

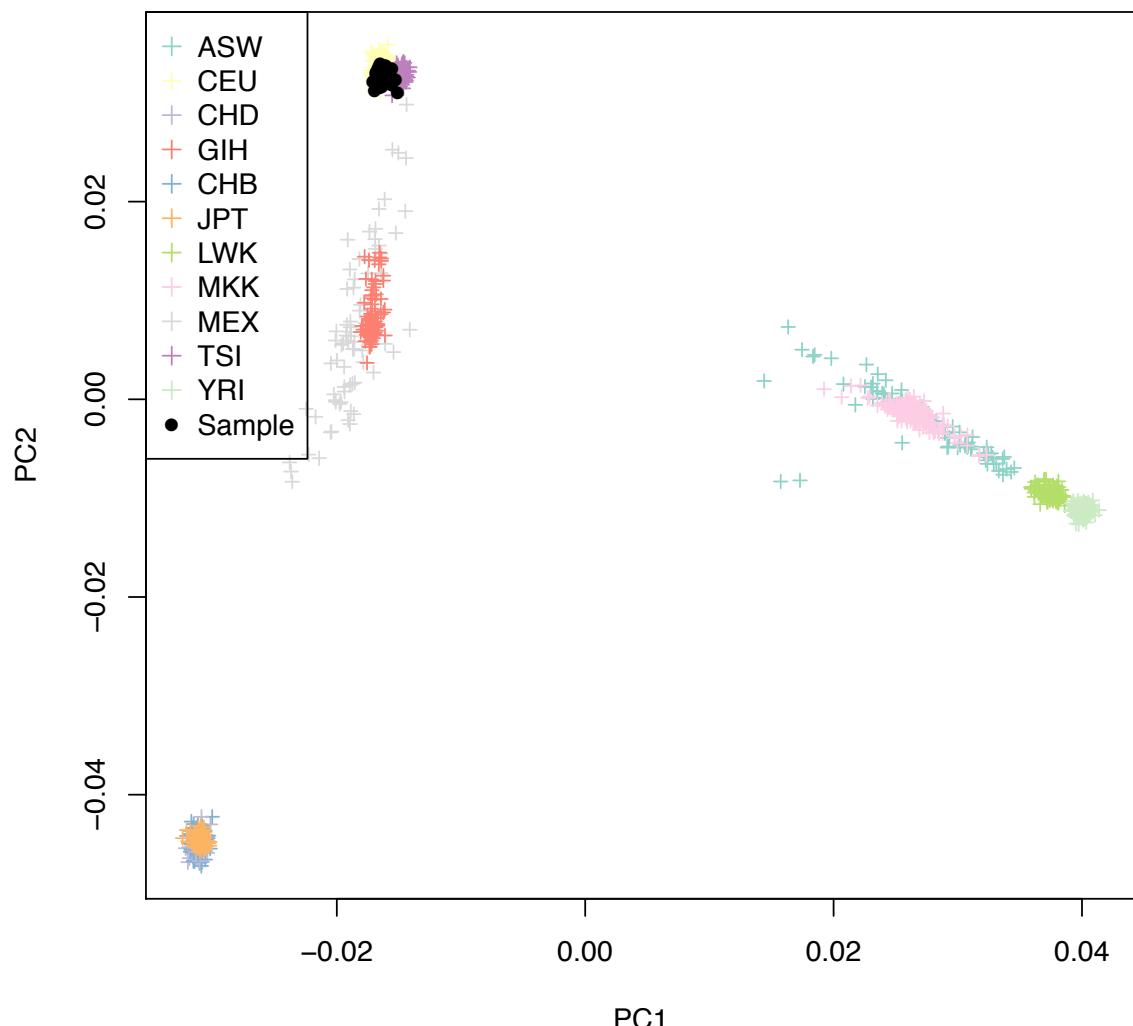
In the format provided by the authors and unedited.

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

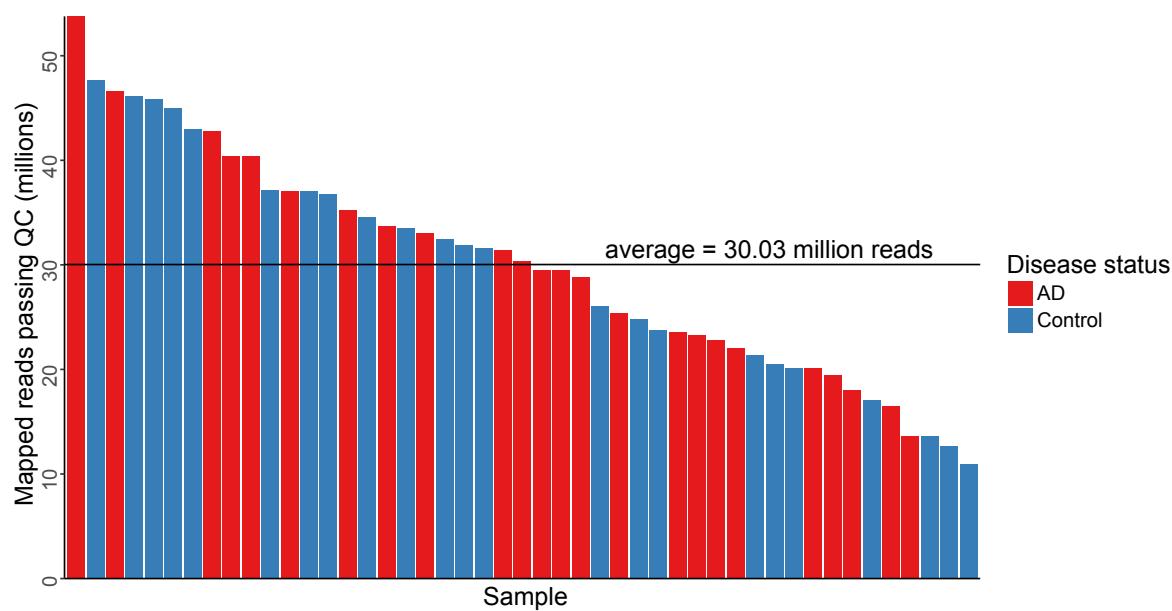
Sarah J. Marzi^{1,2}, Szi Kay Leung^{3,9}, Teodora Ribarska^{4,9}, Eilis Hannon³, Adam R. Smith¹⁰, Ehsan Pishva^{3,5}, Jeremie Poschmann^{3,6}, Karen Moore³, Claire Troakes¹⁰, Safa Al-Sarraj¹, Stephan Beck¹⁰, Stuart Newman⁸, Katie Lunnon³, Leonard C. Schalkwyk^{8,10} and Jonathan Mill^{10*}

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

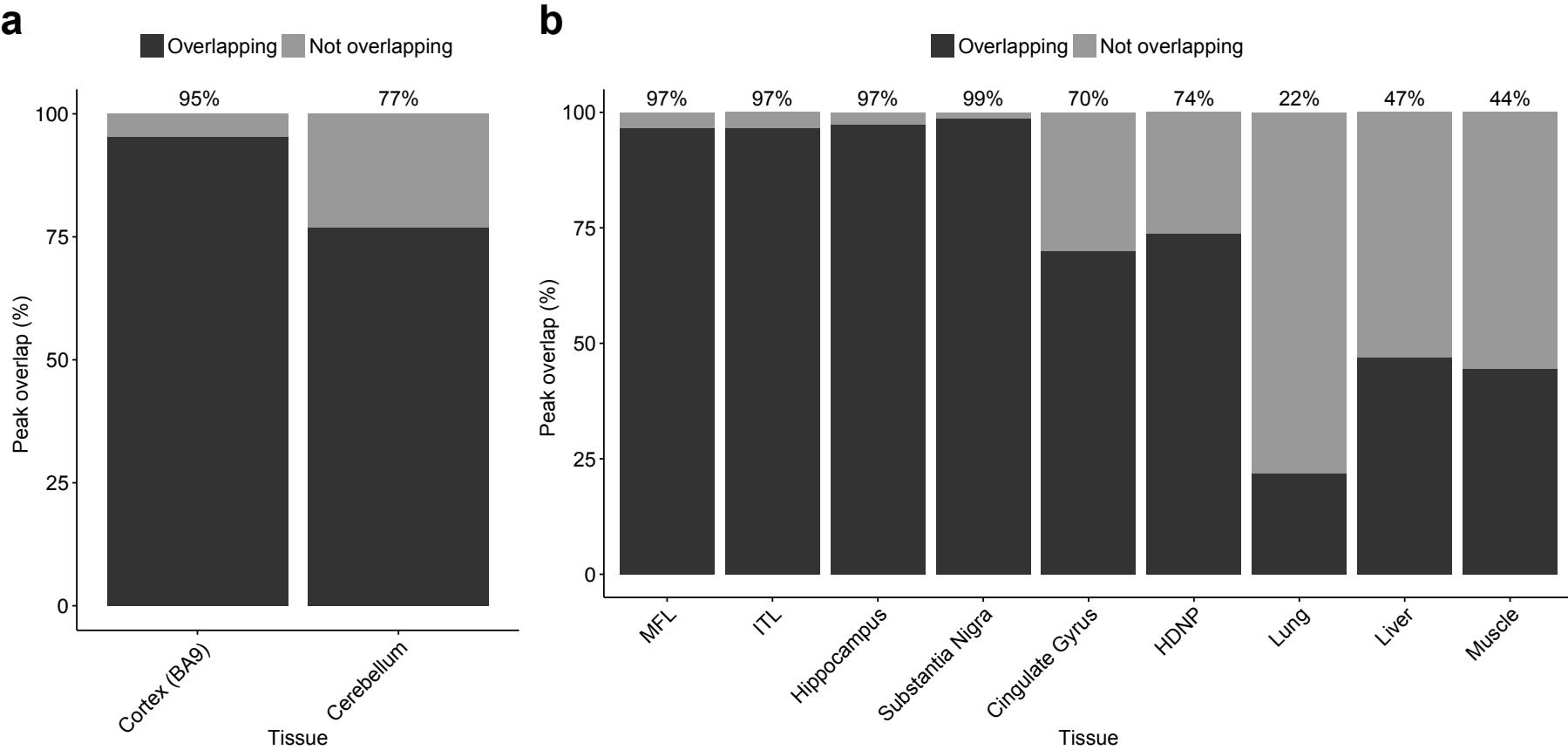
*These authors contributed equally: Szi Kay Leung, Teodora Ribarska. ¹⁰These authors jointly supervised this work: Leonard C. Schalkwyk, Jonathan Mill.
*e-mail: J.Mill@exeter.ac.uk



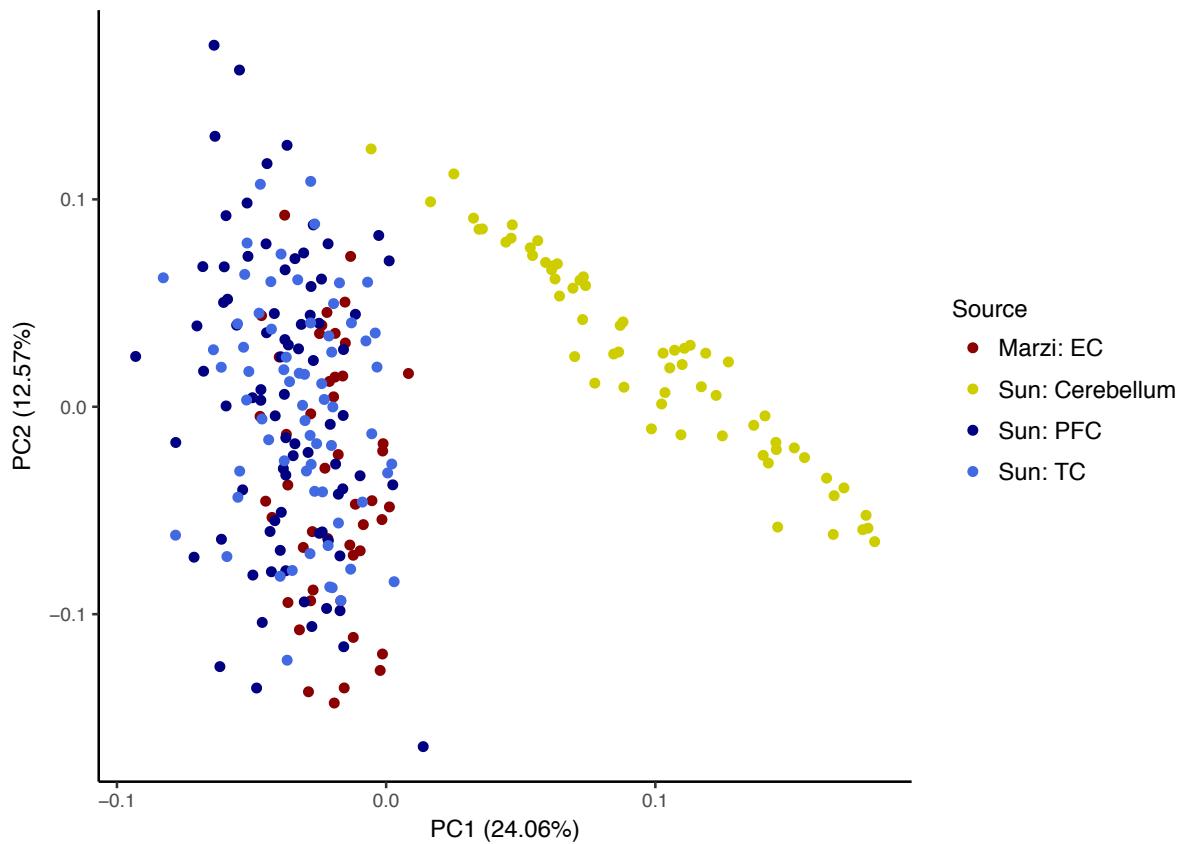
Supplementary Figure 1. Genome-wide SNP data was used to confirm that each of the samples included in our H3K27ac ChIP-seq analysis was of Western European ancestry. Genotype data from samples used in the current ChIP-seq study ($n = 47$ biologically independent samples) underwent principal component analysis with genotype data released from HapMap Phase 3 (<https://www.sanger.ac.uk/resources/downloads/human/hapmap3.html>). Shown are the loads on the first two principal components for our ChIP-seq sample compared to HapMap.



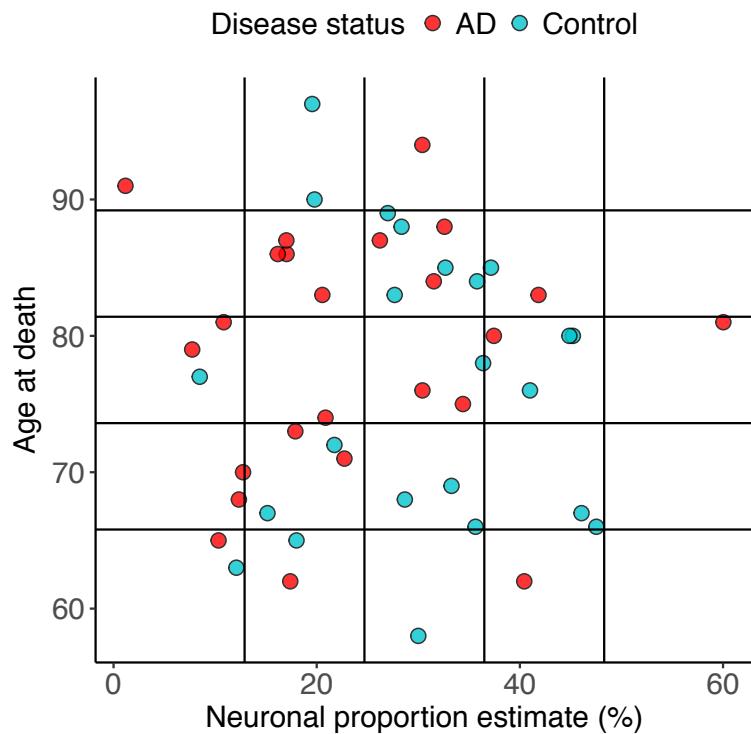
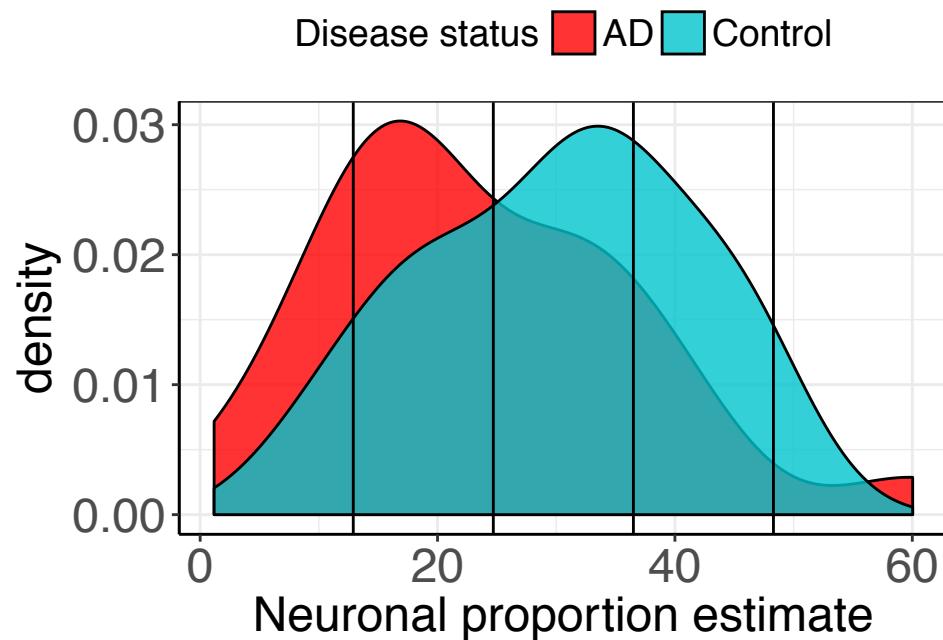
Supplementary Figure 2. The number of ChIP-seq reads passing stringent quality-control (QC) for each human entorhinal cortex sample. On average, we obtained 30,032,623 reads per sample ($n = 47$ biologically independent samples, $SD = 10,638,091$; range = 10,910,000–53,770,000) after filtering and data quality control (QC). There was no significant difference in the number of sequencing reads obtained for AD cases and controls (Welch 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, $t(43.85) = -0.08$).



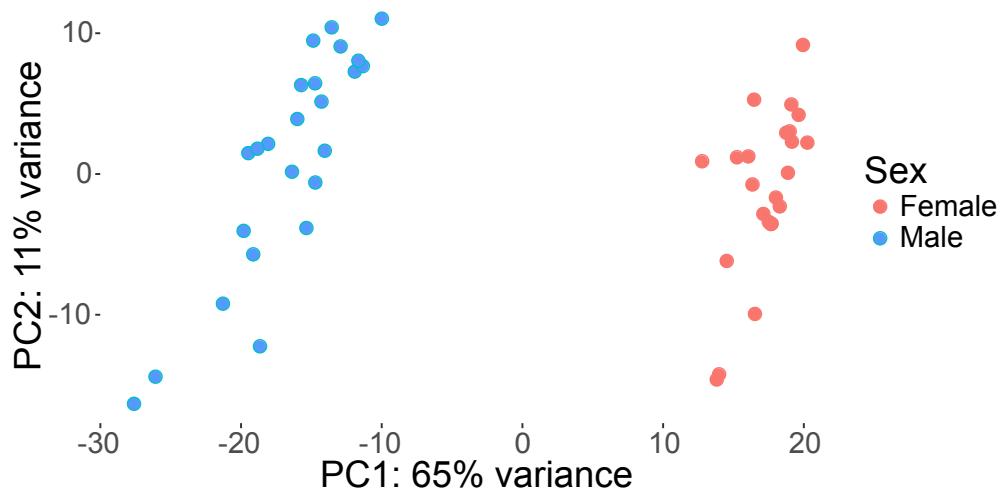
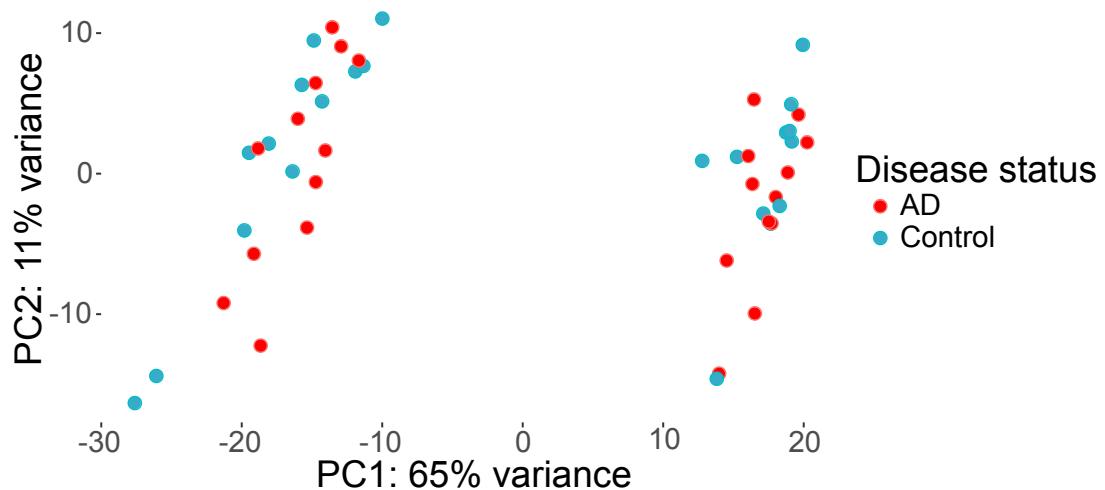
Supplementary Figure 3. H3K27ac peaks identified in our study overlap substantially with those identified in other brain ChIP-seq datasets. Shown is the percentage of H3K27ac peaks from published datasets overlapping with peaks in our entorhinal cortex (EC; n = 182,065) ChIP-seq data. **(a)** 53,882 (95%) of the 56,503 peaks identified by Sun et al (2016) in BA9 overlap our EC peaks, compared to 29,253 (77%) of the 38,069 cerebellum peaks. **(b)** Samples profiled by the Epigenomics Roadmap Consortium overlap our peaks in a tissue-specific manner with ~97% overlap observed for cortical and much lower overlaps for H3K27ac profiles derived from non-brain tissues (22-47%).



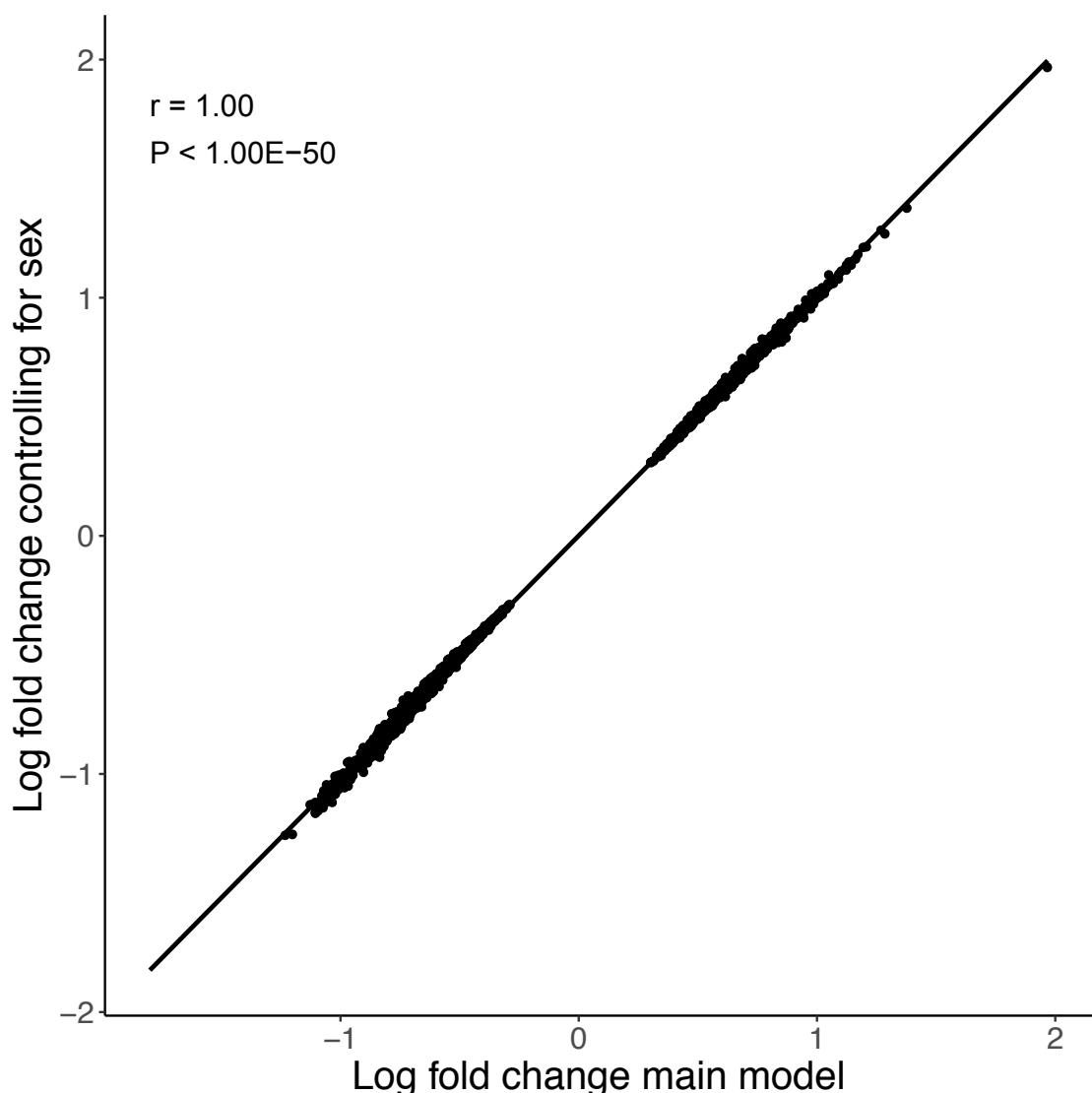
Supplementary Figure 4. H3K27ac peaks from our study cluster with cortical H3K27ac peaks identified by Sun et al (2016). Using the 182,065 entorhinal cortex (EC) peak regions identified in this study (Marzi: EC, n = 47 biologically independent samples), we generated read counts on raw H3K27ac data from cerebellum (n = 62 biologically independent samples), prefrontal cortex (PFC, n = 81 biologically independent samples) and temporal cortex (TC, n = 66 biologically independent samples) published by Sun et al. (2016). Counts per million (CPM) at all peaks were scaled and centered prior to principal components analysis. The figure shows the load of each sample on the first two principal components, confirming that our dataset clusters closely with all cortical samples from Sun et al., but differentially to the cerebellum samples from Sun et al.

a**b**

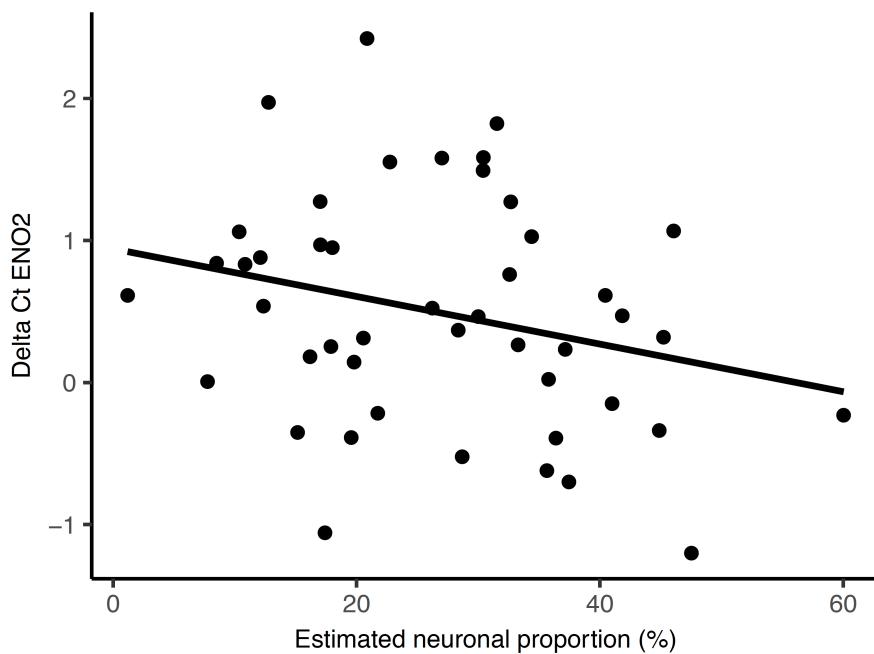
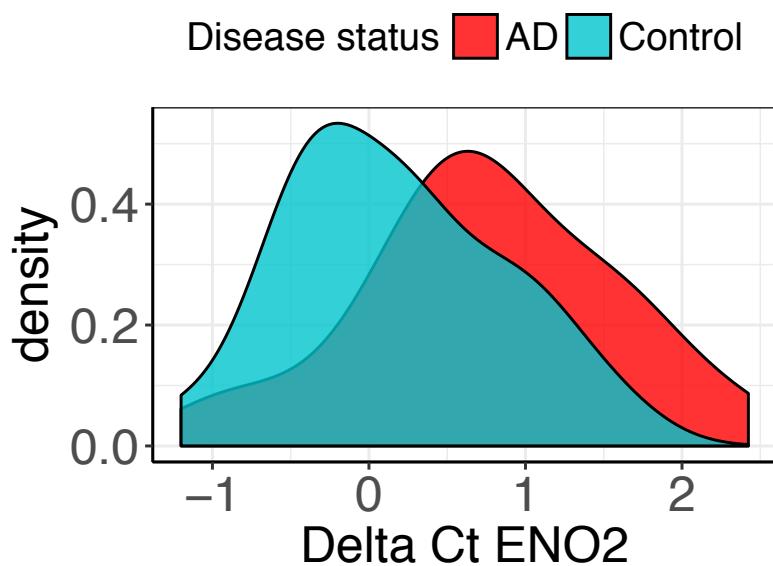
Supplementary Figure 5. Factor-conversion of age and neuronal proportion estimates. Neuronal cell proportion estimates (derived from DNA methylation data using CETS) and age at death were factorised, to adapt for use in the EdgeR statistical model. **(a)** The distribution of the samples' age at death and neuronal proportion estimates based on CETS are plotted together with the four breaks generated for the respective factor versions of the variable by the `cut()` function in R. **(b)** The distribution of the CETS variable highlights decreased neuronal proportion estimates in cases (AD) (average = 23.8%) compared to controls (average = 30.2%).

a**b**

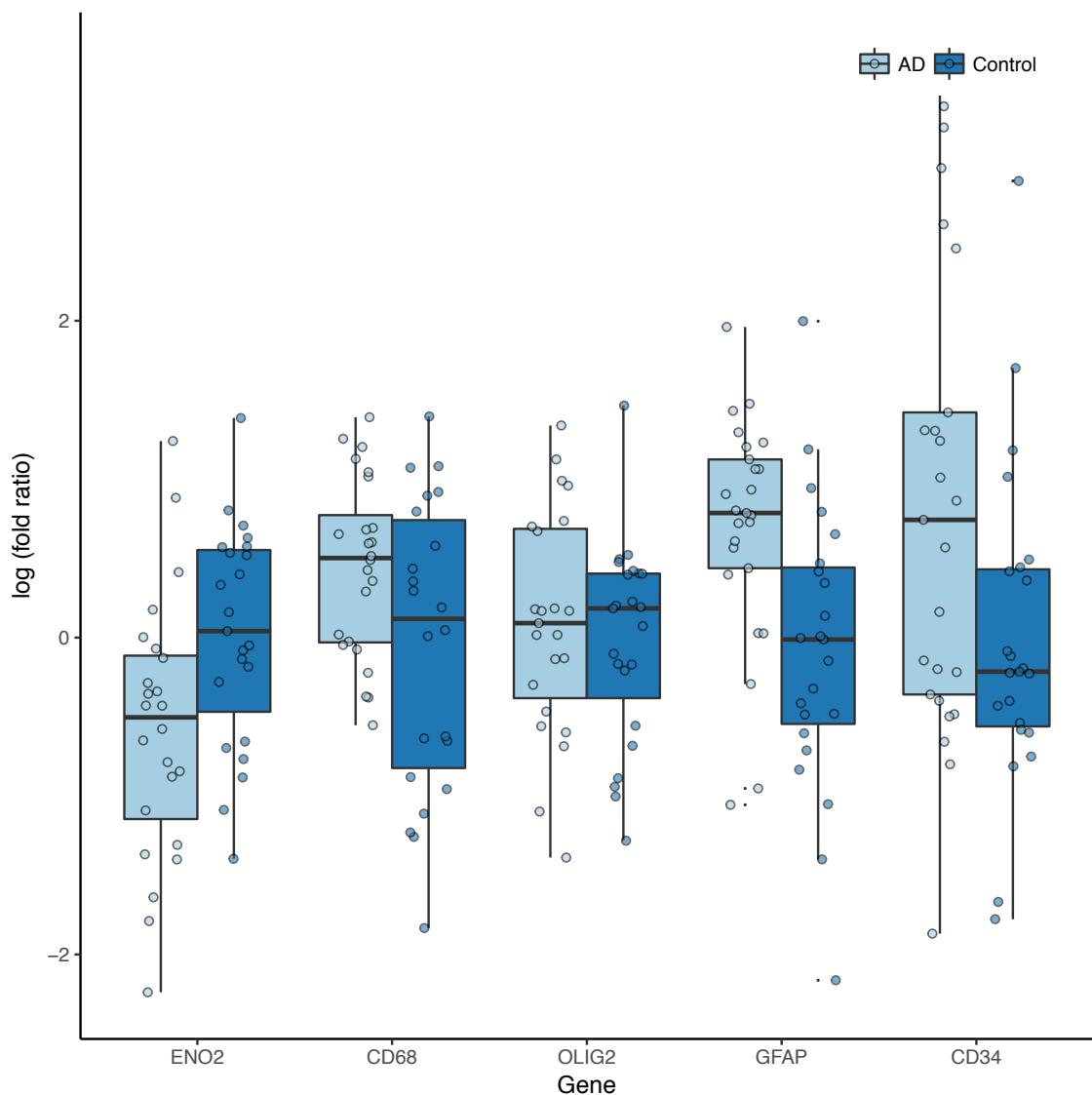
Supplementary Figure 6. Principal component analysis indicates that sex is the primary factor influencing H3K27ac in our dataset. Normalized read counts across our H3K27ac dataset ($n = 47$ biologically independent samples) were subjected to principal component analysis. **(a)** Load on the first principal component showed perfect overlap with recorded sex. **(b)** Variance captured by the same first PCs is not associated with disease status.



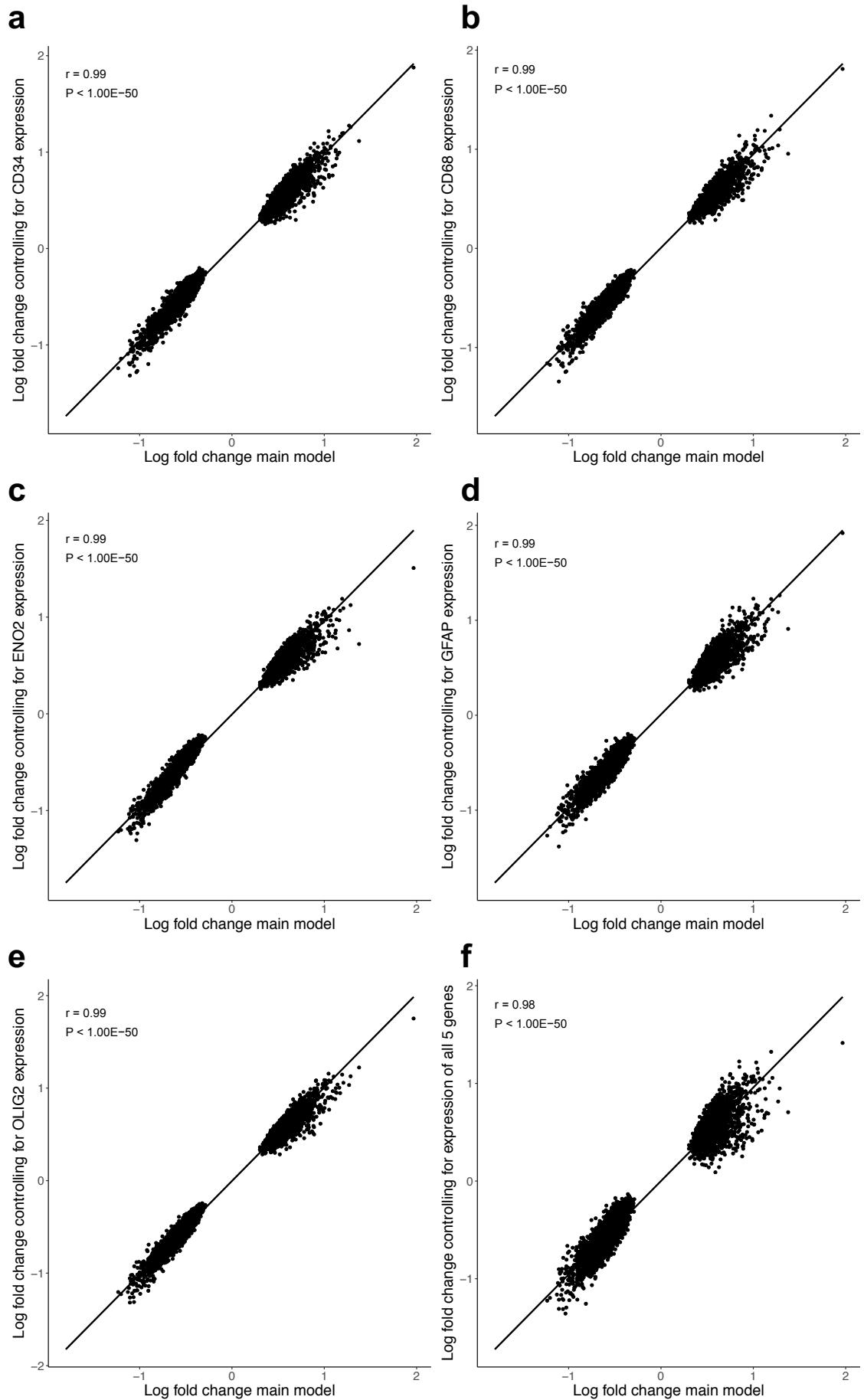
Supplementary Figure 7. Covarying for sex does not influence effect-sizes at AD-associated differentially-acetylated peaks. 4,157 of the 4,162 (99.9%) differentially acetylated peaks (FDR < 0.05) identified in the main model (controlling for age at death and neuronal proportion estimates derived from DNA methylation data) remain significantly differentially acetylated (FDR < 0.05) when additionally controlling for sex. The correlation of effects (log fold change) at the differentially acetylated peaks is near perfect (Pearson's product-moment correlation, n = 4,162 peaks, $r = 1.00$, $P < 1.00E-50$).

a**b**

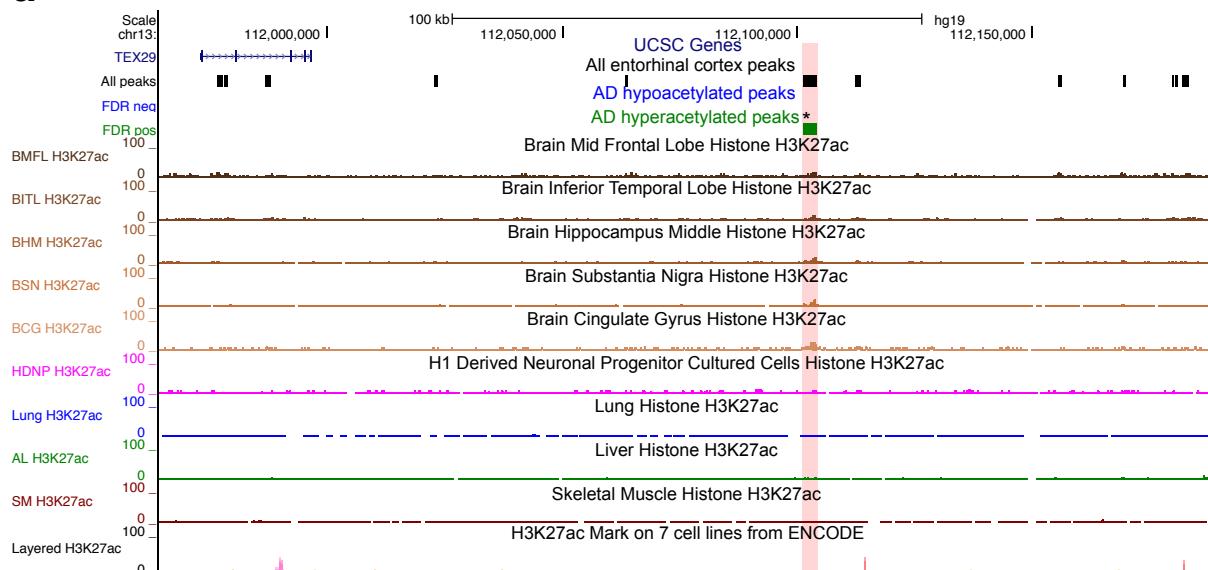
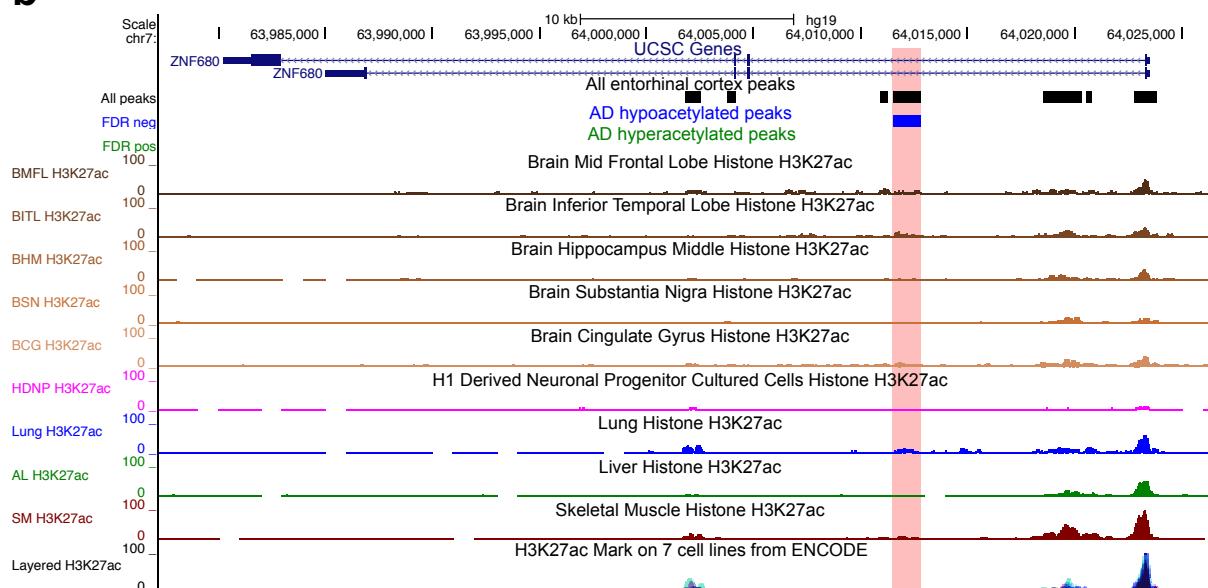
Supplementary Figure 8. The expression of ENO is correlated with the estimated neuronal cell proportion and downregulated in AD cases. (a) There is negative correlation (Pearson's product-moment correlation, $n = 47$ biologically independent samples, $r = -0.26$) between the estimated neuronal cell proportion, derived using DNA methylation data, and delta Ct levels for ENO mRNA, quantified using qPCR and normalized to the geometric mean of the expression of five housekeeping control genes. Reduced delta Ct corresponds to increased gene expression. (b) Levels of ENO are significantly lower in AD cases compared to low pathology samples (linear regression, $n = 47$ biologically independent samples, $P = 0.023$, $\beta = -0.54$, $F(3,43) = 3.23$).



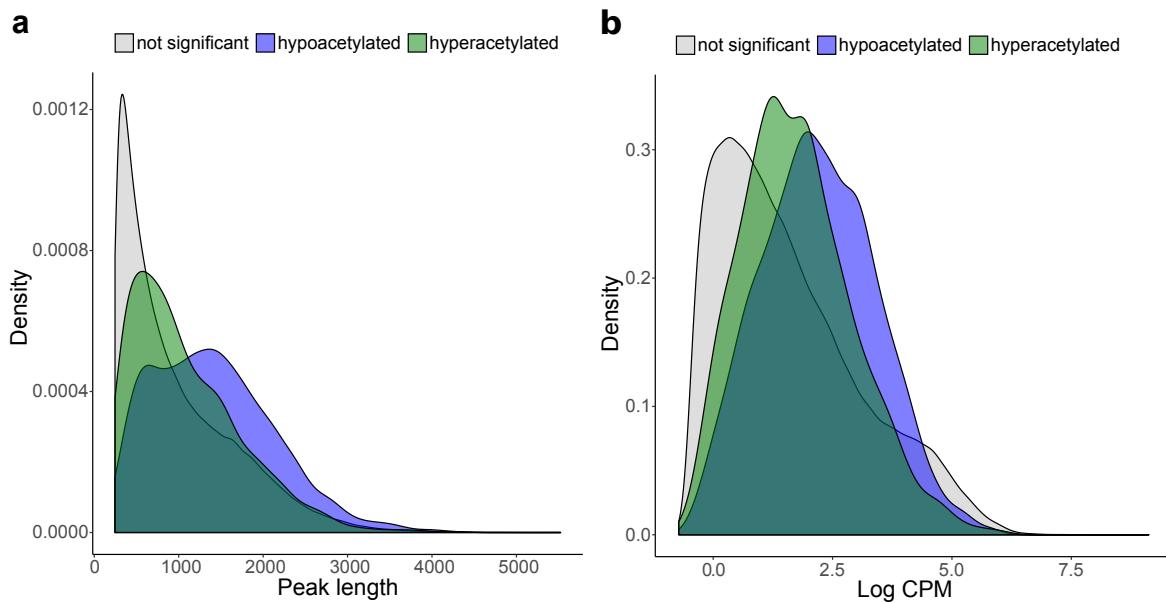
Supplementary Figure 9. Expression of cell-type-specific markers in AD. Given the neurodegeneration and neural cell changes associated with AD pathology, we quantified the levels of transcripts associated with five major brain cell types (*ENO2* (neurons), *OLIG2* (oligodendrocytes), *GFAP* (astrocytes), *CD68* (microglia) and *CD34* (endothelial cells)) in our samples using qPCR relative to the geometric mean of five housekeeping control genes ($n = 47$ biologically independent samples, AD = 24, control = 23). We analysed the association between gene-expression and disease status with a linear model, covarying for the RNA isolation batch and RIN score. Expression of *ENO2* was correlated with the neuronal cell proportion estimates derived from DNA methylation data (**Supplementary Fig. 8**) and found to be significantly reduced in AD samples (linear regression, $n = 47$ biologically independent samples, $P = 0.023$, $\beta = -0.54$, $F(3,43) = 3.23$). In contrast, we found elevated expression of *CD34* ($n = 47$ biologically independent samples, $P = 0.047$, $\beta = 0.69$, $F(3,43) = 6.09$), *CD68* ($n = 46$ biologically independent samples, $P = 0.012$, $\beta = 0.53$, $F(3,42) = 7.25$), and *GFAP* ($n = 47$ biologically independent samples, $P = 0.004$, $\beta = 0.75$, $F(3,43) = 3.27$) in our AD cases. No difference in the expression of *OLIG2* was observed ($n = 46$ biologically independent samples, $P = 0.565$, $\beta = 0.12$, $F(3,42) = 0.32$). 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.



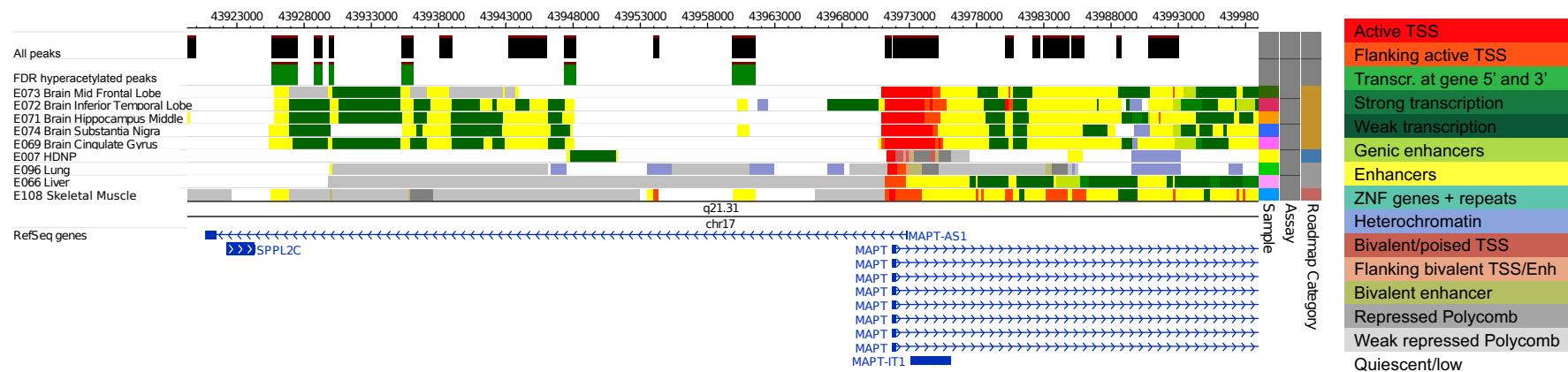
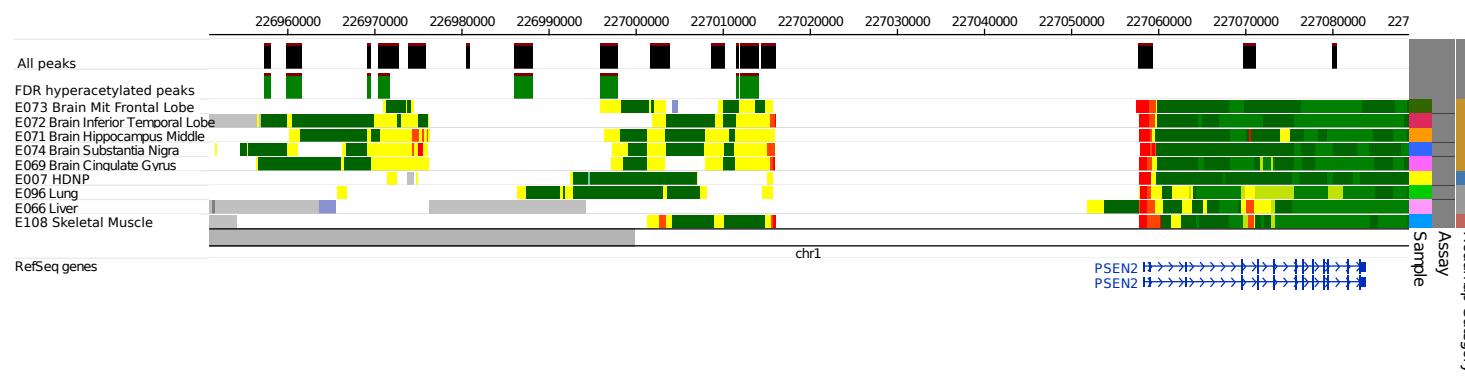
Supplementary Figure 10. Covarying for the expression of cell-type-specific genes does not influence effect-sizes at AD-associated differentially-acetylated peaks. Shown is the correlation of effect-sizes (log fold change) at the 4,162 differentially acetylated peaks (FDR < 0.05) identified in our primary model with those from an analysis additionally including expression of (a) *CD34*, (b) *CD68*, (c) *ENO2*, (d) *GFAP*, (e) *OLIG2* and (f) all five cell-type genes as covariates. Effect-sizes are highly correlated (Pearson's product-moment correlation, n = 4,162 peaks, r = 0.98-0.99, P < 1.00E-50) in each post-hoc analysis.

a**b**

Supplementary Figure 11. The top AD-hyperacetylated peak is characterized by brain-specific H3K27ac signals. H3K27ac ChIP-seq profiles for NIH Epigenomics Roadmap Consortium samples are shown: mid frontal lobe (BMFL), inferior temporal lobe (BITL), middle hippocampus (BHM), substantia nigra (BSN), cingulate gyrus (BCG), H1-derived neuronal progenitor cells (HDNP), lung, liver (AL), skeletal muscle (SM) and the layered H3K27ac track based on multiple cell lines displayed by the UCSC Genome Browser in the default setting. **(a)** A region surrounding the most significantly upregulated peak (quasi-likelihood F test, two-sided, $n = 47$ biologically independent samples, $P = 2.04E-08$, FDR = 0.002; highlighted by an asterisk) on chromosome 13 is presented. The peak and its immediate neighbour are characterized by brain-specific H3K27ac profiles. **(b)** A region around the most significantly downregulated peak (quasi-likelihood F test, two-sided, $n = 47$ biologically independent samples, $P = 1.66E-08$, FDR = 0.002), located in an intron of ZNF680 on chromosome 7, is shown.

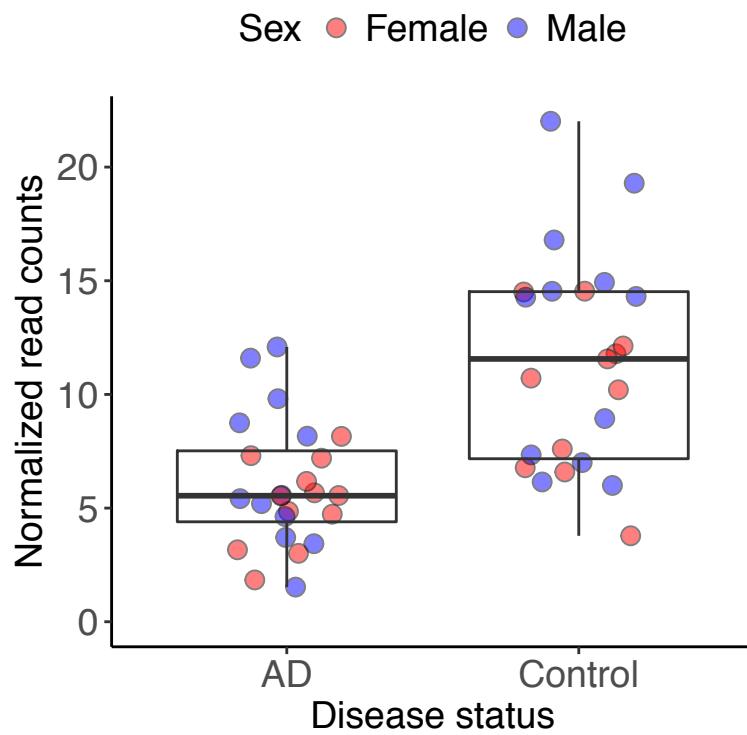


Supplementary Figure 12. AD-associated differentially acetylated peaks are larger and have higher average relative read counts than non-significant ones. On average FDR significant differentially acetylated peaks ($n = 4,162$ peaks) (a) are longer (Welch two-sample t-test, two-sided, $P < 1.00E-50$, average difference in length = 320.02bp, 95%-CI: 298-342bp, $t(4340.5) = 28.70$) and (b) have higher log counts per million (CPM; $P < 1.00E-50$, average difference in log CPM = 0.46, 95%-CI: 0.42-0.50, $t(4463.7) = 24.33$) compared to the non-significant background peaks ($n = 177,903$ peaks). Amongst differentially acetylated peaks, those that are hypoacetylated ($n = 2,867$ peaks) are characterized as (a) being longer ($P = 5.66E-31$, average difference in length = 331.23bp, 95%-CI: 288-374bp, $t(3303.8) = 15.17$) and (b) having higher CPM ($P = 2.69E-50$, average difference in log CPM = 0.45, 95%-CI: 0.37-0.52, $t(3111.4) = 11.70$) than hyperacetylated peaks ($n = 1,475$ peaks).

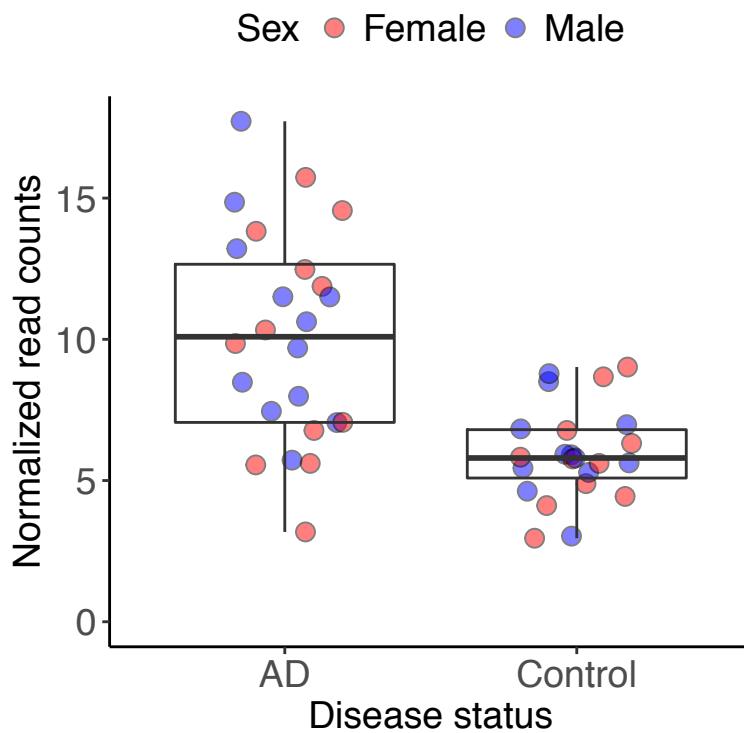
a**b**

Supplementary Figure 13. Chromatin state tracks for the *MAPT* and *PSEN2* hyperacetylated peak clusters. The tracks show the locations of H3K27ac peaks across the plotted window (labeled “All peaks”) as well as those characterized by elevated H3K27ac in AD (quasi-likelihood F test, $n = 47$ biologically independent samples, FDR < 0.05 ; labeled “FDR hyperacetylated peaks”). Also shown are the ChromHMM chromatin states from the Roadmap Epigenomics Consortium for a range of brain tissues (mid frontal lobe, inferior temporal gyrus, hippocampus middle, substantia nigra and cingulate gyrus) as well as non-brain tissue (H1-derived neuronal progenitor cells, lung, liver and skeletal muscle). Plots were produced on the WashU Epigenome Browser v42 using the core 15-state ChromHMM model. (a) A cluster of six AD-hyperacetylated H3K27ac peaks was identified upstream of *MAPT*. Chromatin state annotation shows that

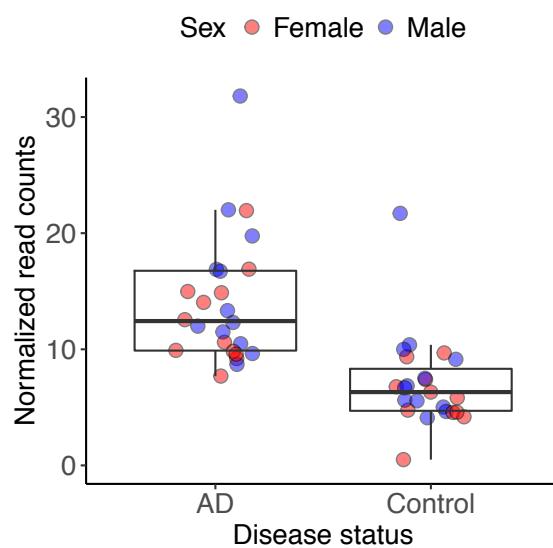
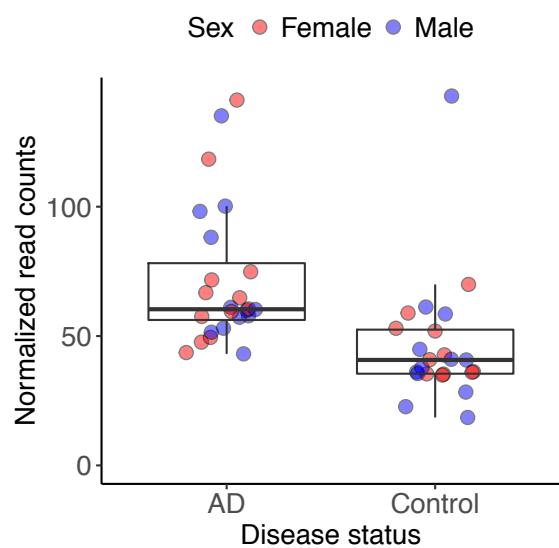
this region is characterized primarily by brain-specific enhancers and blocks of weak transcription. (b) A second cluster spanning nine significantly hyperacetylated peaks was identified upstream of *PSEN2*. *ChromHMM* identified two predominantly brain-specific active chromatin domains, comprised of enhancers and blocks of weak transcription.



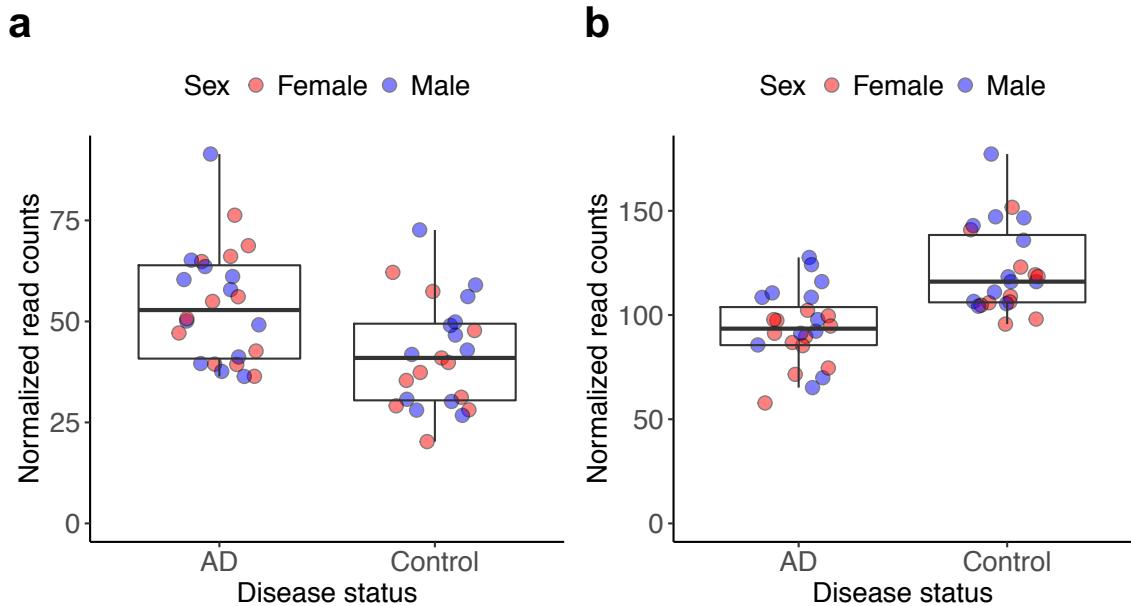
Supplementary Figure 14. A H3K27ac peak located downstream of APP is characterized by significant hypoacetylation in AD. (a) We identified one significantly hypoacetylated peak (FDR < 0.05) on chromosome 21, located ~100 kb downstream of APP (chr21:27160993-27161475; quasi-likelihood F test, two-sided, n = 47 biologically independent samples; P = 3.94E-04, FDR = 0.033, log fold change = -0.72). 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.



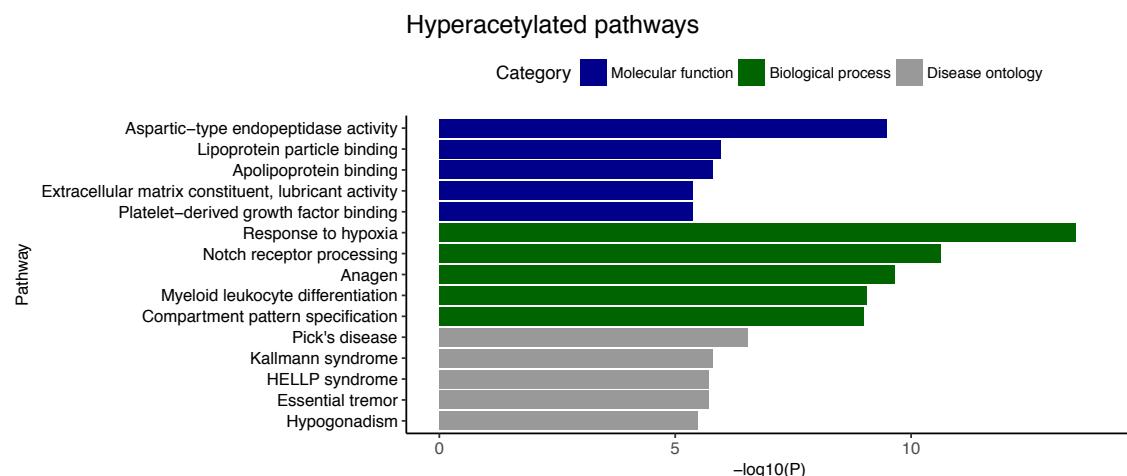
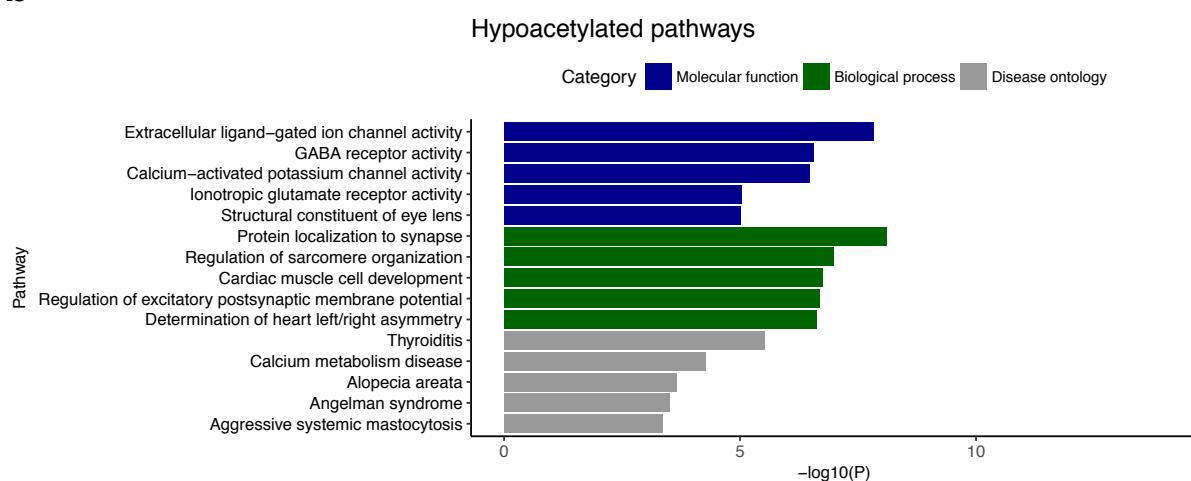
Supplementary Figure 15. A H3K27ac peak located in an intron of *PSEN1* is characterized by significant hyperacetylation in AD. (a) We identified one significantly hyperacetylated peak (FDR < 0.05) on chromosome 14, located in intron 6 of *PSEN1* (chr14:73656445-73656860; quasi-likelihood F test, two-sided, n = 47 biologically independent samples; P = 3.44E-04, log fold change = 0.68). 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.

a**b**

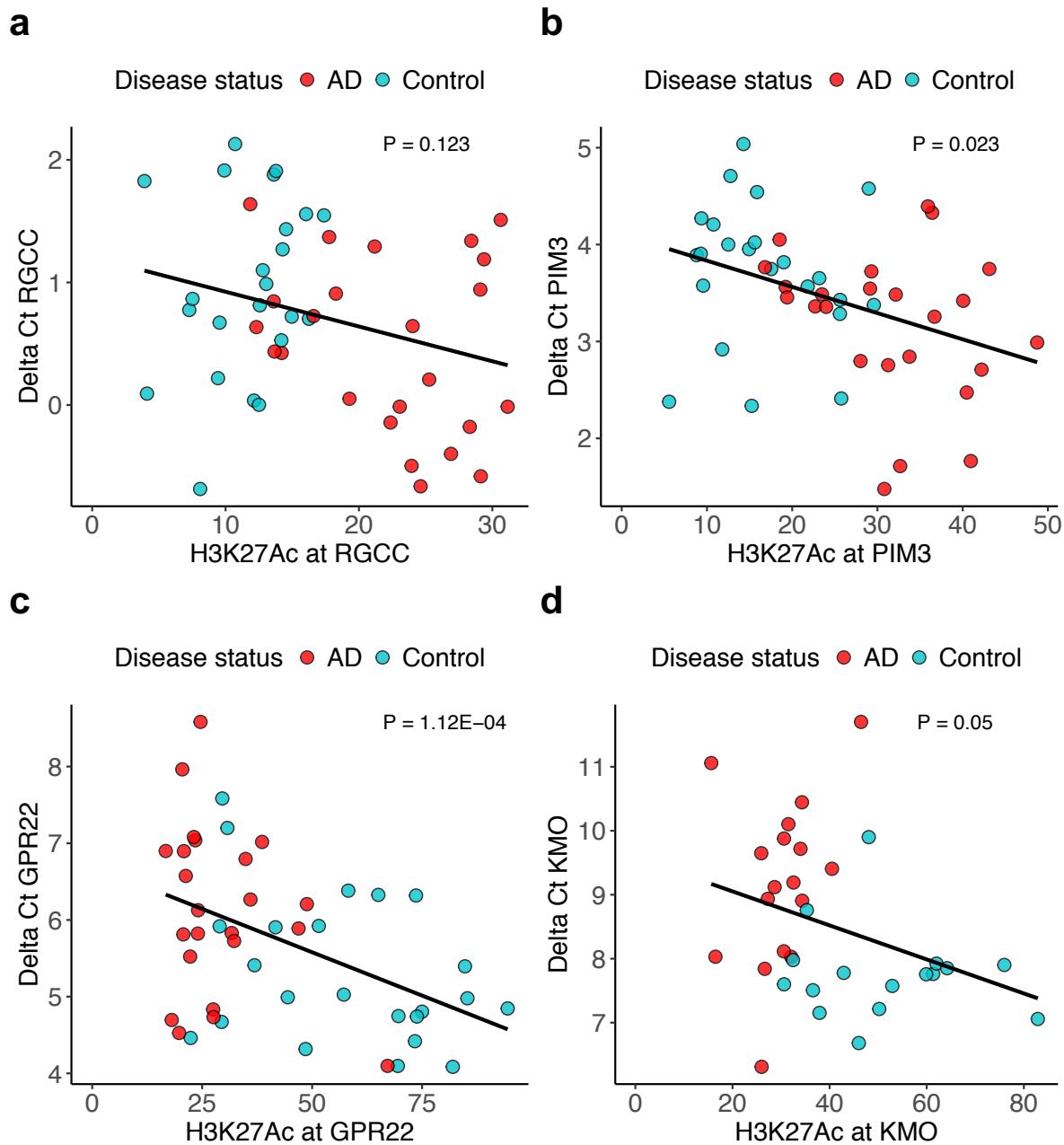
Supplementary Figure 16. Two intragenic H3K27ac peaks located across CR1 are characterized by significant hyperacetylation in AD. We identified two significantly hyperacetylated peaks (FDR < 0.05) on chromosome 1, located within the gene body of CR1: (a) chr1: 207753457-207753813; quasi-likelihood F test, two-sided, n = 47 biologically independent samples; P = 1.15E-06, log fold change = 0.99 and (b) chr1: 207754916-207756572; quasi-likelihood F test, two-sided, n = 47 biologically independent samples; P = 5.40E-04, log fold change = 0.56. 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.



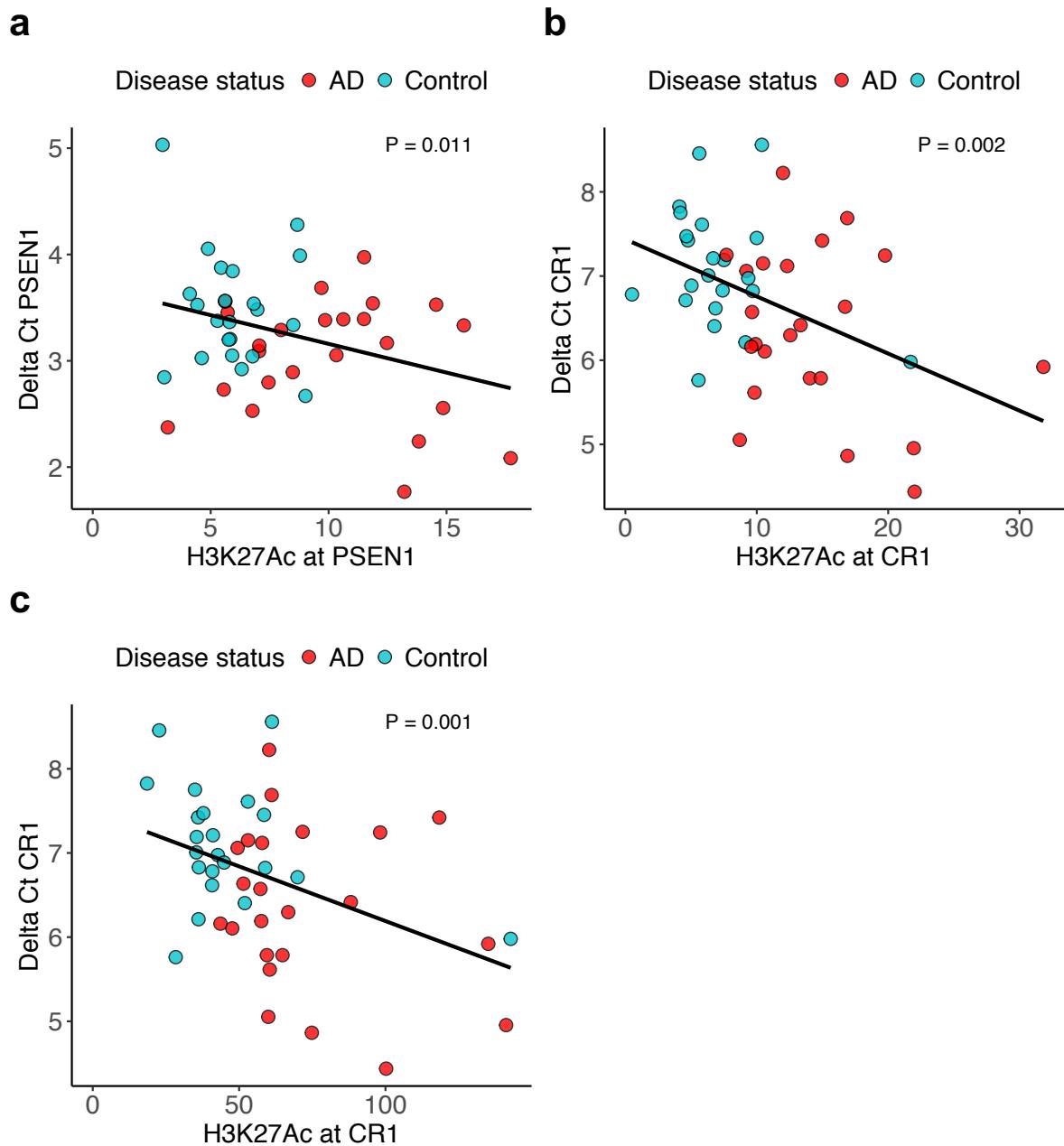
Supplementary Figure 17. Two differentially acetylated peaks overlap the GWAS LD block on chromosome 19. (a) We identified one significantly hyperacetylated peak ($FDR < 0.05$) on chromosome 19, mapping to the gene body of *TOMM40* and < 15 kb upstream of *APOE* (chr19:45394441- 45395396; quasi-likelihood F test, two-sided, $n = 47$ biologically independent samples; $P = 2.13E-04$, log fold change = 0.48). (b) A peak located in intron 1 of *PPP1R37* was hypoacetylated in AD (chr19: 45639588- 45641733; quasi-likelihood F test, two-sided, $n = 47$ biologically independent samples; $P = 7.65E-04$, log fold change = -0.33). 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.

a**b**

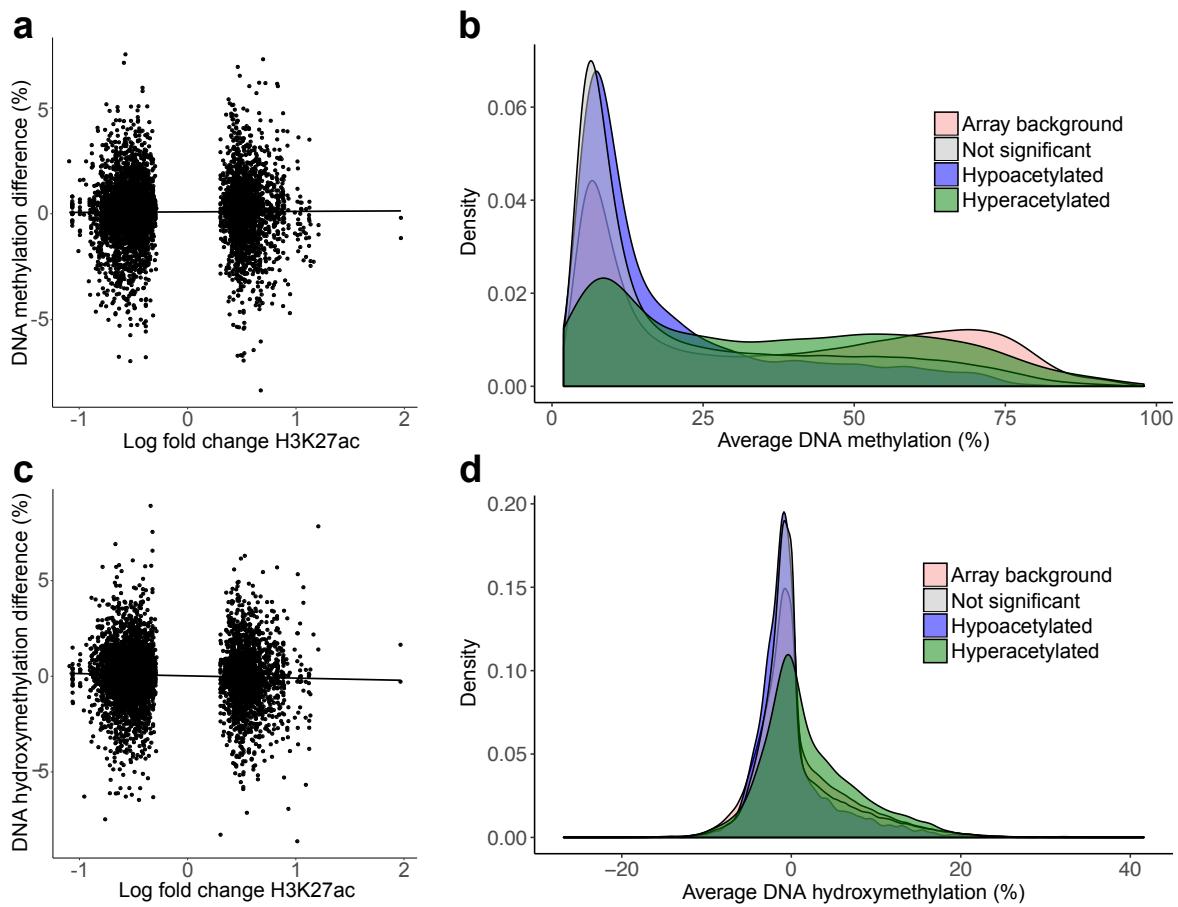
Supplementary Figure 18. Neurobiological and disease-related pathways are enriched in hyper- and hypoacetylated regions. Using all significantly hyper- and hypoacetylated peaks ($n = 4,162$ peaks, FDR <0.05), we conducted functional enrichment analyses against the background set of all peaks ($n = 182,065$ peaks) using *GREAT*. Shown are the top five independent enrichments in the categories Molecular function, biological process and disease ontology (see **Methods**). **(a)** AD-related pathways enriched in the hyperacetylated peaks include “lipoprotein particle binding” ($P = 1.10E-06$) and “response to hypoxia” ($P = 3.17E-14$) as well as “Pick’s disease” ($P = 2.93E-07$), a form of fronto-temporal dementia. **(b)** Amongst hypoacetylated pathways we identified neuronal transmission pathways, including “protein location to synapse” ($P = 7.86E-09$) and “GABA receptor activity” ($P = 2.70E-07$). The *GREAT* pathway analysis implements a hypergeometric test and only multiple-testing corrected significant results (FDR < 0.05) are reported as output.



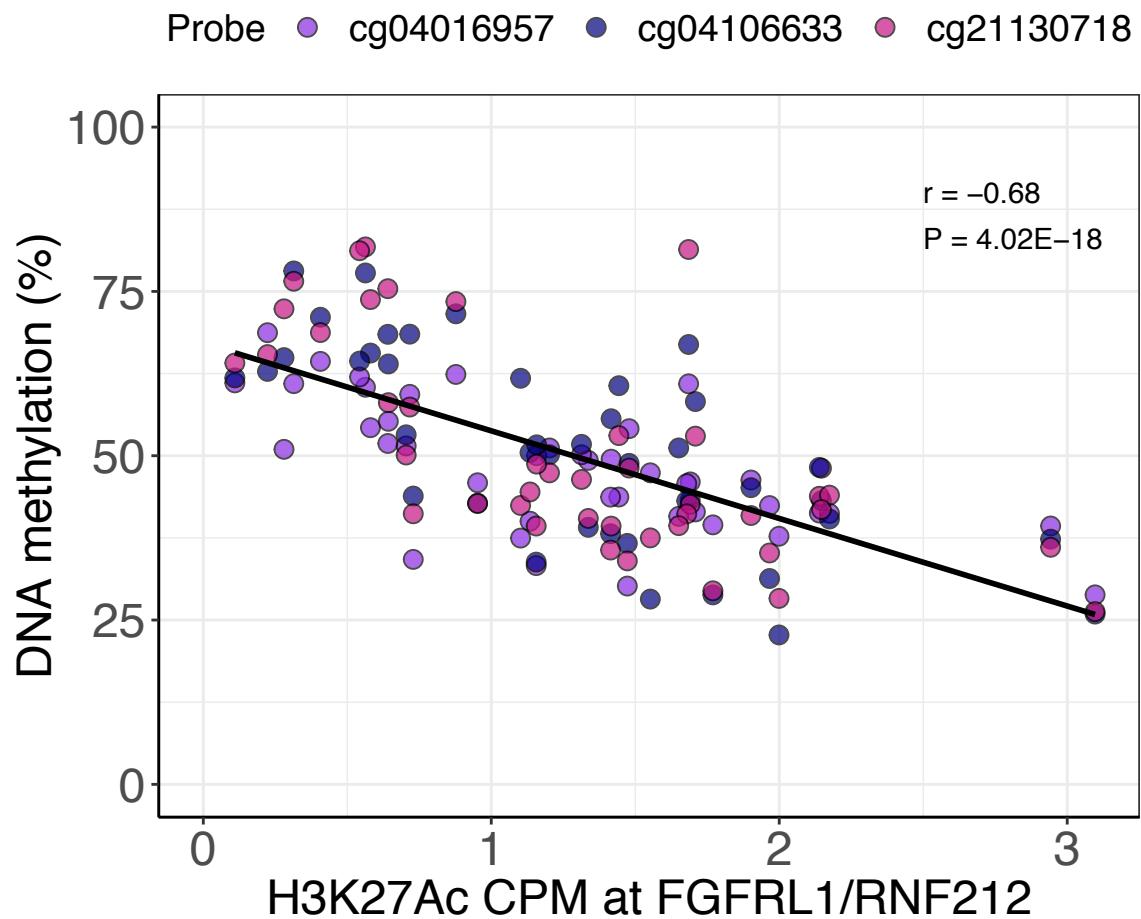
Supplementary Figure 19. The expression of *RGCC*, *PIM3*, *GPR22*, and *KMO* is associated with levels of H3K27ac across proximal AD-associated differentially acetylated peaks in samples included in our ChIP-seq dataset. Shown is the relationship between gene expression and H3K27ac for (a) *RGCC* (peak: chr13: 42094789-42095919; quasi-likelihood F test, two-sided, $n = 47$ biologically independent samples; $P = 0.123$, log fold change = -0.22), (b) *PIM3* (peak: chr22: 50342521-50343567; quasi-likelihood F test, two-sided, $n = 47$ biologically independent samples; $P = 0.023$, log fold change = -0.35), (c) *GPR22* (peak: chr7: 107111795-107113029; quasi-likelihood F test, two-sided, $n = 46$ biologically independent samples; $P = 1.1E-04$, log fold change = -0.38), and (d) *KMO* (peak: chr1: 241694436-241695782; quasi-likelihood F test, two-sided, $n = 33$ biologically independent samples; $P = 0.050$, log fold change = -0.15).



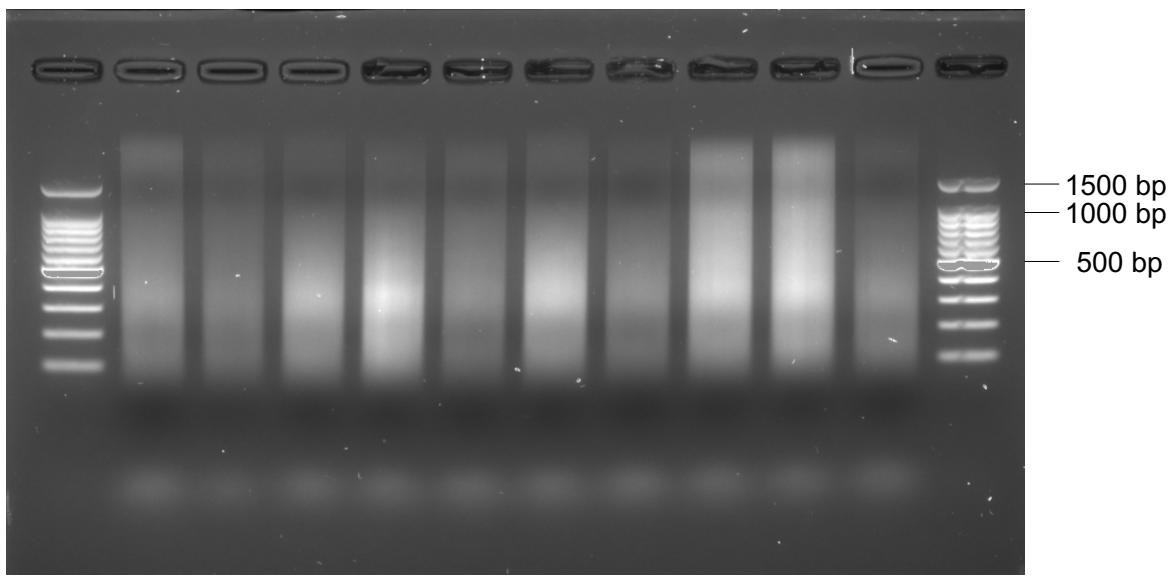
Supplementary Figure 20. The expression of *PSEN1* and *CR1* is associated with levels of H3K27ac across proximal AD-associated differentially acetylated peaks in samples included in our ChIP-seq dataset. Shown is the relationship between gene expression and H3K27ac for (a) *PSEN1* (peak: chr14: 73656445 – 73656860; quasi-likelihood F test, two-sided, $n = 47$ biologically independent samples; $P = 0.011$, log fold change = -0.42), (b) *CR1* (peak 1: chr1: 207753457-207753813; quasi-likelihood F test, two-sided, $n = 45$ biologically independent samples; $P = 0.002$, log fold change = -0.37), and (c) *CR1* (peak 2: chr1: 207754916-207756572; quasi-likelihood F test, two-sided, $n = 45$ biologically independent samples; $P = 0.001$, log fold change = -0.31).



Supplementary Figure 21. Divergent DNA methylation patterns are observed around AD hyperacetylated peaks. (a) Effect sizes for association between AD and DNA methylation at CpG probes within 1kb of FDR significant differentially acetylated peaks are not correlated with log fold change in H3K27ac at these peaks (Pearson's product-moment correlation, $n = 42$ biologically independent samples, $n = 268,477$ peak-probe pairs, $r = 0.009$, $P = 0.443$). (b) Probes in vicinity of entorhinal cortex H3K27ac peaks ($n = 232,233$ probes) show a skew towards lower DNA methylation compared to the array-background ($n = 135,182$ probes) (Welsh two-sample t-test, two-sided, $P < 1.00E-50$, average beta difference = 12.47%, 95%-CI: 12.35-12.59%, $t(551560) = 195.94$). (c) Effect sizes for association between AD and DNA hydroxymethylation at CpG probes within 1kb of FDR significant peaks are minimally, but significantly, negatively correlated with log fold change in H3K27ac at these peaks (Pearson's product-moment correlation, $n = 42$ biologically independent samples, $n = 268,477$ probe-peak pairs, $r = -0.045$, $P = 1.63E-04$). (d) DNA hydroxymethylation is significantly lower at sites near H3K27ac peaks ($n = 232,233$ probes), compared to the array background ($n = 135,182$ probes) (Welsh two-sample t-test, two-sided $P = 3.61E-30$, average beta difference = 0.16%, 95%-CI: 0.13-0.19%, $t(494170) = 10.99$).



Supplementary Figure 22. H3K27ac at a differentially acetylated peak annotated to *FGFRL1* and *RNF212* is negatively correlated with 5mC at three CpG sites. We identified an AD-associated differentially acetylated peak (chr4: 1044452 - 1044737) annotated to *FGFRL1* and *RNF212* at which H3K27ac is negatively correlated with 5mC at three specific 450K array probe (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, $r = -0.70$, $P = 2.98E-07$; combined across the three probes, $P = 4.02E-18$, $r = -0.68$; $n = 42$ biologically independent samples).



Supplementary Figure 23. Chromatin shearing resulted in an average size distribution of 100-1000 bp. 10 representative example agarose gel electrophoresis profiles ($n = 10$ biologically independent samples), generated from DNA extracted after sonication, are shown. Chromatin shearing efficiency was checked for all 47 samples.

Supplementary Table 1. Phenotypic information across all samples.

Sample	Post-mortem diagnosis	Braak stage	Age at death	Sex	Neuronal proportion estimate*	Post-mortem delay (mins)	Experiment
1	Control	0	90	M	20%	2700	ChIP-seq; Gene Expression
2	Control	2	76	F	41%	1320	ChIP-seq; Gene Expression
3	Control	1	72	F	22%	2820	ChIP-seq; Gene Expression
4	Control	2	68	M	29%	3600	ChIP-seq; Gene Expression
5	Control	0	77	F	8%	1260	ChIP-seq; Gene Expression
6	Control	2	89	F	27%	2580	ChIP-seq; Gene Expression
7	Control	0	66	F	36%	3000	ChIP-seq; Gene Expression
8	Control	3	88	F	28%	2340	ChIP-seq; Gene Expression
9	Control	3	97	M	20%	2460	ChIP-seq; Gene Expression
10	Control	2	80	M	45%	1860	ChIP-seq; Gene Expression
11	Control	2	67	F	15%	3720	ChIP-seq; Gene Expression
12	Control	0	84	M	36%	3180	ChIP-seq; Gene Expression
13	Control	2	80	M	45%	3300	ChIP-seq; Gene Expression
14	Control	0	58	M	30%	3480	ChIP-seq; Gene Expression
15	Control	3	78	M	36%	1440	ChIP-seq; Gene Expression
16	Control	2	83	F	-	2340	ChIP-seq; Gene Expression
17	Control	0	69	F	33%	2880	ChIP-seq; Gene Expression
18	Control	0	63	M	12%	1380	ChIP-seq; Gene Expression
19	Control	1	67	M	46%	1500	ChIP-seq; Gene Expression
20	Control	0	66	F	48%	4680	ChIP-seq; Gene Expression
21	Control	2	85	M	33%	3300	ChIP-seq; Gene Expression
22	Control	2	85	F	37%	2700	ChIP-seq; Gene Expression

23	Control	1	65	M	18%	1560	ChIP-seq; Gene Expression
24	Control	3	91	M	-	4980	Gene Expression
25	Control	3	99	F	-	1920	Gene Expression
26	Control	2	90	F	-	4980	Gene Expression
27	Control	2	91	M	-	2820	Gene Expression
28	Control	2	80	F	-	1320	Gene Expression
29	AD	6	94	F	30%	2520	ChIP-seq; Gene Expression
30	AD	6	88	F	33%	1080	ChIP-seq; Gene Expression
31	AD	6	76	M	30%	3300	ChIP-seq; Gene Expression
32	AD	6	74	M	21%	930	ChIP-seq; Gene Expression
33	AD	6	65	F	10%	4200	ChIP-seq; Gene Expression
34	AD	6	71	M	23%	1260	ChIP-seq; Gene Expression
35	AD	6	86	F	17%	1260	ChIP-seq; Gene Expression
36	AD	6	73	F	18%	5460	ChIP-seq; Gene Expression
37	AD	6	68	M	12%	1620	ChIP-seq; Gene Expression
38	AD	6	83	M	42%	4260	ChIP-seq; Gene Expression
39	AD	6	84	M	32%	4020	ChIP-seq; Gene Expression
40	AD	6	83	M	21%	5760	ChIP-seq; Gene Expression
41	AD	6	87	F	17%	2400	ChIP-seq; Gene Expression
42	AD	6	62	M	40%	2940	ChIP-seq; Gene Expression
43	AD	6	81	F	60%	4800	ChIP-seq; Gene Expression
44	AD	6	79	F	8%	3780	ChIP-seq; Gene Expression
45	AD	6	91	M	1%	4080	ChIP-seq; Gene Expression
46	AD	6	86	F	16%	660	ChIP-seq; Gene Expression
47	AD	6	75	F	34%	720	ChIP-seq; Gene Expression
48	AD	6	80	M	37%	2700	ChIP-seq; Gene Expression

49	AD	6	81	F	11%	1200	ChIP-seq; Gene Expression
50	AD	6	70	M	13%	5400	ChIP-seq; Gene Expression
51	AD	6	62	M	17%	1380	ChIP-seq; Gene Expression
52	AD	6	87	F	26%	1860	ChIP-seq; Gene Expression
53	AD	5	85	F	-	2100	Gene Expression
54	AD	N/A	91	F	-	1650	Gene Expression
55	AD	5	89	F	-	4320	Gene Expression
56	AD	5	80	M	-	1080	Gene Expression
57	AD	5/6	70	F	-	870	Gene Expression
58	AD	4	86	M	-	3150	Gene Expression
59	AD	6	92	M	-	1920	Gene Expression
60	AD	4	87	M	-	1140	Gene Expression
61	AD	4	82	M	-	1680	Gene Expression
62	AD	4	79	M	-	>1440	Gene Expression
63	AD	5/6	93	F	-	3900	Gene Expression
64	AD	5	88	M	-	5460	Gene Expression
65	AD	4	91	M	-	1980	Gene Expression
66	AD	6	67	M	-	1320	Gene Expression
67	AD	5	94	F	-	900	Gene Expression
68	AD	6	78	F	-	2160	Gene Expression
69	AD	6	76	M	-	1620	Gene Expression
70	AD	6	85	M	-	3420	Gene Expression
71	AD	3	92	F	-	1740	Gene Expression
72	AD	5/6	80	M	-	2880	Gene Expression
73	AD	6	78	M	-	1320	Gene Expression
74	AD	5	85	F	-	1380	Gene Expression

75	AD	6	77	M	-	4200	Gene Expression
76	AD	4	88	M	-	4440	Gene Expression
77	AD	5	88	M	-	1980	Gene Expression
78	AD	4	81	M	-	3060	Gene Expression
79	AD	5	65	M	-	2280	Gene Expression
80	AD	5	85	F	-	4020	Gene Expression
81	AD	4	79	M	-	1740	Gene Expression
82	AD	6	85	M	-	3960	Gene Expression
83	AD	5	89	M	-	1560	Gene Expression
84	AD	5	92	F	-	960	Gene Expression
85	AD	4	92	M	-	4200	Gene Expression
86	AD	5	90	F	-	600	Gene Expression
87	AD	5	97	F	-	2220	Gene Expression
88	AD	5/6	76	M	-	660	Gene Expression
89	AD	6	83	M	-	1320	Gene Expression
90	AD	6	80	F	-	1320	Gene Expression
91	AD	6	78	F	-	2520	Gene Expression
92	AD	4	91	M	-	2280	Gene Expression
93	AD	6	87	M	-	3060	Gene Expression
94	AD	5	89	F	-	>3000	Gene Expression
95	AD	6	89	M	-	4560	Gene Expression

* Neuronal proportion estimates are derived from DNA methylation data generated on the same samples using the Illumina 450K HumanMethylation Array, calculated using the *CETS* R package (see **Methods**)¹⁰.

Supplementary Table 2. Distribution of H3K27ac peaks across the 24 chromosomes.

Chromosome	Number of peaks
1	17420
2	14739
3	11596
4	8136
5	10476
6	9783
7	9843
8	7921
9	8473
10	9064
11	9324
12	8670
13	4646
14	5927
15	6342
16	6972
17	8139
18	4008
19	5768
20	4885
21	2449
22	3873
X	3257
Y	354

Supplementary Table 3. Expression differences for cell-type-specific genes between AD cases and controls.

Gene	Beta coefficient (Delta Ct)	P
<i>ENO2</i>	-0.535	0.023
<i>CD68</i>	0.527	0.012
<i>OLIG2</i>	0.119	0.565
<i>GFAP</i>	0.749	0.004
<i>CD34</i>	0.685	0.047

Supplementary Table 4. Hyperacetylated AD-associated H3K27ac peaks.

See attached xlsx file SupplementaryTable4_ADHyperacetylatedPeaks.xlsx

Supplementary Table 5. Hypoacetylated AD-associated H3K27ac peaks.

See attached xlsx file SupplementaryTable5_ADHypoacetylatedPeaks.xlsx

Supplementary Table 6. Association statistics for the 9 peaks in the differentially acetylated region on chromosome 17, annotated to *MAPT* and *SPPL2C*.

Start - end	P	P FDR	Log FC
43925717-43927482	7.01E-07	0.005	0.71
43928851-43929350	7.68E-04	0.043	0.60
43929961-43930229	1.08E-03	0.049	0.59
43935330-43936111	1.72E-04	0.024	0.54
43938147-43939070	0.842	0.941	0.02
43943325-43946054	0.165	0.454	0.20
43947431-43948194	8.84E-04	0.046	0.45
43954021-43954371	0.052	0.263	0.46
43959954-43961546	8.94E-05	0.020	0.59

Supplementary Table 7. Association statistics for the 14 peaks in the differentially acetylated region on chromosome 1, annotated to *PSEN2* and *ITPKB*.

Start - end	P	P FDR	Log FC
226957424-226958029	2.19E-05	0.013	0.84
226959991-226961572	8.10E-04	0.044	0.46
226969195-226969518	2.57E-04	0.029	0.66
226970654-226971677	3.40E-04	0.032	0.56
226971888-226972730	0.014	0.144	0.39
226974106-226975826	0.014	0.144	0.37
226980648-226980894	0.144	0.426	0.34
226986193-226988116	1.07E-06	0.005	0.60
226996028-226997864	6.10E-05	0.018	0.53
227001743-227003902	0.012	0.134	0.28
227008727-227010081	2.48E-03	0.068	0.42
227011597-227011851	3.47E-04	0.032	0.60
227012069-227012797	1.67E-04	0.024	0.65
227013015-227014019	7.90E-04	0.043	0.53

Supplementary Table 8. Results from partitioned heritability LD Score regression analysis highlighting that AD GWAS variants are enriched in regions of regulatory/enhancer function in the brain. 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 highly-significant enrichment (enrichment = 1.10 (95%-CI: 1.05 - 1.15)).

		H3K27ac Peaks	Outside Peaks
Observed scale h^2	Estimate	0.0302	0.049
	SE	0.00900	0.00930
	95%-LCI	0.0126	0.0305
	95%-UCI	0.0478	0.0669
Proportion of h^2g		0.383	0.618
Proportion of SNPs		0.349	0.651
Enrichment	Estimate	1.10	0.948
	SE	0.0258	0.0143
	95%-LCI	1.05	0.920
	95%-UCI	1.15	0.976

Supplementary Table 9. Locations, GWAS SNPs and *P* values for the 11 LD blocks constructed from the Lambert et al AD GWAS.

CHR	Start	End	Number of SNPs in GWAS	Min P SNP	Location	<i>P</i>
1	207372941	207875537	1144	rs1752684	207747296	3.65E-15
2	127826533	127889932	255	rs7561528	127889637	6.54E-18
6	47318885	47662484	981	rs9381563	47432637	5.30E-09
7	143083661	143198249	278	rs10808026	143099133	1.42E-11
8	27170788	27489959	869	rs7982	27462481	2.48E-17
11	59826677	60105199	733	rs72924659	60103385	5.35E-13
11	85636074	85874322	648	rs10792832	85867875	6.53E-16
11	121423640	121517613	176	rs11218343	121435587	4.98E-11
14	92926952	92941096	95	rs12590654	92938855	4.10E-08
18	29088958	29088958	1	rs8093731	29088958	4.63E-08
19	45050747	45691126	1504	rs12972156	45387459	<1.00E-50

Supplementary Table 10. Neurobiological and disease-related pathways are enriched in hyper- and hypoacetylated regions. Using all significantly hyper- and hypoacetylated peaks (FDR <0.05), we conducted functional enrichment analyses using GREAT. Shown are the top five independent enrichments in the categories molecular function, biological process and disease ontology (see **Methods**). AD-related pathways enriched in the hyperacetylated peaks include “lipoprotein particle binding” ($P = 1.10\text{E-}06$) and “response to hypoxia” ($P = 3.17\text{E-}14$) as well as “Pick’s disease” ($P = 2.93\text{E-}07$), a form of fronto-temporal dementia. Amongst hypoacetylated pathways we identified neuronal transmission pathways, including “protein location to synapse” ($P = 7.86\text{E-}09$) and “GABA receptor activity” ($P = 2.70\text{E-}07$). DA = differentially acetylated. Gene names in **bold** are represented in multiple pathways within the same ontology.

DA peak type	Ontology	Pathway	Rank	P	Enrichment	Total genes	Genes annotated to DA peaks
Hyperacetylated	Molecular function	aspartic-type endopeptidase activity	1	3.19E-10	5.68	26	<i>BACE2, CTSD, HM13, NRIP2, PSEN1, PSEN2, REN, SPPL2C</i>
		lipoprotein particle binding	2	1.10E-06	3.28	23	<i>APOA1, CDH13, COLEC12, LIPC, MAPT, SCARB1, SORL1</i>
		apolipoprotein binding	3	1.61E-06	4.07	16	<i>LCAT, LIPC, MAPT, SCARB1</i>
		extracellular matrix constituent, lubricant activity	4	4.32E-06	5.25	11	<i>MUC3A, MUC4, MUC5AC</i>
		platelet-derived growth factor binding	5	1.25E-05	4.37	14	<i>COL6A1, PDGFA, PDGFB</i>
	Biological process	response to hypoxia	1	3.17E-14	2.40	221	<i>ACTN4, ADS, AGTRAP, ALDH3A1, ALKBH5, ANGPTL4, ATP1B1, BMP7, CD38, CHRNA4, CHRNA7, CREBBP, CRHR1, CRYAA, CXCR4, DDIT4, DRD2, ENG, FABP1, FNDC1, HIF3A, HIPK2, IRAK1, MDM2, MDM4, MECP2, MT3, NDRG1, NF1, NOTCH1, PDGFA, PDGFB, PGF, PML, PSEN2, RAMP2, RGCC, RPS27A, S100B, SFRP1, SMAD9, STAT5B, STC2, TGFB2</i>
		Notch receptor processing	2	2.40E-11	5.36	20	<i>ADAM10, DLL1, DLL4, DNER, NOTCH1, PSEN1, PSEN2, RPS27A</i>
		anagen	3	2.21E-10	6.99	11	<i>NOTCH1, PPARD, PSEN1, PSEN2</i>
		myeloid leukocyte differentiation	4	8.88E-10	2.77	82	<i>BATF3, CEBPA, CSF1, CSF3, EPHA2, FAM20C, GAB2, GATA2, IRF4, KDM1A, MEF2A, MITF, MT1G, NFATC1, OCSTAMP, PSEN1, PSEN2</i>
		compartment pattern specification	5	1.00E-09	11.31	5	<i>DLL1, LFNG, NOTCH1, PBX3</i>

	Disease ontology	Pick's disease	1	2.93E-07	6.79	11	<i>CHRNA7, MAPT, PSEN1</i>
		Kallmann syndrome	2	1.61E-06	3.88	11	<i>AKAP2, GNRH1, NSMF, PALM2-AKAP2, PROK2, WDR11</i>
		HELLP syndrome	3	1.93E-06	4.50	21	<i>ADAM12, CDKN1C, ENG, NOD2</i>
		essential tremor	4	1.96E-06	3.66	18	<i>HS1BP3, LINGO1, MAPT, PARK2, PPP2R2B</i>
		hypogonadism	5	3.26E-06	3.07	32	
	Molecular function	extracellular ligand-gated ion channel activity	1	1.46E-08	2.27	72	<i>CHRNA1, GABRA1, GABRA3, GABRA4, GABRA5, GABRB1, GABRB3, GABRG2, GABRG3, GABRP, GLRB, GRIA1, GRIA2, GRIK1, GRIK2, GRIN2A, GRIN2B, GRIN3A</i>
		GABA receptor activity	2	2.70E-07	3.43	21	<i>GABBR1, GABRA1, GABRA3, GABRA4, GABRA5, GABRB1, GABRB3, GABRG2, GABRG3, GABRP</i>
		calcium-activated potassium channel activity	3	3.33E-07	2.97	17	<i>DKK1, KCNMA1, KCNMB2, KCNN2, KCNU1</i>
		ionotropic glutamate receptor activity	4	9.06E-06	2.38	18	<i>GRIA1, GRIA2, GRIK, GRIN2A, GRIN2B, GRIN3A</i>
		structural constituent of eye lens	5	9.61E-06	3.53	19	<i>BFSP1, BFSP2, CRYBA4, CRYBB3, CRYGA, CRYGD</i>
	Biological process	protein localization to synapse	1	7.86E-09	4.15	14	<i>ASIC2, CEP112, LRRTM1, NLGN1, NRXN1</i>
		regulation of sarcomere organization	2	1.31E-07	3.69	8	<i>BMP10, EDN1, MEF2A, MEF2C, MYLK3, PROX1</i>
		cardiac muscle cell development	3	1.81E-07	2.35	45	<i>AGTR2, BMP10, CDK1, CXADR, MEF2A, MEF2C, MYH10, MYLK3, MYO18B, MYOCD, NEBL, PROX1, SLC8A1, TBX3, VEGFA</i>
		regulation of excitatory postsynaptic membrane potential	4	2.02E-07	2.36	40	<i>ATXN1, BDNF, CELF4, GRIK1, GRIN2A, GRIN2B, NLGN1, NLGN4X, NRXN1, SNCA</i>
		determination of left/right asymmetry	5	2.36E-07	2.46	54	<i>ARL6, CITED2, KIF3B, MEF2A, MEF2C, MIB1, MKKS, NME7, RFX3, SHH, SOX17, T, TBX3, TGIF, WNTSA, ZIC3</i>
	Disease ontology	thyroiditis	1	2.97E-06	2.52	54	<i>CTLA4, EDN1, FAS, FASLG, GC, IL6, MBL2, PDGFC, PTGS2, PTPRC, SLC26A4, TG, THRIB, TPO, VEGFA</i>

		calcium metabolism disease	2	5.49E-05	4.03	20	<i>CASR, GALNT3, HLF</i>
		alopecia areata	3	2.17E-04	2.65	25	<i>CCL2, CTLA4, FAS, FASLG, FCGR1A, FCGR1B</i>
		Angelman syndrome	4	3.14E-04	4.23	8	<i>ASPM, GABRA1, GABRA3, GABRA5, SLC9A6, UBE3A</i>
		aggressive systemic mastocytosis	5	4.34E-04	2.31	28	<i>BDNF, CHRM3, CXCR3, IL6, NTRK3, PIK3CG, TLR4, VEGFA</i>

Supplementary Table 11. Enriched pathways (FDR < 0.05) for the ontology “Biological Process” amongst AD-hyperacetylated H3K27ac peaks.

See attached xlsx file SupplementaryTable11_GOBiologicalProcess_HyperacetylatedPeaks.xlsx

Supplementary Table 12. Enriched pathways (FDR < 0.05) for the ontology “Molecular Function” amongst AD-hyperacetylated H3K27ac peaks.

See attached xlsx file SupplementaryTable12_GOMolecularFunction_HyperacetylatedPeaks.xlsx

Supplementary Table 13. Enriched pathways (FDR < 0.05) for the ontology “Disease Ontology” amongst AD-hyperacetylated H3K27ac peaks.

See attached xlsx file SupplementaryTable13_GODiseaseOntology_HyperacetylatedPeaks.xlsx

Supplementary Table 14. Enriched pathways (FDR < 0.05) for the ontology “Biological Process” amongst AD-hypoacetylated H3K27ac peaks.

See attached xlsx file SupplementaryTable14_GOBiologicalProcess_HypoacetylatedPeaks.xlsx

Supplementary Table 15. Enriched pathways (FDR < 0.05) for the ontology “Molecular Function” amongst AD-hypoacetylated H3K27ac peaks.

See attached xlsx file SupplementaryTable15_GOMolecularFunction_HypoacetylatedPeaks.xlsx

Supplementary Table 16. Enriched pathways (FDR < 0.05) for the ontology “Disease Ontology” amongst AD-hypoacetylated H3K27ac peaks.

See attached xlsx file SupplementaryTable16_GODiseaseOntology_HypoacetylatedPeaks.xlsx

Supplementary Table 17. Expression differences between AD cases and controls for genes annotated to top differentially acetylated peaks.

Gene	Beta coefficient (Delta Ct)	P
<i>RGCC</i>	0.440	0.002
<i>PIM3</i>	0.560	1.16E-04
<i>ANKRD17</i>	0.044	0.622
<i>ZNF680</i>	0.147	0.344
<i>GPR22</i>	-0.689	0.006
<i>KMO</i>	-0.994	0.004

Supplementary Table 18. Expression differences between AD cases and controls for genes previously implicated in tau and amyloid pathology or from GWAS.

Gene	Beta coefficient (Delta Ct)	P
<i>MAPT</i>	-0.021	0.844
<i>PSEN1</i>	0.441	0.001
<i>PSEN2</i>	0.011	0.913
<i>CR1</i>	0.703	0.001
<i>TOMM40</i>	-0.082	0.329
<i>APP</i>	-0.212	0.087

Supplementary Table 19. Correlation statistics for H3K27ac and DNA methylation at the 439 significantly correlated probe-peak pairs (FDR < 0.05).

See attached xlsx file SupplementaryTable19_AcetylationMethylation_SignificantAssociations.xlsx

Supplementary Table 20. Twelve significant associations between H3K27ac and 5mC involve an AD-associated differentially acetylated peak.
 Shown are the peak and CpG locations, P values for AD case-control differences in H3K27ac (P.peak) and DNA methylation (P.CpG), as well as the correlation of acetylation and methylation (Corr) with P value (P.Corr) for each probe-peak pair.

CHR	Peak location	P.peak	CpG	CpG Location	P.CpG	Corr	P.Corr	Associated Genes
4	1044452-1044737	0.001	cg04106633	1044584	0.028	-0.71	1.36E-07	<i>FGFRL1</i> (+40828); <i>RNF212</i> (+62719)
			cg21130718	1044621	0.205	-0.70	2.98E-07	<i>FGFRL1</i> (+40828); <i>RNF212</i> (+62719)
			cg04016957	1044486	0.421	-0.66	2.30E-06	<i>FGFRL1</i> (+40828); <i>RNF212</i> (+62719)
4	186731733-186733795	3.31E-04	cg19358738	186734530	0.017	-0.59	3.43E-05	
2	91846937-91848659	2.62E-04	cg14828182	91847976	0.613	-0.62	9.94E-06	<i>CRYBA2</i> (-3439)
2	219860440-219862707	0.001	cg03817727	219861653	0.835	-0.61	1.77E-05	<i>PDLIM3</i> (-276102); <i>SORBS2</i> (+145042)
7	2291502-2292303	0.001	cg08027265	2291960	0.043	-0.71	1.66E-07	<i>NUDT1</i> (+10046); <i>SNX8</i> (+62196)
12	10281560-10282904	2.07E-04	cg21250433	10283763	0.080	-0.59	3.93E-05	<i>CLEC7A</i> (+604)
12	108522589-108524396	1.91E-04	cg26373942	108523463	0.026	0.59	4.00E-05	<i>WSCD2</i> (-2035)
14	78869433-78871145	6.69E-05	cg10828316	78869352	0.030	-0.60	2.24E-05	<i>NRXN3</i> (+196)
17	58179068-58180925	3.77E-04	cg24086329	58180849	0.035	-0.58	6.44E-05	<i>RNFT1</i> (-137917); <i>CA4</i> (-47300)
2	17721932-17725300	1.51E-04	cg20176388	17724532	0.056	-0.61	2.20E-05	<i>VSNL1</i> (+1189); <i>SMC6</i> (+211454)

Supplementary Table 21. Pre-optimized TaqMan assays for targeted gene expression analysis (Life Technologies).

Gene Symbol	Gene Name	TaqMan Assay ID
<i>ACTB</i>	actin beta	Hs99999903_m1
<i>ANKRD17</i>	ankyrin repeat domain 17	Hs00289705_m1
<i>APP</i>	amyloid precursor protein	Hs00169098_m1
<i>CD34</i>	CD34 molecule	Hs02576480_m1
<i>CD68</i>	CD68 molecule	Hs00154355_m1
<i>CR1</i>	complement component 3b/4b receptor 1	Hs00559348_m1
<i>EIF4A2</i>	eukaryotic translation initiation factor 4A2	Hs00756996_g1
<i>ENO2</i>	enolase 2	Hs00157360_m1
<i>GAPDH</i>	glyceraldehyde-3-phosphate dehydrogenase	Hs99999905_m1
<i>GFAP</i>	glial fibrillary acidic protein	Hs00909233_m1
<i>GPR22</i>	G protein-coupled receptor 22	Hs01127309_m1
<i>KMO</i>	kynurene 3-monooxygenase (kynurene 3-hydroxylase)	Hs00175738_m1
<i>MAPT</i>	microtubule associated protein tau	Hs00902194_m1
<i>OLIG2</i>	oligodendrocyte lineage transcription factor 2	Hs00377820_m1
<i>PIM3</i>	Pim-3 proto-oncogene, serine/threonine kinase	Hs00420511_g1
<i>PSEN1</i>	presenilin 1	Hs00997789_m1
<i>PSEN2</i>	presenilin 2	Hs01577197_m1
<i>RGCC</i>	regulator of cell cycle	Hs00204129_m1
<i>SF3A1</i>	splicing factor 3a subunit 1	Hs01066327_m1
<i>TOMM40</i>	translocase of outer mitochondrial membrane 40	Hs01587378_mH
<i>UBC</i>	ubiquitin C	Hs00824723_m1
<i>ZNF680</i>	zinc finger protein 680	Hs01691129_g1

References

- Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215-216, doi:10.1038/nmeth.1906 (2012).
- Quintivano, 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).
- 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).
- McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nature biotechnology* **28**, 495-501, doi:10.1038/nbt.1630 (2010).
- Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).
- 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).