



Detecting Brain Structure-Specific Methylation Signatures and Rules for Alzheimer's Disease

ZhanDong Li^{1†}, Wei Guo^{2†}, Tao Zeng^{3†}, Jie Yin^{4,5}, KaiYan Feng⁶, Tao Huang^{3,7*} and Yu-Dong Cai^{8*}

¹ College of Food Engineering, Jilin Engineering Normal University, Changchun, China, ² Key Laboratory of Stem Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai Jiao Tong University School of Medicine, Shanghai, China, ³ Bio-Med Big Data Center, CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Shanghai, China, ⁴ Cancer Institute, Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China, ⁵ Department of Human Genetics, Institute of Genetics, Zhejiang University School of Medicine, Zhejiang University, Hangzhou, China, ⁶ Department of Computer Science, Guangdong AIB Polytechnic College, Guangzhou, China, ⁷ CAS Key Laboratory of Tissue Microenvironment and Tumor, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Shanghai, China, ⁸ School of Life Sciences, Shanghai University, Shanghai, China

OPEN ACCESS

Edited by:

Wei Kong,
Shanghai Maritime University, China

Reviewed by:

Qi Dai,
Zhejiang Sci-Tech University, China
Shuaiqun Wang,
Shanghai Maritime University, China

*Correspondence:

Tao Huang
tohuangtao@126.com
Yu-Dong Cai
cai_yud@126.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Neurodegeneration,
a section of the journal
Frontiers in Neuroscience

Received: 13 March 2022

Accepted: 11 April 2022

Published: 02 May 2022

Citation:

Li Z, Guo W, Zeng T, Yin J,
Feng K, Huang T and Cai Y-D (2022)
Detecting Brain Structure-Specific
Methylation Signatures and Rules
for Alzheimer's Disease.
Front. Neurosci. 16:895181.
doi: 10.3389/fnins.2022.895181

Alzheimer's disease (AD) is a progressive disease that leads to irreversible behavioral changes, erratic emotions, and loss of motor skills. These conditions make people with AD hard or almost impossible to take care of. Multiple internal and external pathological factors may affect or even trigger the initiation and progression of AD. DNA methylation is one of the most effective regulatory roles during AD pathogenesis, and pathological methylation alterations may be potentially different in the various brain structures of people with AD. Although multiple loci associated with AD initiation and progression have been identified, the spatial distribution patterns of AD-associated DNA methylation in the brain have not been clarified. According to the systematic methylation profiles on different structural brain regions, we applied multiple machine learning algorithms to investigate such profiles. First, the profile on each brain region was analyzed by the Boruta feature filtering method. Some important methylation features were extracted and further analyzed by the max-relevance and min-redundancy method, resulting in a feature list. Then, the incremental feature selection method, incorporating some classification algorithms, adopted such list to identify candidate AD-associated loci at methylation with structural specificity, establish a group of quantitative rules for revealing the effects of DNA methylation in various brain regions (i.e., four brain structures) on AD pathogenesis. Furthermore, some efficient classifiers based on essential methylation sites were proposed to identify AD samples. Results revealed that methylation alterations in different brain structures have different contributions to AD pathogenesis. This study further illustrates the complex pathological mechanisms of AD.

Keywords: Alzheimer's disease, brain structure, DNA methylation, machine learning, quantitative rules

INTRODUCTION

Alzheimer's disease (AD) is a progressive disease that initiates the pathogenesis of dementia (Rogers et al., 1996; Karch and Goate, 2015). At the onset, AD symptoms are quite mild, something as seemingly innocuous as mild and gradual memory loss. However, AD can quickly progress to negatively affect not only cognitive ability but also motor skills, thereby reducing the quality of life of patients with this disease, especially the elderly (Reisberg et al., 1987; Weller and Budson, 2018; Atri, 2019). Clinically and by definition, the most common symptom of early-stage AD is a reduction in memory capacity (Weller and Budson, 2018; Atri, 2019). As AD progresses, it may also lead to irreversible behavioral changes, erratic emotions, and loss of motor skills. These conditions make people with AD difficult or almost impossible to take care of.

In 2014, over 5 million people suffered from AD in the United States (Murphy et al., 2015; Taylor et al., 2017). In general, AD symptoms manifest after 65 years of age. Moreover, the incidence of suffering from this disease increases with age, indicating that the sooner the first AD symptoms appear, the more severe the disease would be. Another study estimated that over 200,000 people in the United States are already suffering from advanced-stage AD (Matthews et al., 2019), suggesting that the people who develop this disease are younger than 65 years old and the symptoms are more severe than they should be. These trends make AD one of the most serious public health concerns.

Multiple internal and external pathological factors may affect or even trigger the initiation and progression of AD. DNA methylation is one of the most effective regulatory roles during AD pathogenesis (Mastroeni et al., 2010; De Jager et al., 2014; Qazi et al., 2018). A systematic review summarized the specific roles of epigenetic changes, including methylation alterations, during AD pathogenesis (Mastroeni et al., 2010). Methylation participates in the advancement of multiple progressive diseases, such as cancer and AD (Mastroeni et al., 2010). Multiple demethylation alterations further affect AD pathogenesis *via* regulating downstream proteins, such as *DNMT1*, *MBD2*, and *p66-alpha*, especially in specific brain structures, i.e., the entorhinal cortex (ERC; Mastroeni et al., 2010). This review indicated that methylation contributes to AD pathogenesis in specific brain structures.

Other studies also explored potentially pathological methylation alterations in the different brain structures of people with AD. Methylation status seems to consistently decrease in the hippocampus (HIPPO) during AD progression (Chouliaras et al., 2013). Other researchers focused on the status of methylation in the middle temporal gyrus (Coppieters et al., 2014). Unlike in the HIPPO, DNA modifications as implied by levels of 5-mc and 5-hmc continuously increase in the middle temporal gyrus of patients with AD. Another team identified specific methylation statuses at potentially effective genes, such as *ANK1*, *BIN1*, and *RHBDF2*, which have a specific methylation status during AD initiation (De Jager et al., 2014). Their results implied that some unique genomic loci rather than the entire genome may also be examined to effectively monitor and predict AD progression. The methylation of various candidate genes at specific brain structures may be strongly associated with AD

progression and can be potentially used in clinically predicting AD pathogenesis and progression.

Although multiple loci associated with AD initiation and progression have been identified, the spatial distribution patterns of DNA methylation in the brain of people with AD have not been clarified. A team systematically analyzed the status of methylation in different brain structural regions from more than 60 samples (Semick et al., 2019). The said study provides us four comprehensive datasets for exploring alterations in methylation status in various brain regions at loci that are potentially associated with AD. As the powerful machine learning algorithms can deeply analyze hidden and complicated relationships in a big dataset (Dai et al., 2016; Kong et al., 2020; Yang et al., 2020), we adopted them to investigate above-mentioned comprehensive datasets to identify a group of candidate loci potentially associated with AD that regulate methylation and show structural specificity. In our previous studies, we analyzed tissue-specific methylation that is associated with the pathogenesis of complex diseases, such as cancer (Pan et al., 2019) and viral infections (Chen et al., 2019). Herein, we extend our understanding of methylation in specific tissues by profiling the methylation of brain diseases, which in this case is AD. The above-mentioned four datasets were related to four major brain structures: dorsolateral prefrontal cortex (DLPFC), HIPPO, ERC, and cerebellum (CRB). Previous studies confirmed that these structures are correlated with brain functions associated with AD: decision making and working memory (DLPFC; Heekeren et al., 2006), spatial memory and navigation (HIPPO; Broadbent et al., 2004), neuron information processing and recording (ERC; Keene et al., 2016), and learning (CRB; Raymond et al., 1996). This study focused on these structures because of their functional correlations with AD pathogenesis. On each dataset, we applied several multiple machine learning algorithms to investigate it. The Boruta feature filtering (Kursa and Rudnicki, 2010) and max-relevance and min-redundancy (mRMR; Peng et al., 2005) methods were performed one by one. Some essential methylation features were extracted and ranked in a feature list. This list was fed into the incremental feature selection (IFS; Liu and Setiono, 1998) method that integrated three classification algorithms: random forest (RF; Breiman, 2001), support vector machine (SVM; Cortes and Vapnik, 1995), partial decision tree (PART) (Frank and Witten, 1998), to identify candidate AD-associated methylation features with structural specificity, set up a group of quantitative rules, which can reveal the effects of DNA methylation in various brain regions (i.e., four brain structures) on AD pathogenesis. In addition, on each brain region, some efficient classifiers were built to identify AD samples. Our results revealed that methylation alterations in different brain structures have different contributions to AD pathogenesis. This study further illustrates the complex pathological mechanisms of AD.

MATERIALS AND METHODS

Data

The DNA methylation profiles of AD and control samples from postmortem brain donors' four brain regions: CRB, DLPFC, ERC, and HIPPO, were downloaded from Gene Expression Omnibus

under accession number of GSE125895 (Semick et al., 2019). In CRB dataset, there were 24 AD and 43 control samples. In DLPFC dataset, there were 21 AD and 47 control samples. In ERC dataset, there were 20 AD and 49 control samples. In HIPPO dataset, there were 17 AD and 48 control samples. The DNA methylation levels were measured with Illumina HumanMethylation450 BeadChip and represented with normalized beta values processed with the R/Bioconductor package `minfi`¹. The processed data of 420,852 methylation probes in 269 samples were used for further analysis.

Boruta Feature Filtering

As mentioned in section “Data,” all samples in four datasets were represented by a large number of methylation features. Evidently, not all of these features give contributions for identification of AD samples. It is necessary to exclude irrelevant features first.

To quick complete this task, we employed the Boruta feature filtering (Kursa and Rudnicki, 2010). It is a powerful feature selection method based on RF. The Boruta method involves several steps. First, shuffled data from copies of the original data are created. Second, the RF is trained on the original and shuffled data to obtain feature importance. Third, the *Z* score for each feature is computed on the basis of this feature’s importance score. Fourth, a feature is tagged as important if and only if its *Z* score is greater than that of shadow features. Finally, these steps are repeated for all features. The features tagged as important are outputted as the outcome of this method, which are deemed to be relevant features.

In this study, we adopted the Boruta program obtained from https://github.com/scikit-learn-contrib/boruta_py, which was performed on each dataset with its default parameters. Selected features would be analyzed by the following mRMR method.

Max-Relevance and Min-Redundancy Feature Selection

For the remaining methylation features, we further used mRMR method (Peng et al., 2005) to analyze them. This method calculates the mutual information (MI) between features and class labels for evaluating feature relevance and MI between features to assess the redundancy between features. Its original purpose was to find out a feature subset that had maximum relevance to class labels and minimum redundancies between features in this set. However, such problem is NP-hard. In view of this, it adopts a heuristic way. In this way, investigated features are sorted in a feature list, where important features are assigned high ranks. In the beginning, this list is empty. The method repeatedly selects one features from the remaining features such that it has maximum relevance to class labels and minimum redundancies to features already in the list. Such procedure stops until all features are in the list. This list is called mRMR feature list.

The present study used the mRMR program retrieved from <http://penglab.janelia.org/proj/mRMR/>. It was applied on each dataset, in which samples were represented by features selected

by Boruta. Accordingly, one mRMR feature list was obtained on each dataset.

Incremental Feature Selection

Although mRMR method deeply analyzed the importance of each methylation feature, it was still a problem which features can be selected. Thus, we employed the IFS method (Liu and Setiono, 1998) to determine the optimal number of features according to the integrating supervised classification algorithm. In brief, a series of feature subsets are extracted from the mRMR feature list yielded by the mRMR method at a step interval of 1. For instance, the first feature subset has the top ranked feature, and the second feature subset has the top 2 ranked features, and so on. For each feature subset, a classifier is trained based on the training data induced from this feature subset and the given classification algorithm. All classifiers are evaluated by 10-fold cross-validation (Kohavi, 1995). When all classifiers have been assessed, the best one can be discovered. This classifier is called the optimal classifier and the features used in such classifier are termed as the optimal features.

Classification Algorithm

To execute the IFS method, one classification algorithm is necessary. In this study, we used the following three classification algorithms: RF (Breiman, 2001), SVM (Cortes and Vapnik, 1995), and PART (Frank and Witten, 1998).

Random Forest

Random forest is widely applied in analyzing biological and biomedical data. Several previous studies indicate the satisfactory performance of RF (Pan et al., 2010; Zhao et al., 2018; Jia et al., 2020; Chen et al., 2021, 2022; Ding et al., 2022; Li Z. et al., 2022; Wu and Chen, 2022; Zhou et al., 2022). RF is a meta-classifier because it consists of numerous decision trees. Each decision tree in RF is learned from a bootstrap training data induced from a randomly selected feature subset. For an input sample, each decision tree provides its prediction. RF integrates these predictions by majority voting. To quickly implement RF, the tool “RandomForest” in Weka (Frank et al., 2004) was adopted. For convenience, such tool was executed using default parameters.

Support Vector Machine

Support vector machine is a classic and powerful classification algorithm. At present, it is always an important candidate for constructing efficient classifiers (Chen et al., 2017; Zhou et al., 2020a,b; Liu et al., 2021; Zhu et al., 2021; Li X. et al., 2022). In mathematical terms, the main task of SVM is to transform non-linear data from the original low-dimensional data space to a new high-dimensional data space, where the new data would be linear as guaranteed by a certain kernel trick. SVM then determines support vectors on the data margin between samples from two classes, which can represent a hyperplane in high-dimensional data space to predict the class type of new samples. Likewise, we also used the tool “SMO” in Weka (Frank et al., 2004) to implement SVM. Such SVM is optimized by the sequence minimization optimization algorithm (Platt, 1998).

¹<https://bioconductor.org/packages/minfi/>

PART Rule Learning

Based on one of above two algorithms, investigators can build efficient classifiers. However, they are absolute black-box algorithms. Their classification principles are difficult to be understood. Few clues can be extracted from them to uncover differences between AD and control samples at the methylation level. In view of this, a rule learning algorithm, PART (Frank and Witten, 1998) was employed. Different from other rule learning algorithms, PART can learn rules at a time. Hence, PART is a simpler and more efficient approach without global optimization. In general, PART can produce many partial decision trees and combine the rules from these decision trees *via* the separate-and-conquer technique. Similar to RF and SVM, we also used one tool in Weka to implement PART. The tool was named as “PART.”

Synthetic Minority Oversampling Technique

In each dataset, control samples were about twice as many as AD samples. It is better to balance samples in these two classes. To this end, the widely used method, synthetic minority oversampling technique (SMOTE; Chawla et al., 2002) was adopted, which is an over-sampling method. It iteratively produces new samples for the minor class until the minor class has equal number of samples in major class. In detail, it first randomly selects a sample, say x , in the minor class. Then, a number of samples in this class are picked up such that these samples have smallest distance to x . The randomly selected sample, say y , from above samples and x are used to generate the new sample z , which is defined as the linear combination of x and y . The tool “SMOTE” in Weka was applied in this study. The SMOTE was only used in evaluation of performance of classifiers. The samples generated by it did not participate in evaluating the importance of methylation features.

Performance Evaluation

Matthew correlation coefficients (MCC) is a widely used measurement for binary classification (Matthews, 1975; Zhao et al., 2019; Liang et al., 2020; Zhang et al., 2021b; Yang and Chen, 2022), which was adopted herein to assess the performance of the different classifiers within a 10-fold cross-validation (Kohavi, 1995; Zhao et al., 2019; Zhang et al., 2021a; Tang and Chen, 2022). The MCC can be computed as follows:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}, \quad (1)$$

where TP , TN , FP , and FN indicate the sample number of true-positive, true-negative, false-positive, and false-negative predictions, respectively. MCC values range from -1 to +1 and achieves +1 when the classification model has the best performance.

In addition, we also calculated other five measurements: sensitivity (SN), specificity (SP), prediction accuracy (ACC), precision and F1-measure, to fully display the performance of different classifiers. They can be computed by

$$SN = \frac{TP}{TP + FN} \quad (2)$$

$$SP = \frac{TN}{TN + FP} \quad (3)$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$F1 - measure = \frac{2 \times TP}{2 \times TP + FN + FP} \quad (6)$$

RESULTS

In this study, the DNA methylation dataset on four brain structures of AD and control samples was deeply analyzed. The entire procedures are illustrated in **Figure 1**. This section gave the detailed results produced by such procedures.

Results of Boruta and Max-Relevance and Min-Redundancy Methods

For each brain structure, one methylation dataset was collected. As a large number of methylation features were involved in each dataset, the Boruta method was first applied on each dataset to exclude irrelevant features. Important methylation features remained for each dataset, which are provided in **Supplementary Table 1**. The number of remaining features on each dataset is listed in **Table 1**. It can be observed that less than 200 features were kept on each dataset, greatly reducing the scope of investigated methylation features. These features were deemed to be essential for distinguishing AD patients from control samples on corresponding brain structure.

The features selected by Boruta method were further analyzed by the mRMR method according to **Figure 1**. One mRMR feature list was produced on each dataset, which is also provided in **Supplementary Table 1**. Such list would be investigated in the following analysis.

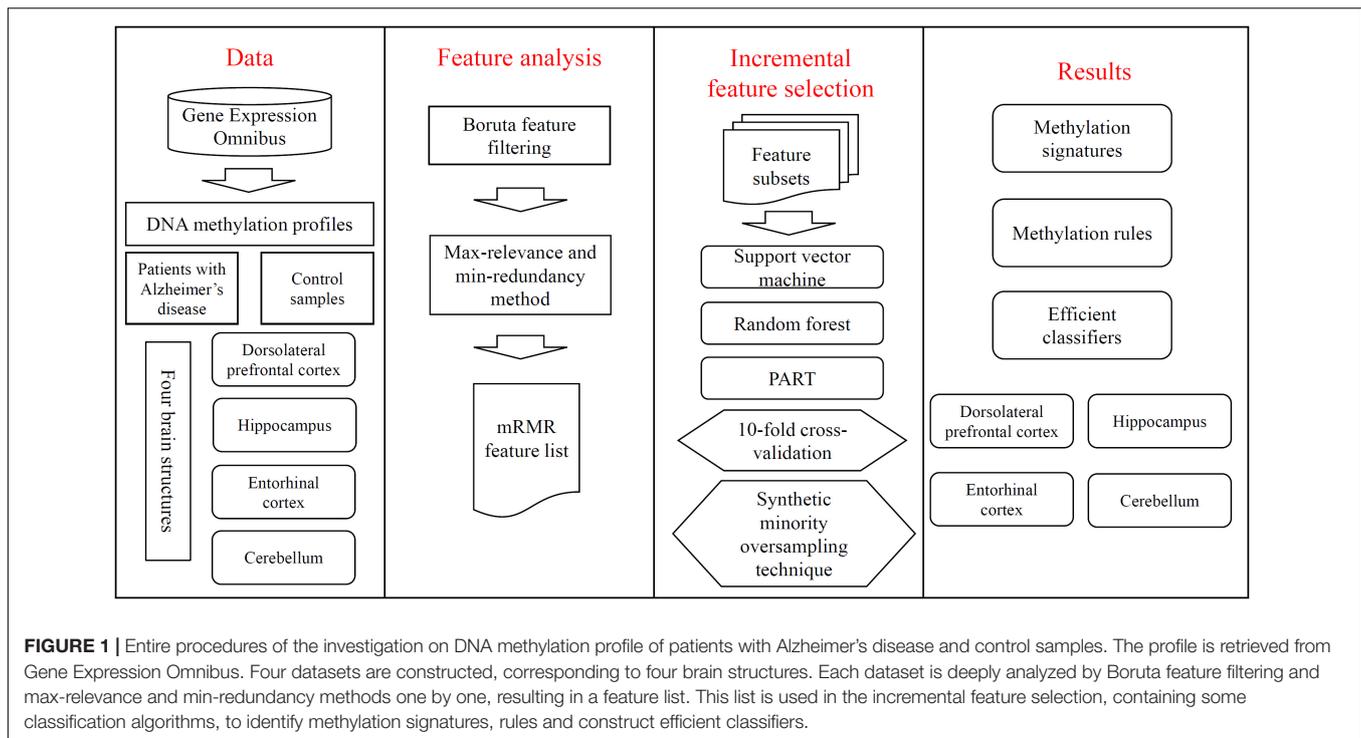
Incremental Feature Selection Results on Four Brain Regions

The mRMR method produced an mRMR feature list on each dataset, corresponding to one brain structure. Then, the IFS method was applied on such list, which integrated three classification algorithms.

For the CRB dataset, the mRMR feature list contained 124 methylation features. Accordingly, 124 feature subsets were

TABLE 1 | Number of methylation features selected by Boruta method on each dataset.

Dataset on brain structure	Number of methylation features
CRB dataset	124
ERC dataset	158
DLPFC dataset	149
HIPPO dataset	132



generated in the IFS method by using step one. On each subset, a classifier was built with one of the three classification algorithms, which was further assessed by 10-fold cross-validation. The predicted results were counted as measurements listed in section “Performance Evaluation,” which are provided in **Supplementary Table 2**. To clearly display the performance of the classifiers based on one classification algorithm and different feature subsets, one IFS curve was plotted for each classification algorithm by setting MCC as Y-axis and number of used features as X-axis, as shown in **Figure 2**. It can be observed that the RF, SVM, and PART produced the highest MCC values of 0.968, 0.936, and 0.846, respectively. These values were obtained by using top 38, 16, and 6, respectively, features in the mRMR feature list. These features were the optimal ones for corresponding classification algorithm. Accordingly, the optimal RF/SVM/PART classifier can be built using corresponding optimal features. The ACC and F1-measure of these classifiers are listed in **Table 2**, whereas the other three measurements: SN, SP, and precision, are illustrated in **Figure 3A**. Evidently, the optimal RF classifier was best, followed by the optimal SVM classifier and optimal PART classifier.

For the other three datasets (ERC, DLPFC, and HIPPO datasets), the same IFS procedures were conducted. Detailed IFS results are provided in **Supplementary Tables 3–5**. Three IFS curves, corresponding to three classification algorithms, were plotted, as shown in **Figures 4–6**. From **Figure 4** for ERC dataset, we can see that RF/SVM/PART yielded the highest MCC of 0.966/0.966/0.934, which was based on top 1/4/2 features in the list. Accordingly, we can build three optimal classifiers with corresponding features. Other measurements of these classifiers are provided in **Table 2** and **Figure 3B**. Clearly, the optimal RF and SVM classifiers gave equal performance,

whereas they were all superior to the optimal PART classifier. For the IFS results on DLPFC dataset, it can be observed from **Figure 5** that three classification algorithms generated the highest MCC values of 0.967, 1.000, and 0.868, respectively. Top 6, 20, and 2, respectively, features in the list were adopted to produce these MCC values, which comprised the optimal features for the corresponding classification algorithm. With these optimal features, three optimal classifiers were built. Their detailed performance is listed in **Table 2** and **Figure 3C**. The optimal SVM classifier gave the best performance, followed by the optimal RF and PART classifiers. As for the last dataset (HIPPO dataset), the highest MCC value for RF/SVM/PART was 0.926/1.000/0.920, which can be observed from **Figure 6**. This MCC was obtained by using top 15/18/34 features in the list. They were optimal features of corresponding classification

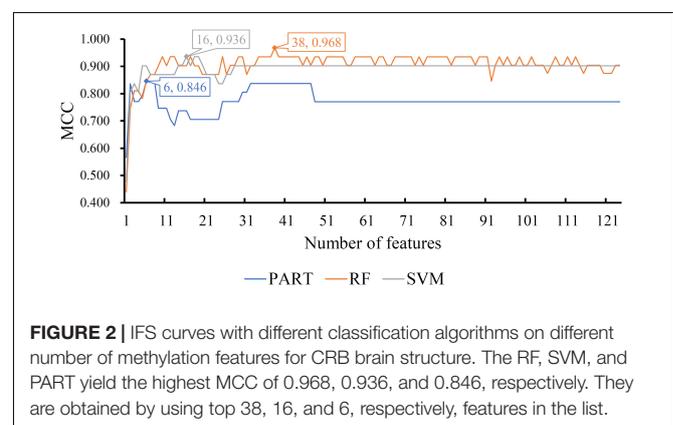


TABLE 2 | Performance of the optimal classifiers based on different classification algorithms on four brain structures.

Dataset of brain structure	Classification algorithm	Number of features	ACC	MCC	F1-measure
CRB dataset	RF	38	0.985	0.968	0.979
	SVM	16	0.970	0.936	0.957
	PART	6	0.925	0.846	0.902
ERC dataset	RF	1	0.986	0.966	0.976
	SVM	4	0.986	0.966	0.976
	PART	2	0.971	0.934	0.952
DLPFC dataset	RF	6	0.985	0.967	0.977
	SVM	20	1.000	1.000	1.000
	PART	2	0.941	0.868	0.909
HIPPO dataset	RF	15	0.969	0.926	0.944
	SVM	18	1.000	1.000	1.000
	PART	34	0.969	0.920	0.938

algorithm. The optimal RF/SVM/PART classifier was constructed with its optimal features. **Table 2** and **Figure 3D** summarizes the detailed performance of these three optimal classifiers. Evidently, the optimal SVM classifier was best, whereas other two optimal classifiers were almost at the same level.

Based on the above IFS results, we can select a best classifier on each dataset, which can be an efficient tool for identification of AD samples based on a certain brain structure.

Classification Rules on Four Brain Regions

On each brain structure, the optimal PART classifier always provided the lowest performance. However, it can provide more clues than other two optimal classifiers to uncover different methylation patterns within a certain brain structure

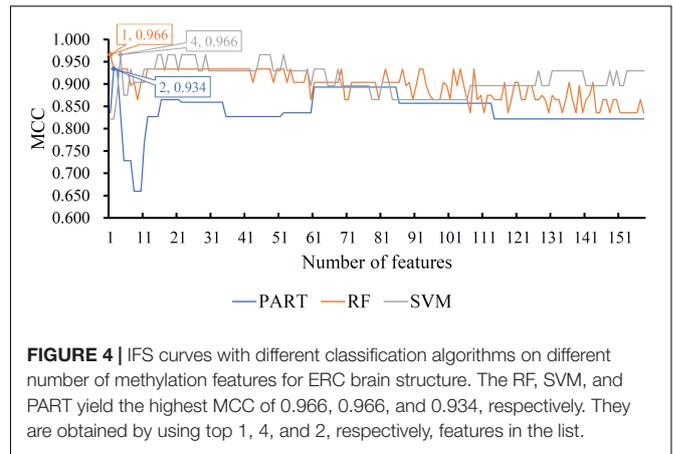


FIGURE 4 | IFS curves with different classification algorithms on different number of methylation features for ERC brain structure. The RF, SVM, and PART yield the highest MCC of 0.966, 0.966, and 0.934, respectively. They are obtained by using top 1, 4, and 2, respectively, features in the list.

of AD patients and control samples. By employing the optimal features used in the optimal PART classifier, PART learned rules based on all AD and control samples. These rules are listed in **Table 3**. Two rules were obtained for CRB brain structure. For other three brain structures, four, two and two, respectively, rules were accessed. These rules clearly displayed the different patterns on methylation levels between AD and control samples, which were helpful to improve our understanding on AD.

DISCUSSION

In this study, several computational methods were adopted to analyze the DNA methylation profile of AD and control samples. As pathological methylation alterations may occur in the different brain structures, the methylation profile on four brain structures was investigated. Some key methylations and rules

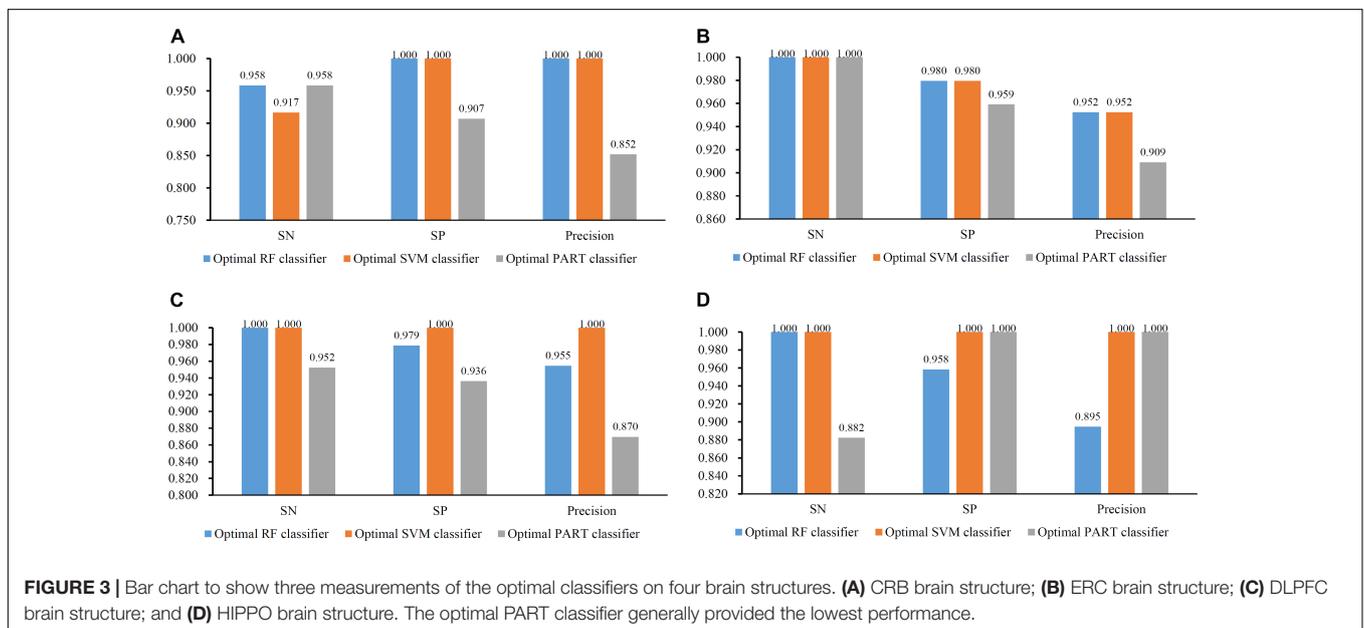


FIGURE 3 | Bar chart to show three measurements of the optimal classifiers on four brain structures. **(A)** CRB brain structure; **(B)** ERC brain structure; **(C)** DLPFC brain structure; and **(D)** HIPPO brain structure. The optimal PART classifier generally provided the lowest performance.

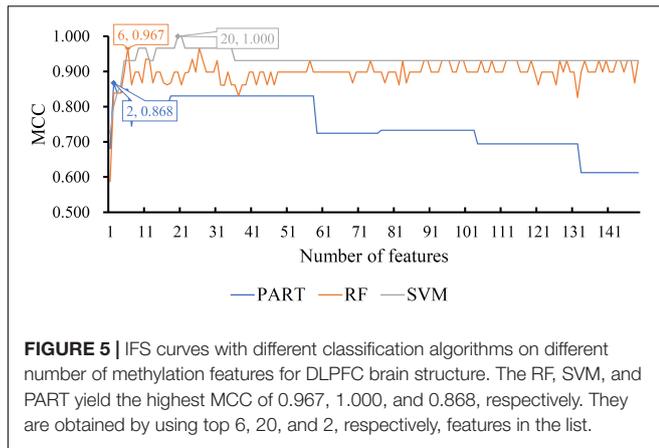


FIGURE 5 | IFS curves with different classification algorithms on different number of methylation features for DLPFC brain structure. The RF, SVM, and PART yield the highest MCC of 0.967, 1.000, and 0.868, respectively. They are obtained by using top 6, 20, and 2, respectively, features in the list.

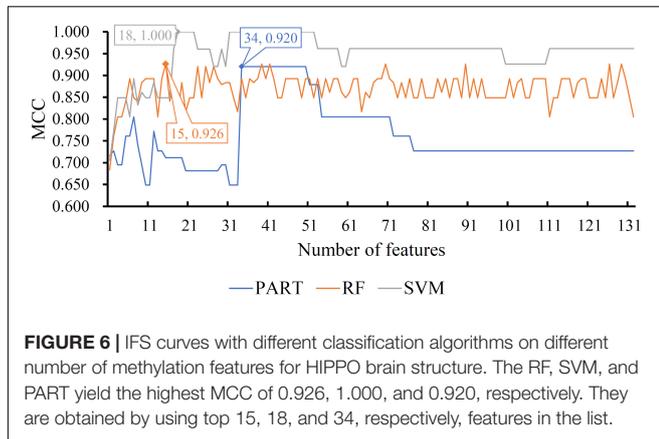


FIGURE 6 | IFS curves with different classification algorithms on different number of methylation features for HIPPO brain structure. The RF, SVM, and PART yield the highest MCC of 0.926, 1.000, and 0.920, respectively. They are obtained by using top 15, 18, and 34, respectively, features in the list.

TABLE 3 | Classification rules generated by PART for four brain structures.

Brain structure	Rule index	Condition	Result
CRB	Rule-1	(cg15926585 > 0.2202) AND (cg08105590 > 0.1467) AND (cg07094785 > 0.7870)	Control
	Rule-2	Others	AD
ERC	Rule-1	cg05810363 ≤ 0.8365	Control
	Rule-2	cg05810363 > 0.8793	AD
	Rule-3	cg08462501 > 0.8234	AD
	Rule-4	Others	Control
DLPFC	Rule-1	(cg14622549 ≤ 0.8740) AND (cg21097354 ≤ 0.2427)	Control
	Rule-2	Others	AD
HIPPO	Rule-1	(cg13076843 > 0.8321) AND (cg06639320 > 0.3468)	AD
	Rule-2	Others	Control

were obtained, as listed in section “Results.” Here, the analyses on some methylations and rules were conducted.

Methylation Signatures and Rules From the Cerebellum

Cerebellum mainly contributes to learning and the establishment of theories and computational models. Given that a reduction

in learning capacity is a typical symptom of AD (Sultana et al., 2006), identifying the methylation alterations during AD pathogenesis in the CRB is necessary. In our prediction list, the top methylation locus was **cg07094785**, which targets *HNRNPUL2*. Although only a few studies directly examined the function of this gene, it reportedly participates in AD pathogenesis, for example, by regulating the accumulation of tau protein in brain (Vanderweyde et al., 2016). Given that CRB tissues are highly specific, this gene was found to effectively function during nerve system degeneration due to age in multiple animal models, including mice and rats (Jasien et al., 2014; Kalinin et al., 2017). The other top-ranked locus was **cg08105590**, which targets *FAM38A*. This locus can also be associated with AD pathogenesis in the CRB at the methylation level. This gene might be correlated with AD by regulating *Piezo1* channels (Velasco-Estevéz et al., 2018), which has been shown to have a tight correlation with AD initiation and progression. Given that such correlations were validated using cerebellar cells (Velasco-Estevéz et al., 2018), we can reasonably speculate that they can be observed in the CRB *in vivo*.

The first parameter for quantitative rules is **cg15926585**, which targets the intermediate region of *COMT* and *TXNRD2*. These two genes are both highly negatively correlated with the abnormal behavioral symptoms of AD (Borroni et al., 2006; Jin et al., 2017). Methylation of this region may inhibit the expressions of these genes and further trigger AD initiation and progression. Two other parameters that target *HNRNPUL2* and *FAM38A* are functionally correlated with AD. Hypermethylation of these genes has also been demonstrated to participate in the development of diseases associated with nerve disorders (Pathak et al., 2019; Gürkan et al., 2020) but not with AD. Therefore, these quantitative rules in the CRB were explained according to publications above, demonstrating the effectiveness of our analysis.

Methylation Signatures and Rules From the Entorhinal Cortex

Entorhinal cortex is a region that controls memory, navigation, and time perception (Goyal et al., 2018; Montchal et al., 2019). ERC is reportedly correlated with AD initiation and progression (Marzi et al., 2018; Grubman et al., 2019; Petrache et al., 2019). In our predicted feature list, the first locus is **cg05810363**, which targets *RHBDF2*. This gene is hypermethylated in HIPPO cells during AD pathogenesis (De Jager et al., 2014). Similar methylation patterns were recently identified in the ERC *via* interacting with *H3K27AC* (Marzi et al., 2017; Marzi et al., 2018). The methylation status of another locus, namely, **cg08462501**, which interacts with *TBC1D1*, is predicted to be downregulated during AD pathogenesis. Given that the ERC is correlated with a specific biomarker called TDP-43, this region could be hypomethylated during AD pathogenesis (Sun et al., 2017). Therefore, *TBC1D1* may also present methylation status in the ERC similar to that in HIPPO.

Three quantitative rules, including the two features discussed in this section, i.e., the hypermethylated gene *RHBDF2* and

the hypomethylated gene *TBC1D1*, may contribute to AD pathogenesis. This contention was supported by previous reports.

Methylation Signatures and Rules From Dorsolateral Prefrontal Cortex

Dorsolateral prefrontal cortex mainly controls the decision-making and working memory functions of the human brain. Given that people with AD show a decrease in the ability to render judgment (Heekeren et al., 2006) and a decline in working memory at late stages (Mars and Grol, 2007), we can reasonably regard this region as a potential target to identify methylation effects associated with AD pathogenesis. A locus predicted by our computational analysis is **cg14622549**, which targets the gene body region of *EP400*. As early as 2013, *EP400* has already been regarded as a potential biomarker for detecting AD at the transcriptomics level (Desire et al., 2013; Trifonova et al., 2022). Methylation of *EP400* is reportedly functionally correlated with autism, another disease of the nervous system that affects DLPFC. This supposition suggests the potential pathological effects of *EP400* and supports our contention that methylation affects specific brain structures. The next locus is **cg21097354**, which targets the intermediate region of *TTYH3* and *AMZ1*. *AMZ1* has been shown to be a potential regulator for nerve system, especially in the elderly (Swingler et al., 2009; Panossian et al., 2018). Given that AD is a typical disease of the nervous system that is associated with aging, we can reasonably predict the potential relationships between DLPFC and AD. On the basis of a monozygotic twin model, a study suggested that the demethylation of *AMZ1* in DLPFC is associated with dyskinesia, another potential symptom of AD (Soerensen et al., 2019). This report further confirmed the pathogenesis and downstream pathogenic regions of this methylation alteration.

The two parameters in the only classification rule were the two aforementioned loci. Their demethylation would aid in the identification of patients with AD according to this rule.

Methylation Signatures and Rules From Hippocampus

Numerous methylation loci can help in the identification of specific methylation alterations associated with AD in HIPPO, a region that regulates memory and navigation. The first locus is **cg08056778**, which targets the 3'UTR region of *TBC1D1* and may further affect the expression of this gene in HIPPO. *TBC1D1* is correlated with cerebral palsy (Mohandas et al., 2018). Moreover, *TBC1D1* may be precisely regulated by effective microRNAs, thereby contributing to the development and aging of the central nervous system (Sephton et al., 2012). An *in vivo* study based on the model organism *Caenorhabditis elegans* confirmed that *TBC1D1* in HIPPO may be functionally correlated with aging-related diseases (Escoubas-GüNey, 2018). Although this study did not cover AD, it nevertheless confirmed the tissue specificity of this gene. Thus, we can reasonably argue that pathological methylation alterations in HIPPO may affect AD pathogenesis, thereby validating our prediction. Similarly, *SDK1*, *ANK1*, and *CHRN4*, which are targeted by **cg04680535**, **cg05066959**, and **cg17179314**, respectively, also have either a

certain gene expression level regulated by a specific methylation status (De Jager et al., 2014; Kanno et al., 2014; Yan et al., 2019) or an epigenomic change during AD pathogenesis (Li et al., 2010; De Jager et al., 2014).

Two parameters associated with the extracted quantitative rules were predicted to aid in the identification of patients with AD. The first parameter is *RHBDF2* (**cg13076843**), which has already been confirmed to be associated with AD (De Jager et al., 2014). The other parameter, that is, **cg06639320**, was also screened to be hypermethylated in patients with AD. The targeting of *FHL2* by this locus is reportedly correlated with AD progression (Bennett et al., 2017; Altuna et al., 2019).

CONCLUSION

This study is the first to identify a group of brain structure-specific AD signatures and quantitative rules at the methylation level. According to previous studies, the identified signatures and the parameters associated with this rule are correlated with AD pathogenesis. Therefore, the present study successfully screened a group of effective molecules with pathological potentials. This study also established a systematic workflow for the identification of effective tissue-specific biomarkers for complex diseases.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE125895>.

AUTHOR CONTRIBUTIONS

TH and Y-DC designed the study. TZ and KF performed the experiments. ZL, WG, and JY analyzed the results. ZL, WG, and TZ wrote the manuscript. All authors contributed to the research and reviewed the manuscript.

FUNDING

This work was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (XDA26040304 and XDB38050200), National Key R&D Program of China (2018YFC0910403), and the Fund of the Key Laboratory of Tissue Microenvironment and Tumor of Chinese Academy of Sciences (202002).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2022.895181/full#supplementary-material>

Supplementary Table 1 | Lists of genes ranked by mRMR for four brain structures.

Supplementary Table 2 | Performance of IFS for CRB brain structure.

Supplementary Table 3 | Performance of IFS for ERC brain structure.

Supplementary Table 4 | Performance of IFS for DLPFC brain structure.

Supplementary Table 5 | Performance of IFS for HIPPO brain structure.

REFERENCES

- Altuna, M., Urdáñez-Casado, A., De Gordo, J. S.-R., Zelaya, M. V., Labarga, A., and Lepasant, J. M. (2019). DNA methylation signature of human hippocampus in Alzheimer's disease is linked to neurogenesis. *Clin. Epigenet.* 11:91. doi: 10.1186/s13148-019-0672-7
- Atri, A. (2019). The Alzheimer's disease clinical spectrum: diagnosis and management. *Med. Clin.* 103, 263–293. doi: 10.1016/j.mcna.2018.10.009
- Bennett, D. A., Yu, L., Yang, J., Klein, H.-U., and De Jager, P. (2017). Epigenomics of Alzheimer's disease. *Transl. Epigenet. Clin.* 227–278. doi: 10.1016/j.trsl.2014.05.006
- Borroni, B., Grassi, M., Agosti, C., Costanzi, C., Archetti, S., Franzoni, S., et al. (2006). Genetic correlates of behavioral endophenotypes in Alzheimer disease: role of COMT, 5-HTTLPR and APOE polymorphisms. *Neurobiol. Aging* 27, 1595–1603. doi: 10.1016/j.neurobiolaging.2005.09.029
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32.
- Broadbent, N. J., Squire, L. R., and Clark, R. E. (2004). Spatial memory, recognition memory, and the hippocampus. *Proc. Natl. Acad. Sci.* 101, 14515–14520. doi: 10.1073/pnas.0406344101
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16, 321–357. doi: 10.1186/1756-0381-6-16
- Chen, L., Li, Z., Zhang, S., Zhang, Y.-H., Huang, T., and Cai, Y.-D. (2022). Predicting RNA 5-methylcytosine sites by using essential sequence features and distributions. *BioMed. Res. Int.* 2022:4035462. doi: 10.1155/2022/4035462
- Chen, L., Wang, S., Zhang, Y. H., Li, J., Xing, Z. H., Yang, J., et al. (2017). Identify key sequence features to improve CRISPR sgRNA efficacy. *IEEE Access* 5, 26582–26590. doi: 10.1109/access.2017.2775703
- Chen, L., Zhang, S., Pan, X., Hu, X., Zhang, Y.-H., Yuan, F., et al. (2019). HIV infection alters the human epigenetic landscape. *Gene Ther.* 26, 29–39. doi: 10.1038/s41434-018-0051-6
- Chen, W., Chen, L., and Dai, Q. (2021). iMPT-FDNPL: identification of membrane protein types with functional domains and a natural language processing approach. *Comput. Mathemat. Methods Med.* 2021:7681497. doi: 10.1155/2021/7681497
- Chouliaras, L., Mastroeni, D., Delvaux, E., Grover, A., Kenis, G., Hof, P. R., et al. (2013). Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiol. Aging* 34, 2091–2099. doi: 10.1016/j.neurobiolaging.2013.02.021
- Coppieters, N., Dieriks, B. V., Lill, C., Faull, R. L., Curtis, M. A., and Dragunow, M. (2014). Global changes in DNA methylation and hydroxymethylation in Alzheimer's disease human brain. *Neurobiol. Aging* 35, 1334–1344. doi: 10.1016/j.neurobiolaging.2013.11.031
- Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297.
- Dai, Q., Bao, C., Hai, Y., Ma, S., Zhou, T., Wang, C., et al. (2016). MTGipick allows robust identification of genomic islands from a single genome. *Briefin. Bioinform.* 19, 361–373. doi: 10.1093/bib/bbw118
- De Jager, P. L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L. C., Yu, L., et al. (2014). Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat. Neurosci.* 17, 1156–1163. doi: 10.1038/nn.3786
- Desire, L., Blondiaux, E., Carriere, J., Haddad, R., Sol, O., Fehlbaum-Beurdeley, P., et al. (2013). Blood transcriptomic biomarkers of Alzheimer's disease patients treated with EHT 0202. *J. Alzheim. Dis.* 34, 469–483. doi: 10.3233/JAD-121501
- Ding, S., Wang, D., Zhou, X., Chen, L., Feng, K., Xu, X., et al. (2022). Predicting Heart Cell Types by Using Transcriptome Profiles and a Machine Learning Method. *Life* 12:228. doi: 10.3390/life12020228
- Escoubas-GüNey, C. (2018). *Role of AMPK in Aging and Age-related Loss of Behavioral Plasticity in C. elegans*. France: Health Université Côte d'Azur.
- Frank, E., Hall, M., Trigg, L., Holmes, G., and Witten, I. H. (2004). Data mining in bioinformatics using Weka. *Bioinformatics* 20, 2479–2481. doi: 10.1093/bioinformatics/bth261
- Frank, E., and Witten, I. H. (1998). “Generating Accurate Rule Sets Without Global Optimization,” in *15th International Conference on Machine Learning*, (New Zealand: University of Waikato). doi: 10.1109/TCBB.2014.2307325
- Goyal, A., Miller, J., Watrous, A. J., Lee, S. A., Coffey, T., Sperling, M. R., et al. (2018). Electrical stimulation in hippocampus and entorhinal cortex impairs spatial and temporal memory. *J. Neurosci.* 38, 4471–4481. doi: 10.1523/JNEUROSCI.3049-17.2018
- Grubman, A., Chew, G., Ouyang, J. F., Sun, G., Choo, X. Y., Mclean, C., et al. (2019). A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat. Neurosci.* 22, 2087–2097. doi: 10.1038/s41593-019-0539-4
- Gürkan, H., Atli, E., Atli, E., Bozatlı, L., Altay, M. A., Yağcıntepe, S., et al. (2020). Chromosomal Microarray Analysis in Turkish Patients with Unexplained Developmental Delay and Intellectual Developmental Disorders. *Noro. Psikiyat. Ars.* 57, 177–191. doi: 10.29399/npa.24890
- Heekeren, H. R., Marrett, S., Ruff, D. A., Bandettini, P., and Ungerleider, L. G. (2006). Involvement of human left dorsolateral prefrontal cortex in perceptual decision making is independent of response modality. *Proc. Natl. Acad. Sci.* 103, 10023–10028. doi: 10.1073/pnas.0603949103
- Jasien, J. M., Daimon, C. M., Wang, R., Shapiro, B. K., Martin, B., and Maudsley, S. (2014). The effects of aging on the BTBR mouse model of autism spectrum disorder. *Front. Aging Neurosci.* 6:225. doi: 10.3389/fnagi.2014.00225
- Jia, Y., Zhao, R., and Chen, L. (2020). Similarity-Based Machine Learning Model for Predicting the Metabolic Pathways of Compounds. *IEEE Access* 8, 130687–130696. doi: 10.1109/access.2020.3009439
- Jin, N., Zhu, H., Liang, X., Huang, W., Xie, Q., Xiao, P., et al. (2017). Sodium selenate activated Wnt/ β -catenin signaling and repressed amyloid- β formation in a triple transgenic mouse model of Alzheimer's disease. *Exp. Neurol.* 297, 36–49. doi: 10.1016/j.expneurol.2017.07.006
- Kalinin, S., Marangoni, N., Kowal, K., Dey, A., Lis, K., Brodsky, S., et al. (2017). The Long-Lasting Rodenticide Brodifacoum Induces Neuropathology in Adult Male Rats. *Toxicol. Sci.* 159, 224–237. doi: 10.1093/toxsci/kfx134
- Kanno, T., Gotoh, A., and Nishizaki, T. (2014). Sphingosine arrests the cell cycle and induces apoptosis by targeting sphingosine-dependent protein kinase and protein kinase C δ in vitro. *Personal. Med. Univer.* 3, 22–27. doi: 10.1016/j.pmu.2014.03.003
- Karch, C. M., and Goate, A. M. (2015). Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biol. Psychiatr.* 77, 43–51. doi: 10.1016/j.biopsych.2014.05.006
- Keene, C. S., Bladon, J., Mckenzie, S., Liu, C. D., O'keefe, J., and Eichenbaum, H. (2016). Complementary functional organization of neuronal activity patterns in the perirhinal, lateral entorhinal, and medial entorhinal cortices. *J. Neurosci.* 36, 3660–3675. doi: 10.1523/JNEUROSCI.4368-15.2016
- Kohavi, R. (1995). “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in *International Joint Conference on artificial intelligence*: Lawrence Erlbaum Associates Ltd, (Netherlands: IEEE Computer Society press), 1137–1145.
- Kong, R., Xu, X., Liu, X., He, P., Zhang, M. Q., and Dai, Q. (2020). 2SigFinder: the combined use of small-scale and large-scale statistical testing for genomic island detection from a single genome. *BMC Bioinform.* 21:159. doi: 10.1186/s12859-020-3501-2.
- Kursa, M. B., and Rudnicki, W. R. (2010). Feature selection with the Boruta package. *J. Stat. Softw.* 36, 1–13.
- Li, M. D., Xu, Q., Lou, X. Y., Payne, T. J., Niu, T., and Ma, J. Z. (2010). Association and interaction analysis of variants in CHRNA5/CHRNA3/CHRNA4 gene cluster with nicotine dependence in African and European Americans. *Am. J. Med. Genet. Neuropsychiatr. Genet.* 153, 745–756. doi: 10.1002/ajmg.b.31043
- Li, X., Lu, L., and Chen, L. (2022). Identification of protein functions in mouse with a label space partition method. *Mathemat. Biosci. Eng.* 19, 3820–3842. doi: 10.3934/mbe.2022176
- Li, Z., Wang, D., Liao, H., Zhang, S., Guo, W., Chen, L., et al. (2022). Exploring the genomic patterns in human and mouse cerebellums via single-cell sequencing

- and machine learning method. *Front. Genet.* 13:857851. doi: 10.3389/fgene.2022.857851
- Liang, H., Chen, L., Zhao, X., and Zhang, X. (2020). Prediction of drug side effects with a refined negative sample selection strategy. *Comput. Mathemat. Methods Med.* 2020:1573543. doi: 10.1155/2020/1573543
- Liu, H., Hu, B., Chen, L., and Lu, L. (2021). Identifying protein subcellular location with embedding features learned from networks. *Curr. Proteom.* 18, 646–660. doi: 10.2174/1570164617999201124142950
- Liu, H. A., and Setiono, R. (1998). Incremental feature selection. *Appl. Intell.* 9, 217–230.
- Mars, R. B., and Grol, M. J. (2007). Dorsolateral prefrontal cortex, working memory, and prospective coding for action. *J. Neurosci.* 27, 1801–1802. doi: 10.1523/jneurosci.5344-06.2007
- Marzi, S. J., Leung, S. K., Ribarska, T., Hannon, E., Smith, A. R., Pishva, E., et al. (2018). A histone acetylome-wide association study of Alzheimer's disease identifies disease-associated H3K27ac differences in the entorhinal cortex. *Nat. Neurosci.* 21, 1618–1627. doi: 10.1038/s41593-018-0253-7
- Marzi, S. J., Ribarska, T., Smith, A. R., Hannon, E., Poschmann, J., Moore, K., et al. (2017). A histone acetylome-wide association study of Alzheimer's disease: neuropathology-associated regulatory variation in the human entorhinal cortex. *BioRxiv* 183541. [preprint] doi: 10.1101/183541
- Mastroeni, D., Grover, A., Delvaux, E., Whiteside, C., Coleman, P. D., and Rogers, J. (2010). Epigenetic changes in Alzheimer's disease: decrements in DNA methylation. *Neurobiol. Aging* 31, 2025–2037. doi: 10.1016/j.neurobiolaging.2008.12.005
- Matthews, B. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. et Biophys. Acta (BBA)-Protein Struct.* 405, 442–451. doi: 10.1016/0005-2795(75)90109-9
- Matthews, K. A., Xu, W., Gaglioti, A. H., Holt, J. B., Croft, J. B., Mack, D., et al. (2019). Racial and ethnic estimates of Alzheimer's disease and related dementias in the United States (2015–2060) in adults aged= 65 years. *Alzheim. Dement.* 15, 17–24. doi: 10.1016/j.jalz.2018.06.3063
- Mohandas, N., Bass-Stringer, S., Maksimovic, J., Crompton, K., Loke, Y. J., Walstab, J., et al. (2018). Epigenome-wide analysis in newborn blood spots from monozygotic twins discordant for cerebral palsy reveals consistent regional differences in DNA methylation. *Clin. Epigenet.* 10:25. doi: 10.1186/s13148-018-0457-4
- Montclaus, M. E., Reagh, Z. M., and Yassa, M. A. (2019). Precise temporal memories are supported by the lateral entorhinal cortex in humans. *Nat. Neurosci.* 22, 284–288. doi: 10.1038/s41593-018-0303-1
- Murphy, S. L., Kochanek, K. D., Xu, J., and Arias, E. (2015). Mortality in the United States. 2014. *NCHS Data Brief* 229, 1–8
- Pan, X., Zeng, T., Yuan, F., Zhang, Y.-H., Chen, L., Zhu, L., et al. (2019). Screening of methylation signature and gene functions associated with the subtypes of isocitrate dehydrogenase-mutation gliomas. *Front. Bioeng. Biotechnol.* 7:339. doi: 10.3389/fbioe.2019.00339
- Pan, X.-Y., Zhang, Y.-N., and Shen, H.-B. (2010). Large-Scale Prediction of Human Protein-Protein Interactions from Amino Acid Sequence Based on Latent Topic Features. *J. Proteom. Res.* 9, 4992–5001. doi: 10.1021/pr100618t
- Panosian, A., Seo, E.-J., and Efferth, T. (2018). Synergy assessments of plant extracts used in the treatment of stress and aging-related disorders. *Synergy* 7, 39–48. doi: 10.1016/j.synres.2018.10.001
- Pathak, D., Srivastava, A. K., Padma, M. V., Gulati, S., and Rajeswari, M. R. (2019). Quantitative Proteomic and Network Analysis of Differentially Expressed Proteins in PBMC of Friedreich's Ataxia (FRDA) Patients. *Front. Neurosci.* 13:1054. doi: 10.3389/fnins.2019.01054
- Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transac. Patt. Anal. Mach. Intell.* 27, 1226–1238. doi: 10.1109/TPAMI.2005.159
- Petrache, A. L., Rajulawalla, A., Shi, A., Wetzel, A., Saito, T., Saido, T. C., et al. (2019). Aberrant excitatory-inhibitory synaptic mechanisms in entorhinal cortex microcircuits during the pathogenesis of Alzheimer's disease. *Cereb. Cortex* 29, 1834–1850. doi: 10.1093/cercor/bhz016
- Platt, J. (1998). *Sequential Minimal Optimizaton: A Fast Algorithm for Training Support Vector Machines*. Technical Report MSR-TR-98-14, 21. Alabama: Microsoft
- Qazi, T. J., Quan, Z., Mir, A., and Qing, H. (2018). Epigenetics in Alzheimer's disease: perspective of DNA methylation. *Mol. Neurobiol.* 55, 1026–1044. doi: 10.1007/s12035-016-0357-6
- Raymond, J. L., Lisberger, S. G., and Mauk, M. D. (1996). The cerebellum: a neuronal learning machine? *Science* 272, 1126–1131. doi: 10.1126/science.272.5265.1126
- Reisberg, B., Borenstein, J., Salob, S. P., and Ferris, S. H. (1987). Behavioral symptoms in Alzheimer's disease: phenomenology and treatment. *J. Clin. Psychiatr.* 48, 9–15.
- Rogers, J., Webster, S., Lue, L.-F., Brachova, L., Civin, W. H., Emmerling, M., et al. (1996). Inflammation and Alzheimer's disease pathogenesis. *Neurobiol. Aging* 17, 681–686.
- Semick, S. A., Bharadwaj, R. A., Collado-Torres, L., Tao, R., Shin, J. H., and Deep-Soboslay, A. (2019). Integrated DNA methylation and gene expression profiling across multiple brain regions implicate novel genes in Alzheimer's disease. *Acta Neuropathol.* 137, 557–569. doi: 10.1007/s00401-019-01966-5
- Sephton, C. F., Cenik, B., Cenik, B. K., Herz, J., and Yu, G. (2012). TDP-43 in central nervous system development and function: clues to TDP-43-associated neurodegeneration. *Biol. Chem.* 393, 589–594. doi: 10.1515/hsz-2012-0115
- Soerensen, M., Li, W., Debrabant, B., Nygaard, M., Mengel-From, J., Frost, M., et al. (2019). Epigenome-wide exploratory study of monozygotic twins suggests differentially methylated regions to associate with hand grip strength. *Biogerontology* 20, 627–647. doi: 10.1007/s10522-019-09818-1
- Sultana, R., Boyd-Kimball, D., Poon, H. F., Cai, J., Pierce, W. M., Klein, J. B., et al. (2006). Redox proteomics identification of oxidized proteins in Alzheimer's disease hippocampus and cerebellum: an approach to understand pathological and biochemical alterations in AD. *Neurobiol. Aging* 27, 1564–1576. doi: 10.1016/j.neurobiolaging.2005.09.021
- Sun, M., Bell, W., Laclair, K. D., Ling, J. P., Han, H., Kageyama, Y., et al. (2017). Cryptic exon incorporation occurs in Alzheimer's brain lacking TDP-43 inclusion but exhibiting nuclear clearance of TDP-43. *Acta Neuropathol.* 133, 923–931. doi: 10.1007/s00401-017-1701-2
- Swingler, T. E., Waters, J. G., Davidson, R. K., Pennington, C. J., Puente, X. S., and Darrah, C. (2009). Degradome expression profiling in human articular cartilage. *Arthr. Res. Ther.* 11:R96. doi: 10.1186/ar2741
- Tang, S., and Chen, L. (2022). iATC-NFMLP: identifying classes of anatomical therapeutic chemicals based on drug networks, fingerprints and multilayer perceptron. *Curr. Bioinform.* 17
- Taylor, C. A., Greenlund, S. F., Mcguire, L. C., Lu, H., and Croft, J. B. (2017). Deaths from Alzheimer's Disease—United States, 1999–2014. *Morbid. Mortal. Weekly Rep.* 66:521. doi: 10.15585/mmwr.mm6620a1
- Trifonova, E. A., Mustafin, Z. S., Lashin, S. A., and Kochetov, A. V. (2022). Abnormal mTOR activity in pediatric autoimmune neuropsychiatric and MIA-associated autism spectrum disorders. *Int. J. Mol. Sci.* 23:967. doi: 10.3390/ijms23020967
- Vanderweyde, T., Apicco, D. J., Youmans-Kidder, K., Ash, P. E. A., Cook, C., Lummertz, et al. (2016). Interaction of tau with the RNA-Binding Protein TIA1 Regulates tau Pathophysiology and Toxicity. *Cell Rep.* 15, 1455–1466. doi: 10.1016/j.celrep.2016.04.045
- Velasco-Estevéz, M., Mampay, M., Boutin, H., Chaney, A., Warn, P., Sharp, A., et al. (2018). Infection Augments Expression of Mechanosensing Piezo1 Channels in Amyloid Plaque-Reactive Astrocytes. *Front. Aging Neurosci.* 10:332. doi: 10.3389/fnagi.2018.00332
- Weller, J., and Budson, A. (2018). Current understanding of Alzheimer's disease diagnosis and treatment. *F1000Research* 7:F1000 Faculty Rev-1161.
- Wu, Z., and Chen, L. (2022). Similarity-based method with multiple-feature sampling for predicting drug side effects. *Comput. Mathemat. Methods Med.* 2022:9547317. doi: 10.1155/2022/9547317
- Yan, T., Ding, F., and Zhao, Y. (2019). Integrated identification of key genes and pathways in Alzheimer's disease via comprehensive bioinformatical analyses. *Hereditas* 156:25. doi: 10.1186/s41065-019-0101-0
- Yang, S., Wang, Y., Chen, Y., and Dai, Q. (2020). MASQC: next Generation Sequencing Assists Third Generation Sequencing for Quality Control in N6-Methyladenine DNA Identification. *Front. Genet.* 11:269. doi: 10.3389/fgene.2020.00269
- Yang, Y., and Chen, L. (2022). Identification of drug-disease associations by using multiple drug and disease networks. *Curr. Bioinform.* 17, 48–59. doi: 10.2174/1574893616666210825115406

- Zhang, Y.-H., Zeng, T., Chen, L., Huang, T., and Cai, Y.-D. (2021b). Determining protein–protein functional associations by functional rules based on gene ontology and KEGG pathway. *Biochim. et Biophys. Acta Proteins Proteom.* 1869:140621. doi: 10.1016/j.bbapap.2021.140621
- Zhang, Y.-H., Li, H., Zeng, T., Chen, L., Li, Z., Huang, T., et al. (2021a). Identifying transcriptomic signatures and rules for SARS-CoV-2 infection. *Front. Cell Dev. Biol.* 8:627302. doi: 10.3389/fcell.2020.627302
- Zhao, X., Chen, L., Guo, Z.-H., and Liu, T. (2019). Predicting drug side effects with compact integration of heterogeneous networks. *Curr. Bioinform.* 14, 709–720. doi: 10.2174/1574893614666190220114644
- Zhao, X., Chen, L., and Lu, J. (2018). A similarity-based method for prediction of drug side effects with heterogeneous information. *Mathemat. Biosci.* 306, 136–144. doi: 10.1016/j.mbs.2018.09.010
- Zhou, J.-P., Chen, L., and Guo, Z.-H. (2020a). iATC-NRAKEL: an efficient multi-label classifier for recognizing anatomical therapeutic chemical classes of drugs. *Bioinformatics* 36, 1391–1396. doi: 10.1093/bioinformatics/btz757
- Zhou, J.-P., Chen, L., Wang, T., and Liu, M. (2020b). iATC-FRAKEL: a simple multi-label web-server for recognizing anatomical therapeutic chemical classes of drugs with their fingerprints only. *Bioinformatics* 36, 3568–3569. doi: 10.1093/bioinformatics/btaa166
- Zhou, X., Ding, S., Wang, D., Chen, L., Feng, K., Huang, T., et al. (2022). Identification of cell markers and their expression patterns in skin based on single-cell RNA-sequencing profiles. *Life*.12:550 doi: 10.1186/s40246-016-0063-5
- Zhu, Y., Hu, B., Chen, L., and Dai, Q. (2021). iMPTCE-Hnetwork: a multi-label classifier for identifying metabolic pathway types of chemicals and enzymes with a heterogeneous network. *Comput. Mathemat. Methods Med.* 2021:6683051. doi: 10.1155/2021/6683051
- Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.
- Copyright © 2022 Li, Guo, Zeng, Yin, Feng, Huang and Cai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.