PM20D1 is a quantitative trait locus associated with Alzheimer's disease

Jose V. Sanchez-Mut [1], Holger Heyn [2,3], Bianca A. Silva[1], Lucie Dixsaut[1], Paula Garcia-Esparcia[4], Enrique Vidal [5], Sergi Sayols [6], Liliane Glauser[1], Ana Monteagudo-Sánchez [7], Jordi Perez-Tur [8], Isidre Ferrer[4], David Monk[7], Bernard Schneider [9], Manel Esteller[10] and Johannes Gräff[1]*

The chances to develop Alzheimer's disease (AD) result from a combination of genetic and non-genetic risk factors[1], the latter likely being mediated by epigenetic mechanisms[2]. In the past, genome-wide association studies (GWAS) have identified an important number of risk loci associated with AD pathology[3], but a causal relationship remains difficult to establish. In contrast, locus-specific or epigenome-wide association studies (EWAS) have revealed site-specific epigenetic alterations, which provide mechanistic insights for a particular risk gene but often lack the statistical power of GWAS[4]. Here, combining both approaches, we report a previously unidentified association of the peptidase M20-domain-containing protein 1 (*PM20D1*) with AD. We find that *PM20D1* is a methylation and expression quantitative trait locus coupled to an AD-risk associated haplotype, which displays enhancer-like characteristics and contacts the *PM20D1* promoter via a haplotype-dependent, CCCTC-binding-factor-mediated chromatin loop. Furthermore, PM20D1 is increased following AD-related neurotoxic insults at symptomatic stages in the APP/PS1 mouse model of AD and in human patients with AD who are carriers of the non-risk haplotype. In line, genetically increasing or decreasing the expression of PM20D1 reduces and aggravates AD-related pathologies, respectively. These findings suggest that in a particular genetic background, PM20D1 contributes to neuroprotection against AD.

Over the last decade, the number of epigenomic studies for AD has rapidly increased, which has been paralleled by the discovery of an increasing number of epigenetically dysregulated genes[5]. Despite that, only few such genes have been reported by independent studies in humans (i.e., ankyrin 1 (*ANK1*)[6, 7], sorbin and SH3-domain-containing 3 (*SORBS3*)[8, 9] and histone deacetylase 2 (*HDAC2*)[10, 11]). Contributing to this discrepancy is the complex heterogeneity of the nervous system, in which intra-individual epigenetic, as well as inter-individual genetic and epigenetic variability, call for more refined and integrative studies[5]. By combining epigenetic with genetic approaches, it has recently become possible to identify single-nucleotide polymorphisms (SNPs) that correlate with alterations in DNA methylation levels—so-called methylation

quantitative trait loci (mQTLs)—the importance of which has just started to be recognized for complex diseases[4]. mQTLs have been reported for several neurological disorders, including schizophrenia[12], obsessive–compulsive[13] and bipolar disorders[14], but (as of now) not for neurodegenerative diseases such as AD.

In two recent EWAS[15, 16] we investigated DNA methylome changes in a panel of samples from patients with neurodegenerative diseases, including Parkinson's disease (PD), Lewy body dementia, as well as AD at different disease stages, and in two major affected brain areas, the hippocampus and the frontal cortex. When we re-analyzed these biologically and technically independent datasets by focusing exclusively on the comparison between samples from healthy controls and patients with advanced-stage AD, we noted that only one gene, *PM20D1*, which was previously described as a mQTL[17], consistently displayed promoter hypermethylation in patients with AD (hereafter referred to as AD samples) than in healthy controls in both studies (Fig. 1a and Supplementary Tables 1 and 2). Confirming this observation, a meta-analysis of publicly available DNA methylation datasets (GSE45775, GSE57361, GSE59685, GSE76105 and GSE80970) found an enrichment of *PM20D1* hypermethylation in AD samples (Fig. 1b,c).

To further investigate the association between the *PM20D1* mQTL and AD, we first selected SNPs that significantly correlated with *PM20D1* DNA methylation, by using a recently described mQTL mapping tool[12] (Supplementary Table 3), and examined their relation with AD by using a publicly available database from a National Center for Biotechnology Information (NCBI) GWAS (phs000168.v1.p1, National Institute on Aging (NIA)). We detected an allele-dose-dependent association with AD for most *PM20D1* mQTL-associated SNPs (Supplementary Table 3). The highest significance was observed for SNPs rs708727 ($P < 0.0001$) and rs960603 ($P = 0.0005$), which co-segregated as a haplotype in nearly 85% of cases according to the 1000 Genomes project (phase-3-phased haplotypes for CEU population (Utah residents with Northern and Western European ancestry); rs708727 A frequency 0.40; rs960603 A frequency 0.45)[18]. Because the *PM20D1* region is poorly represented and in low linkage disequilibrium with the SNPs interrogated in most of the GWAS arrays, which can differ between

[1]Laboratory of Neuroepigenetics, Brain Mind Institute, Faculty of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. [2]Centro Nacional de Análisis Genómico (CNAG), Center for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain. [3]Universitat Pompeu Fabra (UPF), Barcelona, Spain. [4]Department of Pathology and Experimental Therapeutics, University of Barcelona, Institut d'Investigació Biomedica de Bellvitge (IDIBELL), Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Barcelona, Spain. [5]Center for Genomic Regulation, Barcelona Institute for Science and Technology, Barcelona, Spain. [6]Bioinformatics Core Facility, Institute of Molecular Biology, Mainz, Germany. [7]Genomic Imprinting and Cancer Group, Cancer Epigenetics and Biology Program, Institut d'Investigació Biomedica de Bellvitge, Barcelona, Spain. [8]Institut de Biomedicina de València–CSIC, Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas, Instituto de Investigación Sanitaria 'La Fe', València, Spain. [9]Neurodegenerative Studies Laboratory, Brain Mind Institute, Faculty of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. [10]Cancer Epigenetics Group, Cancer Epigenetics and Biology Program, Institut d'Investigació Biomedica de Bellvitge, Barcelona, Spain. *e-mail: johannes.graeff@epfl.ch
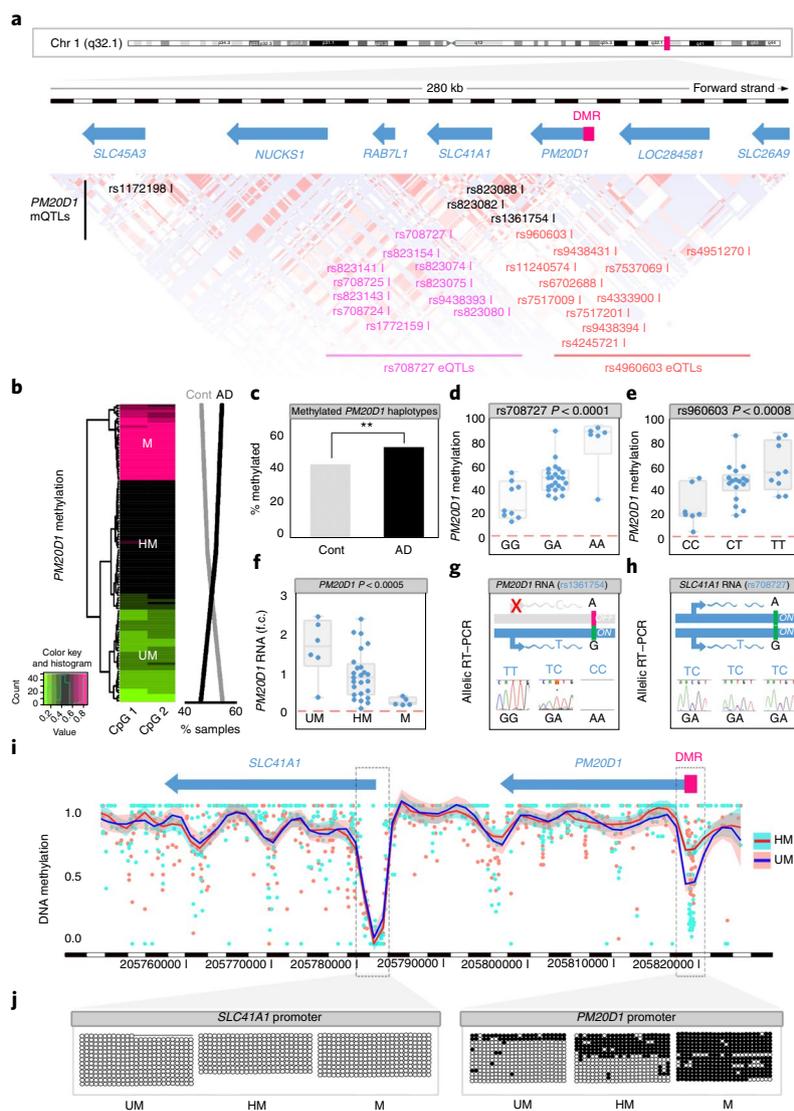
Fig. 1 | Overview of the *PM20D1* locus and its relation to AD in human frontal cortex. a, The region comprises several genes (blue arrows) and is in partial linkage disequilibrium (HapMap CEU Recombination map is indicated below the genes). *PM20D1* mQTLs reflected in Supplementary Table 3 are indicated in black, of which the mQTLs most strongly associated with AD are in bold magenta (rs708727) and bold red (rs960603). The top ten eQTLs in linkage disequilibrium with rs708727 (magenta) and rs960603 (red) obtained from GTEx database (GTEx consortium, neuronal origin tissues—hippocampus and tibial nerve) are also indicated. The differentially methylated region (DMR) on *PM20D1* promoter is represented in magenta. **b**, Left, representative heat map (GSE80970) showing the presence of three main methylation groups in publically available datasets from control and AD IV–VI brain samples using two independent CpG probes—cg1165913 (CpG1) and cg14893161 (CpG2)—targeting sequences in the *PM20D1* promoter. Magenta and green indicate high and low levels of DNA methylation in human frontal cortex samples, respectively. Right, centroid-based clustering (*K*-means) show three main groups— unmethylated (UM), heterozygous methylated (HM) and methylated (M) samples—with a higher prevalence of AD samples (*n* = 197; black line) in the methylated groups than control samples (Cont, *n* = 170; gray line). **c**, Cumulative M haplotype distribution of the data in **b** among control and AD groups (*n* = 734 haplotypes; *P* = 0.0083 by two-sided Chi-square test). **P < 0.01. **d**, rs708727 correlates with the levels of *PM20D1* DNA methylation in human frontal cortex, as measured by pyrosequencing (*n* = 38 biologically independent human brain samples). Data are presented as box plots with minimum, first quartile, median, third quartile and maximum. Single values are represented by circles. *P* < 0.0001 by Pearson correlation. **e**, rs960603 correlates with the levels of *PM20D1* DNA methylation in human frontal cortex, as measured by pyrosequencing (*n* = 34 biologically independent human brain samples). Data are presented as box plots with minimum, first quartile, median, third quartile and maximum. Single values are represented by circles. *P* = 0.0008 by Pearson correlation. **f**, *PM20D1* RNA expression is inversely correlated with *PM20D1* promoter DNA methylation in human frontal cortex samples, as measured by quantitative real-time PCR and pyrosequencing, respectively (*n* = 36 biologically independent human brain samples). Data are presented as box plots with minimum, first quartile, median, third quartile and maximum. Single values are represented by circles. *P* = 0.0005 by Pearson correlation. **g**, rs1361754 is located in the coding region of *PM20D1* and is in linkage disequilibrium with rs708727. RT–PCR Sanger sequencing of rs1361754 heterozygous samples (*n* = 3) detected rs1361754 (T) but not rs1361754 (C) RNA transcripts, indicating that only the rs708727 (G) chromosome was active. **h**, rs708727 is located in the coding region of *SLC41A1*. RT–PCR Sanger sequencing of rs708727 heterozygous samples (*n* = 3) detected both alleles in RNA transcripts, indicating that both chromosomes were active. **i**, Representative whole-genome bisulfite sequencing analysis of DNA methylation in one GG (UM: GSM1173772) and one GA (HM: HM GSM1380998) rs707827 human frontal cortex sample. Differences in DNA methylation are restricted to the DMR region detected in the *PM20D1* promoter. Circles represent CpG sequencing reads. Methylation average is represented by smoothed lines, and 95% confidence intervals are represented by shadows. **j**, Locus-specific bisulfite sequencing of *PM20D1* region confirming the whole-genome bisulfite sequencing and pyrosequencing data (*n* = 3).

different populations[19], and impinge on SNP imputation accuracy[20], the rs708727–rs960603 haplotype has likely escaped other AD-related GWAS.

We then assessed the relationship between the rs708727–rs960603 haplotype and *PM20D1* DNA methylation in human frontal cortex brain samples. We observed an allele-dependent correlation with *PM20D1* promoter methylation (Fig. 1d,e), and further, that *PM20D1* expression was inversely correlated with the methylation of its promoter (Fig. 1f). Notably, *PM20D1* expression was only detected from rs708727–rs960603 G–C chromosomes in heterozygous samples indicating that *PM20D1* was mono-allelically expressed (Fig. 1g). Supporting these observations, rs708727 and rs960603, as well as other mQTL SNPs and the SNPs in linkage disequilibrium with them (Fig. 1a), have previously been described as *PM20D1* expression quantitative trait loci (eQTLs)[21] (Supplementary Table 3). Of note, although rs708727 is located in the coding region of solute carrier family 41 member 1 (*SLC41A1*), rs708727 does not change its protein sequence, and its expression can be detected from both chromosomes (Fig. 1h). Accordingly, *SLC41A1* does not show differences in DNA methylation either (Fig. 1i,j). Lastly, the observed pattern in *PM20D1* methylation between AD and control samples could not be explained by an altered neuron–glia ratio typical for AD, as gray- and white-matter samples from the same samples did not show DNA methylation differences (Supplementary Fig. 1). Taken together, these data point to an association between *PM20D1* hypermethylation and AD and further show that *PM20D1* methylation and expression are dependent on the rs708727–rs960603 haplotype.

Proper control of gene expression often relies on distal cis-regulatory elements that come into physical contact with a gene's promoter region[22]. To better understand the molecular relationship between this rs798727-rs960603 haplotype and *PM20D1* expression, we immortalized B cells from different genetic haplotypes (hereafter referred to as unmethylated (UM) and methylated (M) cells for the GG–CC and AA–TT rs708727–rs960603 samples), which show similar SNP-dependent DNA methylation and expression levels of *PM20D1* as the human brain samples (Supplementary Fig. 2a,b) and conducted chromatin-conformation capture assays (3C). We found that *PM20D1* and its putative regulatory region rs708727-rs960603 interact, but that this interaction was haplotype-dependent, i.e., it was weakened in the *PM20D1* transcriptionally silent M cells (Fig. 2a). Next we used chromatin immunoprecipitation (ChIP) assays to probe for the binding of CCCTC-binding factor (CTCF), which is responsible for more than 90% of DNA loops in mammalian genomes[23]. Consistent with the 3C results, CTCF showed reduced binding to both 3C anchors in M cells (Fig. 2b,c). We also detected increased binding of methyl-CpG-binding protein 2 (MeCP2) and decreased histone acetylation—two hallmarks of a compacted, transcriptionally silent chromatin state in DNA-methylated regions[24]—in the *PM20D1* promoter in M cells (Fig. 2d,e and Supplementary Fig. 2c,d). These epigenetic modifications were functionally important, as treatment with the DNA demethylating agent 5-azacytidine-dC in combination with the HDAC inhibitor valproic acid not only restored the DNA methylation and histone acetylation changes (Supplementary Fig. 3a–d) but also the chromatin loop (Supplementary Fig. 3e) and *PM20D1* RNA expression (Supplementary Fig. 3f).

Recent evidence suggests that disease-associated SNPs often affect the proper functioning of enhancers[25]. Using the Dragon ENhancers database (DENdb)[26], we discovered a predicted enhancer region in proximity to the 3C 5′ locus (Supplementary Fig. 4). Consistent with these in silico data, we found increased levels of the enhancer-enriched epigenetic marks monomethylated Lys4 on histone H3 (H3K4me1) and acetylated Lys27 on histone H3 (H3K27ac)[27] in this region, but only in UM cells, which are not methylated at the *PM20D1* promoter (Fig. 2f,g). Of note,

this finding was reminiscent of cell lines that displayed differential DNA methylation at the *PM20D1* promoter, which also showed an accumulation for H3K4me1 and H3K27ac in the enhancer region[28] (Supplementary Fig. 4). Lastly, to test whether this region has transcription-altering properties, we performed luciferase reporter assays and observed a significant increase in luciferase activity in the presence of the predicted enhancer (Fig. 2h). Taken together, these experiments testify to a haplotype-dependent three-dimensional chromatin interaction between the 3C 5′ locus, with enhancer-like characteristics, and the promoter region of *PM20D1*, which displays loop-dependent differential expression.

On the basis of the newly described differential expression of *PM20D1* according to haplotype and AD risk we showed here, we next investigated whether *PM20D1* might be related to the pathophysiology of AD, by using a combination of in vitro, mouse in vivo and postmortem human studies. In SH-SY5Y neuroblastoma cells, *PM20D1* expression was increased after treatment with AD-related neurotoxic insults, such as reactive oxygen species (ROS) and amyloid-β (Aβ)[1] (Fig. 3a). In the APP/PS1 mouse model[29] of AD, which develops AD-related pathologies such as amyloid plaques, astrogliosis and learning deficits starting at seven months of age (and all of which increase with age)[30], *PM20D1* expression was higher in the frontal cortex at symptomatic stages, as compared to that in the pre-symptomatic stages and in age-matched control mice (Fig. 3b). In human frontal cortex brain samples, we analyzed *PM20D1* expression after correcting for the rs708727–rs960603 haplotype and found that *PM20D1* expression was increased in non-risk haplotype carriers with AD compared to non-risk haplotype carriers without AD (Fig. 3c). Consistent with this observation, a meta-analysis of publicly available RNA expression datasets (GSE33000, GSE15222 and GSE36980) showed an increase in *PM20D1* expression in rs708727–rs960603-haplotype-corrected samples (Supplementary Fig. 5a). In contrast, *PM20D1* expression was decreased in non-haplotype-corrected samples (Supplementary Fig. 5b), which also show an enrichment of *PM20D1*-repressed samples in AD (Supplementary Fig. 5c). Notably, PM20D1 expression, as measured by immunohistochemistry (IHC), was associated with glia-like cells surrounding neuritic plaques in APP/PS1 mice (Supplementary Fig. 6) and in human AD samples (Fig. 3d).

Finally, we genetically manipulated PM20D1 levels by using both virus-mediated *PM20D1* overexpression and antisense oligonucleotides (ASOs) to mimic elevated and reduced PM20D1 levels, respectively. In vitro, we overexpressed *PM20D1* in SH-SY5Y cells and in primary hippocampal cultures from APP/PS1 mice by using lentiviral constructs and assessed ROS-induced cell death and soluble Aβ levels, respectively. *PM20D1* overexpression decreased ROS-induced cell death (Supplementary Fig. 7a), and in primary cultures, in which we found PM20D1 to be actively secreted (Supplementary Fig. 7b), its overexpression reduced Aβ levels as compared to that of control-virus-infected cultures (Supplementary Fig. 7c). In vivo, we overexpressed *PM20D1* by using adeno-associated virus (AAV) vectors that were stereotaxically delivered to the hippocampus and repressed PM20D1 levels by using intracerebroventricular (ICV) injections of ASOs. AAV vectors and ASOs were injected in asymptomatic 3-month-old APP/PS1 mice, and memory performance and amyloid-related pathology were assayed 3 months later by the novel object recognition (NOR) test and by IHC (Fig. 3e). We found that *PM20D1* overexpression (AAV.Mock: $n = 8$; 0.18 ± 0.05 versus AAV.PM20D1: $n = 10$; 544.91 ± 216.48, mean ± s.e.m. for *PM20D1* RNA levels; $P < 0.05$ by one-tailed Student's *t*-test) reduced amyloid plaque load (Fig. 3f,g) and improved NOR performance (Fig. 3h).Conversely, PM20D1 repression (ASO.Scramble: $n = 12$; 1.00 ± 0.11 versus AAV.PM20D1: $n = 12$; 0.76 ± 0.05, mean ± s.e.m. for PM20D1 RNA levels; $P < 0.05$ by one-tailed Student's *t*-test) increased amyloid plaque load (Fig. 3i,j) and impaired NOR performance (Fig. 3k). Taken together, these data describe a
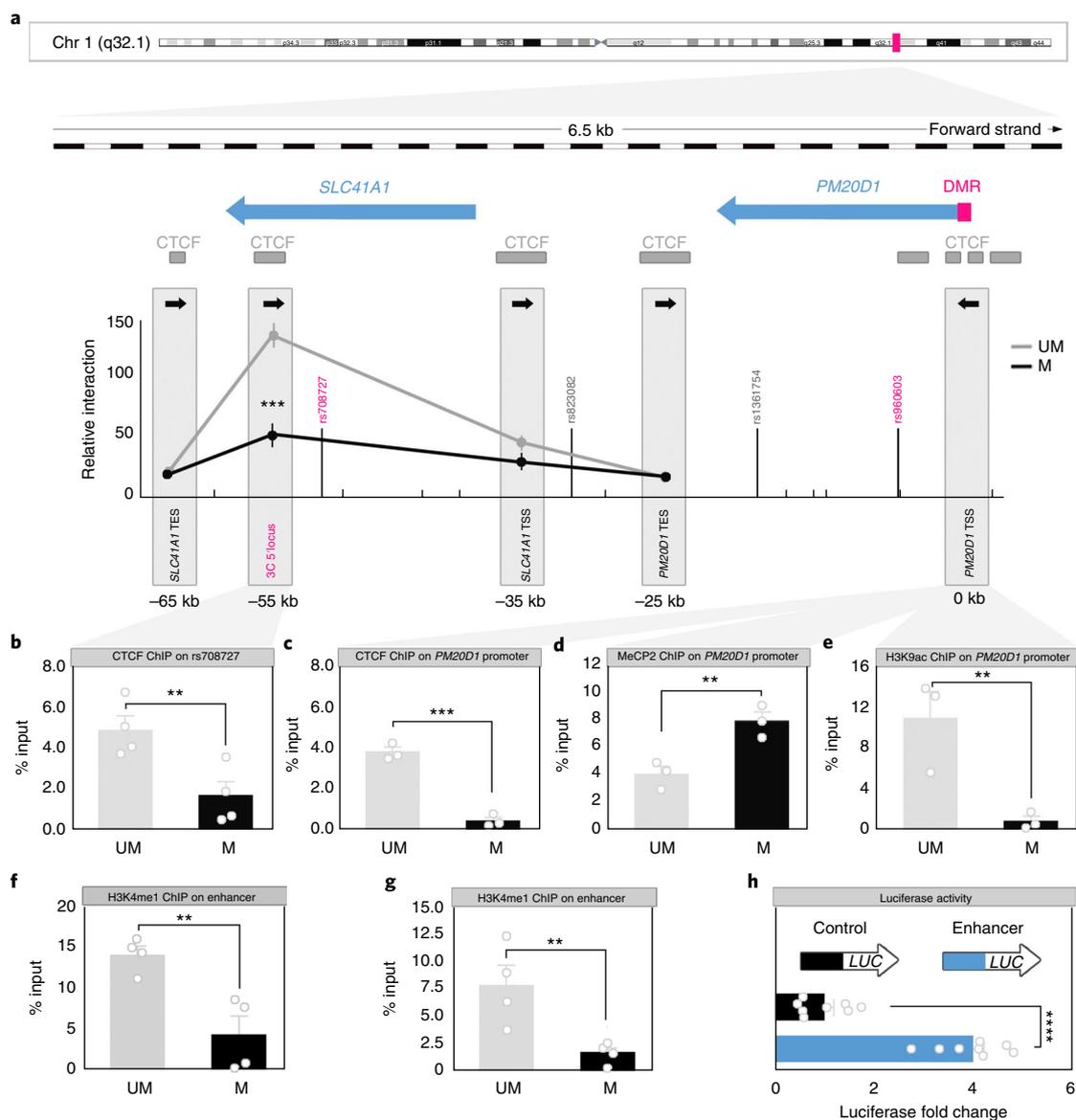
Fig. 2 | Long-range chromatin interaction of the rs708727–PM20D1 locus. a, Genes (blue arrows), *PM20D1* DMR (magenta box) and CTCF-binding sites from UCSC ENCODE tracks (dark gray boxes) are represented. Long-range interactions between CTCF-positive regions have been assayed by 3C (light gray boxes) with different primer combinations (black arrows). The graph represents the measured relative interaction frequencies between regions (anchor or HindIII sites) as a function of the genomic distances (in kb) and methylation profile. Ticks in the *x* axis indicate HindIII target sites. *PM20D1*-related mQTLs' relative positions are also indicated (*n* = 6 independent experiments; *P* = 0.0001 by two-sided Student's *t*-test). TSS, transcription start site; TES, transcription end site. **b**, ChIP assays showing higher binding of CTCF in UM samples on the 3C 5′ locus (*n* = 4 independent experiments; *P* = 0.0143 by one-sided Student's *t*-test). **c**, ChIP assays showing higher binding of CTCF in UM samples on the *PM20D1* promoter region (*n* = 3 independent experiments; *P* = 0.0002 by one-sided Student's *t*-test). **d**, ChIP assays showing higher binding of MeCP2 in M samples on the *PM20D1* promoter region (*n* = 3 independent experiments; *P* = 0.0062 by one-sided Student's *t*-test). **e**, ChIP assays showing that the *PM20D1* promoter region of M samples is depleted of H3K9ac (*n* = 3 independent experiments; *P* = 0.0097 by one-sided Student's *t*-test). **f**, ChIP assays showing that H3K4me1 is enriched in UM samples in the 3C 5′ locus (*n* = 4 independent experiments; *P* = 0.0037 by one-sided Student's *t*-test). **g**, ChIP assays showing that H3K27ac is enriched in UM samples in the 3C 5′ locus (*n* = 4 independent experiments; *P* = 0.0081 by one-sided Student's *t*-test). **h**, DENdb predicted enhancer region capacity was assayed by cloning the region into a pGL4 promoter plasmid and measuring luciferase activity (*n* = 8 independent experiments; *P* < 0.0001 by one-sided Student's *t*-test). Throughout, 3C, ChIP and luciferase reporter assay data are presented as mean ± s.e.m. (**a**–**h**); single values are represented by circles (**b**–**h**). **P* < 0.05, ***P* < 0.01, ****P* < 0.001 and *****P* < 0.0001.

previously unknown, protective role for PM20D1 in the progression of AD-related pathologies.

In sum, we provide evidence for a mQTL–eQTL centered on *PM20D1* that is implicated in AD. Independent analyses of DNA methylation and RNA expression show a significant enrichment of *PM20D1* repression in AD, which we found to be linked via a

CTCF-mediated chromatin loop to an AD-associated haplotype. Functionally, *PM20D1* expression was stimulated by neurotoxic insults both in vitro and in vivo, whereas its overexpression reduced cell death, decreased Aβ levels and improved cognitive performance. Therefore, elevated levels of PM20D1 might provide a potential cellular defense mechanism for AD non-risk-haplotype carriers
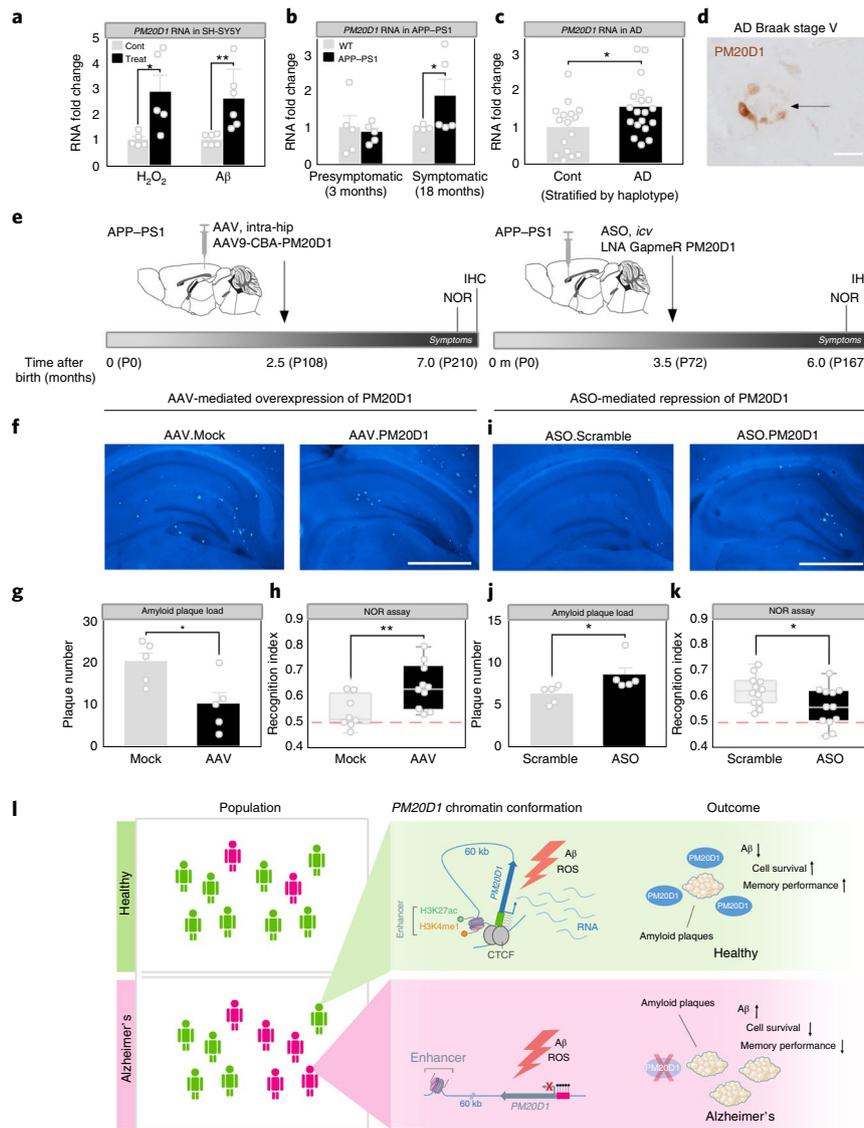
Fig. 3 | Functional relevance of PM20D1 in AD. a, AD-related stressors, such as ROS, produced by H_2O_2 and Aβ treatments increase *PM20D1* RNA expression in SH-SY5Y cells. Data are presented as mean + s.e.m. (H_2O_2: $n = 5$ independent experiments; $P = 0.0274$; Aβ: $n = 6$ independent experiments; $P = 0.0086$ by two-sided Student's *t*-test). **b**, *PM20D1* expression is increased in APP/PS1 mice at the symptomatic stages of pathology ($n = 5$ per genotype and age). Data are presented as means + s.e.m. ($n = 5$ biologically independent mice brain samples; $P = 0.0481$ by one-sided Student's *t*-test). **c**, Haplotype-corrected *PM20D1* RNA levels, showing an upregulation of *PM20D1* expression in frontal cortex of human AD samples. Data represent the mean + s.e.m. of the levels of *PM20D1* RNA expression according to the number of rs708727 G chromosomes of 16 control and 19 AD samples ($P = 0.0148$ by one-sided Student's *t*-test). **d**, Representative IHC image showing PM20D1 localized in the cytoplasm of neurons and glial cells of AD Braak Stage V samples, with glial cells in the vicinity of β-amyloid plaque cores (indicated by the arrow) showing the strongest immunoreactivity ($n = 10$ biologically independent AD Braak Stage V samples; two sections, frontal cortex and hippocampus, per case). Scale bar, 25 μm. **e**, Timeline and schematic of the genetic manipulations of PM20D1 levels using AAV vectors and ASOs. "P" indicates postnatal day. **f**, Representative images showing methoxy-X04 IHC amyloid plaque staining in brains of AAV.Mock-injected and AAV.PM20D1-injected mice. Three slides per mouse and five mice per group were analyzed. Scale bar, 1,000 μm. **g**, Amyloid plaque quantification of samples in **f** ($n = 5$ per group). Data are presented as mean + s.e.m. $P = 0.0123$ by one-sided Student's *t*-test. **h**, NOR test performance of AAV.Mock-injected ($n = 8$) and AAV.PM20D1-injected ($n = 10$) mice. Data are presented as box plots with minimum, first quartile, median, third quartile and maximum. $P = 0.0080$ by one-sided Student's *t*-test. **i**, Representative IHC images for methoxy-X04 amyloid plaque staining of ASO.Scramble-injected and ASO. PM20D1-injected mice. Three slides per mouse and five mice per group were analyzed. Scale bar, 1,000 μm. **j**, Amyloid plaque quantification of samples in **i** ($n = 5$ per group). Data are presented as mean + s.e.m. $P = 0.0296$ by one-sided Student's *t*-test. **k**, NOR test performance of ASO.Scramble-injected and ASO.PM20D1-injected mice ($n = 12$ mice per group). Data are presented as box plots with minimum, first quartile, median, third quartile and maximum. $P = 0.0167$ by one-sided Student's *t*-test. **l**, Model for *PM20D1* function in AD. Left, carriers of methylated *PM20D1* (shown in magenta) are more frequent among people at risk for AD (bottom) than in healthy control subjects (top), in whom *PM20D1* is unmethylated. Right, in samples from individuals with unmethylated *PM20D1* (green box, left; white open circles represent nonmethylated CpG sites), an enhancer region (depicted by the nucleosome) 60 kb downstream of *PM20D1* physically interacts with *PM20D1* promoter via CTCF binding (gray ovals) and favors *PM20D1* transcription. In the presence of AD-related stress (red lightning bolt), such as the presence of ROS and Aβ, PM20D1 expression is enhanced and found to surround amyloid plaques, which reduces ROS-induced cell death and Aβ levels and prevents memory impairment (green box, right). In contrast, in samples with hypermethylated *PM20D1* (red box; left; black circles represent methylated CpG sites), the promoter region of *PM20D1* is not contacted by the enhancer region, *PM20D1* transcription does not occur, and there is no protective effect against Aβ. Throughout, single values are represented by circles (**a**–**c**,**g**–**k**). *$P < 0.05$ and **$P < 0.01$.

(Fig. 3l). Of note, PM20D1 has recently been associated with obesity[31–33] and diabetes[31, 33], which are themselves risk factors for AD[34]. Furthermore, *PM20D1* has been reported to be differentially methylated in individuals with obesity[32] or multiple sclerosis[35], and it lies within the PD 16 (susceptibility) (*PARK16*) locus on chromosome 1, which has previously been associated with Parkinson's disease[36]. Whether *PM20D1* behaves as an epistatic and/or pleiotropic trait across these conditions, as well as its precise mode of action, now require further investigation.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41591-018-0013-y.

References

1. Holtzman, D. M., Morris, J. C. & Goate, A. M. Alzheimer's disease: the challenge of the second century. *Sci. Transl. Med.* **3**, 77sr1 (2011).
2. Chouliaras, L. et al. Epigenetic regulation in the pathophysiology of Alzheimer's disease. *Prog. Neurobiol.* **90**, 498–510 (2010).
3. Scheltens, P. et al. Alzheimer's disease. *Lancet* **388**, 505–517 (2016).
4. Rakyan, V. K., Down, T. A., Balding, D. J. & Beck, S. Epigenome-wide association studies for common human diseases. *Nat. Rev. Genet.* **12**, 529–541 (2011).
5. Sanchez-Mut, J. V. & Gräff, J. Epigenetic alterations in Alzheimer's disease. *Front. Behav. Neurosci.* **9**, 347 (2015).
6. Lunnon, K. et al. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat. Neurosci.* **17**, 1164–1170 (2014).
7. De Jager, P. L. et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat. Neurosci.* **17**, 1156–1163 (2014).
8. Siegmund, K. D. et al. DNA methylation in the human cerebral cortex is dynamically regulated throughout the life span and involves differentiated neurons. *PLoS One* **2**, e895 (2007).
9. Sanchez-Mut, J. V. et al. DNA methylation map of mouse and human brain identifies target genes in Alzheimer's disease. *Brain* **136**, 3018–3027 (2013).
10. Gräff, J. et al. An epigenetic blockade of cognitive functions in the neurodegenerating brain. *Nature* **483**, 222–226 (2012).
11. D'Addario, C. et al. Transcriptional and epigenetic phenomena in peripheral blood cells of monozygotic twins discordant for Alzheimer's disease, a case report. *J. Neurol. Sci.* **372**, 211–216 (2017).
12. Hannon, E. et al. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. *Nat. Neurosci.* **19**, 48–54 (2016).
13. Stewart, S. E. et al. Genome-wide association study of obsessive–compulsive disorder. *Mol. Psychiatry* **18**, 788–798 (2013).
14. Gamazon, E. R. et al. Enrichment of cis-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. *Mol. Psychiatry* **18**, 340–346 (2013).
15. Sanchez-Mut, J. V. et al. Promoter hypermethylation of the phosphatase DUSP22 mediates PKA-dependent TAU phosphorylation and CREB activation in Alzheimer's disease. *Hippocampus* **24**, 363–368 (2014).
16. Sanchez-Mut, J. V. et al. Human DNA methylomes of neurodegenerative diseases show common epigenomic patterns. *Transl. Psychiatry* **6**, e718 (2016).
17. Heyn, H. et al. DNA methylation contributes to natural human variation. *Genome Res.* **23**, 1363–1372 (2013).
18. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
19. Li, H., Teo, Y. Y. & Tan, E. K. Patterns of linkage disequilibrium at *PARK16* may explain variances in genetic-association studies. *Mov. Disord.* **30**, 1335–1342 (2015).
20. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11**, 499–511 (2010).
21. GTEx Consortium. Human genomics. The Genotype–Tissue Expression (GTEx) pilot analysis: multi-tissue gene regulation in humans. *Science* **348**, 648–660 (2015).
22. Bulger, M. & Groudine, M. Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144**, 327–339 (2011).
23. Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194–1211 (2009).
24. Suzuki, M., Yamada, T., Kihara-Negishi, F., Sakurai, T. & Oikawa, T. Direct association between PU.1 and MeCP2 that recruits mSin3A–HDAC complex for PU.1-mediated transcriptional repression. *Oncogene* **22**, 8688–8698 (2003).
25. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
26. Ashoor, H., Kleftogiannis, D., Radovanovic, A. & Bajic, V. B. DENdb: database of integrated human enhancers. *Database (Oxford)* **2015**, bav085 (2015).
27. Smith, E. & Shilatifard, A. Enhancer biology and enhanceropathies. *Nat. Struct. Mol. Biol.* **21**, 210–219 (2014).
28. Creyghton, M. P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* **107**, 21931–21936 (2010).
29. Borchelt, D. R. et al. Accelerated amyloid deposition in the brains of transgenic mice coexpressing mutant presenilin 1 and amyloid precursor proteins. *Neuron* **19**, 939–945 (1997).
30. Jankowsky, J. L. et al. Mutant presenilins specifically elevate the levels of the 42 residue β-amyloid peptide in vivo: evidence for augmentation of a 42-specific γ-secretase. *Hum. Mol. Genet.* **13**, 159–170 (2004).
31. Long, J. Z. et al. The secreted enzyme PM20D1 regulates lipidated amino acid uncouplers of mitochondria. *Cell* **166**, 424–435 (2016).
32. Feinberg, A. P. et al. Personalized epigenomic signatures that are stable over time and co-vary with body mass index. *Sci. Transl. Med.* **2**, 49ra67 (2010).
33. Larrick, J. W., Larrick, J. W. & Mendelsohn, A. R. Uncoupling mitochondrial respiration for diabesity. *Rejuvenation Res.* **19**, 337–340 (2016).
34. Kivipelto, M. & Mangialasche, F. Alzheimer's disease: to what extent can Alzheimer's disease be prevented? *Nat. Rev. Neurol.* **10**, 552–553 (2014).
35. Maltby, V. E. et al. Differential methylation at *MHC* in CD4+ T cells is associated with multiple sclerosis independently of *HLA-DRB1*. *Clin. Epigenetics* **9**, 71 (2017).
36. Simón-Sánchez, J. et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat. Genet.* **41**, 1308–1312 (2009).

Acknowledgements

Author contributions

J.V.S.-M., H.H., M.E. and J.G. conceived the project and designed the experiments; J.V.S.-M., B.A.S., L.D., P.G.-E., L.G., A.M.-S. and D.M. performed the experiments; J.V.S.-M., E.V. and S.S. performed the bioinformatics analysis of the data; J.P.-T., I.F., B.S., D.M. and M.E. contributed to the interpretation of the results; and J.V.S.-M. and J.G. wrote the paper, with input and comments by all of the authors.

Competing interests

A provisional patent application has been filed on the use of PM20D1 methylation and haplotype as biomarkers for Alzheimer's disease (International Application Number PCT/EP2017/067848), with J.V.S.-M., H.H., M.E. and J.G. listed as inventors.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41591-018-0013-y.

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

Correspondence and requests for materials should be addressed to J.G.

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

Methods

EWAS data. EWAS public data were obtained from the NCBI Gene Expression Omnibus (GEO) database (GSE45775 and GSE57361) and analyzed using R software (http://www.R-project.org), as previously described[15, 16]. Briefly, GSE45775 Infinium HumanMethylation27 BeadArray 27K (Illumina) data were quantile-normalized using the 'lumi' package and analyzed with the 'genefilter' package (Bioconductor). DMRs were defined by two or more probes that consistently reported changes in DNA methylation. GSE57361 whole-genome bisulfite sequencing data were aligned using Bismark software[37] and analyzed with the SAMtools[38], bedtools[39], Tabix[40] and ggplot2 packages. DMRs were identified by seeking regions with more than five consecutive CpG sites that were consistently located outside the 95% confidence interval of the smoothed methylation profile. For the meta-analysis, available NCBI GEO DNA methylation data from control and AD brain samples (GSE45775[15], GSE57361[16], GSE59685[6], GSE76105[41] and GSE80970[27]) were further investigated. Samples were stratified in three segments according to the corresponding methylated (M), heterozygous methylated (HM) and nonmethylated (UM) samples—using the centroid-based clustering (K-means) method—for each of the studies independently. Then the cumulative distribution of the segments between the control and AD groups was analyzed by a Chi-square test. Second-order and two-neighbors smoothing was used to show the differential distribution of methylated samples between the controls and AD groups. Source codes are available upon request.

GWAS data. GWAS public data was obtained from NCBI GEO database (phs000168. v1.p1 NIA "Late-onset AD and National Cell Repository for AD Family Study", consisting of 607 families—1,516 affected and 1,306 unaffected). The GWAS NIALOAD dataset was selected based on the high frequency of methylated SNPs in the population and the direct interrogation of *PM20D1* mQTL SNPs without imputation. Selected mQTL SNPs were extracted using PLIN v1.07 software[42] and the relation with AD by fitting a logistic regression via generalized estimation equation (GEE) in a family data under additive genetic model with phenotype as outcome[43]. Hardy–Weinberg equilibrium and allelic test were performed, as previously reported[44]. A linkage disequilibrium plot was obtained from the UCSC track for HapMap Release 24 CEU recombination map and HapMap Linkage Disequilibrium (phase 2) from Phased Genotypes. Haplotypes were obtained from a phase 3 1000 Genomes CEU population (http://phase3browser.1000genomes.org). Source codes are available upon request.

Brain samples. Postmortem tissues were obtained from the IDIBELL Biobank, which is part of the eBrainNet Europe Bank (http://www.brainnet-europe. org/) 'Network of Excellence' funded by the European Commission in the 6th Framework Program 'Life Science' (LSHM-CT-2004-503039). Informed consent was obtained from all participants. The collection of all samples conformed to the relevant regulations, ethical considerations and legislation, as defined by the European Union. Samples were dissected and characterized for Braak stage[45] before further examination. DNA and RNA from gray matter samples of frontal cortex were extracted for subsequent experiments. Only samples with a RNA integrity number (RIN) >6.5, according to the RNA quality test on Agilent's 2100 bioanalyzer, were included in the study. These filtered samples were DNA and RNA from gray matter of frontal cortex samples (Brodmann area 9) of 22 controls (Braak 0–II; 32% female; age 64 ± 3 years, mean ± s.e.m.) and 23 with AD (Braak V–VI; 43% female; age 77 ± 2 years, mean ± s.e.m.), matched for age and gender.

Cell culture. Immortalized B cells[46] were cultured in RPMI 1640 medium (Gibco) supplemented with 10% FBS (Gibco), 100 µ/ml penicillin (Invitrogen) and 100 mg/ml streptomycin (Invitrogen) at 37 °C in a humidified atmosphere of 5% CO_2. Three independent clones per genotype were used. SH-SY5Y neuroblastoma cells (ATCC) were cultured in Dulbecco's modified Eagle's medium (DMEM; Gibco) supplemented with 20% FBS, 100 µ/ml penicillin and 100 mg/ml streptomycin at 37 °C in a humidified atmosphere of 5% CO_2. SH-SY5Y cells were treated with 0.2% hydrogen peroxide (Merck) and with synthetic Aβ (amino acids 1–42)-derived diffusible ligands (ADDLs) (Abcam), prepared as previously described[47], during 6 h H_2O_2 and 24 h ADDL treatment, respectively.

Mice. APP/PS1 mice were maintained under standard animal housing conditions in a normal 12-h light–dark cycle with ad libitum access to food and water. All animal procedures were conducted according to the EPFL's and Switzerland's guidelines on animal welfare (cantonal animal experimentation authorization numbers VD2875 and VD3169).

Stereotaxic injections were performed using a glass pipette (intraMARK, 10- to 20-µm tip diameter, Blaubrand, injection flow: 0.1 µl/min) that was connected to a syringe and a stereotaxic micromanipulator (Kopf Instruments) in deeply anesthetized 3-month-old APP/PS1 male mice. After injection the capillary was left at the injection site for 5 min before slow withdrawal at 0.01 mm/s. Deep anesthesia was induced by subcutaneous injection of a mixture of fentanyl (0.05 mg/kg, Sintetica), midazolam (5 mg/kg, Actavis) and medetomidin (0.5 mg/kg, Orion Pharma). At the end of the surgical procedure an anesthesia reversal mix containing naloxone (1.2 mg/kg, Swissmedic), flumazenil (0.5 mg/kg, Actavis) and atipamezole (2.5 mg/kg, Orion Pharma) was injected subcutaneously, and the mice

were kept warm for an additional 2 h and administered paracetamol (500 mg per 250 ml) in the drinking water for 5 d.

AAV (serotype 9) vectors encoding human PM20D1 (AAV.PM20D1) or a mock (AAV.Mock; empty vector) were bilaterally injected the AAV-containing saline solution (0.5 µl per side) in the dorsal hippocampus (posterior: –2.00 mm; depth: –2.25 mm; lateral: ±1.50 mm; coordinates adapted from the Paxinos Atlas).

ASO LNA GapmeR (Exiqon) 5′-CGGAGACCGTAGCGAA-3′ (ASO.PM20D1) and ASO LNA GampeR 5′-AACACGTCTATACGC-3′ (ASO.Scramble) were used to repress *PM20D1* expression in APP/PS1 mice. We unilaterally injected 1 µl of the ASO-containing Ca^{2+}- and Mg^{2+}-free PBS solution (2.5 µg of ASO per animal) in the right ventricle (posterior: –0.3 mm; depth: –2.5 mm; lateral: 0.9 mm; coordinates adapted from the Paxinos Atlas). For both AAV and ASO treatments, all of the mice were randomly assigned to the treatment groups and included in the analysis, except for animals with clearly misplaced injection sites, as determined by reporter gene fluorescence and/or needle track localization.

Three months after the surgeries, the mice were handled 5 min/d during five consecutive days the week before the NOR test. On test day 1, mice were habituated for 10 min to the open arena. On the second day, mice were placed for 10 min in the arena with two identical objects, and the time exploring each object was automatically recorded (Ethovision, Noldus). On the third day, one of the two familiar objects was replaced by a novel object. The time exploring each of the objects was recorded, and an object recognition index was calculated as the time spent exploring the novel object divided by the total time spent exploring the two objects ($RI = time_{novel}/(time_{novel} + time_{familiar\ object})$).

DNA methylation. DNA was isolated from postmortem human brain samples by phenol–chloroform extraction and bisulphite-converted using the EZ DNA methylation kit (Zymo Research) according to published modifications[48]. Primers for bisulfite cloning and sequencing were designed using the Methyl Primer Express, version 1.0 (Applied Biosystems), and the PCR products were cloned in the pGEMT-easy sequencing vector (Promega) and Sanger-sequenced as previously described[49]. Primers for pyrosequencing were designed using the PyroMark assay design program, version 2.0.01.15 (Qiagen), and pyrosequencing reactions were analyzed with the PyroMark Q24 System version 2.0.6 (Qiagen), following the manufacturer's instructions. Primers for bisulfite cloning and sequencing and for pyrosequencing are listed in Supplementary Table 4.

RNA expression. Total RNA purification and DNase treatment were performed using TRIzol (Invitrogen) and the Turbo DNA-Free kit (Ambion). RNA was reverse-transcribed by using the Thermoscript RT-PCR system (Invitrogen), and the PCR samples were run using the StepOnePlus Real-Time PCR System (Applied Biosystems) and SYBR Green PCR MasterMaster Mix (Applied Biosystems). PCR efficiencies were calculated by using standard dilutions and the LinReg software[50], and three housekeeping genes (*GUSB*, *RPL38* and *TBP*) were used for normalizing PCR signals. Primers are listed in Supplementary Table 4. For the meta-analysis, available NCBI GEO RNA expression data from frontal cortex brain samples of control individuals and those with AD (GSE33000[51], GSE15222[52] and GSE36980[53]) were further investigated. Samples were stratified in three segments according to the corresponding levels of expression using the centroid-based clustering (K-means) method for each of the studies independently. Then, the cumulative distribution of the segments between the control and AD groups was analyzed by a Chi-square test. z-scores were used for comparing overall and stratified levels of expression between control subjects and those with AD. RNA expression datasets were selected based on the number of samples and the alienated frequency of observed/expected frequencies of alleles.

Chromatin conformation capture (3C) protocol. 3C was carried out as described elsewhere[54]. In brief, cells were cross-linked for 5 min with 2% formaldehyde, blocked for 5 min with 125 mM glycine and lysed using a hypotonic buffer solution (0.25 M sucrose, 25 mM KCl, 5 mM $MgCl_2$, 20 mM Tris-HCl, pH 7.5). Nuclei were resuspended in CutSmart Buffer (New England Biolabs) and permeabilized with 0.2% Triton-X100. Chromatin was digested overnight with HindIII (New England Biolabs) and ligated overnight in diluted conditions with T4 DNA ligase (New England Biolabs). Cross-links were then reversed overnight at 65 °C with 200 mM NaCl and proteinase K treatment. DNA was phenol–chloroform purified, ethanol precipitated, and resuspended in TE. BAC RRP11-219P13 (CHORI) was used for assaying all potential 3C products and primer PCR efficiencies. DNA concentration and efficiency of proximity ligation were used for normalizing signals. Primers for 3C are listed in Supplementary Table 4.

Chromatin immunoprecipitation (ChIP) assays. ChIP experiments were carried out as described previously[10]. Briefly, cells were cross-linked for 10 min with 1% formaldehyde and blocked for 5 min with 125 mM glycine; the cells were then lysed by using a cell lysis buffer (5 mM HEPES, 85 mM KCl, 0.5% NP40, pH 8). Nuclei were lysed with a nuclear lysis buffer (50 mM Tris-HCl, 10 mM EDTA, 1% SDS, pH 8) and sonicated to obtain DNA fragments ~300 bp in size (on average) with EpiSonic Multi-Functional Bioprocessor 1100 (EpiGentek). A total of 5–20 µg per sample was used, and an overnight ChIP reaction was performed overnight at 4 °C with antibodies specific for H3K4me1 (Abcam), H3K9ac (Abcam), H3K14ac

(Abcam), H3K27ac (Abcam), H4K12ac (Abcam), CTCF (Millipore) or MeCP2 (Sigma) in diluted conditions. Immunoprecipitated chromatin was then isolated using IgG Dynabeads (Life Sciences) and washed twice each with low-salt buffer (150 mM NaCl, 0.1% SDS, 1% NP-40, 1 mM EDTA, 0.5% sodium deoxycholate, 50 mM Tris-HCl, pH 8), high-salt buffer (500 mM NaCl, 0.1% SDS, 1% NP-40, 1 mM EDTA, 0.5% sodium deoxycholate, 50 mM Tris-HCl, pH 8), lithium buffer (250 mM LiCl, 0.1% SDS, 1% NP-40, 1 mM EDTA, 0.5% Deo-Na, 50 mM Tris-HCl, pH 8) and TE buffer (10 mM Tris-HCl, 0.25 mM EDTA, pH 8). Chromatin complexes were eluted with a solution of 0.1 M $NaHCO_3$ and 1% SDS, and these were de-cross-linked overnight at 65 °C with 200 mM NaCl and proteinase K treatment (Promega). DNA was extracted with phenol–chloroform, precipitated with ethanol and resuspended in 50 μl TE. PCR was carried out in triplicate using SYBR Green PCR MasterMaster Mix (Applied Biosystems). Thermocycling conditions were 10 min at 95 °C, then 50 cycles of 15 s at 95 °C and 1 min at 60 °C. Fluorescence signals were acquired by the StepOnePlus Real-Time PCR System (Applied Biosystems). Values were normalized to total DNA input. Primers for ChIP are listed in Supplementary Table 4.

Cloning. Human *PM20D1* (AK057131.1) was cloned from a library of oligo(dT) cDNA derived from SH-SY5Y neuroblastoma cells. rs708727 enhancer and control regions were cloned from the DNA of immortalized B cells. For *PM20D1*, forward 5′-AAAAGAATTCGCCACCATGGCTCAGCGGTGC-3′ and reverse 5′-CAAGGGATCCAGATGTCGGGAGGAAGGG-3′ primers, containing EcoRI and NotI digestion sites respectively, were used. For the rs708727 enhancer, forward 5′-AGTAGGTACCCTTAGCTTAGCCATGTCATAGCCTTC-3′ and reverse 5′-CGACGCGTCGATATTGGTTTTACTCTCGTATCCTTAAAAAG-3′ primers were used, containing KpnI and MluI digestion sites, respectively. For the enhancer control region—a region near the rs708727 enhancer that is not enriched in active enhancers histone marks—forward 5′- CGACGCGTCGGGAATACTCACGGCACTGGT-3′ and reverse 5′-AGTACTCGAGGGGGCCATTTGCTATGTCACT-3′, containing MluI and XhoI digestion sites, respectively. Transgenes were amplified with high-fidelity Prime STAR Max DNA Polymerase (Takara) according to the manufacturer's instructions. PCR products were cloned in pGEMT-easy (Promega) to amplify the yield. Inserts were released by EcoRI and NotI, and by KpnI and MluI (New England Biolabs), digestions and ligated into pLVX-IRES-ZsGreen1vector (Promega) and pGL4-promoter Luciferase Reporter Vector (Promega), respectively. All cloned plasmids were verified by Sanger sequencing.

Luciferase assays. SH-SY5Y neuroblastoma cells (ATCC) were transfected with pGL4-promoter Luciferase reporter vector (Promega) containing mock (unrelated sequence) or predicted enhancer regions using X-tremeGENE HP DNA transfection reagent (Roche). After 2 d, cells were lysed, and enhancer activity was measured using the Luciferase Assay System (Promega), according to the manufacturer's instructions, and the Tecan's Infinite M200 PRO reader.

Cell viability assays. SH-SY5Y neuroblastoma cells were transfected with pLVX-IRES-ZsGreen1vector (Promega) containing mock (empty vector), or the *PM20D1* gene using the X-tremeGENE HP DNA transfection reagent (Roche). Cells were harvested and seeded in 96-well plates 24 h later and treated with 250 μM H_2O_2 during 24 h the next day. Cell viability was measured using the AlamarBlue cell viability assay (Invitrogen), according to the manufacturer's instructions, and a Tecan's Infinite M200 PRO reader.

Lentivirus production. Lentiviral vectors were produced in HEK-293T cells (ATCC) using a third-generation packaging system by calcium phosphate transfection with the following plasmids: pMD2.G (2.5 μg), psPAX2 (7.5 μg) and pLVX-IRES-ZsGreen1vector containing *PM20D1* or a mock sequence (10 μg). After 4 d, medium was collected and centrifuged in a SW32Ti ultracentrifuge rotor at 19,000 r.p.m. for 90 min at 4 °C. The pellet was resuspended in 120 μl of buffer containing 1× PBS pH 7.4 and 0.5% BSA for a 50× concentrated virus stock. Viral titer was determined using the HIV-1 p24 antigen ELISA kit (Zeptometrix Corp).

Amyloid-β assays. Primary hippocampal neuron–glia co-cultures derived from post natal day 0 APP/PS1 mice[29] were cultured in medium consisting of Neurobasal medium (Invitrogen), B27 supplement (Invitrogen), L-glutamine (Invitrogen) and penicillin–streptomycin (Invitrogen) (0.2 ml per well) on 96-well plates (2.5×10^4 cells/well) coated with Cultrex poly-L-lysine (Trevigen). Cells were infected at day in vitro (DIV) 5 with 10×10^3 to 40×10^3 (100–400 ng/well) viral particles containing either a GFP-expressing (mock) or PM20D1-expressing version of the plasmid. At DIV12, medium from 96 wells was collected, and Aβ levels were measured using the Amyloid-β 40 ELISA Kit (Novex), according to the manufacturer's instructions.

Amyloid plaques. Mouse coronal brain slices (20 μm) were mounted on Superfrost Plus slides coated with Vectabond (Thermo Scientific Laboratories). Sections were fixed in 4% paraformaldehyde for 5 min, washed in PBS and then stained for 15 min with a solution of 10 μg/ml Methoxy-X04 (Abcam) in 50% ethanol. Finally, sections were washed in 50% ethanol and in water, dried and

coverslipped with Fluoromount-G (Southern Biotech). Images were acquired with the EVOS FL digital microscope (Thermo Fisher Scientific) and quantified using ImageJ.

Immunohistochemistry (IHC). Immunohistochemical study of PM20D1 was performed in 4-μm-thick dewaxed paraffin sections of the brain cortex of 18-month-old APP/PS1 mice and AD Braak stage V-VI frontal cortex area 8 of human brains. Tissue sections were boiled in citrate buffer (10mM citric acid, 0.05% Tween 20, pH 6.0) for 20 min to retrieve antigenicity, exposing the epitope for a better antibody recognition. Endogenous peroxidases were blocked with Peroxidase-blocking Solution (Dako) followed by 10% normal goat serum. The primary antibody incubated overnight at 4 °C was the rabbit polyclonal anti-PM20D1 (Sigma-Aldrich). Following this, the sections were incubated with super-sensitive link-label IHC Biogenex Kit system at room temperature. The peroxidase reaction was visualized with diaminobenzidine (DAB) and H_2O_2. The omission of the primary antibody in the same section was used as a control of the immunostaining; no signal was obtained with the incubation of the secondary antibody only (PM20D1 (HPA017571, 1:150, Sigma-Aldrich), HRP-conjugated secondary antibody rabbit (P0448, 1:2,000, Dako; data not shown). Sections were slightly counterstained with hematoxylin.

Western blot analysis. Primary hippocampal cultures derived from wild-type mice were cultured on 35-mm-well plates (2.5×10^5 cells/well) coated with Cultrex poly-L-lysine (Trevigen). Cells were infected at DIV6 with 10×10^4 to 40×10^4 (1–4 μg/well) viral particles containing either a GFP-expressing (mock) or PM20D1-expressing version of the plasmid. At DIV14, cells and medium were collected and, mixed with Laemmli buffer, and the protein concentration was normalized using the DC Protein Assay (Bio-Rad). Equal amounts of protein (30 μg/lane) were separated by SDS–PAGE (10%) and transferred onto nitrocellulose membranes (Amersham, GE Healthcare). Ponceau S staining (Sigma-Aldrich) was used for monitoring protein loading and transference. Nonspecific binding was blocked by incubation in 5% nonfat milk in Tris-buffered saline (100 mM NaCl, 10 mM Tris, pH 7.4) containing 0.2% Tween (TTBS) for 1 h at room temperature. Afterwards, membranes were incubated overnight at 4 °C with the polyclonal rabbit anti-PM20D1 (HPA017571, 1:250; Sigma-Aldrich) in TTBS with 3% BSA. For cell lysates, protein loading was also monitored by using a mouse monoclonal antibody against β-actin (A3854, 1:20,000; Sigma-Aldrich). Membranes were then incubated for 1 h in the appropriate horseradish-peroxidase-conjugated secondary antibodies ((211-032-171, 1:5,000; Jackson ImmunoResearch), and immunocomplexes were revealed by an enhanced chemiluminescence reagent (ECL Advance, Amersham Biosciences). Densitometric quantification was carried out with Fusion FX software (Vilber).

Statistical analysis. Statistical analysis was done using the R package and Prism 6.0 (GraphPad) as described in the figure legends. All experimental data points were included in the analysis, unless they were statistical outliers as determined by Grubb's outlier calculations (GraphPad Prism). All in vitro and in vivo experiments were conducted by experimenters blinded to the experimental conditions and were controlled for their normal distribution.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary.

Data availability. Data from these experiments are available from the corresponding author upon reasonable request. Genetic, DNA methylation and RNA expression data are available at GEO (genetics: phs000168.v1.p1; DNA methylation: GSE45775, GSE57361, GSE59685, GSE76105 and GSE80970; RNA expression: GSE33000, GSE15222 and GSE36980).

References

37. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
38. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
39. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
40. Li, H. Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics* **27**, 718–719 (2011).
41. Watson, C. T. et al. Genome-wide DNA methylation profiling in the superior temporal gyrus reveals epigenetic signatures associated with Alzheimer's disease. *Genome Med.* **8**, 5 (2016).
42. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
43. Chen, M. H. & Yang, Q. GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics* **26**, 580–581 (2010).
44. Guedj, M., Wojcik, J., Della-Chiesa, E., Nuel, G. & Forner, K. A fast, unbiased and exact allelic test for case–control association studies. *Hum. Hered.* **61**, 210–221 (2006).

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

46. Heyn, H. et al. Distinct DNA methylomes of newborns and centenarians. *Proc. Natl. Acad. Sci. USA* **109**, 10522–10527 (2012).

47. Klein, W. L. Aβ toxicity in Alzheimer's disease: globular oligomers (ADDLs) as new vaccine and drug targets. *Neurochem. Int.* **41**, 345–352 (2002).

48. Bibikova, M. et al. Genome-wide DNA methylation profiling using Infinium assay. *Epigenomics* **1**, 177–200 (2009).

49. Fernandez, A. F. et al. The dynamic DNA methylomes of double-stranded DNA viruses associated with human cancer. *Genome Res.* **19**, 438–451 (2009).

50. Ruijter, J. M. et al. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res.* **37**, e45 (2009).

51. Narayanan, M. et al. Common dysregulation network in the human prefrontal cortex underlies two neurodegenerative diseases. *Mol. Syst. Biol.* **10**, 743 (2014).

52. Webster, J. A. et al. Genetic control of human brain transcript expression in Alzheimer's disease. *Am. J. Hum. Genet.* **84**, 445–458 (2009).

53. Hokama, M. et al. Altered expression of diabetes-related genes in Alzheimer's disease brains: the Hisayama study. *Cereb. Cortex* **24**, 2476–2488 (2014).

54. Court, F. et al. Long-range chromatin interactions at the mouse *Igf2–H19* locus reveal a novel paternally expressed long noncoding RNA. *Nucleic Acids Res* **39**, 5893–5906 (2011).

Corresponding author(s):   BC82191B

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

❯ Experimental design

1. Sample size

   Describe how sample size was determined.

   For all animal work, appropriate statistical tests were used and outlined in the materials and methods section, as well as in the figure legend. For sample size estimations we proceeded as follows. For behavioral studies, we estimated to need 12 animals per group to reach statistically meaningful conclusions (significance level of p smaller than 0.05) according to the Guidelines for the Care and Use of Mammals in Neuroscience and Behavioral Research (National Academies Press, Washington (DC), 2003). This estimation is based on
   $n = 1 + 2C \left(s/d \right)^2$
   where C is determined by the power (1-β, here 90%) and the significance level (α, here 5%), the standard deviation (s, here 14% for behavioral tests such as the ones being used based on our own experience) and the difference we would like to detect (d, here 20%).
   For biochemical and imaging analyses, 5 animals per treatment group suffice (Guidelines for the Care and Use of Mammals in Neuroscience and Behavioral Research (National Academies Press, Washington (DC), 2003). This estimation is based on the criteria of whether or not the treated animals will display AD-related pathologies. Thus, the effect size (n) is determined by the probability of committing a Type II error (β) and by p, the proportion of animals that do not show any changes. This proportion is estimated at 70% based on our own previous experience with AD-related pathologies and virus-related interventions.
   $n = \log\beta / \log p$

2. Data exclusions

   Describe any data exclusions.

   Mice with bad injection sites, showing stereotypy, or exploring less than 10 seconds the objects in the novel object recognition test (NOR) were excluded. Also, mice in which no increase of PM20D1 for the AAV.PM20D1 overexpression group or in which no decrease of PM20D1 for the ASO.PM20D1 repression group was observed in hippocampus were excluded. Outliers were determined by Grubb's outlier calculation's.

3. Replication

   Describe the measures taken to verify the reproducibility of the experimental findings.

   All experiments were reproduced three or more times using the same experimental approach (unless otherwise stated). Number of replicates and sample sizes are provided for each of the experiments in figure legends.Findings and conclusions in the manuscript represent those that were consistent across independent experiments.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   All animals were randomly assigned to the treatment groups.

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   All in vitro and in vivo experiments were conducted by experimenters blinded to the experimental conditions.

   Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The <u>exact sample size</u> (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	Test values indicating whether an effect is present *Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.*
☐	☒	A clear description of statistics including <u>central tendency</u> (e.g. median, mean) and <u>variation</u> (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars in <u>all</u> relevant figure captions (with explicit mention of central tendency and variation)

See the web collection on statistics for biologists for further resources and guidance.

▶ Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

> Data was analyzed using R Lumi, Genefilter, Ggplot2, Bismark, SAMtools, Bedtools, Tabix, PLIN v1.07, and Prism 6.0 (GraphPad) software. Source codes are available upon request.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

▶ Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.

> No resctrictions.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

> Immunohistochemical analysis of PM20D1 expression in mice and human samples was performed with the PM20D1 antibody from Sigma-Aldrich (HPA017571; dilution 1:150). The omission of the primary antibody in some section was used as a control of the immunostaining; no signal was obtained with the only incubation of the secondary antibody.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

> American Type Culture Collection (ATCC) SH-SY5Y (CRL-2266).

b. Describe the method of cell line authentication used.

> Only SH-SY5Y cells with less than 40 passages were used. Number of passages, usage, and phenotypic characteristics were traced. No further authentication was performed.

c. Report whether the cell lines were tested for mycoplasma contamination.

> Cell lines were tested and negative for mycoplasma contamination.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

> None of the cell lines used for this study are listed in the ICLAC database.

▶ Animals and human research participants

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.

APP/PS1 mice (JAX mice reference number 005864) were maintained under standard animal housing conditions in a normal 12h light-dark cycle with ad libitum access to food and water. All experiments were conducted exclusively on males.

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

This study do not involve human research participants.