Cell. Mol. Biol. (ISSN: 1165-158X)

Journal Homepage: www.cellmolbiol.org

Cellular and Molecular Biology

Original Article

An age classification model based on DNA methylation biomarkers of aging in human peripheral blood using random forest and artificial neural network

Ye Li[1,2], Yukun Yang[2], Jie He[2,3], Jinxiu Hu[2,3], Xiangqing Zhu[2], Chuan Tian[2], Mengdie Chen[1,2], Xiaojuan Zhao[1,2], Li Ye[2], Hang Pan[2], Xinghua Pan[2,*]

[1] Clinical College of the 920th Hospital of Kunming Medical University, Kunming 650032, Yunnan Province, China
[2] The Stem Cells and Immune Cells Biomedical Techniques Integrated Engineering Laboratory of State and Regions, Cell Therapy Technology Transfer Medical Key Laboratory of Yunnan Province, Kunming Key Laboratory of Stem Cell and Regenerative Medicine, Basic Medical Laboratory, 920th Hospital of Joint Logistics Support Force, PLA, Kunming, 650032, Yunnan Province, China
[3] Aochen Biotechnology(Yunnan) Co., Ltd., Kunming 650106, Yunnan Province, China

Article Info

Article history:

Received: November 19, 2023
Accepted: January 20, 2024
Published: January 31, 2024

Use your device to scan and read the article online

Abstract

Recent epigenetic studies have revealed a strong association between DNA methylation and aging and lifespan, which changes (increases or decreases) with age. Based on these, the construction of age prediction models associated with DNA methylation levels can be used to infer biological ages closer to the functional state of the organism. We downloaded methylation data from the Gene Expression Omnibus (GEO) public database for normal peripheral blood samples from people of different ages. We grouped the samples according to age (18-35 years and >50 years), screened the methylation sites that differed between the two groups, identified 44 differentially methylated sites, and subsequently obtained 11 age-related characteristic methylation sites using the random forest method. Then, we constructed an age classification model with these 11 characteristic methylation sites using an artificial neural network and evaluated its efficacy. The age classification model was constructed by an artificial neural network and its efficacy was evaluated. The model predicted an area under the curve (AUC) of 0.97 in the validation set and accurately distinguished between those aged 18-35 and >50 years. Furthermore, the levels of these 11 characteristic methylation sites also differed significantly between the two sets of samples in the validation set, including six newly identified age-related methylation sites (P<0.001). Finally, we constructed a multifactor regulatory network based on the corresponding genes of age-related methylation sites to reveal the transcriptional and post-transcriptional regulation patterns. As a result of the increasing problem of aging, the age classification model we constructed allows us to accurately distinguish different age groups at the molecular level, which will be more predictive than chronological age for assessing individual aging and future health status.

Keywords: DNA methylation biomarkers; **H**uman peripheral blood; **R**andom forest; **A**rtificial neural network

1. Introduction

Aging is an unavoidable biological process, and, as the body ages, it shows progressive structural degeneration and functional decline, with a progressively increasing risk of many chronic noncommunicable diseases, including cancer, diabetes, cardiovascular disease, and neurodegenerative diseases [1]. It is well known that different human bodies age at different rates. Although an individual's chronological age is a simple indicator of aging, the rate of aging may vary among individuals with the same chronological age, and the variability in cognitive function and health status increases with age. Biological age (BA) is the deviation between apparent and chronological age [2]. With the development of the field of molecular biology, many types of candidate biomarkers that may be considered relevant to the aging process have been established. Among these, epigenetic clocks constructed based on DNA methylation levels are a new and valuable biological age predictor [3]. However, their reliability as biomarkers of aging needs to be confirmed longitudinally.

Epigenetic studies have shown a close association between DNA methylation and human age [4], and the methylation levels of some specific CpG sites are strongly correlated with age among individuals of the same species. The age of humans can be accurately predicted by building statistical models based on age-dependent CpG locus methylation levels and age [5,6]. In recent years, researchers have developed different age prediction clocks, also known as "epigenetic age clocks", using genomic DNA methylation data generated by different methylation array technologies or using datasets from public databases, which use the temporal pattern of DNA methylation levels at specific CpG sites in the genome to predict age [7]. In 2013, two classical "epigenetic age clocks", Hannum's clock [8] and Horvath's clock [9], were developed, both using a blood-tissue-based model. These two clocks are

* Corresponding author.
E-mail address: xinghuapan@aliyun.com (X. Pan).
Doi: http://dx.doi.org/10.14715/cmb/2024.70.1.21

age prediction models based on 71 blood tissues and 353 CpG sites based on 51 tissues and cell types, respectively, with a high age correlation of 96%. Since then, several "epigenetic age clocks" have been reported, involving multiple species and various methods of clock construction based on different tissues [10-15]. The establishment of different epigenetic clocks not only provides an accurate measurement of biological age and the rate of aging, but also provides a good indicator for use in the evaluation of various interventions to slow down aging. For example, mice have been shown to live significantly longer and have lower epigenetic ages when subjected to caloric restriction and growth hormone receptor knockout [16]. In a human clinical trial involving the regeneration of the thymus, reversal of immune senescence was accompanied by a decrease in epigenetic age [17]. Interventions such as dietary restriction, exercise, and rapamycin supplementation have all been shown to slow the epigenetic clock of aged mice without modifying the genomic sequence [18,19].

A variety of machine learning approaches have been tested for the purpose of predicting human age based on molecular-level features, mainly the penalized multi-variate regression method (e.g., Elastic Net) [20], support vector regression [21], gradient boosting regression [22], and random forest regression [23]. In the past few years, deep learning has been gradually introduced in the field of machine learning. Deep learning provides great opportunities in the field of human biological aging research. Deep learning algorithms are based on artificial neural networks. Artificial neural network algorithms are a class of pattern-matching algorithms that mimic biological neural networks and are commonly used to solve classification and regression problems. Vidaki et al. [24] compared the accuracy of epigenetic clocks constructed using a multiple regression analysis model and a generalized regression neural network model. The mean absolute error (MAE) of the former was 4.6 years; in the latter, the age prediction was significantly improved (R2 = 0.96) with an MAE of 3.3 years for the training set and 4.4 years for the blind test set out of a total of 231 cases. Subsequently, Galkin et al. [25] constructed a biological age clock (DeepMAge) based on blood DNA methylation profiles using neural networks with an absolute median error of only 2.77 years, providing higher accuracy.

In this study, age-related characteristic methylation sites were obtained using a random forest approach to screen human peripheral blood methylation data from normal samples of different ages. Then, we constructed an age classification model using an artificial neural network algorithm and tested its efficacy. Our analysis revealed that the levels of the characteristic methylation sites were significantly different between the two age samples and that there was a significant correlation between these methylation sites. Finally, based on characteristic methylation sites, we constructed a multifactor regulatory network, and the transcriptional and post-transcriptional regulation patterns of the genes corresponding to these age-related methylation sites were predicted.

2. Materials and Methods

2.1. Data Download and Preprocessing

We used the National Center for Biotechnology Information (NCBI) GEO (https://www.ncbi.nlm.nih.gov/geo/) by keywords (Age (All Fields) AND (methylation (All Fields))) AND "Homo sapiens"(porgn:__txid9606) to search for methylation data from normal human samples. When screening the data, it was required that the tissue source be peripheral blood and that the DNA methylation data be from adult individuals with age information.

The GSE147221 dataset was obtained after the screening, containing normal samples from 327 adult individuals of different ages and peripheral blood source methylation data. Among these, 94 samples were from people aged >50 years and 108 samples were from people aged between 18 and 35 years. A total of 202 samples were available for the analysis conducted in this study. To test the performance of our model, we divided the samples randomly (half of the samples) into a training set and a validation set.

2.2. Differential DNA Methylation Site Screening

We used the R function "DESeq2" to calculate the differential methylation sites of the two age groups' samples (18-35 years old vs. > 50 years old). The screening condition for differential methylation sites was $|\Delta\beta| > 0.10$ and a P-value of <0.05 [26].

2.3. Functional Annotation of Differentially Methylated Sites

The genes corresponding to the differentially methylated sites were downloaded from the GSE147221 dataset. The corresponding genes were compared with those in the GO and KEGG functional databases using the "ClusterProfiler" function to obtain annotations of these genes for gene function analysis.

2.4. Random Forest Model Construction

We used a random forest algorithm to filter age-related characteristic methylation sites. In machine learning, a random forest is a classification model that contains multiple decision trees, the output of which is determined by the plurality of the categories output by the individual trees. Each tree of the random forest algorithm is constructed according to the following algorithm: N is used to denote the number of training samples and M denotes the number of features. The number of input features, m, is used to determine the decision outcome of a node in the decision tree, where m should be much smaller than M. From the N training samples, a training set is formed by sampling N times with put-back sampling (i.e., bootstrap sampling), and the unsampled samples are used for prediction to evaluate the error. For each node, m features are randomly selected, and the decision of each node in the decision tree is determined based on these features. Based on these m features, its optimal split is calculated. Each tree grows intact without pruning, which is likely to be adopted after building a normal tree classification model.

The R function "randomForest" (https://www.stat.berkeley.edu/breiman/RandomForests/) was used here for the construction of random forests. The input data were DNA methylation profiles and differentially methylated sites, and the characteristic methylated sites that contribute to age prediction were screened.

2.5. Artificial Neural Network Model Construction

We used the R function "nnet" (http://www.stats.ox.ac.uk/pub/MASS4/) to construct an artificial neural network model. After calculating the weights of the characteristic methylation sites, an age classification model was

constructed. Age classification was then performed on the validation set based on the characteristic methylation level, and the effectiveness of the classification model was evaluated.

2.6. Analysis of Methylation Sites with Age-Related Characteristics

In the training set and verification set, the gene expression profiles of the characteristic methylation sites were extracted and visualized using the heatmap function in R to draw a heatmap of the distribution of the characteristic methylation sites in the real and predicted samples.

The differences in the methylation levels at the characteristic methylation sites were evaluated for those aged 18-35 years and >50 years in the verification set; a box diagram was drawn, and the difference in the p-values of methylation levels of the characteristic methylation sites between the two different age groups was calculated.

The methylation profiles of characteristic methylation sites were extracted and the methylation level correlations of characteristic methylation sites were analyzed.

2.7. Construction of a Multifactor Regulatory Network Based on Age-Related Characteristic Methylation Sites

Obtaining the Transcription Factor Target Gene Network of the mRNA-TF Network

The transcription factor target gene data were obtained from the integration of several datasets: TRANSFAC (27), ENCODE (28), CHEA (29), and TRRUST (30), where the transcription factor target gene data from the TRANSFAC, ENCODE, and CHEA datasets were downloaded from the Harmonizonme website (31) (http://amp.pharm.mssm.edu/Harmonizome/) and the TRRUST transcription factor target gene data were downloaded from the TRRUST website (www.grnpedia.org/trrust/). The transcription factor target gene data obtained from the above four databases were combined into an integrated transcription factor target gene network to obtain a pooled network of all the transcription factors regulating the genes corresponding to the characteristic methylation sites.

Obtaining mRNA–miRNA Interactions

Using the RNA Interoperability Database online tool (TarBase v.8http://carolina.imis.athenainnovation.gr/diana_tools/web/index.php?r=tarbasev8%2Findex&miRNAs%5B%5D=&genes%5B%5D=FGF17&sources%5B%5D=1&sources%5B%5D=7&sources%5B%5D=9&publication_year=&prediction_score=&sort_field=&sort_type=&query=), we searched for characteristic methylation sites corresponding to genes with low throughput or prediction scores of 9.5 or higher for mRNA–miRNA relationships.

We searched the database for RNA–RNA interactions among the genes corresponding to the characteristic methylation sites, including mRNA–mRNA, microRNA–mRNA, and lncRNA–mRNA interactions (http://www.rna-society.org/rnainter/ ; a prediction score of 9.5 or higher was required).

3. Results

3.1. Differential Methylation Sites

We identified 44 differentially methylated sites in the two groups of samples, corresponding to ages of 18-35 years and >50 years (Figure 1). Among them, the five sites with the largest Δβ were cg04891883, cg09279615, cg16725305, cg21627202, and cg22302772.

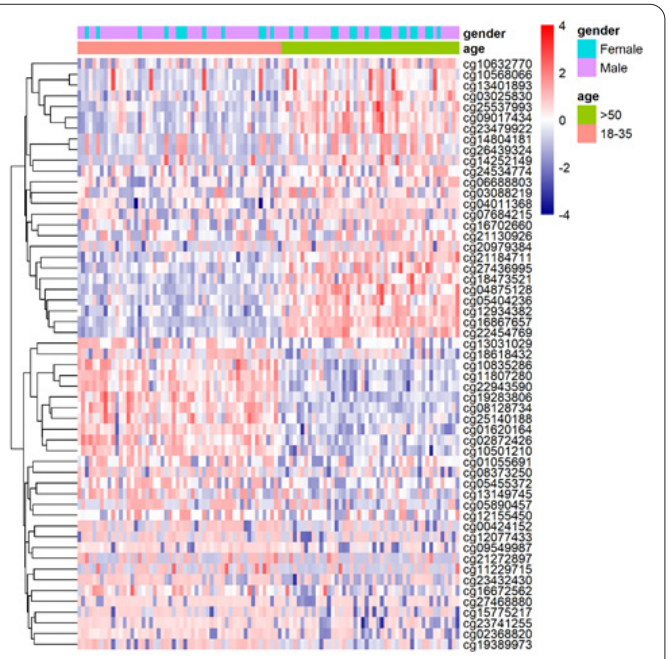

Fig. 1. Heat map of differentially methylated sites.

3.2. Results of Functional Enrichment Analysis of Genes Corresponding to Differentially Methylated Sites

The genes corresponding to the 44 differentially methylated sites were enriched with terms in the GO database such as the neuron-to-neuron synapse, the intrinsic component of the postsynaptic specialization membrane, the integral component of the postsynaptic specialization membrane, and the postsynaptic specialization membrane (Figure 2a, b, and c), as well as the neuroactive ligand-receptor interaction pathway obtained from the KEGG database (Figure 2d).

3.3. Age-Related Characteristic Methylation Sites

When using differential methylation sites to train the random forest classification model, we first explored the parameter "mtry" and found that when it was equal to 9, the error was the smallest (Figure 3a). We then explored the parameter "ntree" and found that when it was equal to 200 or more, the error tended to be stable (figure 3b). The importance of the differential methylation sites was evaluated with the parameter mtry equal to 9 and ntree equal to 200 in the training classification model, and then the age-related characteristic methylation sites were screened.

According to the threshold (importance value cutoff = 0.0075), we screened 11 age-related characteristic methylation sites, which were cg03025830, cg04875128, cg05404236, cg08128734, cg10501210, cg11807280, cg12934382, cg16867657, cg18473521, cg19283806, and cg22454769 (Figure 3c and d).

3.4. Artificial Neural Network Model

We used the age-related characteristic methylation sites identified by the random forest model to construct an artificial neural network (Figure 4a). The receiver operating characteristic (ROC) curves were then plotted in the training and validation sets to evaluate the predictive efficacy of the classification model. The area under the curve (AUC) of the model in the validation set reached 0.97,

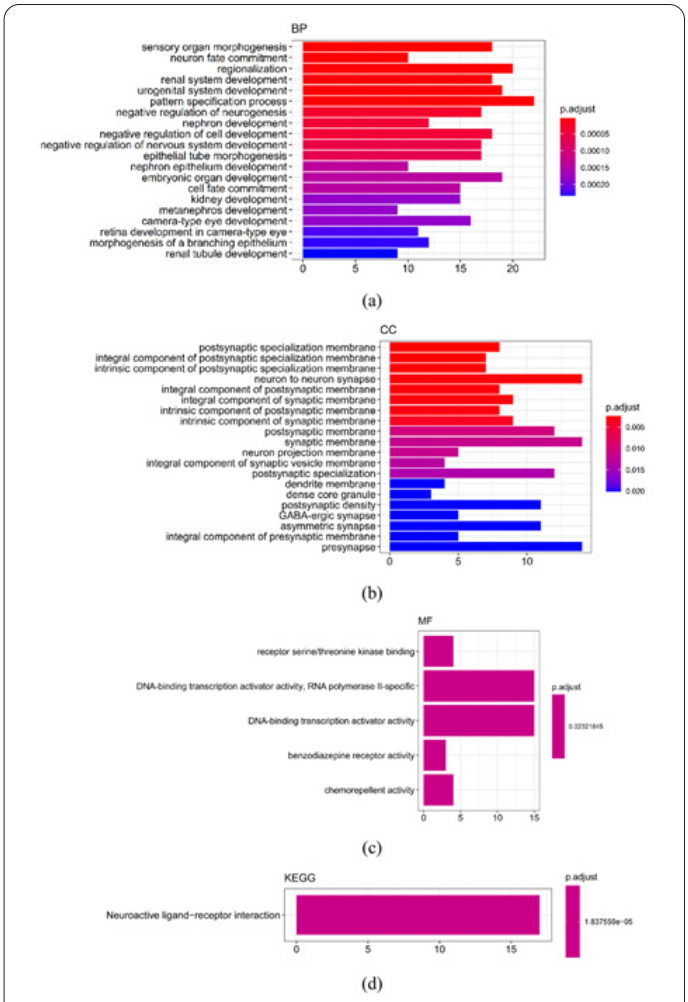

Fig. 2. The differential methylation sites corresponding to gene functional annotations. (a) biological processes from the GO database; (b) cellular components from the GO database; (c) molecular functions from the GO database; and (d) the gene functional enrichment map from the KEGG database.

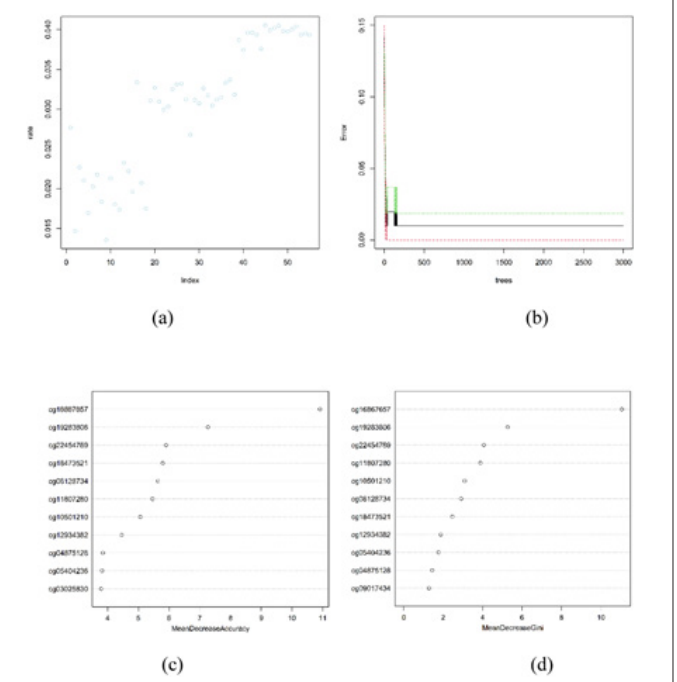

Fig. 3. Construction of a random forest model and screening of age-related characteristic methylation sites. (a) and (b) random forest model parameter training and selection process; (c) and (d) age-related characteristic methylation sites.

indicating that the model was able to accurately assess the age information of the samples (Figure 4b and c).

3.5. Analysis of Methylation Sites with Age-Related Characteristics

3.5.1. Methylation-Level Heatmap of Characteristic Methylation Sites

According to the methylation-level heatmaps of these characteristic methylation sites, we found that the samples from people of different age groups were clustered in their respective classes, and that the methylation levels of the characteristic methylation sites differed significantly between the two age groups. A further illustration of the correlation between the characteristic methylation sites and age is shown in Figure 5. More importantly, the age group predicted by the artificial neural network prediction model was basically consistent with the real age group of the samples.

3.5.2. Differences in the Methylation Levels of the Methylation Sites with Age-Related Characteristics

The figure shows that the methylation levels of the 11 characteristic methylation sites were very significantly and statistically different between the two groups of samples in the validation set, taken from people aged 18-35 and >50 years. The methylation levels of seven of the CpG sites gradually increased with age, while the other four gradually decreased with age (Figure 6). We found nine characteristic methylation sites that corresponded to genes shown in Table 1, among which ELOVL2 and FHL2 have been reported to be associated with biological age in many studies.

3.5.3. Correlation of methylation levels with age-related characteristic methylation sites

The methylation levels of the characteristic meth-

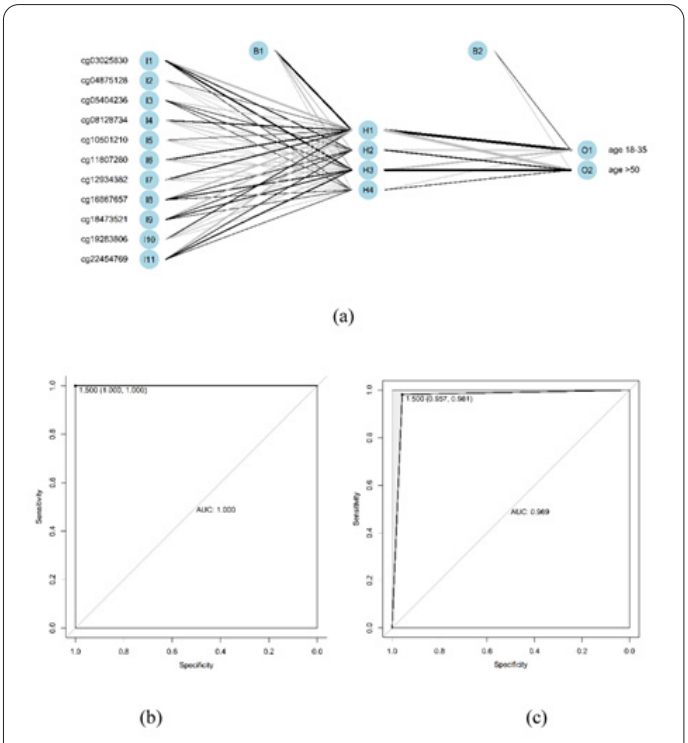

Fig. 4. Artificial neural network model construction and effectiveness evaluation. (a) the construction process of the artificial neural network model, (b) the ROC curve of the training set, and (c) the ROC curve of the verification set.

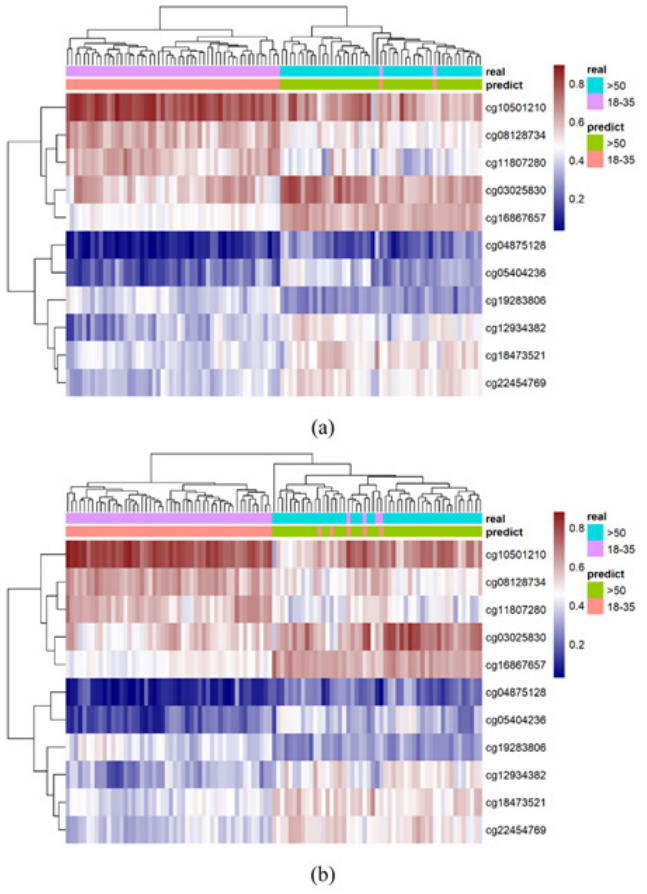

Fig. 5. Heatmap of methylation levels at characteristic methylation sites. (a) the training set and (b) the validation set. The top two rows are annotated with the true and predicted sample distributions, respectively, and the clustering method used for the rows and columns was hierarchical clustering.

ylation sites were all significantly correlated (Figure 7). The strongest positive correlation was found between cg16867657 and cg2245476, with a correlation coefficient of 0.81, while cg16867657 showed a significant negative correlation with cg19283806, with a correlation coefficient of -0.78.

3.6. Multifactor Regulatory Network of Age-Related Characteristic Methylation Sites

A multifactorial regulatory network was constructed based on the regulatory relationships of the RNAs of the genes corresponding to age-related characteristic methylation sites. We identified 894 pairs of interactions between 9 genes and 283 transcription factors. The corresponding gene of cg05404236, IRS2, had 21 pairs of mRNA–miRNA interactions; the corresponding genes of cg08128734 and cg22454769, RASSF5 and FHL2, had one and three pairs of mRNA–miRNA interaction relationships, respectively; and IRS2 and FHL2 had the same lncRNA–mRNA interaction relationship (Figure 8).

4. Discussion

Human aging is influenced by the interaction of many complex factors. Altered DNA methylation status is an important factor, and an increasing number of studies have shown that DNA methylation levels can predict the bio-

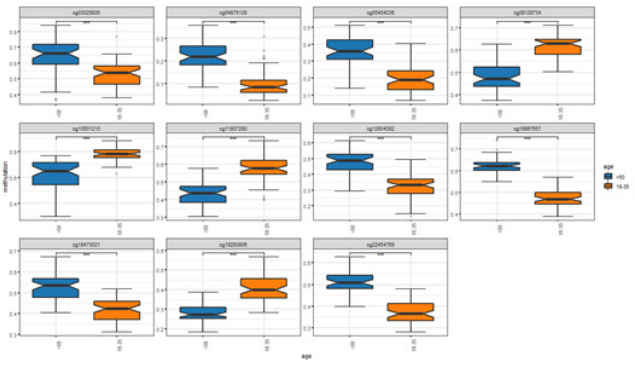

Fig. 6. Methylation levels at the characteristic methylation sites in the validation set. '***'means $P < 0.001$.

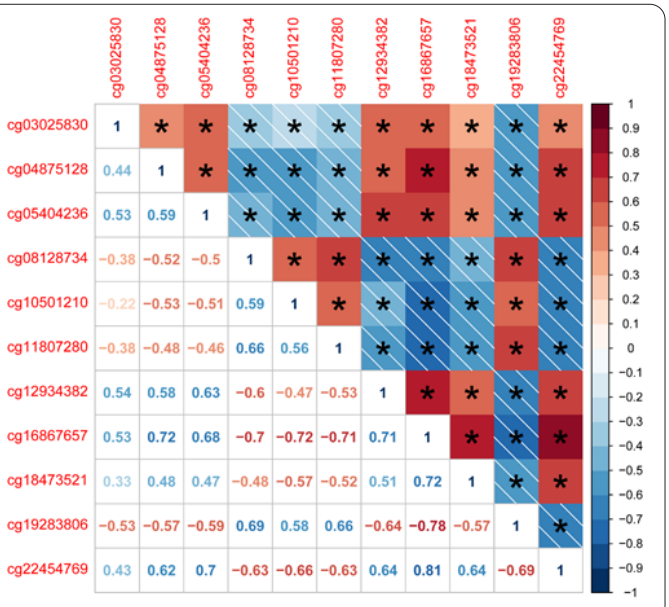

Fig. 7. Significant correlation between age-related characteristic methylation sites. Red represents positive correlations, and blue represents negative correlations. The upper right corner shows the correlation between the characteristic methylation sites according to the color block; "*" indicates a correlation level $P<0.01$; the lower left corner shows the correlation coefficient r value.

Table 1. Genes corresponding to methylation sites with age-related characteristics.

Characteristic Methylation Sites	Corresponding Genes
cg03025830	FGF17
cg04875128	OTUD7A
cg05404236	IRS2
cg08128734	RASSF5
cg12934382	GRM2
cg16867657	ELOVL2
cg18473521	HOXC4
cg19283806	CCDC102B
cg22454769	FHL2

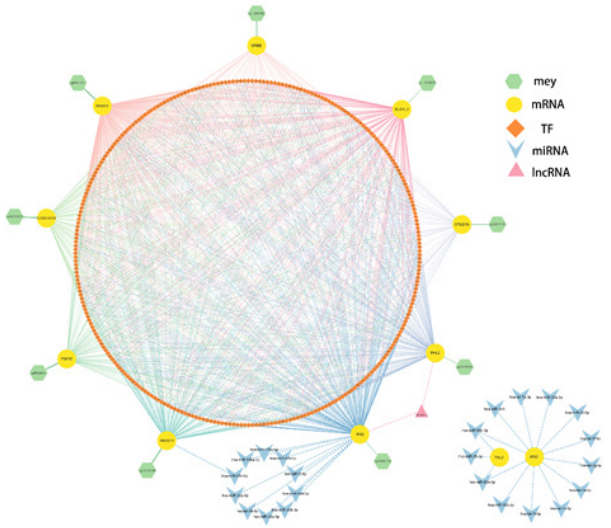

Fig. 8. Multifactor regulatory network of age-related characteristic methylation sites. The different colors and shapes of the nodes in the figure represent different types. Green hexagon: age-associated characteristic methylation sites; yellow ellipse: mRNA; orange diamond: transcription factor (TF); blue V-shape: microRNA; and pink triangle: lncRNA.

logical age of individuals; additionally, changes in DNA methylation levels are chemically and biologically stable and are expected to be biological markers of aging. Our age classification model is derived from peripheral blood, which is advantageous for designing practical diagnostics and testing samples collected from other studies. The advantages of blood tissue have been demonstrated on Hannum's clock, the first epigenetic clock constructed based on blood alone. The clock is also representative of other tissues (e.g., lung, kidney, skin), and its age prediction error is comparable to that in blood. Additionally, this study suggests that age-related changes in DNA methylation modifications are intrinsic to the methylome, rather than being primarily due to cellular heterogeneity, i.e., changes in the cell type composition of whole blood with age [8]. Prior to this, Rakyan et al. also reported age-related epigenetic modifications in purified CD4+ T cells and CD14+ monocytes, similar to the changes observed in whole blood [32].

In this study, 11 characteristic methylation sites screened by random forest had significant age correlation, among which 5 methylation sites (cg04875128, cg10501210, cg16867657, cg18473521, cg19283806) were identical to Hannum's clock. The corresponding genes are OTUD7A, ELOVL2, HOXC4, and CCDC102B. ELOVL2 has been used in many epigenetic age clocks [33-35]. Garali et al. [36] constructed several age prediction models using multiple statistical models based on only seven CpG sites of the peripheral blood ELOVL2 promoter, with a mean absolute deviation (MAD) of 4.41-4.77 and a root mean square error of 6.40-6.73. In addition, we identified six new age-related DNA methylation sites. Among them, cg22454769 corresponds to the gene FHL2, which is frequently reported to be associated with age [37, 38]. Another age-related DNA methylation site, cg03025830, corresponds to the gene fibroblast growth factor 17 (FGF17). A recent study suggested that FGF17 is a key target for restoring oligodendrocyte function in the aging brain. Infusion of FGF17 into the cerebrospinal fluid of aged mice induces proliferation and long-term memory consolidation of oligodendrocyte progenitor cells (OPC) in aged mice, whereas FGF17 blockade impairs cognition in young mice [39].

Our age classification model combines random forest and artificial neural network development, and the AUC of this model in the validation set is as high as 0.97, which can accurately distinguish between the 2 age groups of 18-35 years and >50 years. Current epigenetic age clocks constructed based on DNA methylation sites use mainly machine learning algorithms, and artificial neural network algorithms are less used. Horvath's team has established epigenetic aging clocks for several species (e.g., human [9], rhesus monkey [40], horse [41], dog [42], cat [43], etc.) using penalized regression models. Recently, Zaguia et al. [44] compared and analyzed an age prediction model developed using four different machine learning techniques. Among all the techniques, the model constructed using random forest regression showed the best performance. In addition, Fan et al. [45] screened 34 CpG sites of five genes (FHL2, ELOVL2, C1orf132, TRIM59, and KLF14) from blood samples, used different machine learning algorithms to build their models, and evaluated four age prediction models. They found that the MAD of the model constructed using random forest was only 1.15 years.

Deep learning was founded on the study of artificial neural networks, the core of deep learning. Aging is a complex process. The strength of deep learning in processing nonlinear data holds promise for the future interpretation of more types of complex biological data in the context of aging, with its greater ability to take into account the confounding effects of ethnic, geographic, behavioral, and environmental factors on predicting biological age [46]. In comparison with traditional machine learning algorithms, deep learning offers a paradigm shift for biological age studies. It is able to independently extract meaningful knowledge from complex data and avoid performing feature engineering or feature extraction to obtain robust performance [47]. In 2016, Putin et al. [48] reported a biological age prediction model built based on deep learning algorithms and developed an online system available at http://www.aging.ai. They combined basic blood metrics from more than 56,000 healthy individuals with actual age to design a modular combination of 21 Deep Neural Networks (DNNs). In this combination, the best-performing DNN had an MAE of 6.07 years when predicting age in 10 years, while the MAE for the entire combination was 5.55 years. Two years later, the team combined blood sample data from three populations of aging countries (South Korean, Canadian, and Eastern European) and then used DNNs to construct a biological age clock for a general population with a prediction accuracy of 5.94 years for MAE. The study showed that the accuracy of age prediction for both the combined population and each country's population was improved by using the combined dataset to construct the clock model. Thus, clocks with ethnic diversity have greater accuracy in predicting the biological age of different ethnic populations than general biological age clocks. Deep learning algorithms possess a significant advantage in describing the nonlinear relationship between blood parameters and actual age [49]. Recently, Gialluisi et al. [50] also constructed a biological age clock based on blood parameters including metabolic, cardiac,

renal, and liver function markers by DNNs. The risk relationship between lifestyle, economic status, etc., and the health status of the country's population was significantly predicted by Δage (BA-CA), with positive and negative values representing accelerated and slowed aging of the organism, respectively. Unfortunately, due to the sample size limitation, our age classification model does not use deep learning.

Population aging has become a major trend in human development. As a developing country, the trend of population aging in China is already evident. It is estimated that by 2050, people aged 65 and above will account for about 20% of the total population in China [51]. China is in the process of industrialization and urbanization, which has led to an increasing impact of behavioral changes and environmental factors on human health [52]. The biological age classification model we currently constructed can only classify young and middle-to-old age groups. In the next step, we will expand the sample size, increase the age prediction range, and combine the routine clinical biomarkers' metrics [53,54] to construct a biological age clock applicable to the Chinese population using deep learning.

The traditional view is that the methylation of CpG dinucleotides has a repressive effect on the DNA-binding activity of most transcription factors. However, this view is gradually changing because a number of studies have reported that more than one-third of TFs may preferentially bind to methylated sequences [55,56]. We constructed a multifactor regulatory network of genes corresponding to age-related characteristic methylation sites, which can be used to subsequently investigate the role of transcription factor binding DNA activity in relation to the methylation of CpG, expecting to further reveal the post-transcriptional regulation pattern of age-related genes and explore the aging regulation mechanism of the organism.

5. Conclusions

Our age classification model constructed using random forest and artificial neural network can accurately distinguish between people aged 18-35 years and >50 years, demonstrating the reliability of DNA methylation as a molecular marker of aging. With this model, we can accurately distinguish between different age groups at the molecular level, which will be more predictive than chronological age for assessing individual aging and future health status.

Data Availability

The raw data supporting the conclusions of this article are available from the GEO public database.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Consent for publications

The author read and approved the final manuscript for publication.

Ethics approval and consent to participate

No human or animals were used in the present research.

Informed Consent

The authors declare not used any patients in this research.

Authors' contributions

Ye Li and Xinghua Pan designed the study and performed the experiments, Yukun Yang, Jie He, Jinxiu Hu and Xiangqing Zhu collected the data, Chuan Tian, Mengdie Chen, Xiaojuan Zhao, Li Ye and Hang Pan analyzed the data, Ye Li and Xinghua Pan prepared the manuscript. All authors read and approved the final manuscript.

Funding Statement

This research was supported by the General Project of Yunnan Applied Basic Research Program (No. 202101AT070212); the 920th Hospital of the PLA Joint Logistics Support Force In-Hospital Technology Plan (No. 2020YGD05); and the All-Army Laboratory Animal Project (No. SYDW(2020)19).

Acknowledgments

The authors wish to thank Ph. D. Teng who helped with the evaluation of the data.

References

1. Kong X, Chen H, Li D, Ma L (2022) Effects of imbalance of lipid metabolism through NF-KB pathway on atherosclerosis and vascular aging in rats. Cell Mol Biol 67:144-150. doi: 10.14715/cmb/2021.67.5.20
2. Ryan CP (2021) "Epigenetic clocks": Theory and applications in human biology. Am J Hum Biol 33:e23488. doi: 10.1002/ajhb.23488
3. Jylhava J, Pedersen NL, Hagg S (2017) Biological Age Predictors. Ebiomedicine 21:29-36. doi: 10.1016/j.ebiom.2017.03.046
4. Salameh Y, Bejaoui Y, El HN (2020) DNA Methylation Biomarkers in Aging and Age-Related Diseases. Front Genet 11:171. doi: 10.3389/fgene.2020.00171
5. Lee Y, Haftorn KL, Denault W, Nustad HE, Page CM, Lyle R et al (2020) Blood-based epigenetic estimators of chronological age in human adults using DNA methylation data from the Illumina MethylationEPIC array. Bmc Genomics 21:747. doi: 10.1186/s12864-020-07168-8
6. Ashapkin VV, Kutueva LI, Vanyushin BF (2019) Epigenetic Clock: Just a Convenient Marker or an Active Driver of Aging? Adv Exp Med Biol 1178:175-206. doi: 10.1007/978-3-030-25650-0_10
7. Noroozi R, Ghafouri-Fard S, Pisarek A, Rudnicka J, Spolnicka M, Branicki W et al (2021) DNA methylation-based age clocks: From age prediction to age reversion. Ageing Res Rev 68:101314. doi: 10.1016/j.arr.2021.101314
8. Hannum G, Guinney J, Zhao L, Zhang L, Hughes G, Sadda S et al (2013) Genome-wide methylation profiles reveal quantitative views of human aging rates. Mol Cell 49:359-367. doi: 10.1016/j.molcel.2012.10.016
9. Horvath S (2013) DNA methylation age of human tissues and cell types. Genome Biol 14:R115. doi: 10.1186/gb-2013-14-10-r115
10. Weidner CI, Lin Q, Koch CM, Eisele L, Beier F, Ziegler P et al (2014) Aging of blood can be tracked by DNA methylation changes at just three CpG sites. Genome Biol 15:R24. doi: 10.1186/gb-2014-15-2-r24
11. Yin X, Nie J, Nie X, Nie Y (2022) The relationship between nuclear factor-erythrocyte-related factor 2 and antioxidant enzymes in the placenta of patients with gestational diabetes and the metabolism of umbilical cord endothelial cells. Cell Mol Biol 68:97-103. doi: 10.14715/cmb/2022.68.12.18
12. Marioni RE, Belsky DW, Deary IJ, Wagner W (2018) Association of facial ageing with DNA methylation and epigenetic age predic-

tions. Clin Epigenetics 10:140. doi: 10.1186/s13148-018-0572-2
13. Amiri RM, Mousavi SF, Salek AS, Lopes FB, Momen M, Gianola D et al (2021) Prediction of biological age and evaluation of genome-wide dynamic methylomic changes throughout human aging. G3-Genes Genom Genet 11doi: 10.1093/g3journal/jkab112
14. Han Y, Nikolic M, Gobs M, Franzen J, de Haan G, Geiger H et al (2020) Targeted methods for epigenetic age predictions in mice. Sci Rep-Uk 10:22439. doi: 10.1038/s41598-020-79509-2
15. Thompson MJ, VonHoldt B, Horvath S, Pellegrini M (2017) An epigenetic aging clock for dogs and wolves. Aging (Albany Ny) 9:1055-1068. doi: 10.18632/aging.101211
16. Petkovich DA, Podolskiy DI, Lobanov AV, Lee SG, Miller RA, Gladyshev VN (2017) Using DNA Methylation Profiling to Evaluate Biological Age and Longevity Interventions. Cell Metab 25:954-960. doi: 10.1016/j.cmet.2017.03.016
17. Fahy GM, Brooke RT, Watson JP, Good Z, Vasanawala SS, Maecker H et al (2019) Reversal of epigenetic aging and immunosenescent trends in humans. Aging Cell 18:e13028. doi: 10.1111/acel.13028
18. Rando TA, Chang HY (2012) Aging, rejuvenation, and epigenetic reprogramming: resetting the aging clock. Cell 148:46-57. doi: 10.1016/j.cell.2012.01.003
19. Maegawa S, Lu Y, Tahara T, Lee JT, Madzo J, Liang S et al (2017) Caloric restriction delays age-related methylation drift. Nat Commun 8:539. doi: 10.1038/s41467-017-00607-3
20. Kerepesi C, Meer MV, Ablaeva J, Amoroso VG, Lee SG, Zhang B et al (2022) Epigenetic aging of the demographically non-aging naked mole-rat. Nat Commun 13:355. doi: 10.1038/s41467-022-27959-9
21. Gao X, Liu S, Song H, Feng X, Duan M, Huang L et al (2020) AgeGuess, a Methylomic Prediction Model for Human Ages. Front Bioeng Biotech 8:80. doi: 10.3389/fbioe.2020.00080
22. Xu Y, Li X, Yang Y, Li C, Shao X (2019) Human age prediction based on DNA methylation of non-blood tissues. Comput Meth Prog Bio 171:11-18. doi: 10.1016/j.cmpb.2019.02.010
23. Naue J, Hoefsloot H, Mook O, Rijlaarsdam-Hoekstra L, van der Zwalm M, Henneman P et al (2017) Chronological age prediction based on DNA methylation: Massive parallel sequencing and random forest regression. Forensic Sci Int-Gen 31:19-28. doi: 10.1016/j.fsigen.2017.07.015
24. Vidaki A, Ballard D, Aliferi A, Miller TH, Barron LP, Syndercombe CD (2017) DNA methylation-based forensic age prediction using artificial neural networks and next generation sequencing. Forensic Sci Int-Gen 28:225-236. doi: 10.1016/j.fsigen.2017.02.009
25. Galkin F, Mamoshina P, Kochetov K, Sidorenko D, Zhavoronkov A (2021) DeepMAge: A Methylation Aging Clock Developed with Deep Learning. Aging Dis 12:1252-1262. doi: 10.14336/AD.2020.1202
26. Meng Y, Lin T, Liang S, Gao R, Jiang H, Shao W et al (2018) Value of DNA methylation in predicting curve progression in patients with adolescent idiopathic scoliosis. Ebiomedicine 36:489-496. doi: 10.1016/j.ebiom.2018.09.014
27. Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R et al (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res 31:374-378. doi: 10.1093/nar/gkg108
28. (2011) A user's guide to the encyclopedia of DNA elements (ENCODE). Plos Biol 9:e1001046. doi: 10.1371/journal.pbio.1001046
29. Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom AR, Ma'Ayan A (2010) ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics 26:2438-2444. doi: 10.1093/bioinformatics/btq466
30. Han H, Cho JW, Lee S, Yun A, Kim H, Bae D et al (2018) TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. Nucleic Acids Res 46:D380-D386. doi: 10.1093/nar/gkx1013
31. Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG et al (2016) The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database-Oxford 2016doi: 10.1093/database/baw100
32. Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H et al (2010) Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. Genome Res 20:434-439. doi: 10.1101/gr.103101.109
33. Wozniak A, Heidegger A, Piniewska-Rog D, Pospiech E, Xavier C, Pisarek A et al (2021) Development of the VISAGE enhanced tool and statistical models for epigenetic age estimation in blood, buccal cells and bones. Aging (Albany Ny) 13:6459-6484. doi: 10.18632/aging.202783
34. Correia DH, Cunha E, Corte RF, Manco L (2020) Age prediction in living: Forensic epigenetic age estimation based on blood samples. Legal Med-Tokyo 47:101763. doi: 10.1016/j.legalmed.2020.101763
35. Sukawutthiya P, Sathirapatya T, Vongpaisarnsin K (2021) A minimal number CpGs of ELOVL2 gene for a chronological age estimation using pyrosequencing. Forensic Sci Int 318:110631. doi: 10.1016/j.forsciint.2020.110631
36. Garali I, Sahbatou M, Daunay A, Baudrin LG, Renault V, Bouyacoub Y et al (2020) Improvements and inter-laboratory implementation and optimization of blood-based single-locus age prediction models using DNA methylation of the ELOVL2 promoter. Sci Rep-Uk 10:15652. doi: 10.1038/s41598-020-72567-6
37. Dias HC, Cordeiro C, Pereira J, Pinto C, Real FC, Cunha E et al (2020) DNA methylation age estimation in blood samples of living and deceased individuals using a multiplex SNaPshot assay. Forensic Sci Int 311:110267. doi: 10.1016/j.forsciint.2020.110267
38. Ronn T, Volkov P, Gillberg L, Kokosar M, Perfilyev A, Jacobsen AL et al (2015) Impact of age, BMI and HbA1c levels on the genome-wide DNA methylation and mRNA expression patterns in human adipose tissue and identification of epigenetic biomarkers in blood. Hum Mol Genet 24:3792-3813. doi: 10.1093/hmg/ddv124
39. Iram T, Kern F, Kaur A, Myneni S, Morningstar AR, Shin H et al (2022) Young CSF restores oligodendrogenesis and memory in aged mice via Fgf17. Nature 605:509-515. doi: 10.1038/s41586-022-04722-0
40. Horvath S, Zoller JA, Haghani A, Jasinska AJ, Raj K, Breeze CE et al (2021) Epigenetic clock and methylation studies in the rhesus macaque. Geroscience 43:2441-2453. doi: 10.1007/s11357-021-00429-8
41. Horvath S, Haghani A, Peng S, Hales EN, Zoller JA, Raj K et al (2022) DNA methylation aging and transcriptomic studies in horses. Nat Commun 13:40. doi: 10.1038/s41467-021-27754-y
42. Horvath S, Lu AT, Haghani A, Zoller JA, Li CZ, Lim AR et al (2022) DNA methylation clocks for dogs and humans. P Natl Acad Sci Usa 119:e2120887119. doi: 10.1073/pnas.2120887119
43. Raj K, Szladovits B, Haghani A, Zoller JA, Li CZ, Black P et al (2021) Epigenetic clock and methylation studies in cats. Geroscience 43:2363-2378. doi: 10.1007/s11357-021-00445-8
44. Zaguia A, Pandey D, Painuly S, Pal SK, Garg VK, Goel N (2022) DNA Methylation Biomarkers-Based Human Age Prediction Using Machine Learning. Comput Intel Neurosc 2022:8393498. doi: 10.1155/2022/8393498
45. Fan H, Xie Q, Zhang Z, Wang J, Chen X, Qiu P (2021) Chronological Age Prediction: Developmental Evaluation of DNA Methylation-Based Machine Learning Models. Front Bioeng Biotech

9:819991. doi: 10.3389/fbioe.2021.819991
46. Zhavoronkov A, Mamoshina P, Vanhaelen Q, Scheibye-Knudsen M, Moskalev A, Aliper A (2019) Artificial intelligence for aging and longevity research: Recent advances and perspectives. Ageing Res Rev 49:49-66. doi: 10.1016/j.arr.2018.11.003
47. Ashiqur RS, Giacobbi P, Pyles L, Mullett C, Doretto G, Adjeroh DA (2021) Deep learning for biological age estimation. Brief Bioinform 22:1767-1781. doi: 10.1093/bib/bbaa021
48. Putin E, Mamoshina P, Aliper A, Korzinkin M, Moskalev A, Kolosov A et al (2016) Deep biomarkers of human aging: Application of deep neural networks to biomarker development. Aging (Albany Ny) 8:1021-1033. doi: 10.18632/aging.100968
49. Mamoshina P, Kochetov K, Putin E, Cortese F, Aliper A, Lee WS et al (2018) Population Specific Biomarkers of Human Aging: A Big Data Study Using South Korean, Canadian, and Eastern European Patient Populations. J Gerontol a-Biol 73:1482-1490. doi: 10.1093/gerona/gly005
50. Gialluisi A, Di Castelnuovo A, Costanzo S, Bonaccio M, Persichillo M, Magnacca S et al (2022) Exploring domains, clinical implications and environmental associations of a deep learning marker of biological ageing. Eur J Epidemiol 37:35-48. doi: 10.1007/s10654-021-00797-7
51. Lin CY, Liao HE, Change-Lee SN, Yen YY (2022) Initial and Continuous Effects of Essential Oil Therapy in Relieving Knee Pain Among Older Adults With Osteoarthritis. Altern Ther Health M 28:10-17.
52. Liu J, Jiang B, Fu Y, Wang Z, Liu M, Yin P et al (2022) The Future Challenges of Population Health in China: A Projection of the Major Health Problems and Life Expectancy - China, 2015-2050. China Cdc Weekly 4:276-279. doi: 10.46234/ccdcw2022.069
53. Cao X, Yang G, Jin X, He L, Li X, Zheng Z et al (2021) A Machine Learning-Based Aging Measure Among Middle-Aged and Older Chinese Adults: The China Health and Retirement Longitudinal Study. Front Med-Lausanne 8:698851. doi: 10.3389/fmed.2021.698851
54. Liu Z (2021) Development and Validation of 2 Composite Aging Measures Using Routine Clinical Biomarkers in the Chinese Population: Analyses From 2 Prospective Cohort Studies. J Gerontol a-Biol 76:1627-1632. doi: 10.1093/gerona/glaa238
55. Mann IK, Chatterjee R, Zhao J, He X, Weirauch MT, Hughes TR et al (2013) CG methylated microarrays identify a novel methylated sequence bound by the CEBPB|ATF4 heterodimer that is active in vivo. Genome Res 23:988-997. doi: 10.1101/gr.146654.112
56. Zhu H, Wang G, Qian J (2016) Transcription factors as readers and effectors of DNA methylation. Nat Rev Genet 17:551-565. doi: 10.1038/nrg.2016.83