5-Hydroxymethylcytosine Signatures in Circulating Cell-Free DNA as Diagnostic Biomarkers for Late-Onset Alzheimer’s Disease

Lei Chen
Qingdao University School of Public Health
https://orcid.org/0000-0001-5596-127X

Shunliang Xu
Shandong University School of Medicine: Shandong University Cheeloo College of Medicine

Qianqian Shen
Qingdao University School of Public Health

Hongzhuan Yu
Weifang People’s Hospital

Shengjie Pei
Qingdao University School of Public Health

Yangting Zhang
Qingdao University School of Public Health

Xin He
Qingdao University School of Public Health

QiuZhen Wang (✉ kevin_1971@126.com)
Qingdao University
https://orcid.org/0000-0001-5946-3884

Duo Li
Qingdao University School of Public Health

Research

Keywords: Alzheimer’s disease, 5-hydroxymethylcytosine, Cell-free DNA, Biomarker

Posted Date: August 6th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-774460/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Version of Record: A version of this preprint was published at Journal of Alzheimer's Disease on November 29th, 2021. See the published version at https://doi.org/10.3233/JAD-215217.
Abstract

**Background:** 5-Hydroxymethylcytosine (5hmC) is an epigenetic DNA modification that is highly abundant in nervous system. It has been reported that 5hmC is significantly associated with Alzheimer's disease (AD). Changes in 5hmC signatures can be detected in circulating cell-free DNA (cfDNA), which has shown potential as a non-invasive liquid biopsy material. However, there is no research about genome-wide profiling of 5hmC in cfDNA and its potential for the diagnosis of AD to date.

**Methods:** We carried out a case-control study and used a highly sensitive and selective high-throughput sequencing of chemical labels to detect the genome-wide profiles of 5hmC in human cfDNA and identified differentially hydroxymethylated regions (DhMRs) in AD patients and the control.

**Results:** We detected a significant difference of 5hmC enrichment in gene bodies which were linked to multiple AD pathogenesis-associated signaling pathways in AD patients compared with cognitively normal controls. AD patients can be well distinguished from cognitively normal controls by differentially hydroxymethylated regions (DhMRs) in cfDNA. Specially, we found 7 distinct genes (*RABEP1*, *CPNE4*, *DNAJC15*, *REEP3*, *ROR1*, *CAMK1D*, and *RBFOX1*) had prediction diagnostic potential based on their significant correlations with MMSE and MoCA scores.

**Conclusions:** The present results suggest that 5hmC markers derived from plasma cfDNA can be served as an effective, minimally invasive biomarkers for clinical auxiliary diagnosis of late-onset AD.


1. Introduction

Alzheimer's disease (AD) is the most common cause of dementia in the world [1]. Physicians use a variety of approaches and tools to help make a diagnosis, and a definitive diagnosis of AD can only be made postmortem [2]. Therefore, for clinical diagnosis of AD in living patients, the National Institute on Aging and the Alzheimer's association proposed to incorporate not only the clinical judgment but also biomarker tests. Therefore, developing a noninvasive biomarker test, such as a blood test, to assist in diagnosing AD would be of great significance. Research is underway to develop such biomarker, but to date, no biomarker has been defined based on the sensitivity and specificity needed to diagnose living individuals for routine use in clinical practice [3]. One of the reasons is that the possible biomarkers including beta-amyloid plaques lack specificity for distinguishing dementia caused by AD from cognitive decline that is considered as a part of normal aging [1, 4]. Other major limitations of these biomarkers include that they are difficult to sample, impose a heavy economic burden on the patients, or require a specialist's knowledge for their interpretation [1].

AD is a complex multifactorial neurodegenerative disorder that involves complex interactions between genes and environmental factors [5]. The 21 variants significantly associated with AD susceptibility can
only explain 7.6% of cognitive decline and 2.1% of clinical AD occurrence, highlighting that AD development may be driven by non-genetic factors [6], such as epigenetic modifications. A number of studies have shown aberrant DNA epigenetic modifications in specific brain regions in individuals diagnosed with AD [7]. 5-Methylcytosine (5mC) has been revealed to be modulated by environmental factors and to be involved in the onset and progression of AD [8]. Some studies have focused on the identification of 5mC in peripheral blood DNA to characterize AD in the past decade [9]. However, there are no such studies on other epigenetic modifications. 5-Hydroxymethylcytosine (5hmC) is another major and stable epigenetic marker that is generated from 5mC by the ten-eleven translocation (TET) family of dioxygenases that involves a wide range of biological processes from development to various diseases [10]. More importantly, 5hmC is highly enriched in the central nervous system [11], and dynamically regulated during neural development, being enriched with age across the lifespan [12, 13]. A study of postmortem AD brains indicated that differentially hydroxymethylated genes were significantly enriched in AD-related pathology and clustered in functional gene ontology categories [14]. Therefore, the changes in 5hmC markers may have great potential in AD diagnostics.

There are some intriguing reasons to give consideration to peripheral blood biomarkers for AD. Firstly, it is impossible to obtain brain tissue samples from living patients. Secondly, biopsies in the cerebrospinal fluid (CSF) are very challenging from the perspective of obtaining longitudinal samples due to the risk of the CSF biopsy itself. Cell-free DNA (cfDNA), originating from the DNA of apoptotic and necrotic cells, are found in the blood stream and individually represent a cellular derivative of the original tissue [15]. Elevated cfDNA levels have been detected in patients with severe brain injury which is proof-of-principle that cfDNA is shed from the brain to the blood stream during cell degradation [16]. AD is characterized by brain atrophy associated with apoptotic of synapses and neurons, hence the proportion of the cfDNA in plasma originating from which will increase in AD patients. Therefore, we hypothesized that 5hmC landscape of cfDNA may be a potential diagnostic biomarker of AD and carried out a case control study to test the feasibility.

2. Methods

2.1. Participants

A case control study was implemented from November 2019 to June 2020 in the Second Hospital of Shandong University. Late-onset AD patients (> 60 years) and age-matched (± 3 years) cognitively normal subjects were recruited. The diagnosis of AD was based on the present clinical procedure [1]. Firstly, face-to-face interviews were conducted by trained investigators to screen suspected AD patients using Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MOCA). Next, according to the clinical diagnostic criteria of AD, the suspected AD patients were received a complete set of physical and nervous system examinations, including Alzheimer’s Disease Cooperative Study-Activities of Daily Living Scale (ADCS-ADL), the Neuropsychiatric Inventory-Clinician rating scale (NPI-C), and magnetic resonance imaging. Physicians determine whether the patient’s suffering AD based on these clinical symptoms, and only the subjects who diagnosed as AD by two physicians were included in this study.
Participants with dementia family history, vascular dementia, depressive, psychiatric disease, mental disorder, drug or alcohol abuse were excluded. The study was approved by The Ethics Committee of Qilu Hospital of Shandong University, and registered at the Chinese Clinical Trial Registry (No. ChiCTR2100042537).

2.2. Sample collection and cell-free DNA sample isolation

Peripheral blood (5–10 ml) was collected in EDTA anticoagulant tubes (BD). The samples were placed on ice and sent to the lab within 2 hours. Plasma samples were collected from peripheral blood after centrifugation for 10 min at 1600 g (4°C) and for 10 min at 16,000 g (4°C). The prepared plasma samples were immediately stored at -80°C until purification. cfDNA was extracted from 3–5 ml of plasma using the QIAamp Circulating Nucleic Acid Kit (Qiagen) according the manufacturer's protocol. Concentrations and quality of cfDNA were quantified by Qubit fluorometer (Life Technologies) and 2% agarose gel electrophoresis (Invitrogen).

2.3. 5hmC library preparation and high-throughput sequencing

5hmC libraries were prepared following a previously described selective chemical labeling technique [17]. Briefly, 10 ng cfDNA spiked with control oligos (0.01 pg of unmodified, methylated, or hydroxymethylated oligos per 10 ng cfDNA) was end repaired, 3'-adenylated, ligated to DNA Barcodes (Illumina Compatible), labeled with 5hmC and pulled down, and amplified with 14 cycles of PCR amplification using KAPA Hyper Prep Kit (Kapa Biosystems) according to the manufacturer's instructions. The PCR products were purified using 1X AMPure XP beads according to the manufacturer's instructions. The DNA concentration of each library was measured with a Qubit fluorometer (Life Technologies), and sequencing was performed on the Illumina NextSeq 500 platform.

2.4. Bioinformatics analysis and statistical analysis

All of the de-multiplexed sequencing reads passed filters was first trimmed to remove the low-quality bases and adaptor sequences by Trimmomatic (ver. 0.36) using the following parameters: ILLUMINAFLIP:adapter.f:2:30:7:1:TRUE; LEADING:3; SLIDINGWINDOW:4:15; TRAILING:3; and MINLEN:36. For the 5hmC-captured libraries, the trimmed reads were unique mapped to reference genome UCSC/hg19 by Bowtie2 with default parameters [18]. After filtering the duplicate reads, the mapping information for each reads pair was extracted and output as file of BED format for following analysis. Hydroxymethylated regions (hMRs) over corresponding input of each sample were identified using the peak calling software MACS (version 1.4.9) with default parameters [19]. Differential hMRs (DhMRs) analysis was performed using the PePr (version 1.1.1.18) [20]. We used normalized reads number per 50 bp among the genome of the hMRs to calculate the reads density normalized to one million reads in the library for each genomic position (Wig files). Screeshots of genomic regions were taken using the IGV genome browser. Genomic annotations of differential hMRs were performed by detecting hMRs overlapping each genomic region ≥ 1 bp with bedtools [21]. To functionally annotate putative DhMR genes, we conducted functional enrichment analysis for the identified DhMR genes using
the Fisher's test function in R. GO and KEGG pathways analysis was performed using custom perl scripts with hypergeometric test and WebGestalt (WEB-based GEne SeT AnaLysis Toolkit). The significance (p value) of the overrepresented GO terms or pathway was corrected by the Benjamini–Hochberg procedure. All statistical analysis were performed by SPSS 24.0 software and GraphPad Prism 8.0. Differences between control and AD group were analyzed using Student’s t-test. The correlation of the candidate diagnostic genes for AD with MMSE, MoCA, ADCS-ADL, and NPI-C scores were analyzed using Spearman’s correlation analysis. Receiver operating characteristic curve (ROC) analysis was used to calculate the area under curve (AUC) to preliminarily evaluate the predictive value of gene and AD. \( P<0.05 \) indicates significant.

3. Results

3.1 Basic characteristic of participants

Totally, 18 AD patients and 24 age-matched controls were included in the present study. Table 1 shows the basic characteristics of the study participants. There was no obvious difference in age, gender, total cholesterol (TC), triglyceride (TG), low density lipoprotein cholesterol (LDL-C), and high density lipoprotein cholesterol (HDL-C) of two groups. AD subjects had significant lower levels of cognitive function than control subjects as shown by MMSE and MoCA scores (\( P<0.001 \)).

3.2. Genome-wide 5hmC profiles of cfDNA between control and AD groups

According to the mapped rate and unique mapping rate, there were good sequencing quality observed between the two groups in this study (Table S1). To ask whether or not the cfDNA 5hmC profiles had difference in control and AD groups, we compared the distribution of 5hmC along the gene bodies and found that the overall normalized read density of cfDNA 5hmC was significantly different between the two groups (Figure. 1A). Compared with the controls, AD patients showed higher 5hmC level in gene bodies whereas it was depleted at transcription start site. Then, we analyzed 5hmC enrichment in different genomic characteristic regions and the overall genomic distribution of hMRs in all samples were shown in Figure. 1B. Genome-wide analysis of hMRs in the sequence data found that the majority are intragenic with most enrichment in exons, and depletion in intergenic regions (Figure. 1C), which was consistent with previous studies showing that the majority of 5hmC is enriched in intragenic regions of AD cases [14, 22]. The above results underscored that cfDNA 5hmC profiles of control and AD group indeed displayed significant differences.

3.3. DhMRs involved in AD pathogenesis associated signaling pathways

After compared 5hmC profiles between AD patients and controls, we identified 685 differentially hydroxymethylated regions (DhMRs) between control and AD groups (153 hypo-hydroxymethylated, 532 hyper-hydroxymethylated, \( p<0.05 \)). The list of annotated DhMR genes is shown in supplementary Table S2. To evaluate the classification effects of DhMRs for AD and control samples, we carried out the principal component analysis (PCA) for DhMRs and found that AD samples showed prominent
signatures and could be readily separated from control groups (Figure 2A). Thus, we speculated that DhMRs may have the potential to distinguish AD patients from cognitively normal individuals. Next, we performed KEGG pathway enrichment analysis and found that hyper-hydroxymethylated DhMRs annotated genes in AD samples were mainly distributed in AD pathogenesis-associated pathways, such as tight junction, MAPK signaling pathway, dopaminergic synapse, neurotrophin signaling pathway, and NF-kappa B signaling pathway (Figure 2B). In additional, hypo-hydroxymethylated DhMRs annotated genes in AD samples were also enriched in several AD-related pathways including fatty acid biosynthesis, axon guidance, pyruvate metabolism, and TCA cycle (Figure. 2C). Taken together, these results further indicate that the DhMRs play some roles in the pathogenesis of AD and can be used to separated AD patients from cognitively normal individuals.

3.4 Correlations of DhMRs with the MMSE and MoCA scores

As shown in Table S2, after correction for multiple testing \((q < 0.05)\), we identified 20 DhMRs and annotated to 19 distinct genes (2 hypo-hydroxymethylated, 17 hyper-hydroxymethylated). Of these, many genes are known to be associated with AD-related pathologic processes such as DNA damage, apoptosis, tau phosphorylation, neuronal plasticity, and mitochondrial function. We next assessed the ability of these 19 genes to classify dementia in AD patients with available clinical results. We analyzed the correlation between these 19 genes and MMSE, MoCA, ADCS-ADL, or NPI-C scores, respectively (Table S3). As shown in Table 2, we found 5 genes \((RABEP1, CPNE4, DNAJC15, REEP3, and ROR1)\) were negatively correlated with MMSE and MoCA scores \((r < -0.6, p < 0.001)\). Compared with control samples, the 5hmC level of these 5 genes \((RABEP1, CPNE4, DNAJC15, REEP3, and ROR1)\) in cfDNA were significant higher in AD patients (Figure 3A-E, all \(p < 0.01)\). In additional, we also found that another 4 genes \((CAMK1D, EFNA5, PBRM1, and CCDC141)\) were negatively correlated with the MMSE scores \((r < -0.6, p < 0.005)\) and MoCA scores \((r < -0.6, p < 0.01)\) of male participants. As shown in Table S3, it is interesting to note that \(CAMK1D\) was positively correlated with ADCS-ADL scores in male AD patients \((r = 0.731, p = 0.0396)\) and negatively correlated with the ADCS-ADL scores of female AD patients \((r = -0.632, p = 0.0498)\). An ROC curve was plotted to evaluate the diagnostic performance of genesin participants. As shown in Figure 3F, the AUC of \(RABEP1, CPNE4, DNAJC15, REEP3, and ROR1\) was 0.921 (sensitivity was 94.4%, specificity was 87.5%), 0.898 (sensitivity was 77.8%, specificity was 87.5%), 0.905 (sensitivity was 88.9%, specificity was 83.3%), 0.940 (sensitivity was 77.8%, specificity was 100.0%) and 0.912 (sensitivity was 88.9%, specificity was 83.3%), respectively. These results indicated that abnormal 5hmC level of the 5 genes in cfDNA might have diagnostic accuracy to distinguish AD patients from cognitively normal subjects.

3.5 DhMRs both detected in cfDNA and postmortem brain in AD patients

As we all known, the lesions of AD are in the brain. Therefore, it is important to verify the consistency between the DhMRs detected in cfDNA and postmortem brain of AD patients. There were not many studies on DNA 5hmC associated with AD in patients (mostly studies examined the total level of 5hmC in AD brain by immunohistochemistry methodology), only 1 study utilized the same method as we did to
detected the genome-wide profile of brain 5hmC at specific locus. Therefore, we compared our results with Zhao's results and found 14 DhMRs that were both detected in cfDNA in our study and dorsolateral prefrontal cortex tissue of AD patients [14]. As shown in Table 3, 13 DhMRs were associated with neuritic plaques (NP) and 3 DhMRs were associated with neurofibrillary tangles (NFTs). These DhMRs annotated to 11 NP-related genes and 3 NFTs-related genes, respectively. Only CAMK1D and RBFOX1 were both associated with NP and NFTs. Next, we analyzed the correlation between these 13 genes and MMSE, MoCA, ADCS-ADL, or NPI-C scores, respectively. As shown in Table 4, we found that only GNB1 was negatively correlated with the MMSE scores \(r = -0.682, p < 0.001\) and MoCA scores \(r = -0.732, p < 0.001\) of total participants. However, no significant correlations were observed between GNB1 and ADCS-ADL or NPI-C scores in AD patients \(p > 0.05\), Table S4). Similar to CAMK1D, RBFOX1 was positively correlated with the MMSE scores \(r = 0.714, p < 0.001\) and MoCA scores \(r = 0.635, p < 0.01\) of male participants. However, no significant correlations were observed between RBFOX1 and ADCS-ADL or NPI-C scores in AD patients \(all p > 0.05\). Compared with control samples, the 5hmC level of CAMK1D and GNB1 in cfDNA were significant higher in AD patients \(Figure 4A and B, p < 0.001\), but the 5hmC level of RBFOX1 in cfDNA was significant lower in AD patients \(Figure 4C, p < 0.05\). The AUC of GNB1, CAMK1D, and RBFOX1 was 0.977 (sensitivity was 94.4%, specificity was 95.8%), 0.850 (sensitivity was 100.0%, specificity was 66.7%), and 0.683 (sensitivity was 83.3%, specificity was 58.3%), respectively. These results further indicated that the 5hmC level of some AD-related genes (like CAMK1D) can be detected both in cfDNA of plasma and brain of AD patients, which may have significant potential as a non-invasive liquid biomarker of AD.

4. Discussion

5hmC is chemically stable and can be stored in fragmented cfDNA for potential detection using peripheral blood sampling [17, 23]. Growing evidence showed that a diagnostic model based on 5hmC profile of cfDNA showed the potential to assisted clinical diagnosis in different disease [24-26]. It is worth noting that 5hmC is highly abundant in the nervous system and apoptotic and necrotic synapses and neuros are typical characteristic of AD. These characteristics provide us with a greater opportunity to detect the signal features of cfDNA 5hmC in the blood, which could be reliable biomarkers for AD.

In this study, we utilized hmC-Seal sequencing method to detected cfDNA 5hmC signatures in AD cases and control subjects, so as to try to uncover reliable diagnostic biomarkers for AD. First, we found that cfDNA 5hmC profiles of control and AD group indeed displayed significant differences \(Figure 1\). To date, only 11 studies examining DNA hydroxymethylation associated with AD in postmortem human brain have been published [7]. Most of them examined genome wide levels of 5hmC using immunohistochemistry methodology, but results were inconsistent with some studies reporting decreased [27] whereas others showing increased [28] or no changes [29] in global 5hmC level in AD patients compared with controls, probably due to the sample size and the samples from different stages of AD. Only 2 papers utilized the next generation sequencing to detected the genome-wide profile of brain 5hmC at cortex tissue of AD patients, they found the majority of DhMRs had an increase in 5hmC located in intragenic regions in human AD brain tissue [14, 22], which was consistent with our findings. Second,
our results indicated that AD patients can be well distinguished from cognitively normal individuals by DhMRs (Figure 2A). Especially, we found that DhMRs annotated genes mainly distributed in AD pathogenesis related pathways including tight junction [30], metabolic related pathway (MAPK signaling pathway, pyruvate metabolism, and TCA cycle) [31], fatty acid homeostasis (biosynthesis and degradation) [32], axon guidance, and neurotrophin signaling pathway (Figure 2B and C). Third, we further identified 5 genes (RABEP1, CPNE4, DNAJC15, REEP3, and ROR1), which were negatively correlated with MMSE and MoCA scores, and the AUC of these 5 genes suggesting a good performance on the diagnosis of AD (Table 2 and Figure 3F). Many of the 5 distinct genes have been shown to be related to AD pathology in previous studies [33, 34]. For example, RABEP1 is involved in endocytic membrane fusion and membrane trafficking, and a recent genome-wide association study identified RABEP1 to be associated with increased AD risk [33]. CPNE4 plays a key role in neuroprotection and neurobehavioral development [34]. DNAJC15 is a mitochondrial protein that can regulate mitochondrial metabolism, the dysfunction of which has been associated with several neurologic disorders [35]. ROR1 plays an important role in regulating neurite extension, synapse formation, and synaptic transmission of hippocampal neurons [36]. Furthermore, we found that 12 DhMRs annotated genes were both detected in cfDNA and postmortem AD brain (Table 3), and CAMK1D and RBFOX1 were both associated with NP and NFTs. CAMK1D is involved in modulating neuronal development and plasticity [36] and late-onset AD pathogenesis [37]. RBFOX1 encodes a neuronal RNA-binding protein known to be expressed in neuronal tissues and involved in the pathogenesis of AD. A meta-analysis of amyloid positron emission tomographic imaging data collected on 4314 participants found that RBFOX1 localized around plaques and reduced expression of RBFOX1 was correlated with higher amyloid-β burden and global cognitive decline during life [38]. It is interesting to note that CAMK1D was negatively correlated with the MMSE scores and MoCA scores of male participants, which may be related to calcium regulated hormonal secretion [39]. However, RBFOX1 was positively correlated with the MMSE scores and MoCA scores of male participants. Gender-specific associations might result from gene-gene or gene environment interactions. It is possible that men are exposed to an environment that interacts with the RBFOX1 and/or CAMK1D. Taken together, these results further indicated that the 5hmC level of AD-related genes (like CAMK1D) in cfDNA from plasma present a great potential to clinical auxiliary diagnosis of AD.

To our best knowledge, this is the first research about genome-wide profiling of 5hmC in cfDNA to evaluate the potential for diagnosis of late-onset AD patients. Our findings indicating that AD patients can be well distinguished from cognitively normal subjects by 5hmC signatures in peripheral blood cfDNA. Importantly, we found RABEP1, CPNE4, DNAJC15, REEP3, ROR1, CAMK1D, and RBFOX1 had prediction diagnostic potential based on their significant correlations with MMSE and MoCA scores of participants. In addition, our study provides a new strategy for developing non-invasive peripheral diagnostic biomarkers for late-onset AD.

It is also important to note the limitations of our study. First, similar as most studies that identify AD biomarkers by epigenetic [5, 14, 40], our sample size is relatively small, and thus our results should be considered as a proof of concept rather than conclusive. Second, the origin of cfDNA is complex and remains poorly understood to date. Although the proportion of the cfDNA which released into blood from
apoptotic and necrotic synapses and neurons are increased, it will still be confounded by a lot of other cells from different tissues. Thus, our results could be confounded by cellular heterogeneity inherent. Hence, the profile of brain region-specific and/or neuro-specific cfDNA 5hmC should be investigated in future research. Finally, we cannot make a convincing causal interpretation due to the case-control design. However, the significant correlation between the identified genes and cognitive scale scores, as well as the fine diagnostic performance indicated in the preliminary ROC analysis make us be confident of the clinical auxiliary diagnosis in AD.

Conclusion

In summary, our results suggest that 5hmC signatures of plasma cfDNA can be served as an effective biomarkers for minimally invasive diagnosis of late-onset AD. Building on our study, prospective cohort study in AD population of larger sample size with longitudinal follow-up and sampling over several decades to establish the causality between 5hmC signatures and the onset and progress of AD are warranted in the future.

Abbreviations

5hmC: 5-Hydroxymethylcytosine; AD: Alzheimer’s disease; cfDNA: circulating cell-free DNA; DhMRs: differentially hydroxymethylated regions; MMSE: Mini-Mental State Examination; MoCA: Montreal Cognitive Assessment; ADCS-ADL: Alzheimer’s Disease Cooperative Study-Activities of Daily Living Scale; NPI-C: Neuropsychiatric Inventory-Clinician rating scale; ROC: Receiver operating characteristic curve; AUC: Area under curve

Declarations

Acknowledgements

We thank Dr. Chunxiao Song at Ludwig Institute for Cancer Research, Nuffield Department of Medicine, University of Oxford, for help on cfDNA 5hmC chemical labeling.

Author's contributions

Lei Chen conceived and designed the study. Lei Chen, Shunliang Xu, Qianqian Shen and Hongzhuan Yu acquired, analysed, and interpreted the data. Shunliang Xu, Hongzhuan Yu, Shengjie Pei, Yangting Zhang, and Xin He obtained informed consent and acquired samples and clinical information. Lei Chen and Qiuzhen Wang wrote the draft of the manuscript. All authors critically read and revised the manuscript and are accountable for all aspects of the work. Qiuzhen Wang and Duo Li supervised the study. All the authors read and approved the final manuscript.

Funding
This work was supported by the National Natural Science Foundation of China (NSFC, No. 81472983), Natural Science Foundation of Shandong Province (ZR2020QH294, ZR2015HM024, 2019GSF108066), and Qingdao Postdoctoral Research Grant (RZ2000002906).

**Availability of data and materials**

The datasets used and analysed during the current study are available from the corresponding authors on reasonable request.

**Ethics approval and consent to participate**

The study was approved by The Ethics Committee of Qilu Hospital of Shandong University, and registered at the Chinese Clinical Trial Registry (No. ChiCTR2100042537).

**Consent for publication**

Not applicable

**Conflicts of interest**

The authors have no conflicts of interest.

**References**


**Tables**

Due to technical limitations, table 1-4 is only available as a download in the Supplemental Files section.

**Figures**

**Figure 1**

Genome-wide distribution of 5hmC in cfDNA of AD patients and control individuals. (A) 5hmC densities at TSS (± 2 kb) and TES (± 2 kb) in control and AD subjects. (B) Pie chart shows the overall genomic
Performance of potential cfDNA 5hmC for identification of AD patients and control individuals. (A) Principle component analysis (PCA) plot of 5hmC FPKM in cfDNA from AD and control samples. (B) KEGG enrichment analysis of DhMRs annotated genes with significant 5hmC increase in AD samples. (C) KEGG enrichment analysis of DhMRs annotated genes with significant 5hmC decrease in AD samples.
Figure 3

Performance of distinct genes for diagnosis and prediction of AD. Boxplots of (A) RABEP1, (B) CPNE4, (C) DNAJC15, (D) ROR1, and (E) REEP3 5hmC FPKM in cfDNA samples. (E) A receiver operating characteristic curve for AD patients based on 5hmC level of RABEP1, CPNE4, DNAJC15, ROR1, and REEP3. **p< 0.01, ***p< 0.001, ****p< 0.0001.
Figure 4

Cell-free 5hmC for diagnosis and prediction of AD. Boxplots of (A) CAMK1D, (B) GNB1, and (C) RBFOX1 5hmC FPKM in cfDNA samples. (D) A receiver operating characteristic curve for AD patients based on 5hmC level of CAMK1D, GNB1 and RBFOX1. *p< 0.05, ****p< 0.0001.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.