Epigenetic signatures of methylated DNA cytosine in Alzheimer's disease

Irfete S. Fetahu[1,2]*, Dingailu Ma[1,3]*, Kimberlie Rabidou[1], Christian Argueta[1,2], Michael Smith[1], Hang Liu[3], Feizhen Wu[3,4]†, Yujiang G. Shi[1,2]†

Alzheimer's disease (AD), a progressive neurodegenerative disorder, is the most common untreatable form of dementia. Identifying molecular biomarkers that allow early detection remains a key challenge in the diagnosis, treatment, and prognostic evaluation of the disease. Here, we report a novel experimental and analytical model characterizing epigenetic alterations during AD onset and progression. We generated the first integrated base-resolution genome-wide maps of the distribution of 5-methyl-cytosine (5mC), 5-hydroxymethyl-cytosine (5hmC), and 5-formyl/carboxy-cytosine (5fC/caC) in normal and AD neurons. We identified 27 AD region–specific and 39 CpG site–specific epigenetic signatures that were independently validated across our familial and sporadic AD models, and in an independent clinical cohort. Thus, our work establishes a new model and strategy to study the epigenetic alterations underlying AD onset and progression and provides a set of highly reliable AD-specific epigenetic signatures that may have early diagnostic and prognostic implications.

INTRODUCTION

Alzheimer's disease (AD) is clinically characterized by the accumulation of amyloid-β (Aβ) plaques and neurofibrillary tangles, synaptic and neuronal loss, and cognitive decline (1). Mutations in the *amyloid precursor protein* (*APP*), *presenilin 1* (*PSEN1*), or *PSEN2* genes lead to autosomal dominant AD, with disease onset occurring before the age of 65 (2), while the presence of two copies of *apolipoprotein ε4* (*APOE4*) is associated with late-onset AD (onset >60 years) (3, 4). Recently, large-scale genome-wide association studies have identified several additional genes involved with the onset and progression of AD (5); however, the identified genes have not had a meaningful prognostic value for the onset of AD when compared to the above-mentioned genes. Altogether, these genetic studies provide a framework of AD-associated gene networks and a classic panel for genetic screening of familial AD, which accounts for less than 20% of the AD patient population (6). On the other hand, the identification of effective biomarkers for the early detection and diagnosis of sporadic AD, which accounts for the vast majority of AD cases, remains a major challenge.

DNA methylation at the fifth position of cytosine [5-methyl-cytosine (5mC)] plays an important role in neuronal gene expression and neural development. Aberrant DNA methylation is associated with many neuronal disorders, including AD (7–10). 5mC can be further oxidized to 5-hydroxymethyl-cytosine (5hmC), 5-formyl-cytosine (5fC), and 5-carboxy-cytosine (5caC) by the ten-eleven-translocation (TET) family of dioxygenases. The oxidized products of 5mC accumulate during brain development and early adulthood, suggesting an independent and critical role in neuronal development and func-

tion (11, 12). Several studies have suggested that dysregulated DNA methylation/demethylation is linked to AD onset and progression (13–18); however, the relationship between altered levels of 5mC, 5hmC, 5fC, and 5caC and AD is not known. This is primarily due to the lack of comprehensive (base resolution) integrated reference maps of these epigenetic marks during neural cell differentiation, maturation, and brain development. It is also, in part, due to the lack of experimental AD models and research strategies to map, analyze, and define AD-specific changes of the methylome and their oxidized derivatives. Therefore, illustrating the DNA 5mC, 5hmC, and 5fC/caC methylomes of early- or late-onset AD will likely uncover unique and common epigenetic pathways, and signatures that may apply to all AD pathogenesis.

RESULTS

### Description of the study model and subjects

The overview of our experimental paradigm is outlined in Fig. 1. We used new base-resolution mapping and analytical technologies, such as oxidative bisulfite deep sequencing (OXBS-seq) and methylase-assisted bisulfite deep sequencing (MAB-seq), to characterize dynamic genome-wide distributions of 5mC, 5hmC, and 5fC/5caC modifications. We established cell culture models of neurogenesis by directing the differentiation of induced pluripotent stem cells (iPSCs) derived from healthy control [wild type (WT)] and AD patient–derived cells to neural progenitor cells, and then to cortical neurons, as previously described (19). The cell lines include a normal cell line (WT), two early-onset familial AD (EOAD) cell lines (PSEN1 and PSEN2), and a late-onset familial AD (LOAD) cell line (APOE4), which are the strongest genetic risk factors for developing AD (5). AD cell lines either carrying mutations in *PSEN1*, *PSEN2*, or homozygous *APOE4* served as "AD-in-dish" models to mimic EOAD and LOAD, respectively. To ensure generalizability of the epigenetic features and signatures of healthy controls and AD-associated epigenetic alterations found in cell culture models, we extended our epigenomic studies to postmortem brain tissues of a normal donor and sporadic AD cases. Last, we validated the AD-specific 5mC

[1]Division of Endocrinology, Diabetes and Hypertension, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA. [2]Harvard Medical School, Boston, MA 02115, USA. [3]Laboratory of Epigenetics, Institutes of Biomedical Sciences, Shanghai Medical College of Fudan University, Shanghai 200032, China. [4]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: yujiang_shi@hms.harvard.edu (Y.G.S.); wufz@fudan.edu.cn (F.W.)
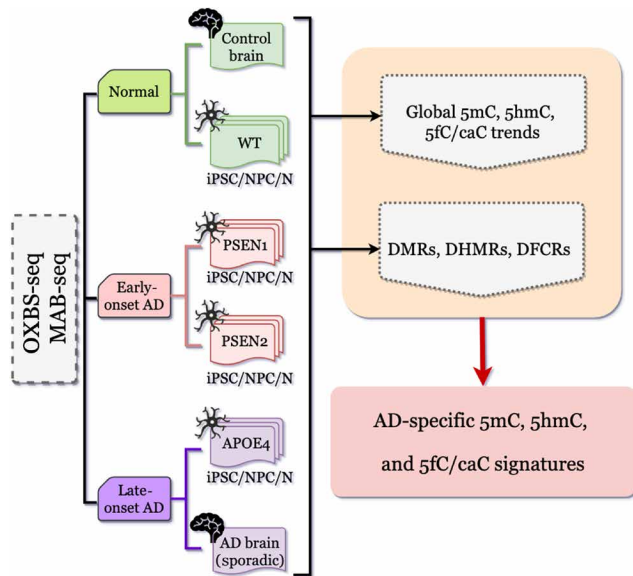
Fig. 1. Schematic of our study model. iPSC, induced pluripotent stem cells; NPC, neural precursor cells; N, neurons; APOE4, apolipoprotein E isoform 4; PSEN1, presenilin 1; PSEN2, presenilin 2; WT, wild type; MAB-seq, methylase-assisted bisulfite sequencing; OXBS-seq, oxidative bisulfite sequencing; DMRs, differentially methylated regions; DHMRs, differentially hydroxymethylated regions; DFCRs, differentially formyl/carboxylated regions.

signatures defined in our study using publicly available normal and AD cohorts.

Identification of 5mC, 5hmC, and 5fC/caC landscapes during neurogenesis

Pluripotency genes, such as *NANOG, OCT4,* and *SOX2,* are expressed in iPSC-WT. Expression of neural precursor cell (NPC) markers, including *FOXG1, NES,* and *TBR2,* confirmed the complete induction of iPSCs to NPCs. Subsequently, successful differentiation of NPCs to neurons was confirmed with the expression of neuronal specific marks, such as *CUX1, MAP2, TBR1,* and *TUJ1* (Fig. 2, A and B). Similarly, all iPSCs derived from AD patients can also be differentiated to NPCs and neurons (fig. S1, A to C).

Until recently, uncovering the exact patterns of 5mC, 5hmC, and 5fC/caC at base resolution has proven to be a challenging task due to the lack of technologies that allow us to distinguish 5mC from its oxidized modifications (5hmC and 5fC/caC) at a given single cytosine base. To circumvent this problem, we developed an in-house protocol and analytical algorithm to identify genome-wide distribution of sites and levels of 5mC, 5hmC, and 5fC/caC based on recently reported OXBS-seq (identifies 5mC and 5hmC, where the combined levels range from 0 to 100%) and MAB-seq (identifies 5fC/caC, where levels range from 0 to 100%) [20, 21].

Globally, the levels of 5mC were higher in both NPCs and neurons (89.6 and 89.3%, respectively; $P < 2.2 \times 10^{-16}$) compared to iPSCs (82.5%), indicating that these 5mC sites tend to be more frequently methylated in the differentiated neural stages than in iPSCs (Fig. 2C, left). These findings are consistent with previously published reports [22, 23]. Levels of 5hmC were lower ($P < 2.2 \times 10^{-16}$) in NPCs (8.3%) and neurons (7.7%) than in iPSCs (9.8%) (Fig. 2C, middle). The levels of 5fC/caC were higher in the iPSCs (36%) than in neurons (34%,

$P < 2.2 \times 10^{-16}$) (Fig. 2C, right). We observed a surge of 5fC/caC level (48%) in NPCs, which is in agreement with previous studies in mice showing that levels of 5fC/5caC transiently increase during NPC stage and rapidly decline as cells commit to neural lineages [24], highlighting the critical roles of this modification in lineage commitment. We should note that due to its technical nature, MAB-seq is unable to distinguish between 5fC and 5caC. Therefore, the identified cytosine modification by MAB-seq reflects the total modification of 5fC/caC. Detailed description and quantification of genome-wide sites of 5mC, 5hmC, and 5fC/caC at base resolution are provided in Supplementary Text and table S1. Region- and gene-specific examples of 5mC, 5hmC, and 5fC/caC at base resolution are shown in Fig. 2D.

Disease and developmental differences in 5mC, 5hmC, and 5fC/caC profiles

Using similar analytical approaches as in WT cells, we have also successfully mapped and analyzed at base resolution 5mC, 5hmC, and 5fC/caC distribution in AD patient–derived cell lines [for technical details and description of each sequencing (SEQ) dataset, see table S1]. Overall, global levels of 5mC and 5hmC in EOAD models and PSEN1 and PSEN2 cell lines followed the same patterns as the WT cells, which were characterized by gain of 5mC and loss of 5hmC in neurons compared to the iPSCs, whereas levels of 5fC/caC decreased in neurons compared to the iPSCs (Fig. 3, A and B). Intriguingly, in both PSEN1 and PSEN2 cell lines, levels of 5fC/caC failed to peak at the NPC stage as opposed to the trend in the WT NPCs. In the LOAD model cell line APOE4, we noted that 5mC levels increased during differentiation; however, the differences between iPSCs and neurons were not as prominent when compared to the WT or EOAD cell lines (Fig. 3C). Changes in 5hmC levels were also marginal between the different stages, while, similar to the EOAD models, 5fC/caC failed to peak in the NPC stage, which appears to be a common feature in all AD cell lines (Fig. 3, A to C). Furthermore, both PSEN1 and PSEN2 cell lines show similar global patterns of 5mC, 5hmC, and 5fC/caC during differentiation of iPSCs to neurons that appear to be independent of *PSEN1* or *PSEN2* gene mutations, but common to the EOAD model. Changes in methylation of cytosine at the base level for 5mC, 5hmC, and 5fC/caC across our different cell line models as well as during the differentiation process in normal and disease settings are depicted in Fig. 3D.

Differentially methylated DNA regions demarcate key AD genes

To determine the loci-specific similarities and differences of epigenetic profiles between WT and AD cells, we next performed more detailed comparative analysis of the 5mC, 5hmC, and 5fC/caC profiles between WT and PSEN2-AD cell models. We focused our initial comparative analysis between WT and PSEN2 lines, as the N141I mutation in *PSEN2* has been well characterized and gives rise to autosomal dominant early-onset AD in patients younger than 65 years old [2]. In both WT and PSEN2, levels of 5mC increased in NPCs and neurons compared to iPSCs (Fig. 4A, top). Gain of 5mC levels was marked by loss of 5hmC levels in both WT and PSEN2 lines during neural differentiation (Fig. 4A, middle). Of note, at any given stage, we observed that both 5mC and 5hmC of PSEN2 were slightly, but significantly, higher ($P < 2.2 \times 10^{-16}$) in iPSCs and neurons, but lower ($P < 2.2 \times 10^{-16}$) in NPCs compared to the WT lines. The dynamic changes of 5fC/caC levels in WT and PSEN2 were
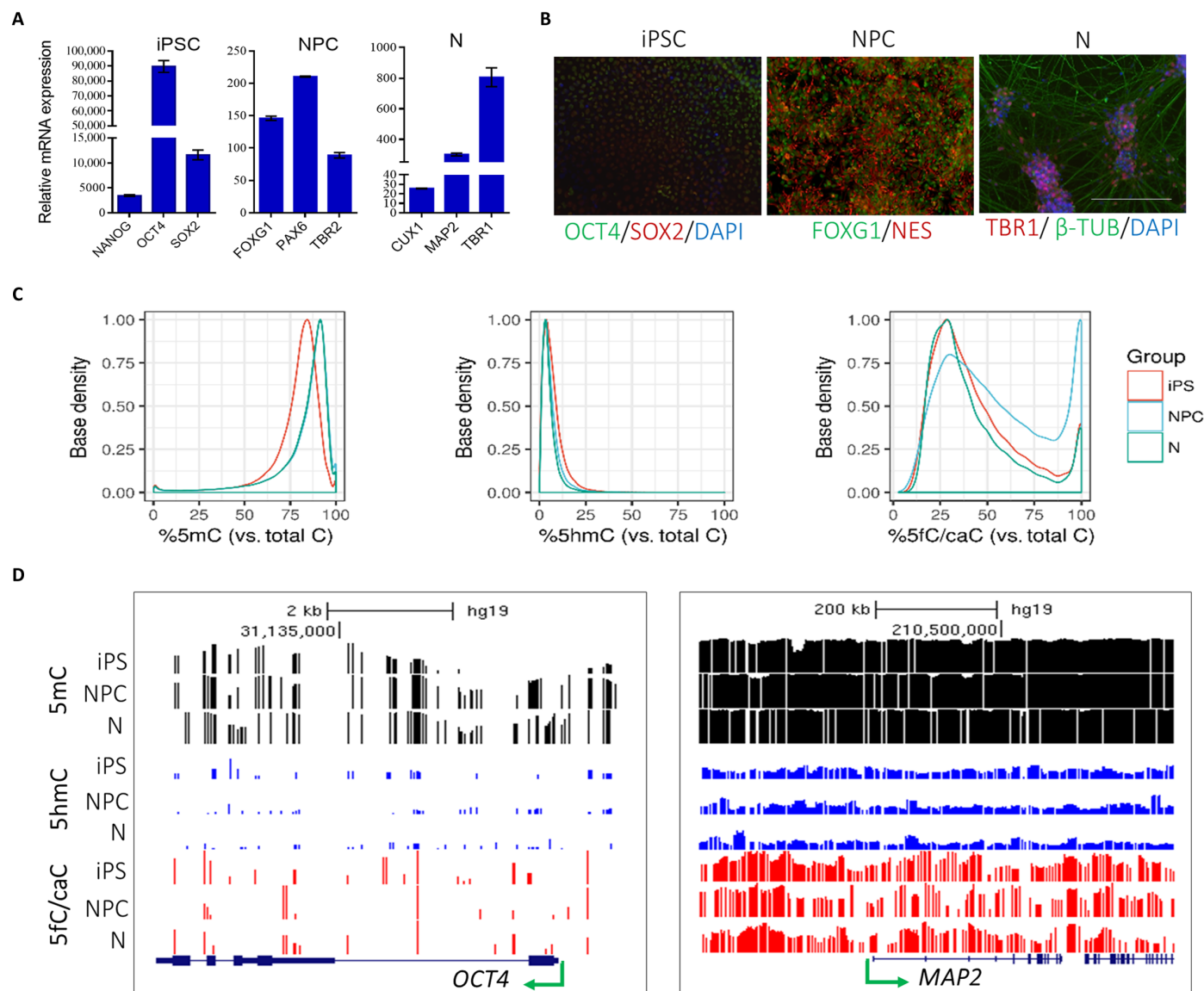
Fig. 2. Global methylation trends in iPSCs and iPSC-derived NPCs and neurons in WT cell lines. (**A** and **B**). Pluripotency, neural stem cell, and neural cell markers determined by quantitative reverse transcription polymerase chain reaction (qRT-PCR) and immunofluorescence. Scale bar, 200 μm. (**C**) Distribution of 5mC/5hmC/5fC/caC levels in WT iPSCs, NPCs, and neurons. The *y* axis indicates the levels of 5mC, 5hmC, or 5fC/caC at each reference cytosine (at least 10 reads required). (**D**) Single-base profiles of 5mC, 5hmC, and 5fC/caC for the pluripotency gene *OCT4* and the neural-specific gene *MAP2* in iPSCs and the consecutive stages of iPSC-derived NPCs and neurons. Genomic coordinates: for *OCT4*, chr6:31,132,042-31,139,528; for *MAP2*, chr2:210,002,875-210,783,411. In all samples, the scale for 5mC, 5hmC, and 5fC/caC was 0 to 100%.

markedly different from those of 5mC and 5hmC during neural differentiation. The peak in 5fC/caC levels in the NPC stage as observed in our WT line appears to be an epigenetic feature that is critical in neurodevelopment (*24*). However, this pattern fails to be established in PSEN2 cells (Fig. 4A, bottom), suggesting an aberrant 5fC/caC methylome in the PSEN2 AD cell lines.

We next analyzed the distribution of 5mC, 5hmC, and 5fC/caC across averaged RefSeq genes in WT and PSEN2 cell lines. Profiles of 5mC in both WT and PSEN2 cells were similar at all differentiation stages, characterized by a large drop around the transcription start site (TSS) and an increase in gene body toward transcription end site (TES), dropping again around TES (Fig. 4B, top row). 5hmC profiles also

showed a dip at TSS, followed by a gradual accumulation after TSS, finally surging at TES at all stages in both cell lines (Fig. 4B, middle row). On the other hand, 5fC/caC profiles are more complex than those of 5mC and 5hmC during differentiation in the two cell lines (Fig. 4B, bottom row). In iPSCs and NPCs, the profiles were characterized by a peak at TSS. Of note, in NPCs, the peak was greatly diminished in PSEN2 cells. In neurons, the peaks observed at TSS in iPSC and NPC disappear and become a dent.

To identify the regions that contain differential levels of 5mC (DMR), 5hmC (DHMR), and 5fC/caC (DFCR) during differentiation (iPSCs to neurons) in both WT and PSEN2 cells, we analyzed stage-specific differences in DMRs, DHMRs, and DFCRs between
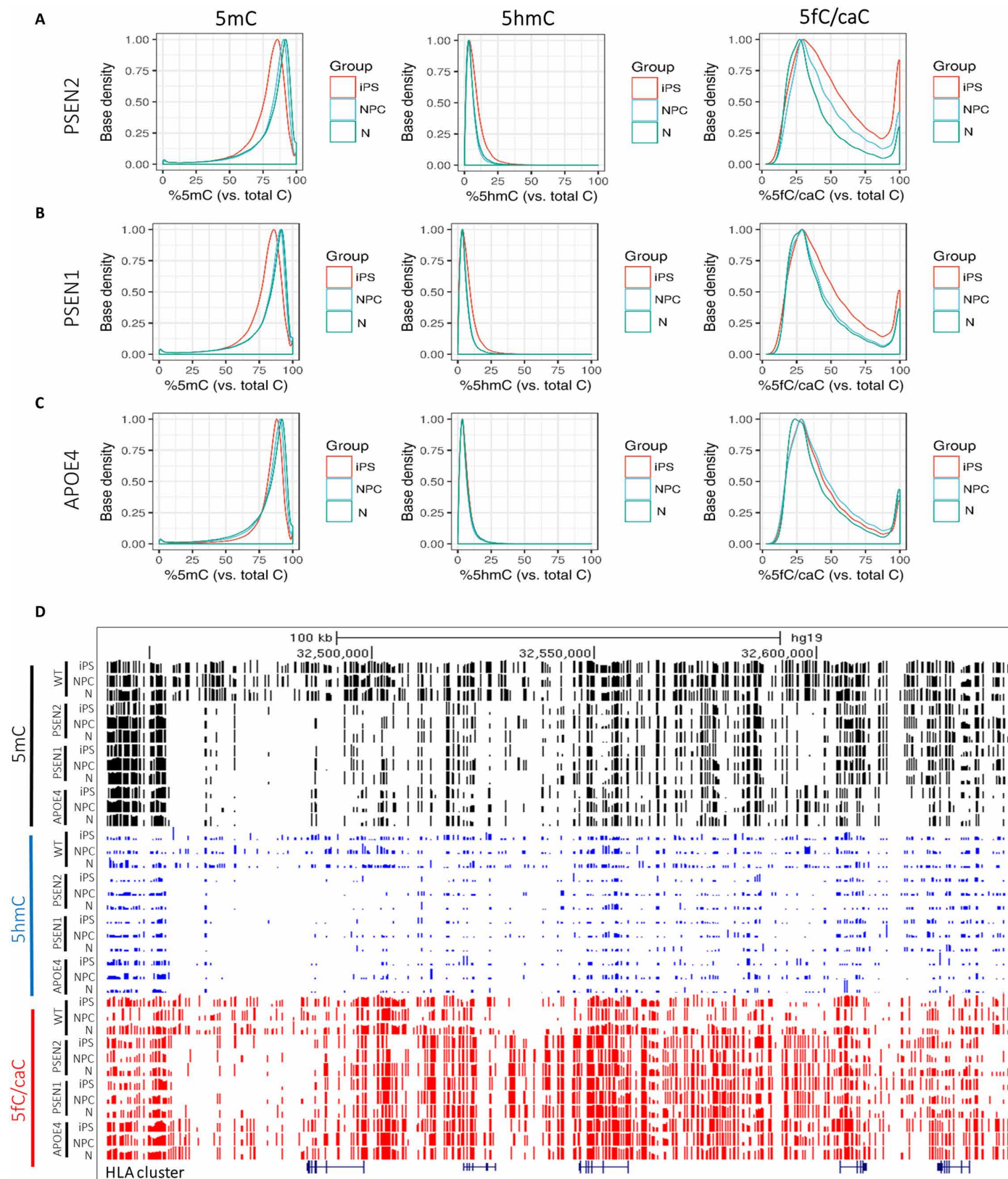
Fig. 3. Disease and developmental differences in WT and AD neurons. (**A** to **C**) Distribution of 5mC/5hmC/5fC/caC levels in PSEN2, PSEN1, and APOE4 patient-derived cell lines. The *y* axis indicates the levels of 5mC, 5hmC, or 5fC/caC at each reference cytosine (at least 10 reads required). (**D**) Representative regions showing 5mC, 5hmC, and 5fC/5caC levels at single-base resolution in WT and AD-patient derived cell lines at the human leukocyte antigen (HLA) gene cluster (chr6:32,440,096-32,643,715).
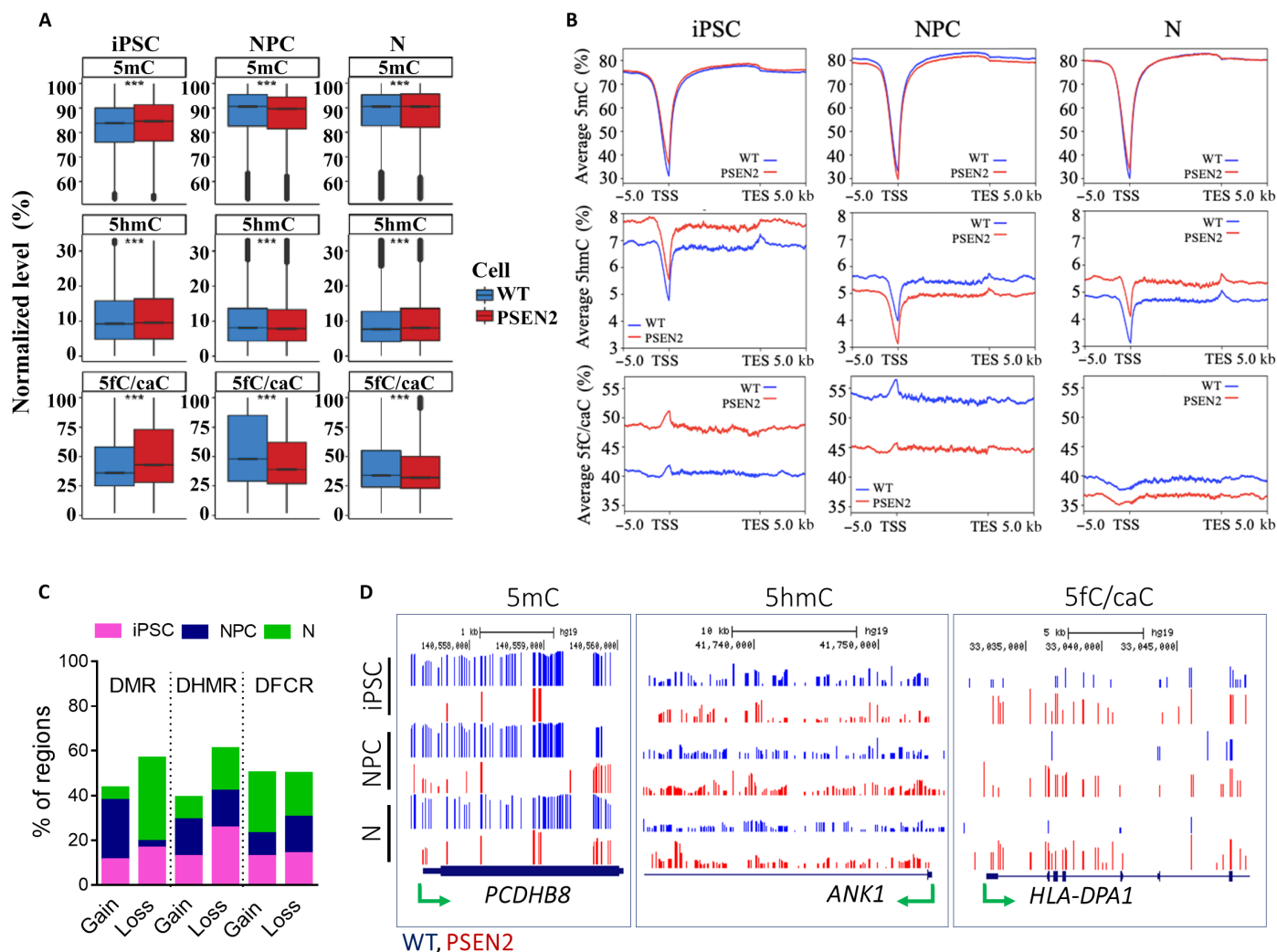
Fig. 4. iPSCs and iPSC-derived NPCs and neurons in PSEN2 lines are characterized by distinct methylomes compared to the WT. (**A**) Total 5mC, 5hmC, and 5fC/caC levels in iPSCs and iPSC-derived NPCs and neurons in WT and PSEN2. The x axis indicates the levels of 5mC, 5hmC, or 5fC/caC at each reference cytosine (at least 10 reads required). ***$P < 2.2 \times 10^{-16}$, determined by Wilcoxon rank sum test. (**B**) Levels of 5mC, 5hmC, and 5fC/caC in WT and PSEN2 cell lines at TSS and gene body in iPSCs, NPCs, and neurons. For each cytosine modification, gene body was normalized to 0 to 100%. Normalized density is plotted from 5 kbp upstream of TSSs to 5 kbp downstream of TESs. (**C**) Stacked bars show the differentially 5mC, 5hmC, and 5fC/caC regions presented as gains and losses during directed differentiation of iPSCs to neurons in WT and PSEN2 lines. (**D**) Representative regions showing 5mC, 5hmC, and 5fC/5caC levels at single-base resolution in WT and PSEN2 cell lines. In all samples, the scale for 5mC and 5fC/caC was 0 to 100%, while for 5hmC, the scale was 0 to 70%. Genome coordinates for the images shown were as follows: *PCDHB8*, chr5:140,556,693-140,560,759; *ANK1*, chr8:41,731,112-41,755,347; *HLA-DPA1*, chr6:33,030,742-33,049,833.

WT and PSEN2 cell lines (Fig. 4C). The detailed characterization of each stage-specific group of DMRs, DHMRs, or DFCRs between WT and PSEN2 lines is depicted in the Supplementary Text. Gene ontology (GO) enrichment analyses for the genes identified in DMRs (8713), DHMRs (1052), and DFCRs (2147) revealed a significant enrichment on neurodevelopmental processes, including forebrain development ($P < 10^{-6}$), synapse organization ($P < 10^{-9}$), and differentiation of neurons ($P < 10^{-8}$), highlighting the link between 5mC, 5hmC, and 5fC/caC methylome changes and disease-critical neural gene programs. The top 10 most highly significant gene sets ranked by P value are presented in fig. S2 (A to C). Focusing only on the pathways identified in the GO enrichment analysis that contained known AD risk genes revealed that changes in the epigenetic component overlap with diverse and critical biological pathways, including

immune response, metabolism, and oxidative stress response, all of which have been shown to be disrupted in AD (table S2, A to C). Other relevant genes identified in DMRs, DHMRs, or DFCRs include *ANK1*, *BACE2*, *BIN1*, *CLU*, *HLA-DPA1*, and *PCDHB8*, which are known AD risk genes and have well-established roles in AD pathology. These findings demonstrate that changes in 5mC, 5hmC, and 5fC/caC epigenetic profiles are directly associated with a network of key AD susceptibility genes in PSEN2 AD cell models. For instance, we observed loss of 5mC at TSS and gene body of *PCDHB8* at all stages in PSEN2 cells compared to the WT cells. This gene codes for neural cadherin–like protein and has been reported as an AD risk gene. *ANK1* is a well-established gene with a critical role in AD pathology, and few studies have reported its epigenetic deregulation in AD (*17*, *25*). *ANK1* is generally characterized by accumulation

of 5hmC in the PSEN2 model in comparison to the WT. An example from the DFCRs is the *HLA-DPA1* gene. The role of *HLA-DPA1* has also been studied in the context of AD, and genetic polymorphisms of this gene increase risk for AD (*26*). We observed stark differences in 5fC/caC levels along the entire *HLA-DPA1* gene in the PSEN2 cell model compared to the WT (Fig. 4D). Collectively, our findings highlight the pivotal role of proper timing and landscape of the three DNA methylation states in regulation of AD-critical genes. Because the patterns of methylation of all states are disrupted in AD patient–derived lines at all differentiation stages, it is postulated that the dysregulation of 5mC, 5hmC, and 5fC/caC on these key AD genes may intrinsically and directly be linked to AD pathology.

Discovery of AD-specific 5mC, 5hmC, and 5fC/caC signatures
To verify the aforementioned global changes in DNA methylation at all three states in additional AD cell lines, we applied the same in-depth analysis approaches to characterize DMRs, DHMRs, and DFCRs in another EOAD cell model, PSEN1, and the LOAD model APOE4. The comparison analysis identified a total of 2197 DMRs, 22 DHMRs, and 128 DFCRs, which were shared among all three AD cell models. To broaden our findings from cell culture models and to ensure that the changes of DNA methylation states are disease specific and not due to the aging process itself or a cell culture artifact, we also mapped and analyzed 5mC, 5hmC, and 5fC/caC of DNA samples derived from postmortem brain tissues of healthy donor and AD patients. Last, we overlapped our two sets of DMRs, DHMRs, and DFCRs (originating from cell lines or brain tissues), which resulted in 1447 DMRs, 7 DHMRs, and 23 DFCRs. We further extracted these shared DMRs/DHMRs/DFCRs to a final number of 71 AD-specific epigenetic signature regions located in autosomes, which were defined according to the following criteria: (i) All signature loci must be consistently presenting the same trends of gain or loss of 5mC, 5hmC, or 5fC/caC across all disease models compared to the control, and (ii) because 5mC is the most abundant modification, we applied more stringent criteria using a methylation difference of <-1 or >1 as a cutoff to reduce confounding background signal, while cutoff for 5hmC and 5fC/caC was <-0.1 or >0.1. These 71 regions were associated with 56 different genes (table S3). Because biological reproducibility is key, we applied another filter to our regional signatures based on methylation variation (using a cutoff of $\leq\pm1$ fold change for 5mC and $\leq\pm1.5$ fold change for 5hmC and 5fC/caC) between the samples we used in our study. This led to 19 signature regions for the 5mC modification, of which 17 were marked by loss of 5mC and 2 were marked by gain of 5mC in AD samples compared to the controls (Fig. 5A). Using the same approach, we narrowed down our 5hmC signature regions to five, of which four were marked by loss of 5hmC and one was characterized by gain of 5hmC in AD samples versus controls (Fig. 5B). Similarly, for the 5fC/caC group of signatures, we identified the final three signatures, where two had loss of 5fC/caC, while one had gain of 5fC/caC in AD samples compared to normal controls (Fig. 5C). Most of the 5mC signatures (8 of 19) were located in intronic regions, followed by 5 of 19 in promoters, 3 of 19 in distal intergenic regions, 2 of 19 in exons, and 1 of 19 in 3′ untranslated regions (3′UTRs). The 5hmC gene signatures were physically located as follows: one of five in distal intergenic regions and the rest were equally distributed in promoters (two of five) and 3′UTRs (two of five). All 5fC/caC (three of three) signatures were located in distal intergenic regions (table S4). The genes harboring epigenetic signatures can be classified into four

major subgroups according to their diversified biological function: (i) neurodevelopment and neuronal transcription factors, (ii) critical cellular processes, (iii) RNA and associated proteins (including noncoding RNAs), and (iv) cell signaling. Additional information for each signature region is provided in table S4 and the Supplementary Text. Collectively, we have identified 27 regional signatures of the methylated DNA cytosine that are specifically associated with AD.

Lastly, because our approaches allow us to investigate these cytosine modifications at single-base resolution, we next aimed to narrow down our regional signatures to individual CpG sites based on the following criteria: (i) identify the CpG sites that overlapped in the signature regions in all study samples, (ii) sort CpG sites that have the same trends of methylation (gain or loss) in AD versus controls, and (iii) apply cutoff of methylation fold-change difference of ≤-0.1 and ≥0.1 to avoid false positives. This led to a total of 39 AD-specific CpG sites that were consistently modified in early-onset, late-onset, familial, and sporadic AD models, which contain 27 5mC-, 4 5hmC-, and 8 5fC/caC-specific sites (Supplementary Text, Fig. 5D, and table S5). Finally, most of our 5mC, 5hmC, and 5fC/caC signatures were directly associated with genes involved in neuronal functions and/or have been implicated in AD.

Validation of AD-specific epigenetic signatures in an independent cohort and their clinical application
To demonstrate that our identified epigenetic signatures are a common molecular feature of AD, we validated our signatures in a replication cohort (ROS/MAP cohort; control, $n = 163$; AD, $n = 371$), where DNA methylation profiles were determined using Illumina 450k array (*17*, *27*). Samples with Braak&Braak score 0-II were defined as controls, while samples with Braak&Braak score IV-VI were considered as AD. Notably, 14 of 18 of the 5mC signature regions were identified in the ROS/MAP cohort. We used methylation levels of 14 signatures that were present in the ROS/MAP validation cohort and the following available general clinical features: age, sex, and race for the diagnosis of AD. To determine the beta coefficients and the significance of each signature region, we performed logistic regression, which allowed us to model data with binomial distribution. Following corrections for multiple hypothesis testing, eight genes (*ACKR3*, *ARHGEF16*, *CASZ1*, *ECE1*, *KIF26A*, *MIR153-2*, *NACC2*, and *NFATC1*) were also found to be significant in the ROS/MAP cohort. In addition, clinical features (age and sex) were also significant and were highlighted as critical in predicting AD (table S6).

Lastly, to determine overall accuracy of our signatures in predicting AD, we used methylation levels in our signature regions along with the clinical features to calculate receiver operating characteristic (ROC) curve and area under the curve (AUC) by applying and comparing four classification models: support vector machine (SVM), decision tree, logistic regression, and quadratic discriminant analysis (QDA). As shown in Fig. 6, the SVM model showed the best predictive capacity with an AUC = 0.9128 ($P < 2.2 \times 10^{-6}$, specificity = 98%, sensitivity = 48%). This was followed by the predictive ability generated by decision tree with an AUC = 0.8235 ($P < 2.2 \times 10^{-6}$, specificity = 96%, sensitivity = 61%), logistic regression with an AUC = 0.7941 ($P = 4.746 \times 10^{-5}$, specificity = 86%, sensitivity = 55%), and QDA with an AUC = 0.7905 ($P = 7.233 \times 10^{-6}$, specificity = 87%, sensitivity = 54%). Thus, as a proof of concept, the newly identified AD-specific epigenetic signatures in this study can be independently validated in another cohort and have a tremendous clinical implication in early detection and diagnosis of AD.
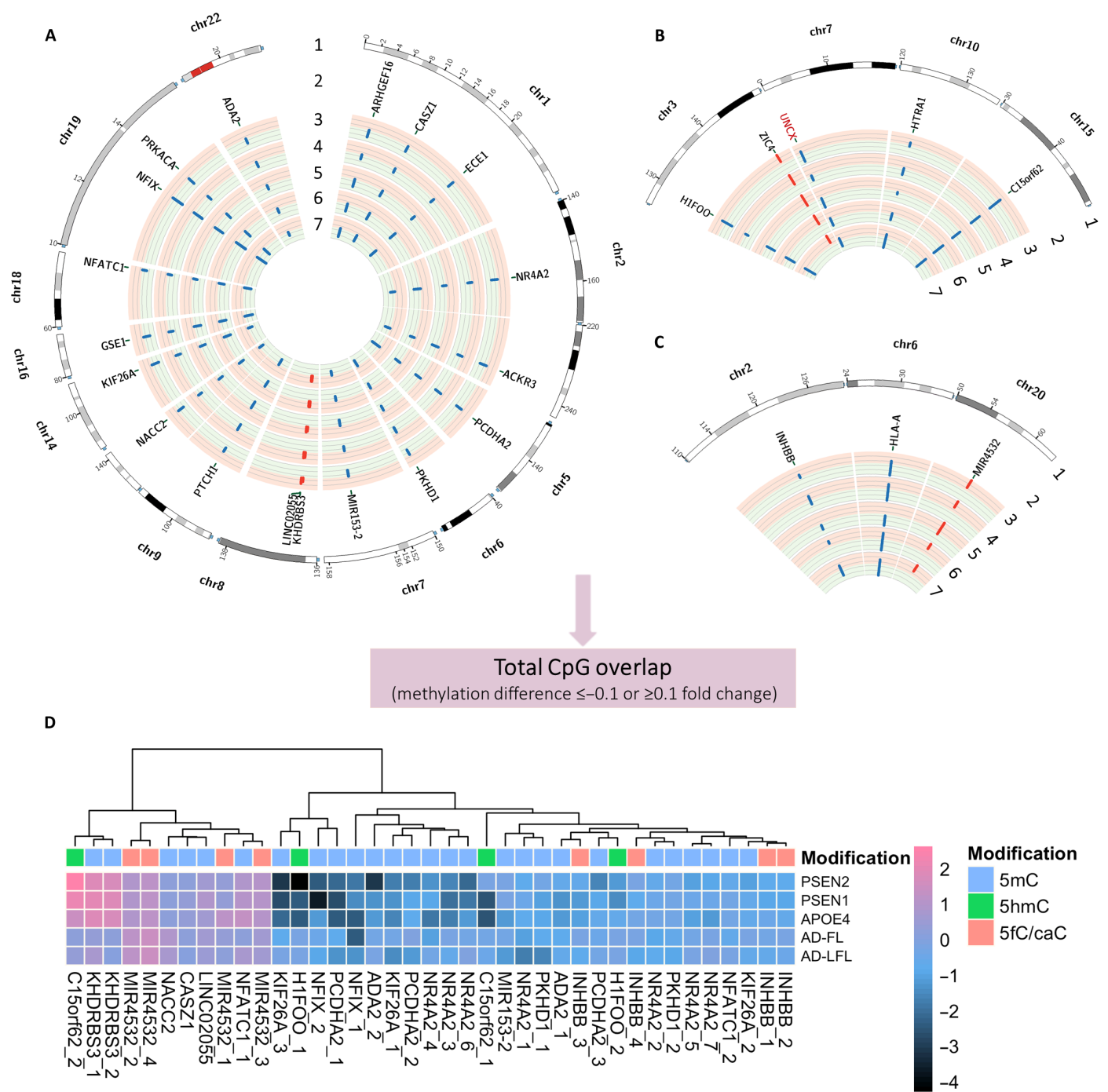
Fig. 5. Epigenetic signatures associated with AD. (A to C). The outermost circle (1) presents chromosome ideograms (in megabases). Changes in 5mC (A), 5hmC (B), or 5fC/caC (C) levels are shown as differences of 5mC, 5hmC, or 5fC/caC levels between normal and AD. Red bars indicate gain of 5mC, 5hmC, or 5fC/caC in AD versus controls, while blue bars mark loss of 5mC, 5hmC, and 5fC/caC in AD versus normal. The second layer of the circle (2) shows genes associated with AD epigenetic signatures. Genes in red letters fall in the ultraconserved regions. Layers 3 to 7 show 5mC, 5hmC, and 5fC/caC ratios, respectively, in AD versus control: (3) PSEN2 neurons/WT neurons, (4) PSEN1 neurons/WT neurons, (5) APOE4 neurons/WT neurons, (6) AD–frontal lobe/normal brain, and (7) AD–left frontal lobe/normal brain. (**D**) 5mC, 5hmC, and 5fC/caC signature regions narrowed down to individual CpG sites. Using methylation variation difference of ≤±1 fold change for 5mC and ≤±1.5 fold change for 5hmC and 5fC/caC, we identified 19 regions for 5mC signatures, 5 regions for 5hmC signatures, and 3 regions for 5fC/caC signatures. We then further filtered our 5mC, 5hmC, and 5fC/caC signatures by using methylation difference of ≤−0.1 or ≥0.1 fold change as a final cutoff. This yielded 27 CpG sites in the 5mC signatures, of which 21 and 6 CpG sites were marked by loss or gain of methylation, respectively, in AD samples versus controls. In the 5hmC group, we identified three CpG sites and one CpG site, which were marked by loss or gain of 5hmC, respectively, in AD samples compared to controls. In the three signature regions of 5fC/caC, we identified an equal number of CpG sites (four in each group) that were marked by gain or loss of 5fC/caC levels in AD versus controls.
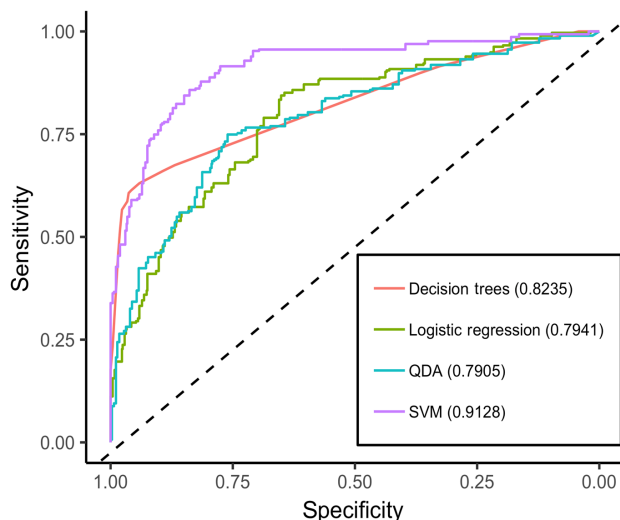
Fig. 6. ROC revealed that SVM and decision tree models have the best predictive ability with an AUC = 0.9128 (specificity = 98% and sensitivity = 48%) and AUC = 0.8235 (specificity = 96% and sensitivity = 61%), respectively. This was followed by predictive ability generated by logistic regression, AUC = 0.7941 (specificity = 86% and sensitivity = 55%), and quadratic discriminant analysis (QDA), AUC = 0.7905 (specificity = 87% and sensitivity = 54%).

DISCUSSION

We have identified a set of AD-specific CpG signatures across all states of DNA cytosine methylation in early-onset, late-onset, familial, and sporadic AD models. The 27 signature regions and the unique 39 CpG sites flagged in the roadmap highlight the key distinguishable and measurable features of AD epigenome from the norm, making them valuable and suitable for early molecular diagnosis of AD. These signatures are AD-specific and age independent, which were further validated in a large clinical cohort. Notably, a major group of the signatures occur on the genes implicated in AD or AD-related pathways, or on genomic regions that are critical for proper neurodevelopment. Last, these 27 signatures and more specifically the 39 CpG sites identified can easily be converted into a minimally invasive diagnostic panel that allows testing of these predictable molecular signatures for detection of AD. This was first evidenced by our study of the PSEN1 cell model, which originated from an asymptomatic patient carrying a mutation in the *PSEN1* gene. These findings strongly suggested that the existing 5mC, 5hmC, and 5fC/caC changes associated with our signatures may precede the disease process, and they are, therefore, not secondary or consequence to the later stages of the disease progression. Furthermore, although epigenetic landscapes have been suggested to change with increasing age, our identified signatures are age independent; rather, our 5mC, 5hmC, and 5fC/caC signatures appeared to be limited to AD condition, regardless of familial or sporadic origin. This will not only provide mechanistic insight into disease etiology but also potentially identify strategies for early detection and therapeutic intervention at preclinical stages.

In summary, the present study provides the first comprehensive and comparative reference maps of genome-wide distribution of all three DNA methylation states at base resolution during neural differentiation from iPSCs to mature neurons and in postmortem brain tissues of both normal donor and AD patients. These maps not only uncovered the dynamic changes of the landscapes of 5mC, 5hmC,

and 5fC/caC during neural differentiation of normal and AD-derived iPSCs but also allowed us to pin down the specific changes at any given cytosine and that of three methylation states. These findings identified that the dynamic and coordinated changes of three DNA methylation states are important epigenetic features specifically associated with neural differentiation process. We found that the observed epigenetic features in normal cells were dysregulated in AD cell models, suggesting that precise regulation, proper establishment, and maintenance of these epigenetic features likely play important roles in regulation of neural lineage commitment, maturation, and function. Following our initial findings, future investigations focusing on continually and thoroughly attesting these AD-specific epigenetic signatures in larger cohorts are warranted to ultimately turn these into even more reliable, reproducible, and verifiable signatures.

MATERIALS AND METHODS
Cell lines and tissue samples

Normal and AD patient–derived iPSCs, neural progenitor cells, and cortical neuronal cells were obtained from Axol Biosciences (Cambridge, UK). Control or disease iPS lines were generated using episomal vector reprogramming of somatic cells (newborn, male). In addition to control lines, in our study, we used iPSCs carrying mutation L286V in *PSEN1* and mutation N141I in *PSEN2*, both of which are associated with EOAD. Age of patients when the skin cells were harvested for reprogramming was 38 years (female) and 81 years (female) for PSEN1 and PSEN2 lines, respectively. The iPSCs derived from a LOAD patient carrying homozygous *APOE4* were generated by reprogramming fibroblasts harvested from a female patient at the age of 87 years. Directed differentiation of iPSCs to cortical neurons was performed as described previously (*19*). Characterization of iPSCs, neural progenitor cells, and cortical neurons in normal and AD patient–derived lines was performed at Axol Biosciences (Cambridge, UK). DNA samples originating from brain tissues of healthy (60 years, female), AD–frontal lobe region (87 years, male), and AD–left frontal lobe region (75 years, male) were purchased from BioChain (CA, USA). All samples were collected and processed according to protocols approved by the Brigham and Women's Hospital.

Oxidative bisulfite and methylase-assisted sequencing at single–base pair resolution

To map and quantify the levels of 5mC and 5hmC genome wide at base resolution, we used OXBS-seq in our samples using TrueMethyl Seq according to the recommended protocol by Cambridge Epigenetix (CEGX; Cambridge, UK) (*20*). Briefly, 1 µg of DNA extracted following the phenol-chloroform protocol was spiked-in with synthetic DNA sequencing controls, and then DNA was sheared to ~800 base pairs (bp) using a Covaris E220 sonicator. Following that, DNA samples were split into two halves for the library module: ~500 ng for bisulfite conversion and ~500 ng for oxidative bisulfite conversion, generating in parallel two libraries for each sample, enabling us to distinguish at base resolution 5mC and 5hmC. MAB-seq was performed according to the protocol described previously by Neri *et al.* (*21*). Briefly, 1 µg of DNA was methylated using M.SssI [New England Biolabs (NEB), MA, USA] and then sheared to 350 bp using a Covaris M220 sonicator. DNA libraries were prepared according to the Illumina protocol (TruSeq DNA PCR-Free Library Preparation Kit, CA, USA). Adaptor ligated libraries were treated with M.SssI (NEB, MA, USA) to methylate

bases introduced through end repair, bisulfite-converted using the EpiTect Bisulfite Kit (Qiagen, CA, USA), and then amplified using KAPA HiFi HotStart Uracil+ Readymix (KAPA Biosystems, MA, USA). Samples were sequenced using an Illumina HiSeq X Ten platform, generating at least 100 GB per sample.

RNA isolation, reverse transcription, and RT-PCR
RNA was isolated using TRIzol reagent following the recommended protocol (Invitrogen, CA, USA). Reverse transcription was performed using iScript reverse transcription reagents from Bio-Rad (CA, USA). Real-time quantitative reverse transcription polymerase chain reaction (qRT-PCR) was performed with iQ SYBR Green Supermix (Bio-Rad, CA, USA). Primer sequences for *NES*, *MAP2*, *PAX6*, and *SOX2* genes were described previously (19, 28, 29). The rest of the primer sequences are listed in table S7.

Immunofluorescence staining
Cells grown on sterile glass cover slips were fixed with 3.7% paraformaldehyde in phosphate-buffered saline (PBS) for 20 min and washed extensively with PBS. Samples were permeabilized with 0.2% Triton X-100 in PBS for 20 min and blocked with 5% goat serum (or 3% bovine serum albumin) in PBS for 45 min. Cells were incubated with primary antibodies for 1 hour at room temperature. Immunoglobulin G (IgG) was used as negative control. After extensive washing with PBS, samples were incubated with respective secondary antibodies. Nuclei were stained with 4′,6-diamidino-2-phenylindole (DAPI) for 10 min, and samples were mounted using Fluoromount-G (Southern Biotech, AL, USA).

Whole-genome sequencing data quality
To assess quality of BS/OXBS-Seq data, spike-in DNA controls were mixed with genomic DNA samples. We used CEGX_QC (https://bitbucket.org/cegx-bfx/cegxqc) package with default parameters to detect the conversion rate of 5mC and 5hmC. The conversion efficiency was determined on the basis of synthetic DNA controls from CEGX and was considered successful if C > 90% for BS and OXBS libraries, while hmC < 5% in BS libraries and >80% in OXBS libraries (https://insidedna.me/tool_page_assets/pdf_manual/cegxqc.pdf) (table S8A). Similarly, to determine the methylation efficiency of M.SssI in MAB-seq, we spiked in unmethylated CG lambda DNA (Promega, WI, USA) and observed almost complete methylation of unmethylated CGs (~97%) and high bisulfite conversion efficiency (~97%) (table S8B). We randomly selected 30 million reads from each trimmed sequence files, mapped them onto lambda genome using bsmap, and extracted methylation signals using command "methratio.py -t 2 -r -m 5 -z." M.SssI enables methylation of Cs with high efficiency in the CG context, instead of the CH context, while the bisulfite treatment converts unmethylated Cs into Ts in both CG and CH contexts. Therefore, the M.SssI methylase efficiency was calculated by the methylated-CG% based on the cytosine methylation level in the CG context, while bisulfite conversion efficiency was identified by unmethylated-CH% based on the lack of cytosine methylation in the CH context. Because of the fact that MAB-seq allows us to confidently determine 5fC/caC only in the CG context (30) and that 5hmC is almost exclusively located in CG dinucleotides (31), we focused our studies of 5mC and 5hmC in CG sites to generate comparable and comprehensive genome-wide maps for all three cytosine modifications.

Identification of 5mC, 5hmC, and 5fC/5caC sites and regions
Raw sequencing data were trimmed using Trim Galore (v0.4.0) (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) to remove low-quality bases and adaptor sequences. Trimmed reads were mapped onto the reference genome (hg19 for samples of human origin and mm9 for samples of mouse origin) using bsmap (v2.74) (32), followed by removal of PCR duplicates. Methylation signals were extracted using methratio.py, a script in the bsmap package (v3.4.2) (33). An R script was generated to maintain chromosome coordinates consistent for methylation sites between each pair of BS and OXBS samples, allowing us to identify 5mC and 5hmC signals in these common sites with the mlml script in methpipe using the default settings. To calculate 5fC/5caC sites, we used the following formula: 5fC/5caC = NT/(NC + NT), where NT and NC are the number of Ts and Cs in the CG context, respectively. M.SssI methylase error rate and bisulfite conversion inefficiency was 1.64%, which was determined in lambda genome in our pilot experiment in mESC WT and mESC Tdg KO cell lines (table S8B). Error rate was used in binomial distribution and q value package to adjust for the 5fC/5caC signal. The number of 5fC/5caC sites was determined by a binomial test, as described previously (34). Only 5fC/5caC sites with cutoff of coverage ≥ 10, $P < 0.01$, and false discovery rate < 0.01 were kept for the downstream analyses. It is possible that our samples may carry mutations different from the reference genome, which could result in higher false positives. To address this issue, we used biscuit package (https://github.com/zwdzwd/biscuit) to identify these mutations from trimmed mapped reads and removed these potential mutations from 5mC, 5hmC, and 5fC/5caC sites. The remaining sites were used for the downstream analysis.

To call for DMRs, DHMRs, and DFCRs between the two groups of methylomes, the methpipe package (http://smithlabresearch.org/software/methpipe/) was used (33). We started by assembling a proportion table containing read proportions for all target methylomes with "merge-methcounts" program included in methpipe. After creating the proportion table and specifying the design matrix, we performed differential methylation analyses, which consists of (i) regression ("radmeth regression" default parameters), (ii) combining significance, and (iii) multiple testing adjustment steps ("radmeth adjust -bins 1:200:1"). All DMRs/DHMRs/DFCRs were filtered by $P < 0.01$ and contain at least five differentially methylated CG sites in each region. To call for DMRs, DHMRs, and DFCRs between a pair of methylomes, we generated an R script that enables us to use 200-bp step size across the entire genome with 2-kb bin and calculate the methylation difference along with P value (Student's t test) in both samples. If adjacent bins had continual methylation differences between samples, which were identified with cutoff of fold change > 2 and $P < 0.05$, then these bins were iteratively merged together, and methylation difference was calculated for the merged region.

Data normalization
To allow comparison of 5mC, 5hmC, and 5fC/caC signals among samples, we normalized 5mC, 5hmC, and 5fC/caC signals across samples using nonoverlapping 1-kb bins spanning the whole genome. For 5mC and 5hmC, the signal in each bin was represented by TNC/(TNC + TNT), where TNT and TNC are the total number of Ts and Cs within a bin, respectively. For 5fC/5caC, the signal in each bin was calculated with average signal within the bin, which was due to the lower density of 5fC/5caC sites across the whole

genome compared to 5mC or 5hmC sites. Last, the signals of 5mC, 5hmC, and 5fC/5caC were smoothed over the whole genome, which allows us to compare the signal differences of the same type of modification between samples.

Pathway analysis

We used the clusterProfiler package in Bioconductor to perform enrichment analysis for identified genes (35). The script enrichGO was used to perform GO pathway enrichment analysis. The enrichGO calculates P values using the hypergeometric distribution.

ROC, specificity, and sensitivity calculation

To determine the beta coefficients and the significance of each signature in predicting AD, we performed logistic regression using R function glm(family = binomial(link = "logit")), in which the parameters allowed us to model data with binomial distribution. Half of the data were used as training dataset, and the other half were used as test dataset. To evaluate the overall accuracy of our signature regions, we calculated ROC curve and AUC by applying and comparing four classification models: SVM, decision tree, logistic regression, and QDA. P values were calculated with an exact binomial test.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/5/8/eaaw2880/DC1

Supplementary Text

Fig. S1. Directed differentiation of human iPSCs to neural progenitors and cortical neurons and expression of stage-specific markers.

Fig. S2. Biological interpretation of genomic regions identified in the DMRs/DHMRs/DFCRs in PSEN2 and WT cell lines.

Table S1. Total 5mC, 5hmC, and 5fC/caC sites in WT and AD cell lines.

Table S2A. GO functional enrichment analysis of genes in DMRs (cutoff $P < 0.05$).

Table S2B. GO functional enrichment analysis of genes in DHMRs (cutoff $P < 0.05$).

Table S2C. GO functional enrichment analysis of genes in DFCRs (cutoff $P < 0.05$).

Table S3A. AD-related 5mC signatures.

Table S3B. AD-related 5hmC signatures.

Table S3C. AD-related 5fC/caC signatures.

Table S4A. AD-related 5mC signatures (cutoff ±1 fold-change methylation variation).

Table S4B. AD-related 5hmC signatures (cutoff ±1.5 fold-change methylation variation).

Table S4C. AD-related 5fC/caC signatures (cutoff ±1.5 fold-change methylation variation).

Table S5A. CpG-specific AD epigenetic signatures (5mC).

Table S5B. CpG-specific AD epigenetic signatures (5hmC).

Table S5C. CpG-specific AD epigenetic signatures (5fC/caC).

Table S6. Validation of AD-related epigenetic signatures in the ROS/MAP replication cohort.

Table S7. qRT-PCR primers.

Table S8A. Summary of whole-genome OXBS-seq experiments.

Table S8B. Summary of whole-genome MAB-seq experiments.

References (36–42)

REFERENCES AND NOTES

1. R. Lardenoije, A. Iatrou, G. Kenis, K. Kompotis, H. W. M. Steinbusch, D. Mastroeni, P. Coleman, C. A. Lemere, P. R. Hof, D. L. A. van den Hove, B. P. F. Rutten, The epigenetics of aging and neurodegeneration. *Prog. Neurobiol.* **131**, 21–64 (2015).

2. H. M. Lanoiselée, G. Nicolas, D. Wallon, A. Rovelet-Lecrux, M. Lacour, S. Rousseau, A.-C. Richard, F. Pasquier, A. Rollin-Sillaire, O. Martinaud, M. Quillard-Muraine, V. de la Sayette, C. Boutoleau-Bretonniere, F. Etcharry-Bouyx, V. Chauviré, M. Sarazin, I. le Ber, S. Epelbaum, T. Jonveaux, O. Rouaud, M. Ceccaldi, O. Félician, O. Godefroy, M. Formaglio, B. Croisile, S. Auriacombe, L. Chamard, J.-L. Vincent, M. Sauvée, C. Marelli-Tosi, A. Gabelle, C. Ozsancak, J. Pariente, C. Paquet, D. Hannequin, D. Campion; collaborators of the CNR-MAJ project, *APP*, *PSEN1*, and *PSEN2* mutations in early-onset Alzheimer disease: A genetic screening study of familial and sporadic cases. *PLOS Med.* **14**, e1002270 (2017).

3. E. H. Corder, A. M. Saunders, W. J. Strittmatter, D. E. Schmechel, P. C. Gaskell, G. W. Small, A. D. Roses, J. L. Haines, M. A. Pericak-Vance, Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* **261**, 921–923 (1993).

4. W. J. Strittmatter, A. M. Saunders, D. Schmechel, M. Pericak-Vance, J. Enghild, G. S. Salvesen, A. D. Roses, Apolipoprotein E: High-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1977–1981 (1993).

5. E. Bagyinszky, Y. C. Youn, S. S. A. An, S. Kim, The genetics of Alzheimer's disease. *Clin. Interv. Aging* **9**, 535–551 (2014).

6. J. Povova, P. Ambroz, M. Bar, V. Pavukova, O. Sery, H. Tomaskova, V. Janout, Epidemiological of and risk factors for Alzheimer's disease: A review. *Biomed. Pap. Med. Fac. Univ. Palacky Olomouc Czech Repub.* **156**, 108–114 (2012).

7. A. Rudenko, M. M. Dawlaty, J. Seo, A. W. Cheng, J. Meng, T. Le, K. F. Faull, R. Jaenisch, L.-H. Tsai, Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. *Neuron* **79**, 1109–1122 (2013).

8. C. J. Klein, M.-V. Botuyan, Y. Wu, C. J. Ward, G. A. Nicholson, S. Hammans, K. Hojo, H. Yamanishi, A. R. Karpf, D. C. Wallace, M. Simon, C. Lander, L. A. Boardman, J. M. Cunningham, G. E. Smith, W. J. Litchy, B. Boes, E. J. Atkinson, S. Middha, P. J. B Dyck, J. E. Parisi, G. Mer, D. I. Smith, P. J. Dyck, Mutations in *DNMT1* cause hereditary sensory neuropathy with dementia and hearing loss. *Nat. Genet.* **43**, 595–600 (2011).

9. G. A. Kaas, C. Zhong, D. E. Eason, D. L. Ross, R. V. Vachhani, G.-l. Ming, J. R. King, H. Song, J. D. Sweatt, TET1 controls CNS 5-methylcytosine hydroxylation, active DNA demethylation, gene transcription, and memory formation. *Neuron* **79**, 1086–1093 (2013).

10. R.-R. Zhang, Q.-Y. Cui, K. Murai, Y. C. Lim, Z. D. Smith, S. Jin, P. Ye, L. Rosa, Y. K. Lee, H.-P. Wu, W. Liu, Z.-M. Xu, L. Yang, Y.-Q. Ding, F. Tang, A. Meissner, C. Ding, Y. Shi, G.-L. Xu, Tet1 regulates adult hippocampal neurogenesis and cognition. *Cell Stem Cell* **13**, 237–245 (2013).

11. C.-X. Song, K. E. Szulwach, Y. Fu, Q. Dai, C. Yi, X. Li, Y. Li, C.-H. Chen, W. Zhang, X. Jian, J. Wang, L. Zhang, T. J. Looney, B. Zhang, L. A. Godley, L. M. Hicks, B. T. Lahn, P. Jin, C. He, Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* **29**, 68–72 (2011).

12. R. Lister, E. A. Mukamel, J. R. Nery, M. Urich, C. A. Puddifoot, N. D. Johnson, J. Lucero, Y. Huang, A. J. Dwork, M. D. Schultz, M. Yu, J. Tonti-Filippini, H. Heyn, S. Hu, J. C. Wu, A. Rao, M. Esteller, C. He, F. G. Haghighi, T. J. Sejnowski, M. M. Behrens, J. R. Ecker, Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).

13. L. Chouliaras, D. Mastroeni, E. Delvaux, A. Grover, G. Kenis, P. R. Hof, H. W. M. Steinbusch, P. D. Coleman, B. P. F. Rutten, D. L. A. van den Hove, Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiol. Aging* **34**, 2091–2099 (2013).

14. M. A. Bradley-Whitman, M. A. Lovell, Epigenetic changes in the progression of Alzheimer's disease. *Mech. Ageing Dev.* **134**, 486–495 (2013).

15. N. Coppieters, B. V. Dieriks, C. Lill, R. L. M. Faull, M. A. Curtis, M. Dragunow, Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. *Neurobiol. Aging* **35**, 1334–1344 (2014).

16. T. Lashley, P. Gami, N. Valizadeh, A. Li, T. Revesz, R. Balazs, Alterations in global DNA methylation and hydroxymethylation are not detected in Alzheimer's disease. *Neuropathol. Appl. Neurobiol.* **41**, 497–506 (2015).

17. P. L. De Jager, G. Srivastava, K. Lunnon, J. Burgess, L. C. Schalkwyk, L. Yu, M. L. Eaton, B. T. Keenan, J. Ernst, C. McCabe, A. Tang, T. Raj, J. Replogle, W. Brodeur, S. Gabriel, H. S. Chai, C. Younkin, S. G. Younkin, F. Zou, M. Szyf, C. B. Epstein, J. A. Schneider, B. E. Bernstein, A. Meissner, N. Ertekin-Taner, L. B. Chibnik, M. Kellis, J. Mill, D. A. Bennett, Alzheimer's disease: Early alterations in brain DNA methylation at *ANK1*, *BIN1*, *RHBDF2* and other loci. *Nat. Neurosci.* **17**, 1156–1163 (2014).

18. J. Zhao, Y. Zhu, J. Yang, L. Li, H. Wu, P. L. De Jager, P. Jin, D. A. Bennett, A genome-wide profiling of brain DNA hydroxymethylation in Alzheimer's disease. *Alzheimers Dement.* **13**, 674–688 (2017).

19. Y. Shi, P. Kirwan, F. J. Livesey, Directed differentiation of human pluripotent stem cells to cerebral cortex neurons and neural networks. *Nat. Protoc.* **7**, 1836–1846 (2012).

20. M. J. Booth, T. W. B. Ost, D. Beraldi, N. M. Bell, M. R. Branco, S. Balasubramanian, S. Reik, Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nat. Protoc.* **8**, 1841–1851 (2013).

21. F. Neri, D. Incarnato, A. Krepelova, S. Rapelli, F. Anselmi, C. Parlato, C. Medana, F. Dal Bello, S. Oliviero, Single-base resolution analysis of 5-formyl and 5-carboxyl cytosine reveals promoter DNA methylation dynamics. *Cell Rep.* **10**, 674–683 (2015).

22. G. Ficz, M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews, W. Reik, Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398–402 (2011).

23. R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q.-M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, J. R. Ecker, Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).

24. L. M. Wheldon, A. Abakir, Z. Ferjentsik, T. Dudnakova, S. Strohbuecker, D. Christie, N. Dai, S. Guan, J. M. Foster, I. R. Corrêa Jr., M. Loose, J. E. Dixon, V. Sottile, A. D. Johnson,

A. Ruzov, Transient accumulation of 5-carboxylcytosine indicates involvement of active demethylation in lineage specification of neural stem cells. *Cell Rep.* **7**, 1353–1361 (2014).

25. K. Lunnon, R. Smith, E. Hannon, P. L. De Jager, G. Srivastava, M. Volta, C. Troakes, S. Al-Sarraj, J. Burrage, R. Macdonald, D. Condliffe, L. W. Harries, P. Katsel, V. Haroutunian, Z. Kaminsky, C. Joachim, J. Powell, S. Lovestone, D. A. Bennett, L. C. Schalkwyk, J. Mill, Methylomic profiling implicates cortical deregulation of *ANK1* in Alzheimer's disease. *Nat. Neurosci.* **17**, 1164–1170 (2014).

26. R. Ferrari, Y. Wang, J. Vandrovcova, S. Guelfi, A. Witeolar, C. M. Karch, A. J. Schork, C. C. Fan, J. B. Brewer; International FTD-Genomics Consortium (IFGC); International Parkinson's Disease Genomics Consortium (IPDGC); International Genomics of Alzheimer's Project (IGAP), P. Momeni, G. D. Schellenberg, W. P. Dillon, L. P. Sugrue, C. P. Hess, J. S. Yokoyama, L. W. Bonham, G. D. Rabinovici, B. L. Miller, O. A. Andreassen, A. M. Dale, J. Hardy, R. S. Desikan, Genetic architecture of sporadic frontotemporal dementia and overlap with Alzheimer's and Parkinson's diseases. *J. Neurol. Neurosurg. Psychiatry* **88**, 152–164 (2017).

27. D. A. Bennett, A. S. Buchman, P. A. Boyle, L. L. Barnes, R. S. Wilson, J. A. Schneider, Religious orders study and rush memory and aging project. *J. Alzheimers Dis.* **64**, S161–S189 (2018).

28. A. Giorgetti, M. C. N. Marchetto, M. Li, D. Yu, R. Fazzina, Y. Mu, A. Adamo, I. Paramonov, J. C. Cardoso, M. B. Monasterio, C. Bardy, R. Cassiani-Ingoni, G.-H. Liu, F. H. Gage, J. C. Izpisua Belmonte, Cord blood–derived neuronal cells by ectopic expression of Sox2 and c-Myc. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 12556–12561 (2012).

29. J. M. Garner, D. W. Ellison, D. Finkelstein, D. Ganguly, Z. Du, M. Sims, C. H. Yang, R. B. Interiano, A. M. Davidoff, L. M. Pfeffer, Correction: Molecular heterogeneity in a patient-derived glioblastoma xenoline is regulated by different cancer stem cell populations. *PLOS ONE* **10**, e0145052 (2015).

30. H. Wu, X. Wu, L. Shen, Y. Zhang, Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nat. Biotechnol.* **32**, 1231–1240 (2014).

31. M. Yu, G. C. Hon, K. E. Szulwach, C.-X. Song, P. Jin, B. Ren, C. He, Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nat. Protoc.* **7**, 2159–2170 (2012).

32. Y. Xi, W. Li, BSMAP: Whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* **10**, 232 (2009).

33. Q. Song, B. Decato, E. E. Hong, M. Zhou, F. Fang, J. Qu, T. Garvin, M. Kessler, J. Zhou, A. D. Smith, A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLOS ONE* **8**, e81148 (2013).

34. M. Yu, G. C. Hon, K. E. Szulwach, C.-X. Song, L. Zhang, A. Kim, X. Li, Q. Dai, Y. Shen, B. Park, J.-H. Min, P. Jin, B. Ren, C. He, Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**, 1368–1380 (2012).

35. G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

36. F. Montarolo, S. Perga, S. Martire, D. N. Navone, A. Marchet, D. Leotta, A. Bertolotto, Altered NR4A subfamily gene expression level in peripheral blood of Parkinson's and Alzheimer's disease patients. *Neurotox. Res.* **30**, 338–344 (2016).

37. Q. Chu, J. K. Diedrich, J. M. Vaughan, C. J. Donaldson, M. F. Nunn, K.-F. Lee, A. Saghatelian, HtrA1 proteolysis of ApoE in vitro is allele selective. *J. Am. Chem. Soc.* **138**, 9473–9478 (2016).

38. S. Grau, A. Baldi, R. Bussani, X. Tian, R. Stefanescu, M. Przybylski, P. Richards, S. A. Jones, V. Shridhar, T. Clausen, M. Ehrmann, Implications of the serine protease HtrA1 in amyloid precursor protein processing. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6021–6026 (2005).

39. E. A. Eckman, M. Watson, L. Marlow, K. Sambamurti, C. B. Eckman, Alzheimer's disease β-amyloid peptide is increased in mice deficient in endothelin-converting enzyme. *J. Biol. Chem.* **278**, 2081–2084 (2003).

40. B. Funalot, T. Ouimet, A. Claperon, C. Fallet, A. Delacourte, J. Epelbaum, T. Subkowski, N. Léonard, V. Codron, J.-P. David, P. Amouyel, J.-C. Schwartz, N. Helbecque, Endothelin-converting enzyme-1 is expressed in human cerebral cortex and protects against Alzheimer's disease. *Mol. Psychiatry* **9**, 1122–1128 (2004).

41. J. M. Long, B. Ray, D. K. Lahiri, MicroRNA-153 physiologically inhibits expression of amyloid-β precursor protein in cultured human fetal brain cells and is dysregulated in a subset of Alzheimer disease patients. *J. Biol. Chem.* **287**, 31298–31310 (2012).

42. S. Trazzi, C. Fuchs, E. Valli, G. Perini, R. Bartesaghi, E. Ciani, The amyloid precursor protein (APP) triplicated gene impairs neuronal precursor differentiation and neurite development through two different domains in the Ts65Dn mouse model for Down syndrome. *J. Biol. Chem.* **288**, 20817–20829 (2013).

Citation: I. S. Fetahu, D. Ma, K. Rabidou, C. Argueta, M. Smith, H. Liu, F. Wu, Y. G. Shi, Epigenetic signatures of methylated DNA cytosine in Alzheimer's disease. *Sci. Adv.* **5**, eaaw2880 (2019).