ancestry (**Supplementary Fig. 1**). After stringent quality control (QC) of the raw H3K27ac ChIP-seq data (see **Methods**), we obtained a mean of 30,032,623 (SD = 10,638,091) sequencing reads per sample, with no difference in read-depth between AD cases and controls (Welsh two-sample t-test, two-sided, n = 47 biologically independent samples, P = 0.93, average read count difference: 260,673, 95%-confidence interval (CI): -6,602,319-6,080,972, t(43.85) = -0.08; **Supplementary Fig. 2**). This represents, to our knowledge, the most extensive analysis of H3K27ac in the human entorhinal cortex yet undertaken. Using combined data from all 47 samples (see **Methods**) we identified 182,065 high confidence H3K27ac peaks; these are distributed across all 24 chromosomes (**Supplementary Table 2**) spanning a mean length of 983bp (SD = 682bp) with a mean distance between neighbouring peaks of 15,536bp (SD = 116,040bp). We validated the identified peaks using two independent ChIP-seq datasets: first, we obtained locations for cortex and cerebellum H3K27ac peaks from a recent analysis of autism and control brains[17]; second, we downloaded peak profiles for

multiple cell- and tissue-types from the NIH Epigenomics Roadmap Consortium[18] (see **Methods**). As expected, there was a very high overlap between H3K27ac peaks called in these previous neocortical datasets and our ChIP-seq data, with a notably lower overlap with H3K27ac data from non-cortical tissues (**Supplementary Fig. 3** and **Supplementary Fig. 4)**. For example, ~95% of BA9 H3K27ac peaks identified by Sun et al[17] overlap our peaks, compared to ~77% of the cerebellum peaks identified in their experiment. Furthermore, samples profiled by the NIH Epigenomics Roadmap Consortium overlap our peaks in a tissue-specific manner with up to 97% overlap observed for H3K27ac profiles derived cortical tissues and much lower overlaps from non-brain tissues (22-47%)**.**

AD-associated differential acetylation in the entorhinal cortex

We quantified read counts across every peak in each of the 47 individual samples included in the ChIP-seq study using *HTSeq* and employed a quasi-likelihood F test, implemented in the Bioconductor package *EdgeR* (see **Methods** for full description), to test for differences in H3K27ac between AD cases and low pathology controls. Our primary analysis model controlled for age at death and neuronal cell proportion estimates derived from DNA methylation data generated on the same samples (**Supplementary Table 1**, **Supplementary Fig. 5**). A total of 4,162 (2.3%) of the 182,065 peaks were characterized by AD-associated differential acetylation at a false discovery rate (FDR) < 0.05 (**Fig. 1**), with a significant enrichment of hypoacetylated AD-associated peaks (2,687 (1.5%)) compared to hyperacetylated AD-associated peaks (1,475 (0.8%)) (exact binomial test, n = 4,162 peaks, *P* < 1.00E-50, percentage of hypoacetylated peaks: 65%, 95%-CI: 63-66%) (**Fig. 1**). Because there are global differences in H3K27ac between males and females – the first principal component of variable H3K27ac was associated with sex in our data (**Supplementary Fig. 6**) – we undertook subsequent sensitivity analyses, confirming that 4,157 (99.9%) of the 4,162 differentially acetylated peaks remain significantly differentially acetylated (FDR < 0.05) when additionally controlling for sex, with a near-perfect correlation in AD-associated H3K27ac differences between models (Pearson's product-moment correlation, r = 1.00, P < 1.00E-50; **Supplementary Fig. 7**). Given the neurodegeneration and neural cell changes associated with AD pathology, we also quantified the levels of transcripts associated with five major brain cell types (*ENO2* (neurons), *OLIG2* (oligodendrocytes), *GFAP* (astrocytes), *CD68* (microglia) and *CD34* (endothelial cells)) in our samples (see **Methods**). As expected, expression of the neuronal gene *ENO2* was correlated with the neuronal cell proportion estimates derived from DNA methylation data (**Supplementary Fig. 8**) and found to be reduced in AD samples (linear regression, n = 47 biologically independent samples, *P* = 0.023, β = -0.54, F(3,43) = 3.23) (**Supplementary Fig. 8, Supplementary Fig. 9** and **Supplementary Table 3**). We also identified elevated

expression of *CD34* (n = 47 biologically independent samples, *P* = 0.029, β = 0.76, F(3,44) = 7.07), *CD68* (n = 46 biologically independent samples, *P* = 0.012, β = 0.53, F(3,42) = 7.25), and *GFAP* (n = 47 biologically independent samples, *P* = 0.003, β = 0.76, F(3,44) = 3.48) in our AD cases (**Supplementary Fig. 9** and **Supplementary Table 3**), reflecting results from previous studies[19,20]. Post-hoc analyses of our AD-associated peaks using models controlling for levels of these five cell-types showed that effects remained highly correlated with those from our original model (Pearson's product-moment correlations with results from models controlling for i) *ENO2*: r = 0.99, *P* < 1.00E−50; ii) *OLIG2*: r = 0.99, *P* < 1.00E−50; iii) *GFAP*: r = 0.99, *P* < 1.00E−50; iv) *CD68*: r = 0.99, *P* < 1.00E−50; v) *CD34*: r = 0.99, *P* < 1.00E−50; and vi) all cell markers combined: r = 0.98, *P* < 1.00E−50), indicating that the AD-associated differences in H3K27ac are robust to cell-type heterogeneity (**Supplementary Fig. 10)**. Finally, for each sample we also calculated standard ENCODE ChIP-seq quality metrics (see **Methods**) and included these as post-hoc covariates in our analyses; again AD-association effect-sizes remained highly correlated with those estimated from our original model (Pearson's product-moment correlation, r = 0.95, *P* < 1.00E−50).

UCSC Genome Browser tracks showing H3K27ac levels in AD cases and controls, in addition to association statistics, across the genome can be accessed at https://epigenetics.essex.ac.uk/AD_H3K27ac/. The ten top-ranked hyper- and hypoacetylated peaks associated with AD are shown in **Table 1,** with a complete list given in **Supplementary Table 4** (hyperacetylated peaks) and **Supplementary Table 5** (hypoacetylated peaks). *P* values for sex, cell-type expression and quality metric controlled models at each of the 4,162 differentially acetylated peaks are reported in **Supplementary Table 4** and **Supplementary Table 5**. Peaks were subsequently annotated to genes using an approach that takes into account the strength of proximal and distal DNA-binding events (see **Methods**). In total, differentially-acetylated peaks were annotated to 4,039 genes (hyperacetylated peaks: 1,728 genes; hypoacetylated peaks: 2,528 genes). The most significant AD-associated hyperacetylated peak (chr13: 112101248-112102698; *P* = 2.04E-08; log fold change = 0.93) is annotated to both *SOX1* and *TEX29* on chromosome 13 (**Fig. 2**, **Table 1**). Of note, H3K27ac data from the Epigenomics Roadmap Consortium show that this region is characterized by brain-specific enhancer activity (**Supplementary Fig. 11**). The most significant AD-associated hypoacetylated peak (chr7: 64011549-64012825; *P* = 1.66E-08; log fold change = -0.86) is located within intron 1 of *ZNF680* on chromosome 7 (**Fig. 3**, **Supplementary Fig. 11**, **Table 1**). Global clustering of samples by normalized read counts across all hyper- and hypoacetylated peaks (FDR < 0.05) indicated that, as expected, samples group primarily by disease status (**Fig. 4**). AD-associated differentially acetylated peaks (FDR < 0.05) are significantly longer (Welsh two-sample t-test, two-sided, *P* < 1.00E-

50, average difference in length = 320.02bp, 95%-CI: 298-342bp, t(4340.5) = 28.70) and characterized by higher read-depths (CPM; P < 1.00E-50, average difference in log CPM = 0.46, 95%-CI: 0.42-0.50, t(4463.7) = 24.33) than non-significant peaks (**Supplementary Fig. 12**). Of note, within AD-associated peaks, hypoacetylated peaks are significantly longer (P = 5.66E-31, average difference in length = 331.23bp, 95%-CI: 288-374bp, t(3303.8) = 15.17) and have higher read-depths (P = 2.69E-50, average difference in log CPM = 0.45, 95%-CI: 0.37-0.52, t(3111.4) = 11.70) compared to hyperacetylated peaks. We used *RSAT* to identify enriched transcription factor binding motifs located within AD-associated differentially acetylated peaks (see **Methods**), observing a significant enrichment of binding motifs for specificity protein 1 (Sp1) (binomial test, P < 1.00E-50), a transcription factor that has been implicated in the immune response, apoptosis and chromatin remodelling, amongst AD-hyperacetylated peaks (FDR < 0.05). Of note, previous publications have reported dysregulated expression of Sp1 and its co-localization with neurofibrillary tangles in AD[21,22].

Differential H3K27ac is observed in regulatory regions annotated to genes previously implicated in both tau and amyloid neuropathology

One of the top-ranked AD-associated hyperacetylated peaks is located proximal to the gene encoding microtubule associated protein tau (MAPT) (chr17: 43925717-43927482; P = 7.01E-07; log fold change = 0.71; **Table 1**), which is widely expressed in the nervous system where it functions to promote microtubule assembly and stability. Tau is believed to play a key role in AD neuropathology, with hyperphosphorylation of the tau protein precipitating the neurofibrillary tangles associated with the pathogenesis of AD[23,24]. Closer inspection of the region around this AD-associated peak highlighted an extended cluster of six hyper-acetylated H3K27ac peaks (FDR < 0.05) spanning 36kb (chr17: 43925717 - 43961546) located within a *MAPT* antisense transcript (*MAPT_AS1*) ~10kb upstream of the *MAPT* transcription start site (**Fig. 5**; **Supplementary Table 6**). H3K27ac ChIP-seq data from the NIH Epigenomics Roadmap Consortium show that this region is characterized by CNS-related H3K27ac signatures (**Fig. 5**), with *ChromHMM*[25] identifying the region as an active chromatin domain in brain comprised of enhancers and blocks of weak transcription (**Supplementary Fig. 13**). Strikingly, AD-associated differentially-acetylated peaks were also found in the vicinity of other genes known to play a direct mechanistic role in AD. We identified a significantly hypoacetylated peak (chr21: 27160993 - 27161475; P = 3.94E-04; log fold change = -0.72) on chromosome 21, located ~100kb downstream of the amyloid precursor protein gene (*APP*), which encodes the precursor molecule to Aβ, the main component of amyloid plaques[26-28] (**Supplementary Fig. 14**). We also identified significant hyperacetylation in the vicinity of the presenilin genes *PSEN1* and *PSEN2*, which encode integral components of the gamma secretase complex and play a key role in generation of

Aβ from APP[29]. In *PSEN1* we found significantly elevated H3K27ac across a peak within intron 6 (chr14: 73656445 - 73656860; *P* = 3.44E-04; log fold change = 0.68; **Supplementary Fig. 15**). In *PSEN2* we identified consistent hyperacetylation in AD cases across nine H3K27ac peaks (FDR < 0.05) spanning a ~57 kb region upstream of the transcription start-site (chr1: 226957424 - 227014019; **Fig. 6, Supplementary Fig. 13** and **Supplementary Table 7**). Of note, highly-penetrant mutations in *APP*, *PSEN1*, and *PSEN2* are associated with familial forms of early-onset AD[30]. We used a hypergeometric test (one-sided) to confirm a significant enrichment of AD-associated differential acetylation associated with i) familial AD genes (*APP*, *PSEN1*, *PSEN2*): 11 FDR-significant peaks from a total of 163 annotated peaks, *P* = 0.001; ii) the *MAPT* locus: six FDR-significant peaks from a total of 36 annotated peaks, *P* = 1.54E-04; and iii) all four AD pathology genes combined: 17 FDR-significant peaks from a total of 199 annotated peaks, *P* = 4.05E-06. The identification of altered regulation of these loci in late-onset sporadic AD brain further supports a key role for altered amyloid processing in the onset of neuropathology.

Genetic risk for AD is enriched in entorhinal cortex H3K27ac peaks, with specific differentially-acetylated peaks overlapping known AD GWAS regions

Using the AD GWAS meta-analysis results from Lambert and colleagues[7] we performed LD Score regression to quantify the proportion of SNP heritability accounted for by variants colocalizing with the entorhinal cortex H3K27ac peaks identified in this study (see **Methods**). Of the total heritability across all SNPs derived from stage 1 of the GWAS meta-analysis results (h^2 = 0.0789), a large proportion (38.3%, h^2 = 0.0302 (95%-CI: 0.0126 - 0.0478)) is accounted for by variants within H3K27ac peaks, representing a significant enrichment (enrichment = 1.10 (95%-CI: 1.05 - 1.15) (**Supplementary Table 8**), and indicating that AD risk alleles are preferentially located in regions of regulatory/enhancer function in the brain. We next identified instances where there is an overlap between AD-associated differential H3K27ac and genomic regions harbouring risk variants. Briefly, we defined linkage-disequilibrium (LD) blocks around the genome-wide significant (*P* < 5.0E-08) GWAS variants identified by the stage 1 meta-analysis by Lambert and colleagues[7] (**Supplementary Table 9**), which contained a total of 292 overlapping entorhinal cortex H2K27ac peaks (see **Methods**). Two of the 11 GWAS LD blocks contained significant AD-associated H3K27ac peaks (FDR < 0.05), although there was no overall enrichment of AD-associated differential acetylation at the 292 peaks (Wilcoxon rank-sum test with continuity correction, n= 182,065 peaks, *P* = 0.364, W = 27354000). Two peaks of AD-associated hyperacetylation were located within a GWAS region on chromosome 1, mapping to the gene body of *CR1* (chr1: 207753457-207753813; *P* = 1.15E-06; log fold change = 0.99 and chr1: 207754916-207756572; *P* = 5.40E-04; log fold change = 0.56; **Supplementary Fig. 16**). *CR1* encodes a

transmembrane glycoprotein expressed in microglia with a role in the innate immune system, promoting phagocytosis of immune complexes and cellular debris, in addition to $A\beta^{31\text{-}33}$. Two other AD-associated differentially acetylated peaks were found to be located within a GWAS region on chromosome 19, including a hyperacetylated peak (chr19:45394441- 45395396; *P* = 2.13E-04; log fold change = 0.48) mapping to the gene body of *TOMM40* in the immediate vicinity of *APOE* (**Supplementary Fig. 17**). Another H3K27ac peak in this LD block was significantly hypoacetylated in AD (chr19: 45639588-45641733; *P* = 7.65E-04; log fold change = -0.33), mapping to intron 1 of *PPP1R37*.

AD-associated differentially-acetylated peaks are enriched for functional processes related to neuropathology

We next calculated statistical enrichments for ontological annotations amongst our AD-associated peaks (see **Methods**), interrogating gene ontologies for molecular function and biological processes as well as human diseases (hypergeometric test; see **Supplementary Table 10** and **Supplementary Fig. 18**, with a full list of significant ontologies in **Supplementary Tables 11-16**). Multiple ontological categories associated with AD progression and pathology were identified as being enriched (FDR < 0.05) amongst hyperacetylated peaks, including "lipoprotein particle binding"[6,34] (*P* = 1.10E-06), "beta-amyloid metabolic process"[23] (*P* = 4.94E-08), "response to hypoxia"[35,36] (*P* = 3.17E-14), and "Pick's disease" (*P* = 2.93E-07), a form of fronto-temporal dementia also characterized by tau pathology[24,37]. Amongst hypoacetylated peaks we observed an enrichment of categories related to neurotransmitter-functions, including "GABA receptor activity"[38] (*P* = 2.70E-07) as well as categories related to neuronal transmission and synapses, such as "protein location to synapse" (*P* = 7.86E-09). Because the observed enrichment for functional processes related to neuropathology might reflect underlying cellular heterogeneity between samples we repeated these analyses using only peaks that were significantly differentially acetylated (FDR < 0.05) in our model controlling for all five neural cell types. Our sensitivity analyses confirmed that most AD-related pathway enrichments were robust to cellular heterogeneity including "lipoprotein particle binding" (*P* = 4.36E-05), "apolipoprotein binding" (*P* = 1.24E-04), "response to hypoxia" (*P* = 5.28E-08), "beta-amyloid metabolic process" (*P* = 1.61E-05) and "Pick's disease" (*P* = 4.04E-05). Although there is some overlap between the genes in each pathway (**Supplementary Table 10**), and the ontological categories we identify are not totally independent, these results indicate a striking enrichment of AD-related pathways in both hyper- and hypoacetylated regions.

AD-associated variable H3K2ac is associated with the expression of nearby genes.

We next quantified the expression of selected cortex-expressed genes located proximal to top-ranked differentially-acetylated peaks (*RGCC*, *PIM3*, *ANKRD17*, *ZNF680*, *GPR22*, and *KMO*) in an extended set of 95 entorhinal cortex samples (average age = 81.3, SD = 9.3, range = 58-99) from both AD cases (n = 67, mean Braak stage = 5.4, SD = 0.8) and age-matched low pathology controls (n = 28, mean Braak stage = 1.5, SD = 1.1) **(Supplementary Table 1**). The expression of *RGCC* and *PIM3*, both located proximal to hyperacetylated peaks, was found to be significantly up-regulated in AD cases (linear regression; *RGCC*: n = 94 biologically independent samples, *P* = 0.002, β = 0.44, F(3,92) = 25.78; *PIM3*: n = 95 biologically independent samples, *P* = 9.52E-05, β = 0.57, F(3,92) = 2.88) (**Fig. 7**). Furthermore, the expression of *GPR22* and *KMO*, both located proximal to hypoacetylated peaks, was found to be significantly down-regulated in AD cases (*GPR22*: n = 95 biologically independent samples, *P* = 0.005, β = -0.71, F(3,90) = 12.09; *KMO*: n = 60 biologically independent samples, *P* = 0.006, β = -0.96, F(3,57) = 2.77) (**Fig. 7**). Of note, the expression of these four transcripts was positively associated with levels of H3K27ac (i.e. higher H3K27ac is associated with a reduced delta Ct value) across their proximal AD-associated differentially acetylated peaks in samples included in our ChIP-seq dataset (quasi-likelihood F test, two-sided; *RGCC*: n = 47 biologically independent samples, *P* = 0.123, log fold change = -0.22; *PIM3*: n = 47 biologically independent samples, *P* = 0.023, log fold change = -0.35; *GPR22*: n = 46 biologically independent samples, *P* = 1.1E-04, log fold change = -0.38; *KMO*: n = 33 biologically independent samples, *P* = 0.050, log fold change = -0.15) (**Supplementary Fig. 19**). Interestingly, previous studies have implicated dysregulation of *RGCC*[39], *GPR22*[40] and *KMO*[41] in AD. In contrast, *ANKRD17* and *ZNF680* were not found to be differentially expressed in AD cases (**Supplementary Table 17**); of note H3K27ac enhancer domains often reside considerable distance from their target gene[42] and effects on more distal transcripts cannot be excluded. We also quantified the expression of six genes - previously implicated in tau and amyloid pathology or from GWAS – that were annotated to differentially acetylated peaks (*MAPT*, *PSEN1*, *PSEN2*, *APP*, *TOMM40* and *CR1*) (**Supplementary Table 18**). We found significantly higher expression of *PSEN1* (linear regression, n = 95 biologically independent samples, *P* = 4.91E-04, β = 0.45, F(3,92) = 4.98) and *CR1* (n = 91 biologically independent samples, *P* = 0.001, β = 0.72, F(3,88) = 5.89) in AD cases (**Fig. 7**), reflecting the hyperacetylated peaks observed in the vicinity of these genes, with the expression of both transcripts being robustly associated with levels of H3K27ac across their proximal AD-associated differentially acetylated peaks in samples included in our ChIP-seq dataset (quasi-likelihood F test, two-sided; *PSEN1*: n = 47 biologically independent samples, *P* = 0.011, log fold change = -0.42; *CR1*: n = 45 biologically independent samples, peak 1: *P* = 0.002, log fold change = -0.37;peak 2: *P* = 0.001, log fold change =-0.31) (**Supplementary Fig. 20**). In contrast, the expression of

MAPT (linear regression, n = 94 biologically independent samples, P = 0.783, β = -0.03, F(3,91) = 17.19), *PSEN2* (n = 88 biologically independent samples, P = 0.913, β = 0.01, F(3,84) = 0.64), *APP* (n = 95 biologically independent samples, P = 0.078, β = -0.22, F(3,92) = 2.02) and *TOMM40* (n = 94 biologically independent samples, P = 0.286, β = -0.09, F(3,90) = 3.14) was not significantly different between AD cases and controls, or associated with proximal levels of H3K27ac (**Supplementary Table 18**).

Integrative analysis of DNA and histone modifications reveal unique distributions of DNA modifications across regions of differential acetylation

Our previous work identified cortex-specific variation in DNA methylation (5mC) robustly associated with AD pathology[9,10]. We were therefore interested in exploring the relationship between H3K27ac and both 5mC and another DNA modification – DNA hydroxymethylation (5hmC), which is enriched in the brain and believed to play an important role in neuronal function, learning and memory[43,44] - in our samples. Both modifications were profiled using DNA isolated from the same entorhinal cortex samples using oxidative bisulfite (oxBS) conversion in conjunction with the Illumina 450K HumanMethylation array ("450K array") (see **Methods**). Focusing on Illumina 450K sites within 1kb of our H3K27ac peaks, we identified 268,477 probe-peak pairs (comprising of 232,233 unique 450K array probes and 62,714 (34.45% of total) unique H3K27ac peaks). 6,838 probes mapped to within 1kb of an AD-associated differentially acetylated peak (FDR < 0.05; n = 1,649 unique peaks (616 hyperacetylated, 1,033 hypoacetylated)). First, we tested for differential 5mC and 5hmC associated with AD at these probes, controlling for age at death and cell-type proportion estimates. None of the differences in 5mC (minimum P = 2.47E-03) or 5hmC (minimum P = 1.53E-03) were significant when correcting for multiple testing (n = 6,838 tests; P < 7.31E-05), indicating that there is little direct overlap in AD-associated variation in H3K27ac and DNA modifications. Comparing effect sizes at these 6,838 peak–probe pairs identified no evidence for an overall correlation between AD-associated H3K27ac and 5mC differences (Pearson's product-moment correlation, r = 0.009, P = 0.443; **Supplementary Fig. 21**) with a small, but significant, negative correlation for 5hmC (Pearson's product-moment correlation r = -0.045, P = 1.63E-04; **Supplementary Fig. 21**). As expected, both DNA modifications are significantly lower in the vicinity of H3K27ac peaks compared to the genome-wide 450K array background (Welsh two-sample t-test, two-sided; 5mC: P < 1.00E-50**,** average beta difference = 12.47%, 95%-CI: 12.35-12.59%, t(551560) = 195.94; 5hmC: P = 3.61E-30, average beta difference = 0.16%, 95%-CI: 0.13-0.19%, t(494170) = 10.99; **Supplementary Fig. 21**), consistent with H3K27ac being localized at active enhancers and promoters. We next explored the relationship between variable 5mC and H3K27ac in our samples, finding that data for 439 probe-peak pairs (corresponding to 419 unique 450K

array probes and 319 unique H3K27ac peaks) were significantly correlated (FDR < 0.05) (**Supplementary Table 19**); 414 (94.31%) of the significant correlations were negative, with higher H3K27ac being associated with lower 5mC. Of note, 12 of the significant associations between H3K27ac and 5mC involve an AD-associated differentially acetylated peak (**Supplementary Table 20** and **Supplementary Fig. 22**). Interestingly, we identified an AD-associated differentially acetylated peak (chr4:1044452- 1044737, quasi-likelihood F test, two-sided, *P* = 0.001, log fold change = 0.83) annotated to *FGFRL1* and *RNF212* at which H3K27ac is correlated with 5mC at three specific 450K array probes (Pearson's product-moment correlation, n = 42 biologically independent samples; cg04016957, r = -0.66, *P* = 1.66E-07; cg04106633, r = -0.71, *P* = 1.36E-07; cg21130718, r = -0.70, *P* = 2.98E-07).

Discussion

We quantified H3K27ac across the genome in post-mortem entorhinal cortex tissue samples, identifying widespread AD-associated acetylomic variation. Strikingly, differentially acetylated peaks were identified in the vicinity of genes implicated in both tau and amyloid neuropathology as well as genomic regions containing variants associated with sporadic late-onset AD. Partitioned heritability analysis highlighted a highly-significant enrichment of AD risk variants in entorhinal cortex H3K27ac peak regions. Finally, targeted gene expression analysis showed that variable H3K27ac is associated with transcriptional variation at proximal genes including *CR1*, *GPR22*, *KMO*, *PIM3*, *PSEN1* and *RGCC*. This is the first study of variable H3K27ac yet undertaken for AD; in addition to identifying molecular pathways associated with AD neuropathology, we introduce a framework for genome-wide studies of this modification in complex disease.

Given its close relationship with transcriptional activation, for example via the mediation of transcription factor binding, the identification of AD-associated variation in H3K27ac highlights potential novel regulatory genomic pathways involved in disease etiology. We find widespread alterations in H3K27ac associated with AD, including in the vicinity of several genes known to be directly involved in the progression of Aβ and tau pathology[23,45] (*APP*, *PSEN1*, *PSEN2*, *MAPT*), supporting the notion that dysregulation of both pathways is involved in the onset of AD. Interestingly, although our study assessed brains from donors affected by sporadic late-onset AD, we identify widespread altered H3K27ac in the vicinity of genes implicated in familial early-onset AD. This indicates that these two forms of the disease may share common pathogenic pathways and mechanisms. Given that histone-acetylation modifiers are amongst the most promising target pharmacological treatments of AD[13,46], the identification of altered H3K27ac in AD is important, giving clues as to which genes and pathways may be involved.

Our study has a number of limitations, which should be considered when interpreting these results. First, we undertook ChIP-seq using bulk entorhinal cortex samples comprising a mix of neuronal and non-neuronal cell-types. This is an important limitation in epigenomic studies of a disease characterized by cortical neuronal loss. However, we were able to control, in part, for variation in neuronal proportions in our samples by i) deriving neuronal proportion estimates for each sample using DNA methylation data generated on the same tissue samples[47] and ii) quantifying the expression of levels of transcripts associated with five major brain cell types (*ENO2* (neurons), *OLIG2* (oligodendrocytes), *GFAP* (astrocytes), *CD68* (microglia) and *CD34* (endothelial cells)) in our samples. Despite these efforts to control for cellular heterogeneity, however, it is plausible that we have not selected the most optimal cell type-specific markers for this purpose, especially because certain markers reflect both the activity and abundance of specific cell types (e.g. CD68 in microglia). Second, our cross-sectional analysis of post-mortem brain tissue makes direct causal inference difficult, and it is likely that many of the changes in H3K27ac we observe result from the AD pathology itself. In this regard, however, it is interesting that we see disease-associated H3K27ac in the vicinity of genes causally implicated in familial forms of AD and we were able to show that genes annotated to specific differentially acetylated peaks showed disease-associated gene expression differences in the same samples. Third, although our targeted gene expression analyses identified differences at a number of genes annotated to differentially-acetylated peaks, we cannot make any conclusions about the relationship between H3K27ac and gene expression at a genome-wide level. Fourth, we have assessed a relatively small number of samples. In this light, it is notable that we identify substantial differences between AD cases and controls, with disease-associated regulatory variation in genes and functional pathways known to play a role in the onset and progression of neuropathology. The clear clustering between patients and controls at our differentially acetylated peaks suggests that, despite a complex and heterogeneous etiology, AD may be characterized by a common molecular pathology in the entorhinal cortex, reflecting neuropathological analyses. Furthermore, our differential gene expression analyses of transcripts associated with selected differentially acetylated peaks highlighted consistent differences in an extended set of samples. Fifth, chromatin architecture and transcriptional regulation is influenced by a multitude of epigenetic mechanisms. Although profiling H3K27ac can provide relatively robust information about transcriptional activity, it represents only one of perhaps ~100 post-translational modifications occurring at > 60 histone amino-acid residues regulating genomic function. A recent study, for example, identified dysregulation of H4K16ac in AD brain[48]. Finally, we annotated genes to H3K27ac peaks by assigning genes to regulatory regions within 5kb upstream and 1kb downstream of

the transcription start site (proximal) or up to 1000kb for distal interactions, although such an approach may not be optimal. Recent chromatin conformation studies suggest that distal enhancers do not necessarily regulate the most proximal gene[49]; although no detailed Hi-C data currently exists for adult human cortex, the generation of these data will enable us to further interrogate the functional consequences of the AD-associated differences reported here.

In summary, we provide compelling evidence for widespread acetylomic dysregulation in the entorhinal cortex in AD. Our data suggest that regulatory variation at multiple loci, including in the vicinity of several known AD risk genes – *APP*, *CR1*, *MAPT*, *PSEN1*, *PSEN2* and *TOMM40* – is robustly associated with disease, supporting the notion of common molecular pathways in both familial and sporadic AD. In addition to identifying molecular pathways associated with AD neuropathology, we present a framework for genome-wide studies of histone modifications in complex disease, integrating our data with results obtained from genome-wide association studies as well as other epigenetic marks profiled on the same samples.

Accession codes

Gene expression omnibus (GEO): accession number GSE102538.

Acknowledgements

This work was funded by US National Institutes of Health grant R01 AG036039 to J.M. S.J.M. and T.R. were funded by the EU-FP7 Marie Curie ITN EpiTrain (REA grant agreement no. 316758). S.K.L. is funded by a Medical Research Council (MRC) CASE PhD studentship. Sequencing infrastructure was supported by a Wellcome Trust Multi User Equipment Award (WT101650MA) and Medical Research Council (MRC) Clinical Infrastructure Funding (MR/M008924/1). Generation of DNA hydroxymethylation data was funded by an Alzheimer's Association US New Investigator Research Grant (grant number NIRG-14-320878) to K.L., and a grant from BRACE (Bristol Research into Alzheimer's and Care of the Elderly) to K.L. The authors acknowledge the help of Konrad Paszkiewicz at the University of Exeter Sequencing Service for advice on ChIP-seq experiments. We also acknowledge the help of Vladimir Teif at the University of Essex in generating the UCSC Genome Browser tracks. Analysis was facilitated by access to the Genome high performance computing cluster at the University of Essex School of Biological Sciences. We acknowledge the help of Ben Lee from the University of Exeter Medical School for advice on gene expression assays. We acknowledge Shyam Prabhakar and Wenjie Sun from the

Genome Institute of Singapore for sharing their brain H3K27ac data for our comparative analyses.

Author contributions

SJM, SKL, TR, EP, and KM conducted laboratory experiments. JM, LCS and SJM designed the study. JM supervised the project and obtained funding. SJM undertook primary data analyses and bioinformatics, with analytical and computational input from LCS, EH and SN. EH undertook the LD Score regression and GWAS enrichment analyses. CT and SA-S provided brain tissue for analysis. KL and AS generated and pre-processed the DNA modification data. JP provided advice for the ChIP-seq analyses. SJM and JM drafted the manuscript. All of the authors read and approved the final submission.

Competing financial interests

The authors declare no competing financial interests.

References

1 Brookmeyer, R., Johnson, E., Ziegler-Graham, K. & Arrighi, H. M. Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement* **3**, 186-191, doi:10.1016/j.jalz.2007.04.381 (2007).

2 Wenk, G. L. Neuropathologic changes in Alzheimer's disease. *J Clin Psychiatry* **64 Suppl 9**, 7-10 (2003).

3 Hardy, J. & Selkoe, D. J. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* **297**, 353-356, doi:10.1126/science.1072994 (2002).

4 Karch, C. M., Cruchaga, C. & Goate, A. M. Alzheimer's disease genetics: from the bench to the clinic. *Neuron* **83**, 11-26, doi:10.1016/j.neuron.2014.05.041 (2014).

5 Reitz, C., Brayne, C. & Mayeux, R. Epidemiology of Alzheimer disease. *Nat Rev Neurol* **7**, 137-152, doi:10.1038/nrneurol.2011.2 (2011).

6 Liu, C. C., Kanekiyo, T., Xu, H. & Bu, G. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat Rev Neurol* **9**, 106-118, doi:10.1038/nrneurol.2012.263 (2013).

7 Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics* **45**, 1452-1458, doi:10.1038/ng.2802 (2013).

8 Lunnon, K. & Mill, J. Epigenetic studies in Alzheimer's disease: current findings, caveats, and considerations for future studies. *Am J Med Genet B Neuropsychiatr Genet* **162B**, 789-799, doi:10.1002/ajmg.b.32201 (2013).

9 Lunnon, K. *et al.* Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat Neurosci* **17**, 1164-1170, doi:10.1038/nn.3782 (2014).

10 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, doi:10.1038/nn.3786 (2014).

11 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).

12 Cuadrado-Tejedor, M. *et al.* A First-in-Class Small-Molecule that Acts as a Dual Inhibitor of HDAC and PDE5 and that Rescues Hippocampal Synaptic Impairment in Alzheimer's Disease Mice. *Neuropsychopharmacology* **42**, 524-539, doi:10.1038/npp.2016.163 (2017).

13 Fischer, A. Targeting histone-modifications in Alzheimer's disease. What is the evidence that this is a promising therapeutic avenue? *Neuropharmacology* **80**, 95-102, doi:10.1016/j.neuropharm.2014.01.038 (2014).

14 Rao, J. S., Keleshian, V. L., Klein, S. & Rapoport, S. I. Epigenetic modifications in frontal cortex from Alzheimer's disease and bipolar disorder patients. *Transl Psychiatry* **2**, e132, doi:10.1038/tp.2012.55 (2012).

15 Zhang, K. *et al.* Targeted proteomics for quantification of histone acetylation in Alzheimer's disease. *Proteomics* **12**, 1261-1268, doi:10.1002/pmic.201200010 (2012).

16 Narayan, P. J., Lill, C., Faull, R., Curtis, M. A. & Dragunow, M. Increased acetyl and total histone levels in post-mortem Alzheimer's disease brain. *Neurobiol Dis* **74**, 281-294, doi:10.1016/j.nbd.2014.11.023 (2015).

17 Sun, W. *et al.* Histone Acetylome-wide Association Study of Autism Spectrum Disorder. *Cell* **167**, 1385-1397 e1311, doi:10.1016/j.cell.2016.10.031 (2016).

18 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).

19 Hopperton, K. E., Mohammad, D., Trepanier, M. O., Giuliano, V. & Bazinet, R. P. Markers of microglia in post-mortem brain samples from patients with Alzheimer's disease: a systematic review. *Molecular psychiatry* **23**, 177-198, doi:10.1038/mp.2017.246 (2018).

20 Kamphuis, W. *et al.* Glial fibrillary acidic protein isoform expression in plaque related astrogliosis in Alzheimer's disease. *Neurobiol Aging* **35**, 492-510, doi:10.1016/j.neurobiolaging.2013.09.035 (2014).
21 Santpere, G., Nieto, M., Puig, B. & Ferrer, I. Abnormal Sp1 transcription factor expression in Alzheimer disease and tauopathies. *Neurosci Lett* **397**, 30-34, doi:10.1016/j.neulet.2005.11.062 (2006).
22 Citron, B. A., Dennis, J. S., Zeitlin, R. S. & Echeverria, V. Transcription factor Sp1 dysregulation in Alzheimer's disease. *J Neurosci Res* **86**, 2499-2504, doi:10.1002/jnr.21695 (2008).
23 Ittner, L. M. & Gotz, J. Amyloid-beta and tau--a toxic pas de deux in Alzheimer's disease. *Nat Rev Neurosci* **12**, 65-72, doi:10.1038/nrn2967 (2011).
24 Spillantini, M. G. & Goedert, M. Tau pathology and neurodegeneration. *Lancet Neurol* **12**, 609-622, doi:10.1016/S1474-4422(13)70090-5 (2013).
25 Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-216, doi:10.1038/nmeth.1906 (2012).
26 Scheuner, D. *et al.* Secreted amyloid beta-protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. *Nat Med* **2**, 864-870 (1996).
27 Goate, A. *et al.* Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature* **349**, 704-706, doi:10.1038/349704a0 (1991).
28 Cruchaga, C. *et al.* Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset Alzheimer's disease families. *PloS one* **7**, e31039, doi:10.1371/journal.pone.0031039 (2012).
29 De Strooper, B. *et al.* Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein. *Nature* **391**, 387-390, doi:10.1038/34910 (1998).
30 Goate, A. & Hardy, J. Twenty years of Alzheimer's disease-causing mutations. *J Neurochem* **120 Suppl 1**, 3-8, doi:10.1111/j.1471-4159.2011.07575.x (2012).
31 Crehan, H. *et al.* Complement receptor 1 (CR1) and Alzheimer's disease. *Immunobiology* **217**, 244-250, doi:10.1016/j.imbio.2011.07.017 (2012).
32 Heppner, F. L., Ransohoff, R. M. & Becher, B. Immune attack: the role of inflammation in Alzheimer disease. *Nat Rev Neurosci* **16**, 358-372, doi:10.1038/nrn3880 (2015).
33 Villegas-Llerena, C., Phillips, A., Garcia-Reitboeck, P., Hardy, J. & Pocock, J. M. Microglial genes regulating neuroinflammation in the progression of Alzheimer's disease. *Curr Opin Neurobiol* **36**, 74-81, doi:10.1016/j.conb.2015.10.004 (2016).
34 Jaeger, S. & Pietrzik, C. U. Functional role of lipoprotein receptors in Alzheimer's disease. *Curr Alzheimer Res* **5**, 15-25 (2008).
35 Sun, X. *et al.* Hypoxia facilitates Alzheimer's disease pathogenesis by up-regulating BACE1 gene expression. *Proc Natl Acad Sci U S A* **103**, 18727-18732, doi:10.1073/pnas.0606298103 (2006).
36 Zlokovic, B. V. Neurovascular pathways to neurodegeneration in Alzheimer's disease and other disorders. *Nat Rev Neurosci* **12**, 723-738, doi:10.1038/nrn3114 (2011).
37 Warren, J. D., Rohrer, J. D. & Rossor, M. N. Frontotemporal dementia. *Bmj* **347**, f4827 (2013).
38 Limon, A., Reyes-Ruiz, J. M. & Miledi, R. Loss of functional GABA(A) receptors in the Alzheimer diseased brain. *Proc Natl Acad Sci U S A* **109**, 10071-10076, doi:10.1073/pnas.1204606109 (2012).
39 Counts, S. E. & Mufson, E. J. Regulator of Cell Cycle (RGCC) Expression During the Progression of Alzheimer's Disease. *Cell Transplant* **26**, 693-702, doi:10.3727/096368916X694184 (2017).
40 Zhao, J., Deng, Y., Jiang, Z. & Qing, H. G Protein-Coupled Receptors (GPCRs) in Alzheimer's Disease: A Focus on BACE1 Related GPCRs. *Front Aging Neurosci* **8**, 58, doi:10.3389/fnagi.2016.00058 (2016).

41 Maddison, D. C. & Giorgini, F. The kynurenine pathway and neurodegenerative disease. *Semin Cell Dev Biol* **40**, 134-141, doi:10.1016/j.semcdb.2015.03.002 (2015).
42 Visel, A., Rubin, E. M. & Pennacchio, L. A. Genomic views of distant-acting enhancers. *Nature* **461**, 199-205, doi:10.1038/nature08451 (2009).
43 Lunnon, K. *et al.* Variation in 5-hydroxymethylcytosine across human cortex and cerebellum. *Genome biology* **17**, 27, doi:10.1186/s13059-016-0871-x (2016).
44 Kinde, B., Gabel, H. W., Gilbert, C. S., Griffith, E. C. & Greenberg, M. E. Reading the unique DNA methylation landscape of the brain: Non-CpG methylation, hydroxymethylation, and MeCP2. *Proc Natl Acad Sci U S A* **112**, 6800-6806, doi:10.1073/pnas.1411269112 (2015).
45 Selkoe, D. J. & Hardy, J. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol Med* **8**, 595-608, doi:10.15252/emmm.201606210 (2016).
46 Graff, J. & Tsai, L. H. Histone acetylation: molecular mnemonics on the chromatin. *Nat Rev Neurosci* **14**, 97-111, doi:10.1038/nrn3427 (2013).
47 Guintivano, J., Aryee, M. J. & Kaminsky, Z. A. A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression. *Epigenetics : official journal of the DNA Methylation Society* **8**, 290-302, doi:10.4161/epi.23924 (2013).
48 Nativio, R. *et al.* Dysregulation of the epigenetic landscape of normal aging in Alzheimer's disease. *Nat Neurosci* **21**, 497-505, doi:10.1038/s41593-018-0101-9 (2018).
49 Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523-527, doi:10.1038/nature19847 (2016).

Figure 1. Variable H3K27ac associated with Alzheimer's disease (AD) in the entorhinal cortex. (a) Manhattan plot showing the raw -$\log_{10}$ *P* value for differential H3K27ac against chromosomal location from the *EdgeR* quasi-likelihood F test (two-sided), controlling for age and derived neuronal proportion (n = 47 biologically independent samples). Variation in H3K27ac at 4,162 peaks was identified as being associated with AD (red line = FDR < 0.05). **(b)** Volcano plot showing the raw -$\log_{10}$ *P* value and log fold change for differential H3K27ac at each entorhinal cortex H3K27ac peak (n = 47 biologically independent samples, red line = FDR < 0.05). Of the AD-associated peaks, 1,475 (35%, 95%-CI: 34-37%) are hyperacetylated (higher H3K27ac) in AD and 2,687 (65%, 95%-CI: 63-66%) are hypoacetylated (lower H3K27ac) in AD (exact binomial test, n = 4,162 peaks, *P* < 1.00E-50).

Figure 2. The top-ranked AD-associated hyperacetylated peak is annotated to *SOX1* and *TEX29* on chromosome 13. Shown are (**a**) normalized read counts and (**b**) a regional track of H3K27ac ChIP-seq data showing weighted mean AD (n = 24 biologically independent samples) and control (n = 23 biologically independent samples) ChIP-seq coverage per million reads in addition to an overlay track highlighting the acetylation differences. (**a**) The most significant AD-hyperacetylated peak is characterized by a consistent increase in H3K27ac in patients (quasi-likelihood F test, two-sided, *P* = 2.04E-08, FDR = 0.002, log fold change = 0.93). The center line of the boxplot shows the median, the outer hinges correspond to the 25th and 75th percentile, respectively, and the whiskers extend to the most extreme observed value within 1.5 times the interquartile range (IQR) from the two hinges. (**b**) This peak is located on chromosome 13 and annotated to both *SOX1* and *TEX29*. Also shown is the location of all entorhinal cortex H3K27ac peaks in this region, hyper- and hypoacetylated peaks (FDR < 0.05), as well as the -log 10 *P* value and log fold change of normalized read count differences for each peak calculated using a quasi-likelihood F test.

Figure 3. The top-ranked AD-associated hypoacetylated peak is located in intron 1 of *ZNF680* on chromosome 7. Shown are (**a**) normalized read counts and (**b**) a regional track of H3K27ac ChIP-seq data showing weighted mean AD (n = 24 biologically independent samples) and control (n = 23 biologically independent samples) ChIP-seq coverage per million reads in addition to an overlay track highlighting the acetylation differences. (**a**) The most significant AD-hypoacetylated peak (quasi-likelihood F test, two-sided, *P* = 1.66E-08, FDR = 0.002) is characterized by a consistent decrease in H3K27ac in cases (log fold change = -0.86). The center line of the boxplot shows the median, the outer hinges correspond to the 25th and 75th percentile, respectively, and the whiskers extend to the most extreme observed value within 1.5 times the IQR from the two hinges. (**b**) This peak is

667 located in intron 1 of *ZNF680* on chromosome 7. Also shown is the location of all entorhinal
668 cortex H3K27ac peaks in this region, hyper- and hypoacetylated peaks (FDR < 0.05), as well
669 as the -log 10 *P* value and log fold change of normalized read count differences for each
670 peak calculated using a quasi-likelihood F test.
671
672 **Figure 4. Clustering of AD and low pathology control samples by H3K27ac levels at**
673 **differentially acetylated peaks. (a)** A heatmap, clustering samples (n = 47 biologically
674 independent samples) by normalized read counts in all 1,475 significant AD hyperacetylated
675 peaks (FDR < 0.05), generates three distinct groups: one comprised of controls only (group
676 1, n = 13 biologically independent samples), a pure group of cases (group 2, n = 20
677 biologically independent samples), and a mixed group containing both cases and controls
678 (group 3, n = 14 biologically independent samples). Controls grouped together with cases in
679 group 3 (n = 10 biologically independent samples) are characterized by significantly
680 decreased neuronal proportion estimates, compared to those in the pure control group 1
681 (Welsh two sample t-test, two-sided, *P* = 7.10E-04, mean reduction in estimated neuronal
682 proportion (%) = 15%, 95%-CI: 7-23%, t(19.81) = 4.00). **(b)** A heatmap, clustering samples
683 by all 2,687 significant AD hypoacetylated peaks (FDR < 0.05), divides the samples into two
684 main groups: group 1 (n = 17 biologically independent samples) is composed mainly of
685 controls, whereas group 2 (n = 30 biologically independent samples) contains more cases
686 than controls. Interestingly, controls classified into group 2 are characterized by lower
687 neuronal proportion estimates than those in group 1 (Welsh two sample t-test, two-sided, n =
688 23 biologically independent samples, *P* = 0.004, mean reduction in neuronal proportion (%)
689 = 14%, 95%-CI: 5-23%, t(15.85) = 3.41). The clustering defined by hyper- or hypoacetylated
690 peaks is not significantly associated with sex (n = 47 biologically independent samples;
691 hyperacetylated cluster: chi-square test, *P* = 0.763, $\chi^2(2) = 0.54$; hypoacetylated cluster: chi-
692 square test with Yates' continuity correction, *P* = 0.269, $\chi^2(1) = 1.22$) or age at death (n = 47
693 biologically independent samples; hyperacetylated cluster: linear regression, *P* = 0.827,
694 F(2,44) = 0.19; hypoacetylated cluster: Welsh two-sample t-test, two-sided, *P* = 0.580, mean
695 age difference = -1.59 years, 95%-CI: -7.36-4.18 years, t(37.20) = -0.56).
696
697 **Figure 5. A region annotated to *MAPT* spanning six H3K27ac peaks is characterized**
698 **by significant hyperacetylation in AD.** A cluster of nine H3K27ac peaks was identified on
699 chromosome 17. All nine peaks are hyperacetylated in cases (quasi-likelihood F test, two-
700 sided, n = 47 biologically independent samples, mean log fold change = 0.46;
701 **Supplementary Table 5**). (**a**) For six of the nine peaks this increase in H3K27ac associated
702 with AD is significant (FDR < 0.05). The center line of the boxplot shows the median, the

outer hinges correspond to the 25th and 75th percentile, respectively, and the whiskers extend to the most extreme observed value within 1.5 times the interquartile range (IQR) from the two hinges. (**b**) The region is located ~10kb upstream of *MAPT* and is (**c**) characterized by brain specific H3K27ac profiles. The boundaries of the significantly differentially acetylated peak region are highlighted in red.

Figure 6. A region annotated to *PSEN2* spanning nine H3K27ac peaks is characterized by significant hyperacetylation in AD. A cluster of 14 H3K27ac peaks was identified on chromosome 1. All 14 peaks are hyperacetylated in cases (quasi-likelihood F test, two-sided, n = 47 biologically independent samples, mean log fold change = 0.52; **Supplementary Table 6**). (**a**) For nine of the 14 peaks this increase in H3K27ac associated with AD is significant (FDR < 0.05). The center line of the boxplot shows the median, the outer hinges correspond to the 25th and 75th percentile, respectively, and the whiskers extend to the most extreme observed value within 1.5 times the interquartile range (IQR) from the two hinges. (**b**) The region is located ~44kb upstream of *PSEN2* and is (**c**) characterized by predominantly brain-specific H3K27ac profiles. The boundaries of the significantly differentially acetylated peak region are highlighted in red.

Figure 7. AD-associated differential expression of transcripts annotated to differentially-acetylated peaks. We quantified the expression of selected cortex-expressed genes located proximal to top-ranked differentially-acetylated peaks in an extended set of 95 entorhinal cortex samples. The abundance of each test gene was determined by relative quantification to the geometric mean of the five housekeeping genes (*ACTB*, *EIF4A2*, *GAPDH*, *SF3A1*, and *UBC*) incorporating experimental variables (RNA isolation batch, RIN score) as covariates. Shown for each gene is the relative expression (log_2 fold ratio) in AD cases (n = 67 biologically independent samples) and controls (n = 28 biologically independent samples). The center line of the boxplot shows the median, the outer hinges correspond to the 25th and 75th percentile, respectively, and the whiskers extend to the most extreme observed value within 1.5 times the interquartile range (IQR) from the two hinges. For each of the genes shown, we identified a significant difference between groups in the direction predicted from our H3K27ac ChIP-seq data. The expression of *RGCC*, *PIM3*, *PSEN1* and *CR1* - located proximal to hyperacetylated peaks - was found to be significantly up-regulated in AD cases (linear regression, *RGCC*: n = 94 biologically independent samples, P = 0.002, β = 0.44, F(3,90) = 12.09; *PIM3*: n = 95 biologically independent samples, P = 9.52E-05, β = 0.57, F(3,92) = 25.78; *PSEN1*: n = 95 biologically independent samples, P = 4.91E-04, β = 0.45, F(3,92) = 4.98; *CR1*: n = 91 biologically independent samples, P = 0.001, β = 0.72, F(3,88) = 5.89). The expression of *GPR22* and *KMO* - located

proximal to hypoacetylated peaks - was found to be significantly down-regulated in AD cases (*GPR22*: n = 95 biologically independent samples, $P = 0.005$, $\beta = -0.71$, $F(3,91) = 2.88$; *KMO*: n = 60 biologically independent samples, $P = 0.006$, $\beta = -0.96$, $F(3,57) = 2.77$).

Table 1. Differential H3K27ac associated with AD. Shown are the ten top-ranked hyper- and hypoacetylated H3K27ac peaks, controlling for age at death and neuronal proportion estimates derived from DNA methylation data. Genes were annotated to each H3K27ac peak using GREAT[50]. The expression of underlined genes was quantified using qPCR.

Rank	Chr	Position (start – end)	*P* value	FDR	Log FC	GREAT annotated genes
Hyperacetylated peaks						
1	13	112101248-112102698	2.04E-08	0.002	0.93	*SOX1, TEX29*
2	13	42094789-42095919	6.31E-08	0.003	0.92	*RGCC, VWA8*
3	22	50342521-50343567	1.02E-07	0.003	0.93	*PIM3, CRELD2*
4	5	640598-642071	1.36E-07	0.003	0.88	*CEP72, TPPP*
5	8	145180336-145181125	2.72E-07	0.004	1.12	*FAM203A, MAF1*
6	17	19665361-19666514	3.86E-07	0.004	0.77	*ALDH3A1, ULK2*
7	1	9392591-9393233	5.25E-07	0.004	0.83	*SLC25A33, SPSB1*
8	17	19619421-19620832	5.43E-07	0.004	0.80	*SLC47A2*
9	17	43925717-43927482	7.01E-07	0.005	0.71	*MAPT, SPPL2C*
10	1	9341867-9342320	8.55E-07	0.005	1.05	*SPSB1, H6PD*
Hypoacetylated peaks						
1	7	64011549-64012825	1.66E-08	0.002	-0.86	*ZNF680, ZNF736*
2	21	29827289-29828201	5.70E-08	0.003	-0.85	*N6AMT1*
3	1	179175226-179176637	7.03E-08	0.003	-0.70	*ABL2, TOR3A*
4	1	241397411-241399621	9.73E-08	0.003	-0.75	*GREM2, RGS7*
5	12	13627258-13629064	1.46E-07	0.003	-0.80	*EMP1, GRIN2B*
6	8	3964265-3966191	2.44E-07	0.004	-0.57	*CSMD1*
7	4	74088063-74089559	2.52E-07	0.004	-0.68	*COX18, ANKRD17*
8	6	166401119-166402753	2.85E-07	0.004	-1.06	*SDIM1, T*
9	7	107111795-107113029	2.88E-07	0.004	-0.90	*DUS4L, GPR22*
10	1	241694436-241695782	3.37E-07	0.004	-0.71	*KMO*

Online Methods

Samples

Post-mortem brain samples from 95 individuals - 67 with advanced AD neuropathology and 28 neuropathology-free brain samples - were provided by the MRC London Neurodegenerative Disease Brain Bank (http://www.kcl.ac.uk/ioppn/depts/cn/research/MRC-London-Neurodegenerative-Diseases-Brain-Bank/MRC-London-Neurodegenerative-Diseases-Brain-Bank.aspx). No statistical methods were used to pre-determine sample sizes but our sample sizes are similar to, or larger than, those reported in previous brain ChIP-seq analyses[17,48]. Ethical approval for the study was provided by the NHS South East London Research Ethics Committee (REC) 3. Subjects were approached in life for written consent for brain banking, and all tissue donations were collected and stored following legal and ethical guidelines (NHS reference number 08/MRE09/38; the HTA license number for the LBBND brain bank is 12293). Samples for this ChIP-seq study were selected from a larger collection of post-mortem entorhinal cortex (Brodmann area (BA) 28/34) samples, based on Braak staging, a standardized measure of neurofibrillary tangle burden determined at autopsy[51]. We prioritized cases with high Braak staging and controls with lower Braak scores (**Supplementary Table 1**). All samples were dissected by trained specialists, snap-frozen and stored at −80 °C. SNP array data from each donor generated using the Illumina Infinium HumanCore-24 SNP array was merged with HapMap Phase 3 data and genetic principal components (PCs) were calculated with GCTA[52] to confirm the ethnicity of each sample as European (**Supplementary Fig. 1**). A detailed list of demographic and sample data for each individual included in the final analyses is provided in **Supplementary Table 1**.

Chromatin immunoprecipitation (ChIP)

Tissue from a subset of 27 individuals with advanced AD neuropathology and 27 individuals with minimal neuropathology were selected for our ChIP-seq analyses. Samples were randomized at all experimental stages, with processing batches comprising an equal number of AD cases and controls. Samples were labeled with anonymized ID codes and processed in batches, blinding disease status from the experimenter/analyst for individual samples. Chromatin immunoprecipitation was performed using the iDeal ChIP-Seq kit for Histones (Cat# C01010051, Diagenode, Seraing, Belgium) as detailed below, using the standard kit components unless otherwise stated. 30 mg of entorhinal cortex tissue was homogenized with a dounce homogenizer in 1 mL ice-cold phosphate buffered saline (PBS) buffer with protease inhibitor cocktail (PIC). The suspension was centrifuged at 4,000 rpm for 5 minutes at 4°C, discarding the supernatant. The pellets were resuspended in 1 mL PBS containing 1% formaldehyde, rotating at room temperature for 8 minutes. The cross-linking process

was terminated by adding 100 µL glycine solution, followed by 5 minutes of rotation. After 5 minutes of centrifugation at 4,000 rpm and 4°C, the pellet was washed twice with ice-cold PBS (suspending the pellet in 1 mL PBS with PIC, centrifuging for 5 minutes at 4,000 rpm and 4°C, and discarding the supernatant), then lysed in 10 mL ice-cold lysis buffer iL1 and iL2, sequentially (re-suspending the pellet in 10 mL lysis buffer, mixing gently for 10 minutes at 4°C, centrifuging for 5 minutes at 4,000 rpm and 4°C, and discarding the supernatant). The cross-linked lysate was suspended in 1.8 mL shearing buffer iS1 containing PIC and sonicated in aliquots of 300 µL for 10 cycles (30 seconds on/off each cycle) on a Bioruptor Pico (Diagenode, Seraing, Belgium). After shearing, samples were transferred to 1.5 mL microcentrifuge tubes and centrifuged at 14,000 rpm for 10 minutes, collecting the supernatant, containing the soluble sheared chromatin with fragments of an average size range of 100-1000bp as visualized by agarose gel electrophoresis (**Supplementary Fig. 23**).

Immunoprecipitation was performed on the SX-8G IP-Star robot (Diagenode, Seraing, Belgium), following the manufacturer's protocol. All samples were immunoprecipitated with H3K27ac polyclonal antibody (Diagenode, Seraing, Belgium) (Cat #C15410196, lot number: A1723-0041D). This antibody has been validated for ChIP-seq applications (validation data from Diagenode available at: https://www.diagenode.com/en/p/h3k27ac-polyclonal-antibody-premium-50-mg-18-ml). In addition, a randomly selected subgroup of 12 samples – 6 cases and 6 controls – were immunoprecipitated with rabbit IgG antibody (iDeal ChIP-seq kit) as negative control. 1-1.5 µL of H3K27ac or IgG antibody were first mixed with 98.5-99 µL ChIP buffer iC1, 0.5 µL PIC and 4 µL of 5% bovine serum albumin (BSA), which was incubated with magnetic beads for 3 hours at 4°C. Next the antibody conjugate was added to 180 µL chromatin for overnight (15h) immunoprecipitation at 4°C in an immunoprecipitation mix also containing 20 µL ChIP buffer iC1, 1 µL PIC and 4 µL of 5% BSA. After immunoprecipitation, the beads were re-suspended in in 100 µL elution buffer iE1, to which 4 µL elution buffer iE2 was added. Cross-link reversal was performed on a PCR thermoblock for 4 hours at 65°C. DNA was extracted using Micro ChIP DiaPure columns (Diagenode, Cat No C03040001, Seraing, Belgium) according to the manufacturers protocol, eluting the DNA from the column matrix in 30 µL DNA elution buffer (MicroChIP DiaPure columns; Diagenode, Cat #C03040001, Denville, NJ, USA). Quantitative PCR, using 1% input DNA, was used to confirm specific enrichment of H3K27ac at positive control genes (*IEF4A2* and *GAPDH*) but not at negative control genes (*MBex2* and *TSH2b*; all primers were provided by Diagenode).

Illumina short-read sequencing

Libraries were prepared using the MicroPlex Library Preparation kit v2 (Diagenode, Cat #C05010013, Seraing, Belgium) on a SX-8G IP-Star robot according to the manufacturer's protocol. DNA concentrations were measured with Qubit dsDNA HS Assay Kits (Invitrogen, Cat# Q32851, Carlsbad, CA, USA) on the Qubit 2.0 Fluorometer (Invitrogen) and library fragment profiles generated on the Agilent 2100 BioAnalyzer using Agilent High Sensitivity DNA kits (Agilent Technologies, Cat# 5067-4626, Santa Clara, CA, USA). Following our stringent quality-control filtering, 7 samples were excluded from sequencing based on poor qPCR results after immunoprecipitation or low library concentration. The remaining 47 samples (from 24 cases and 23 controls) were sequenced on an Illumina HiSeq-2500 using single-end sequencing and a read length of 50bp. ChIP-seq data are available to download from GEO (accession number GSE102538).

Data pre-processing and quality control

Global sample anomalies were ruled out using *fastqc*[53] summary measures. All fastq files were aligned to the *Homo sapiens* reference genome (hg19, Broad Institute) using *Bowtie*[54]. The output SAM files were converted to binary (BAM) format. All BAM files were sorted and indexed using *samtools*[55]. PCR duplicates were removed using *Picard* (http://broadinstitute.github.io/picard/). *Samtools* was used to additionally remove non-uniquely mapped reads as well as reads with a sequencing quality score $q < 30$. Final read counts after QC for all 47 samples are shown in **Supplementary Fig. 2.** On average, we obtained 30,032,623 reads per sample (SD = 10,638,091; range = 10,910,000-53,770,000) and individual read counts did not associate with disease status (Welsh two-sample t-test, two-sided, n = 47 biologically independent samples, $P = 0.93$, average read count difference: 260,673, 95%-CI: -6,602,319-6,080,972, $t(43.85) = -0.08$). For each sample we also calculated standard ENCODE ChIP-seq quality metrics (see https://genome.ucsc.edu/ENCODE/qualityMetrics.html#definitions): uniquely mappable reads (UMR), normalized stand cross-correlation (NSC), relative strand cross-correlation (RSC) and nonredundancy fraction (NrF) as well as the percentage of reads in peaks. These quality metrics were included as post-hoc covariates in our analyses.

Peak calling and read counts

All filtered BAM files were merged into one grouped file and converted to *tagAlign* format using *bedtools*[56]. Peaks were called on this merged file using *MACS2*[57], keeping all duplicates, since duplicates were removed from each sample previously and any remaining duplicates would result from the same read occurring in more than one sample. From the

resulting peaks those located in unmapped contigs and mitochondrial DNA were filtered out as well as peaks that did not meet a significance threshold of $P < 1.00E-07$ for peak calling. The bed file of peaks was converted to gff format using *awk* and *R*, and reads for each individual sample were generated using *HTSeq*[58]. Final filtering was performed using the Bioconductor package *EdgeR*[59], excluding peaks with fewer than 2 samples showing at least 1 read per million, resulting in a total of 182,065 peaks to be tested. Principal components analysis (PCA) in *R* using *DESeq2*[60] confirmed that the epigenetically predicted gender was identical to the recorded one (**Supplementary Fig. 6**), with load on the first two principal components not related to disease status. Analysis scripts related to this project are available to download from: https://epigenetics.essex.ac.uk/AD_H3K27ac/.

Peak validation

We validated the 182,065 union peaks in two ways. First, we obtained the locations of H3K27ac peaks called in the cortex (BA9) and cerebellum from a recent paper by Sun and colleagues[17]. Second, we downloaded H3K27ac profiles produced by the NIH Roadmap Epigenomics Consortium[18] from the Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo) for multiple cell-/tissue-types including several brain regions (mid frontal lobe (GSM773015), inferior temporal gyrus (GSM772995), middle hippocampus (GSM773020), substantia nigra (GSM997258), cingulate gyrus (GSM773011), H1-derived neuronal progenitor cells (HDNPs, GSM753429), lung (GSM906395), liver (GSM1112808) and skeletal muscle (GSM916064)). The downloaded files were in bed format, on which we performed peak calling using *MACS2* and the same specifications as described for our own samples, discounting any duplicate reads. We calculated the overlap between each peak set and our peaks by quantifying the percentage of peaks from the external sample overlapping our peaks using the Bioconductor package *GenomicRanges*[61]. In addition, using the 182,065 entorhinal cortex (EC) peak regions identified in this study, we generated read counts on raw H3K27ac data from cerebellum, prefrontal cortex (PFC) and temporal cortex (TC) published by Sun et al. (2016)[17]. Counts per million (CPM) at all peaks were scaled and centered prior to principal components analysis.

Differential peak calling

We used the quasi-likelihood F test[62] in *EdgeR*[59] to analyse peak differences between AD-cases and controls, allowing us to correct for potential confounders in the analysis of differential peaks. Our analyses accounted for additional phenotypic variation across the samples, including age at death and neuronal proportion estimates based on DNA methylation profiles from the Illumina 450K HumanMethylation Array from the same samples, which were calculated using the *CETS R* package[47]. We imputed the median

CETS estimate for one individual with missing DNA methylation data. Age at death and CETS estimates were converted to five-level factors using the *R* function *cut()* specifying five breaks prior to being included as covariates in the *EdgeR* differential peak calling method. This function divides a numerical variable into five bins of equal length, determined by the range of the variable. The distribution of the age and CETS variable (including the imputed individual) with the respective bins of the factor variables are shown in **Supplementary Fig. 5**. We next calculated normalization factors based on sample-specific library compositions and estimated both sample and peak-specific dispersions, specifically for a generalized linear model controlling for factorized CETS estimates and age at death. The quasi-likelihood F-test was conducted after fitting a quasi-likelihood model[62] using the *glmQLFit()* and *glmQLFTest()* functions respectively. Effect sizes are reported as log fold change, a standard measure for quantifying sequencing read count differences between different conditions. Log fold change refers to the $\log_2$-transformed ratio of normalized read counts between cases and controls, with positive values indicating higher normalized read counts in AD samples. As further sensitivity analyses, we repeated the differential peak calling model, covarying additionally for i) sex, ii) expression of cell-type specific marker genes for the five major brain cell-types (*CD34*, *CD68*, *ENO2*, *GFAP* and *OLIG2* – see Online Methods: Gene expression analyses) individually and combined, and iii) five ChIP-seq quality metrics (UMR, RSC, NSC, NrF and percent reads in peaks – see Online Methods: Data pre-processing and quality control). *P* values in each of the additional control models at all 4,162 differentially acetylated peaks from our main model are reported in **Supplementary Tables 4** and **5**. The *bedtools* program *genomecov* was used to generate coverage value scaled by library size and the number of samples per group, for each sample. These were then joined using *unionbedg* and summed using a *Perl* script to produce a weighted mean for each variable sized interval defined by read overlaps and used to generate UCSC genome browser tracks (accessible at https://epigenetics.essex.ac.uk/AD_H3K27ac/).

Genomic annotation and enrichment analyses

Peaks were annotated to genes using the *Genomic Region Enrichment and Annotation Tools* (*GREAT*)[50], using the basal plus extension option and assigning genes to regulatory regions within 5kb upstream and 1kb downstream of the transcription start site (proximal) or up to 1000kb for distal interactions. In addition, we performed enrichment analyses calculating statistical enrichments for ontological annotation (gene ontologies for molecular function, biological processes[63] and human diseases[64]). Functional enrichment analyses were conducted for significantly hyper- and hypoacetylated peaks (FDR < 0.05) separately, using the basal plus extension option. Significance in the enrichment test is based on a hypergeometric test of genes annotated to the test set (hyper-/ hypoacetylated peaks)

compared to the background set of genes annotated to all 182,065 peaks called across all samples. Results presented in **Supplementary Fig. 18** are restricted to the top five non-redundant enrichments (separated by at least two nodes in the local directed acyclic graph visualizing the hierarchy of enriched terms from a single ontology) associated with at least three genes in the test set for the ontology categories biological process, molecular function, and disease ontology and we show full enrichments across all categories in **Supplementary Tables 11-16**. Enrichments for familial AD and AD pathology genes were calculated using a one-tailed hypergeometric test.

Motif enrichment analysis

Motif analysis was performed using the *Regulatory Sequence Analysis Tools suite* (*RSAT*)[65,66], available at http://rsat.sb-roscoff.fr. Peak sequences were reduced to 1500bp on each side of the peak centre, and motif discovery was conducted on 6 and 7mer oligonucleotides, comparing the statistically enriched sequences with known transcription factor motifs from *JASPAR*[67] (core nonredundant vertebrates) and *Homer*[68] (Human TF motifs). Enrichments were computed using a binomial test, relative to the background peak sequences (n = 182,065 peaks) for significantly hyper- and hypoacetylated peaks (FDR < 0.05).

Integration of H3K27ac ChIP-seq data with results from AD genome-wide association studies (GWAS)

The summary statistics for the stage 1 GWAS from Lambert and colleagues[7] were downloaded from http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php. These results were clumped (p1 = 1e-4; p2 = 1e-4, r2 =0.1, window = 3000kb) using *plink*[69], which collapses multiple correlated signals (due to linkage disequilibrium (LD)) into regions which contain independent signals. LD relationships were inferred from a reference genotype dataset (Phase 1) from another study[70]. Neighbouring regions located within 250kb of each other on the same chromosome were subsequently merged. After clumping, each region was assigned the minimum *P* value for all SNPs contained in the region (from Lambert et al), and regions were then filtered to the genome-wide significance threshold (*P* < 5.0E-08). This yielded 11 LD blocks for the genome-wide significant findings from Lambert et al., which were then overlapped with our AD-associated differentially acetylated peaks using the Bioconductor package *GenomicRanges*[61]. To estimate the proportion of AD heritability attributable to H3K27ac peaks in adult brain, we performed partitioned heritability analysis using the LD Score regression software (https://github.com/bulik/ldsc)[71,72]. LD scores were generated based on custom annotations derived from our ChIP-Seq data and 1000 genomes reference data (downloaded alongside the software from

https://data.broadinstitute.org/alkesgroup/LDSCORE/). Genetic variants were annotated to our ChIP-seq peaks and heritability statistics calculated using the publicly available GWAS results from Lambert et al[7]. Enrichment statistics were calculated as the proportion of heritably divided by the proportion of SNPs for variants annotated to H3K27ac peaks.

Gene expression analysis

30-50mg of frozen entorhinal cortex tissue from 95 individuals [67 high neuropathology; 28 low neuropathology] was homogenized with Qiazol (Qiagen, Valencia, CA, USA) as per the manufacturer's instructions before being run through a QIAshredder (Qiagen, Valencia, CA, USA). Total RNA was extracted using the Qiagen RNeasy column purification system and treated with DNase I. The Agilent 2200 TapeStation was used to check the quality and concentration of the extracted RNA samples. Complementary DNA (cDNA) was reverse transcribed using the Invitrogen VILO cDNA synthesis kit (Life Technologies) in 20µL reactions according to manufacturer's instructions. After stringent QC, quantitative RT–PCR was performed in duplicate using the QuantStudio 12K Flex (Applied Biosystems) in conjunction with the TaqMan low-density array (TLDA) platform using off the shelf pre-optimized assays targeting i) genes located proximal to top-ranked AD-associated H3K27ac peaks (*ANKRD17*, *GPR22*, *KMO*, *RGCC*, *PIM3*, and *ZNF680*), ii) additional genes previously implicated in AD etiology and/or neuropathology and also annotated to significant differentially-acetylated peaks (*APP*, *CR1*, *MAPT*, *PSEN1*, *PSEN2*, and *TOMM40*), iii) genes expressed in the major brain cell types to control for cell-type heterogeneity (*ENO2* (neurons), *OLIG2* (oligodendrocytes), *GFAP* (astrocytes), *CD68* (microglia) and *CD34* (endothelial cells), and iv) five house-keeping genes (*ACTB*, *EIF4A2*, *GAPDH*, *SF3A1*, and *UBC*) identified as being most stably expressed in the brain using GeNORM (Primer Design, Southampton, UK). A full list of qPCR assays used is given in **Supplementary Table 21**. PCR cycling conditions were 50°C for 2 min, 94.5°C for 10 min and 45 cycles of 97°C for 15s and 60°C for 1 min. We undertook stringent QC of raw qPCR data, removing samples where there was high variability between duplicates (Ct > 0.5). The abundance of each test gene was determined by the comparative Ct method[73], expressed relative to the geometric mean of the five house-keeping genes quantified in parallel. Data were $\log_2$transformed to ensure normal distribution, and presented as a fold-difference in expression of AD cases relative to controls. Associations of gene expression were assessed by linear regression of deltaCt against disease status, incorporating experimental variables (RNA isolation batch, RIN score) as covariates. We tested for an association between H3K27ac and gene expression using *EdgeR* as described above, including RNA isolation batch and RIN score as covariates.

Integrative analysis with DNA methylation and hydroxymethylation

DNA methylation and hydroxymethylation data was available (A. Smith et al, unpublished) from entorhinal cortex DNA for 42 of the samples profiled using ChIP-seq in this study. DNA methylation and hydroxymethylation profiles were generated on the Illumina Infinium HumanMethylation450 BeadChip (Illumina Inc., CA, USA) ("Illumina 450K array") using the TrueMethyl Array kit (Cambridge Epigenetix, Cambridge, UK). Profiles for both modifications were pre-processed, normalized and filtered according to a stringent standardised quality control pipeline, as described previously[43] using the *wateRmelon*[74] package in *R*. We identified probes on the array within 1kb of differentially acetylated peaks (FDR < 0.05) using the Bioconductor package *GenomicRanges*[61]. 268,477 peak-probe pairs (comprising of 232,233 unique 450K array probes and 62,714 (34.45% of total) unique H3K27ac peaks). Of these, a total of 1,649 of the 4,162 FDR significant differentially acetylated peaks were located within 1kb of at least one CpG probe on the array, with a total of 6,838 probes mapping to the 1kb neighbourhood of these 1,649 peaks. For each CpG-peak pair we correlated the log fold change in H3K27ac between AD cases and controls to the difference in DNA methylation or hydroxymethylation between AD cases and controls estimated from a linear model controlling for the same covariates as in the differential acetylation analysis. We examined patterns of DNA methylation and hydroxymethylation across probes in the vicinity of AD hyper- and hypoacetylated peaks, as well as those in vicinity of all background peaks and the whole microarray background using Welsh two sample t-tests. Finally, we analysed the correlation of acetylation and DNA methylation at all peak-probe pairs using a Pearson's product-moment correlation between H3K27ac counts per million and DNA methylation (normalized betas).

Statistical analysis

For the ChIP-seq analysis we used the quasi-likelihood F test[62] in *EdgeR*[59]. Our analyses accounted for additional phenotypic variation across the samples, including age at death and neuronal proportion estimates based on DNA methylation profiles from the Illumina 450K HumanMethylation Array from the same samples, which were calculated using the *CETS* R package[47]. Peaks were considered differentially acetylated at a false discovery rate (FDR) < 0.05 (controlled by Benjamini-Hochberg for n = 182,065 tests). ChIP-seq data is summarized as read counts per peak and sample. *EdgeR* assumes a negative binomial distribution, which is the most appropriate distribution for overdispersed count data, such as sequencing read counts in features. The data distribution at each of the 182,065 peaks was not formally tested. Associations of gene expression were assessed by linear regression of deltaCt against disease status, incorporating experimental variables (RNA isolation batch, RIN score) as covariates. We tested for an association between H3K27ac and gene expression

using *EdgeR* as described above, including RNA isolation batch and RIN score as covariates. Given that the statistical software *R* cannot report arbitrarily small *P* values due to computational memory constraints, we report $P < 1.00E-50$ whenever the software output showed $P = 0$ or a *P* value less than 1.00E-50.

Life Sciences Reporting Summary

Further information on experimental design is available in the **Life Sciences Reporting Summary**.

Data availability

Raw data has been deposited in GEO under accession number GSE102538. Browsable UCSC genome browser tracks of our processed H3K27ac ChIP-seq data are available as a resource at: https://epigenetics.essex.ac.uk/AD_H3K27ac/.

Code availability

Analysis code is given in **Supplementary Software** and also available to download from https://epigenetics.essex.ac.uk/AD_H3K27ac/code/index.html.

Methods-only References

50 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nature biotechnology* **28**, 495-501, doi:10.1038/nbt.1630 (2010).

51 Braak, H. & Braak, E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol* **82**, 239-259 (1991).

52 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics* **88**, 76-82, doi:10.1016/j.ajhg.2010.11.011 (2011).

53 Andrews, S. FastQC: a quality control tool for high throughput sequence data (2010). http://www.bioinformatics.babraham.ac.uk/projects/fastqc

54 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).

55 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

56 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

57 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

58 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169, doi:10.1093/bioinformatics/btu638 (2015).

59 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).

60 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

61 Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS Comput Biol* **9**, e1003118, doi:10.1371/journal.pcbi.1003118 (2013).

62 Lun, A. T., Chen, Y. & Smyth, G. K. It's DE-licious: A Recipe for Differential Expression Analyses of RNA-seq Experiments Using Quasi-Likelihood Methods in edgeR. *Methods Mol Biol* **1418**, 391-416, doi:10.1007/978-1-4939-3578-9_19 (2016).

63 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* **25**, 25-29, doi:10.1038/75556 (2000).

64 Osborne, J. D. *et al.* Annotating the human genome with Disease Ontology. *BMC Genomics* **10 Suppl 1**, S6, doi:10.1186/1471-2164-10-S1-S6 (2009).

65 Thomas-Chollier, M. *et al.* RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. *Nucleic Acids Res* **40**, e31, doi:10.1093/nar/gkr1104 (2012).

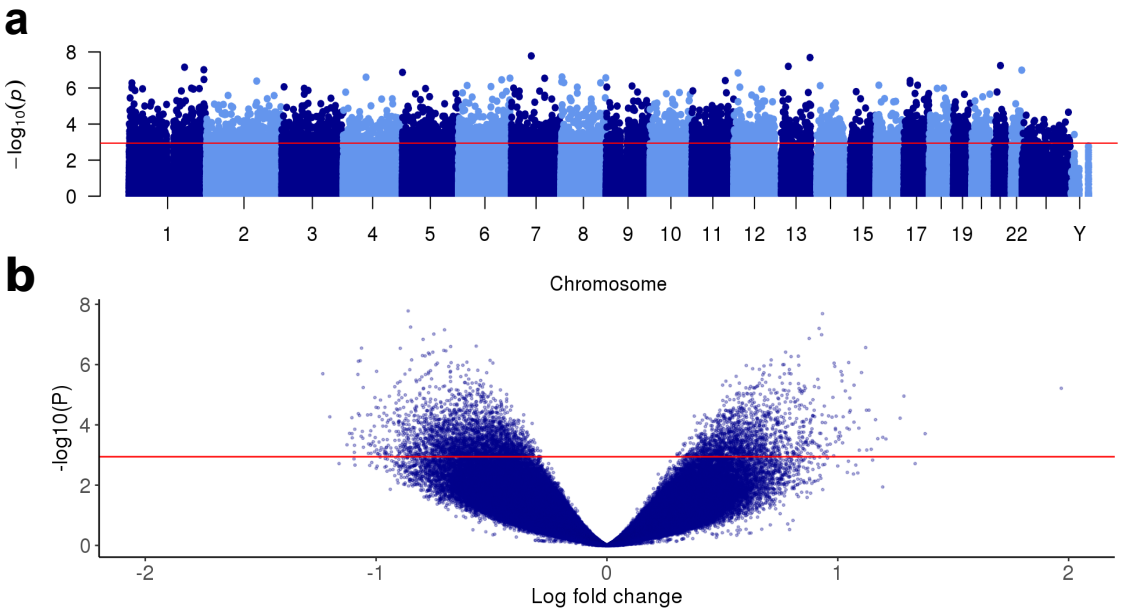
66 Thomas-Chollier, M. *et al.* A complete workflow for the analysis of full-size ChIP-seq (and similar) data sets using peak-motifs. *Nat Protoc* **7**, 1551-1568, doi:10.1038/nprot.2012.088 (2012).

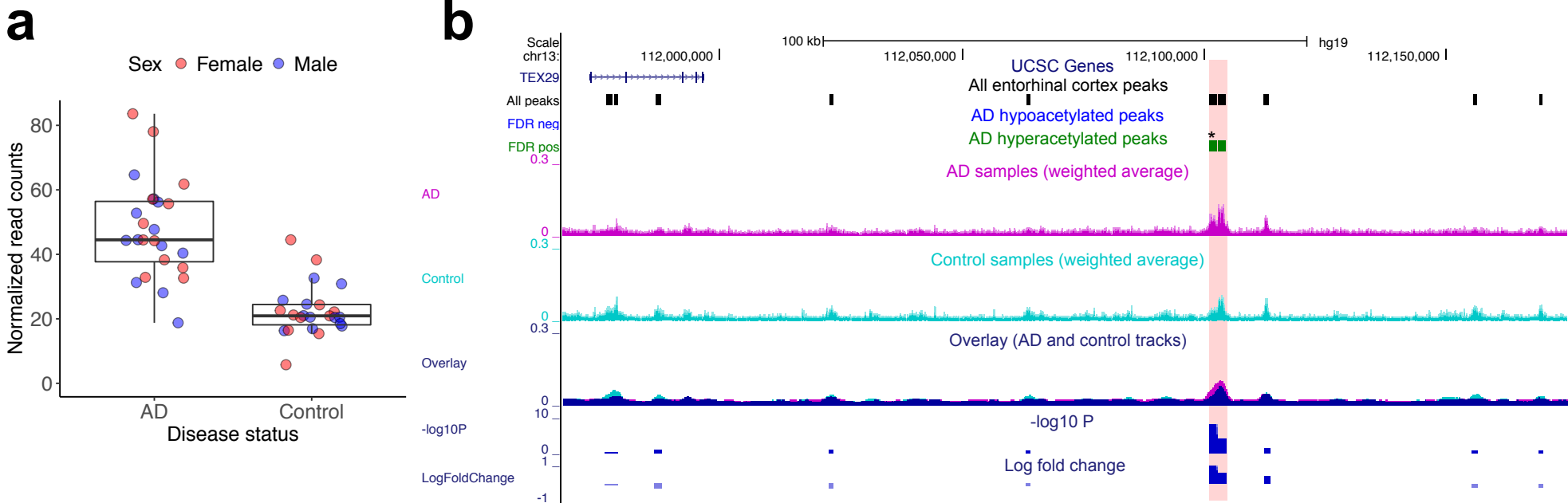
67 Mathelier, A. *et al.* JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **44**, D110-115, doi:10.1093/nar/gkv1176 (2016).

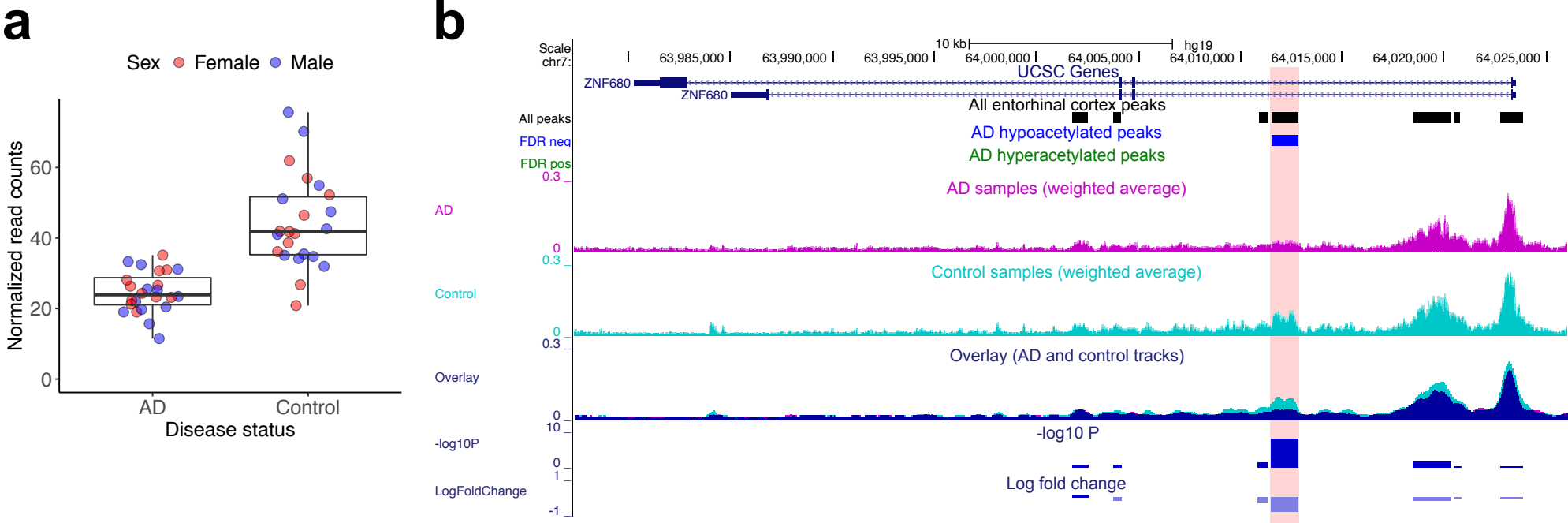
68 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).

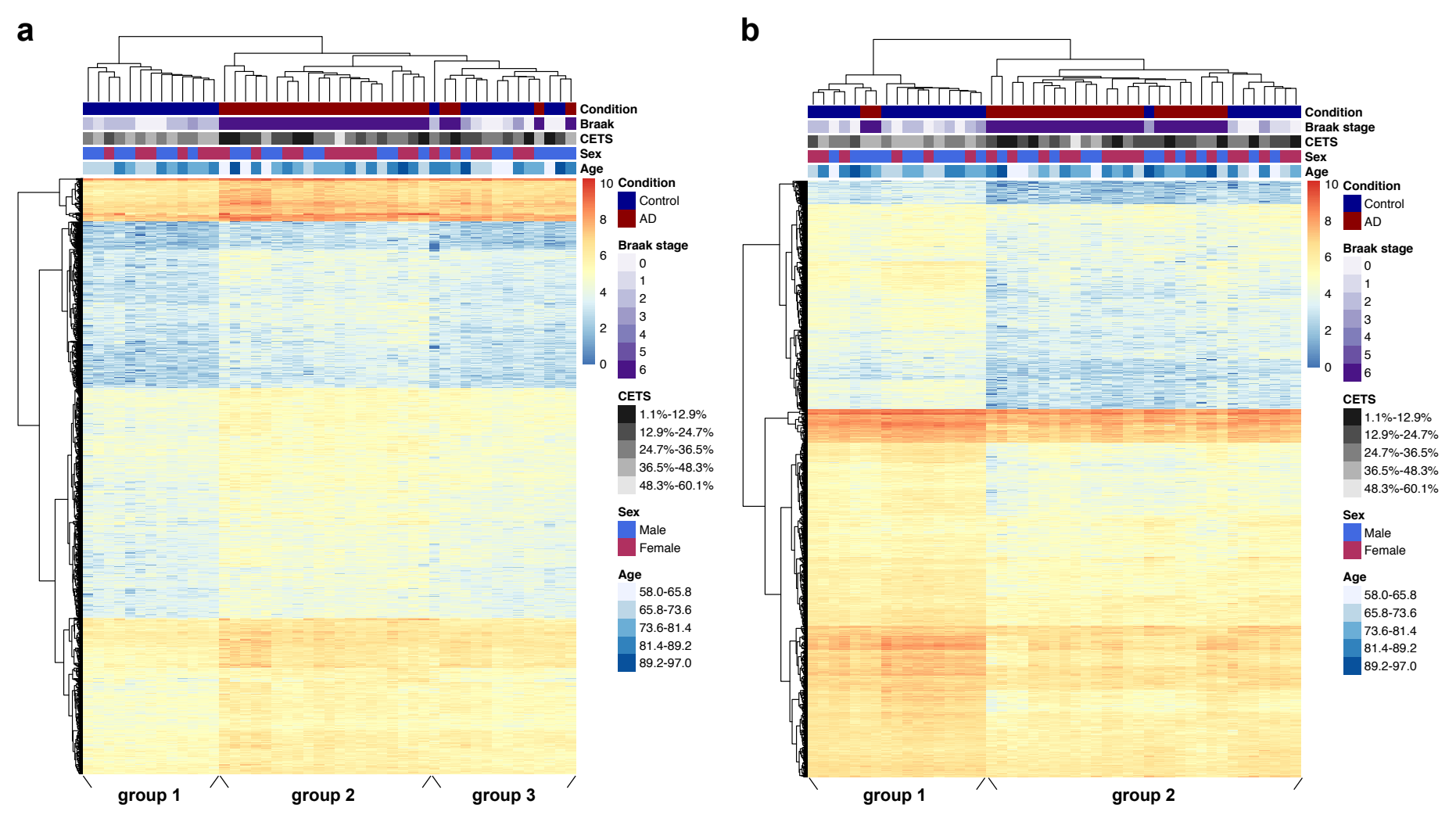
69 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).

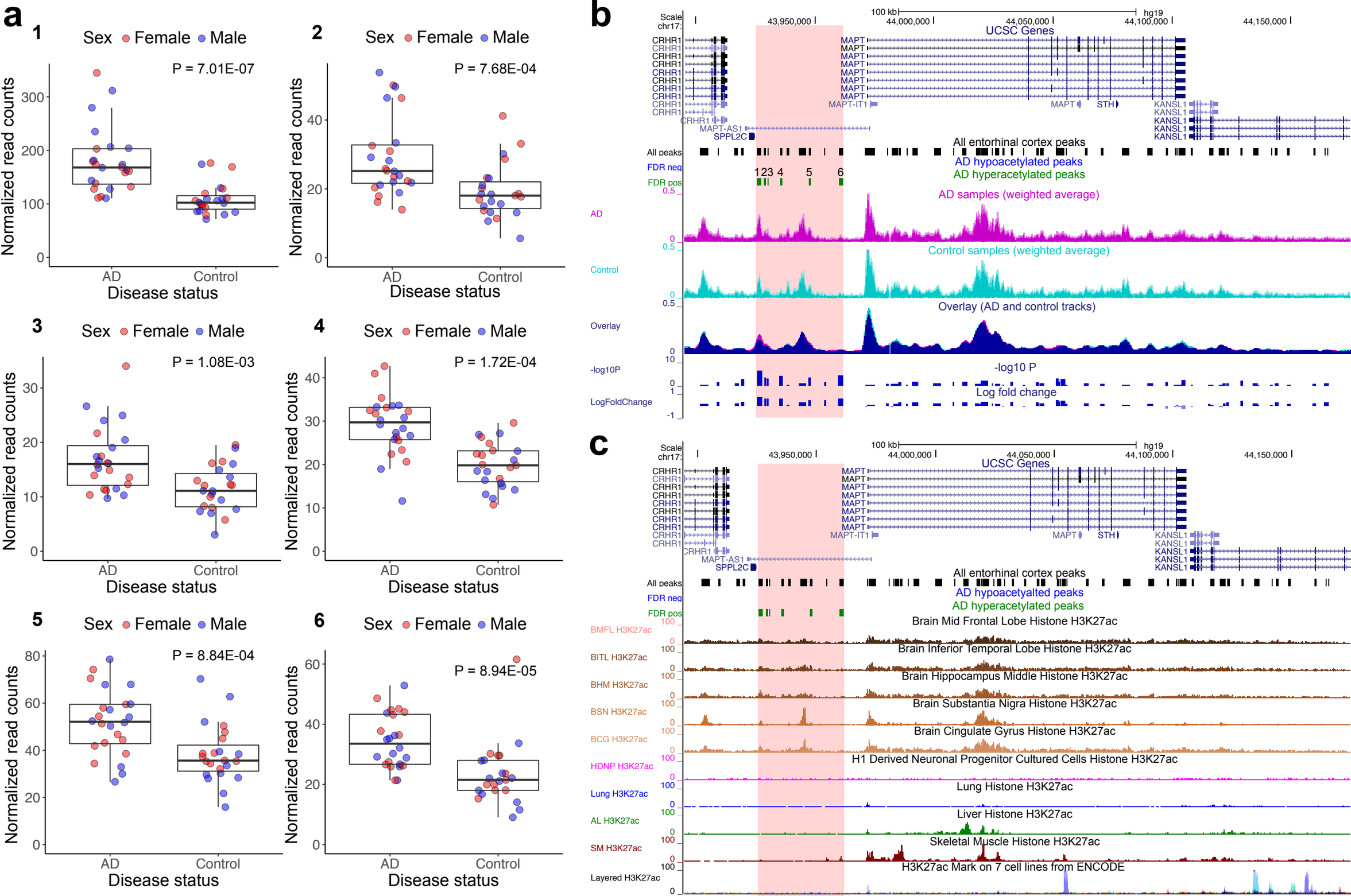
70 Hannon, E. *et al.* An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. *Genome biology* **17**, 176, doi:10.1186/s13059-016-1041-x (2016).
71 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295, doi:10.1038/ng.3211 (2015).
72 Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-1235, doi:10.1038/ng.3404 (2015).
73 Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**, e45 (2001).
74 Pidsley, R. *et al.* A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* **14**, 293, doi:10.1186/1471-2164-14-293 (2013).

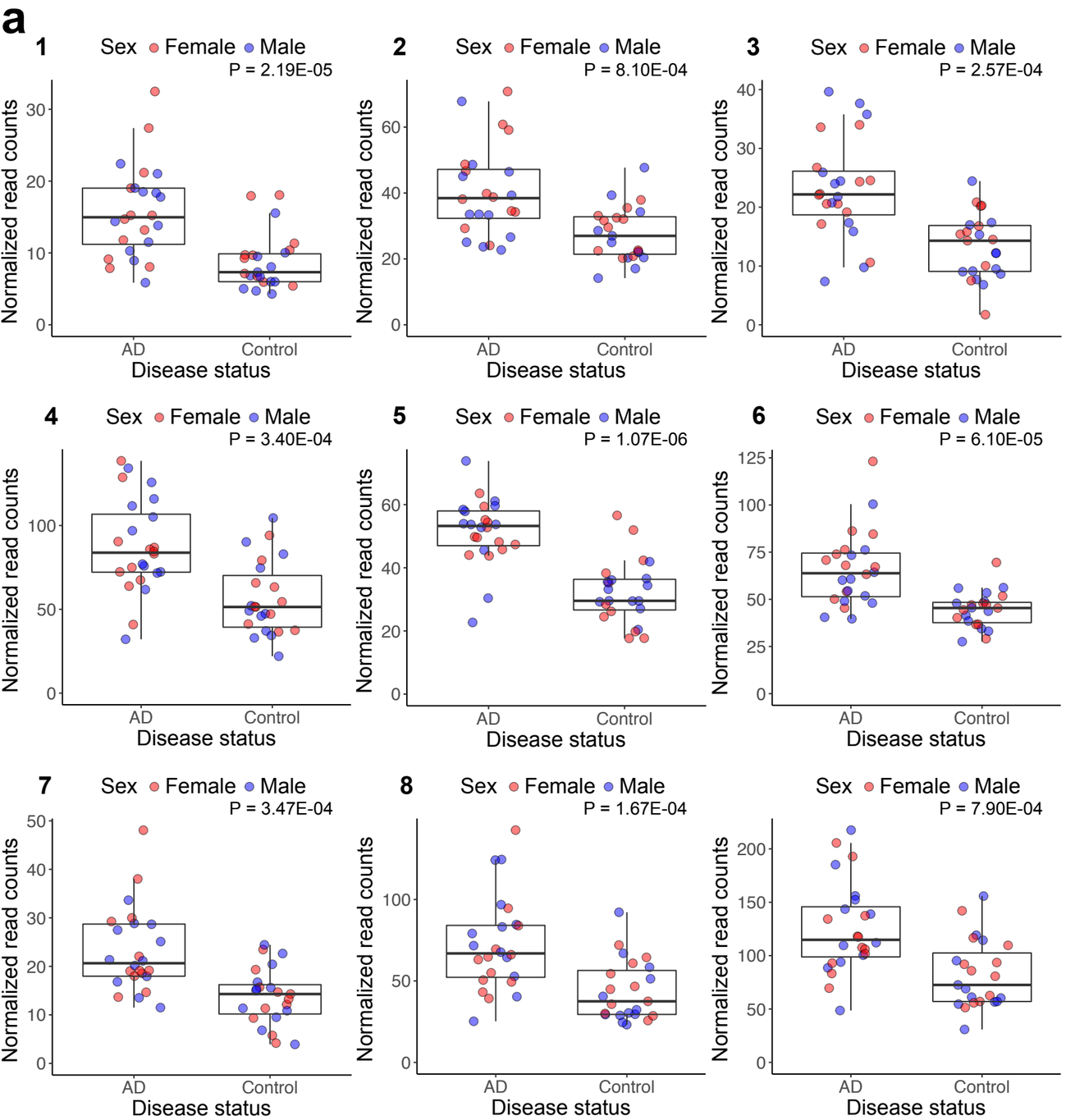

a
$-\log_{10}(p)$
8
6
4
2
0
1
2
3
4
5
6
7
8
9
10
11
12
13
15
17
19
22
Y
b
Chromosome
-log10(P)
8
6
4
2
0
-2
-1
0
1
2
Log fold change

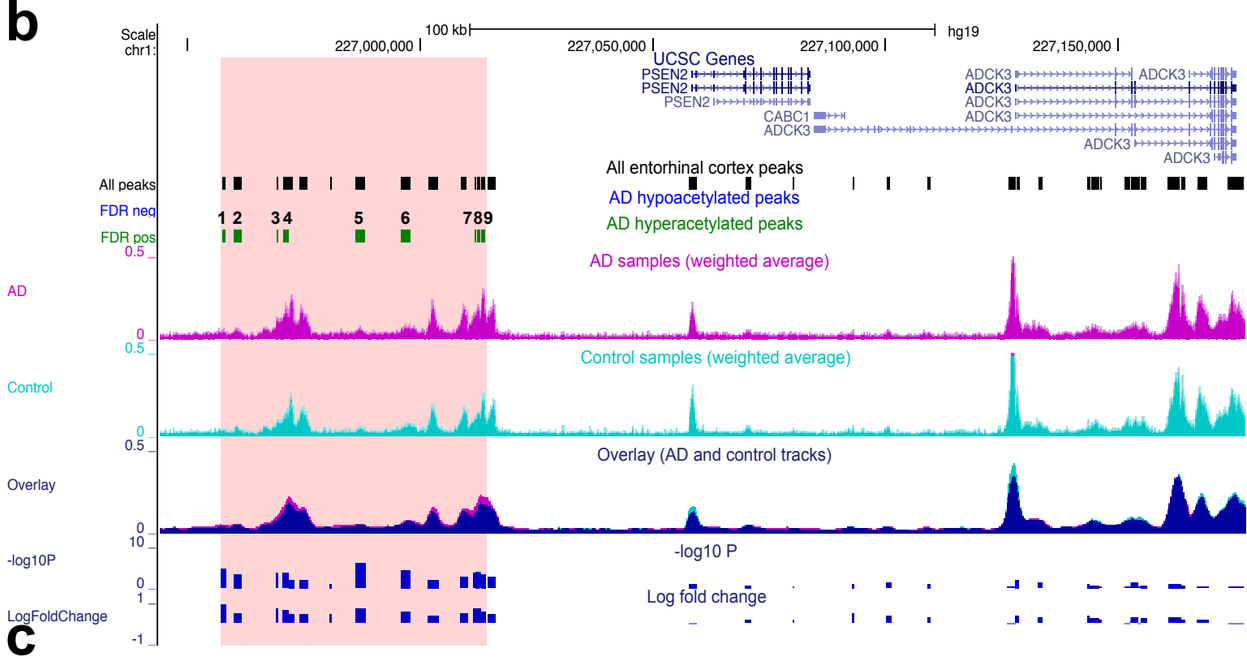

a
Sex
Female
Male
Normalized read counts
80
60
40
20
0
AD
Control
Disease status
b
Scale
chr13:
100 kb
hg19
112,000,000
112,050,000
112,100,000
112,150,000
UCSC Genes
TEX29
All entorhinal cortex peaks
All peaks
AD hypoacetylated peaks
FDR neg
AD hyperacetylated peaks
FDR pos
AD samples (weighted average)
AD
Control samples (weighted average)
Control
Overlay (AD and control tracks)
Overlay
-log10 P
-log10P
Log fold change
LogFoldChange

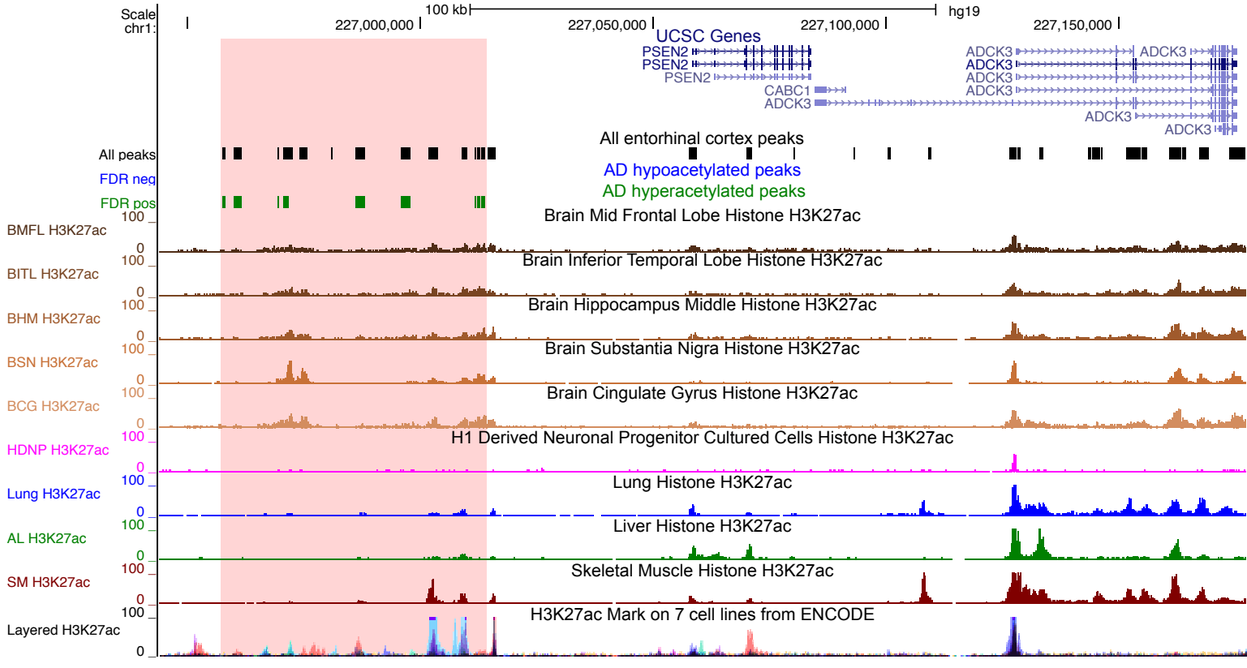

a
Sex
Female
Male
Normalized read counts
60
40
20
0
AD
Control
Disease status
b
Scale
chr7:
10 kb
hg19
63,985,000
63,990,000
63,995,000
64,000,000
64,005,000
64,010,000
64,015,000
64,020,000
64,025,000
UCSC Genes
ZNF680
ZNF680
All entorhinal cortex peaks
All peaks
AD hypoacetylated peaks
FDR neg
AD hyperacetylated peaks
FDR pos
AD samples (weighted average)
0.3
AD
0
0.3
Control samples (weighted average)
Control
0
0.3
Overlay (AD and control tracks)
Overlay
0
10
-log10 P
-log10P
0
1
Log fold change
LogFoldChange
-1

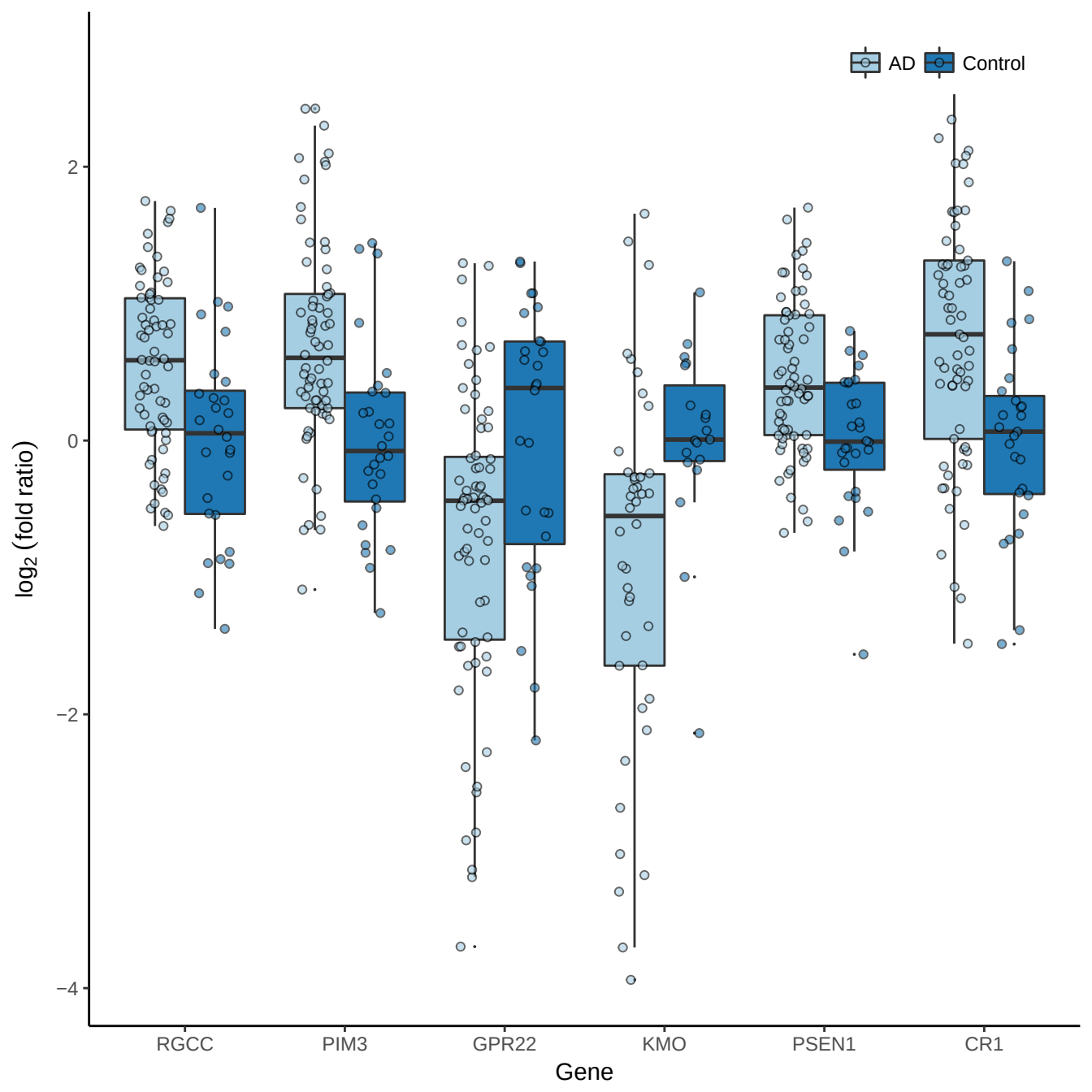

a
Condition
Braak
CETS
Sex
Age
10
8
6
4
2
0
Condition
Control
AD
Braak stage
0
1
2
3
4
5
6
CETS
1.1%-12.9%
12.9%-24.7%
24.7%-36.5%
36.5%-48.3%
48.3%-60.1%
Sex
Male
Female
Age
58.0-65.8
65.8-73.6
73.6-81.4
81.4-89.2
89.2-97.0
group 1
group 2
group 3
b
Condition
Braak stage
CETS
Sex
Age
10
8
6
4
2
0
Condition
Control
AD
Braak stage
0
1
2
3
4
5
6
CETS
1.1%-12.9%
12.9%-24.7%
24.7%-36.5%
36.5%-48.3%
48.3%-60.1%
Sex
Male
Female
Age
58.0-65.8
65.8-73.6
73.6-81.4
81.4-89.2
89.2-97.0
group 1
group 2


a
1
Sex Female Male
P = 7.01E-07
Normalized read counts
AD
Control
Disease status
2
Sex Female Male
P = 7.68E-04
Normalized read counts
AD
Control
Disease status
3
Sex Female Male
P = 1.08E-03
Normalized read counts
AD
Control
Disease status
4
Sex Female Male
P = 1.72E-04
Normalized read counts
AD
Control
Disease status
5
Sex Female Male
P = 8.84E-04
Normalized read counts
AD
Control
Disease status
6
Sex Female Male
P = 8.94E-05
Normalized read counts
AD
Control
Disease status
b
Scale
chr17:
43,950,000
100 kb
hg19
44,000,000
44,050,000
44,100,000
44,150,000
UCSC Genes
CRHR1
MAPT
MAPT-IT1
STH
KANSL1
MAPT-AS1
SPPL2C
All entorhinal cortex peaks
All peaks
FDR neg
AD hypoacetylated peaks
FDR pos
AD hyperacetylated peaks
1 2 3 4 5 6
AD samples (weighted average)
AD
Control samples (weighted average)
Control
Overlay (AD and control tracks)
Overlay
-log10P
-log10 P
LogFoldChange
Log fold change
c
Scale
chr17:
43,950,000
100 kb
hg19
44,000,000
44,050,000
44,100,000
44,150,000
UCSC Genes
CRHR1
MAPT
MAPT-IT1
STH
KANSL1
MAPT-AS1
SPPL2C
All entorhinal cortex peaks
All peaks
FDR neg
AD hypoacetylated peaks
AD hyperacetylated peaks
FDR pos
BMFL H3K27ac
Brain Mid Frontal Lobe Histone H3K27ac
BITL H3K27ac
Brain Inferior Temporal Lobe Histone H3K27ac
BHM H3K27ac
Brain Hippocampus Middle Histone H3K27ac
BSN H3K27ac
Brain Substantia Nigra Histone H3K27ac
BCG H3K27ac
Brain Cingulate Gyrus Histone H3K27ac
HDNP H3K27ac
H1 Derived Neuronal Progenitor Cultured Cells Histone H3K27ac
Lung H3K27ac
Lung Histone H3K27ac
AL H3K27ac
Liver Histone H3K27ac
SM H3K27ac
Skeletal Muscle Histone H3K27ac
Layered H3K27ac
H3K27ac Mark on 7 cell lines from ENCODE


a
1
Sex Female Male
P = 2.19E-05
2
Sex Female Male
P = 8.10E-04
3
Sex Female Male
P = 2.57E-04
4
Sex Female Male
P = 3.40E-04
5
Sex Female Male
P = 1.07E-06
6
Sex Female Male
P = 6.10E-05
7
Sex Female Male
P = 3.47E-04
8
Sex Female Male
P = 1.67E-04
Sex Female Male
P = 7.90E-04
Normalized read counts
AD
Control
Disease status


b
Scale
chr1:
100 kb
hg19
227,000,000
227,050,000
227,100,000
227,150,000
UCSC Genes
PSEN2
ADCK3
CABC1
All entorhinal cortex peaks
AD hypoacetylated peaks
AD hyperacetylated peaks
All peaks
FDR neg
FDR pos
1 2 3 4 5 6 7 8 9
AD samples (weighted average)
AD
Control samples (weighted average)
Control
Overlay (AD and control tracks)
Overlay
-log10P
-log10 P
LogFoldChange
Log fold change


c
Scale
chr1:
100 kb
hg19
227,000,000
227,050,000
227,100,000
227,150,000
UCSC Genes
PSEN2
ADCK3
CABC1
All entorhinal cortex peaks
AD hypoacetylated peaks
AD hyperacetylated peaks
All peaks
FDR neg
FDR pos
BMFL H3K27ac
Brain Mid Frontal Lobe Histone H3K27ac
BITL H3K27ac
Brain Inferior Temporal Lobe Histone H3K27ac
BHM H3K27ac
Brain Hippocampus Middle Histone H3K27ac
BSN H3K27ac
Brain Substantia Nigra Histone H3K27ac
BCG H3K27ac
Brain Cingulate Gyrus Histone H3K27ac
HDNP H3K27ac
H1 Derived Neuronal Progenitor Cultured Cells Histone H3K27ac
Lung H3K27ac
Lung Histone H3K27ac
AL H3K27ac
Liver Histone H3K27ac
SM H3K27ac
Skeletal Muscle Histone H3K27ac
Layered H3K27ac
H3K27ac Mark on 7 cell lines from ENCODE


AD
Control
2
0
−2
−4
log2 (fold ratio)
RGCC
PIM3
GPR22
KMO
PSEN1
CR1
Gene