

Original Research

Identification of DNA Methylation Signature and Rules for SARS-CoV-2 Associated with Age

Lei Chen^{1,2,†}, Huiping Liao^{3,†}, Guohua Huang^{4,†}, Shijian Ding¹, Wei Guo⁵, Tao Huang^{6,7,*}, Yudong Cai^{1,*}

¹School of Life Sciences, Shanghai University, 200444 Shanghai, China

²College of Information Engineering, Shanghai Maritime University, 201306 Shanghai, China

³Ophthalmology & Optometry Medical School, Shandong University of Traditional Chinese Medicine, 250002 Jinan, Shandong, China

⁴School of Electrical Engineering, Shaoyang University, 422000 Shaoyang, Hunan, China

⁵Key Laboratory of Stem Cell Biology, Shanghai Jiao Tong University School of Medicine (SJTUSM) & Shanghai Institutes for Biological Sciences (SIBS), Chinese Academy of Sciences (CAS), 200031 Shanghai, China

⁶Bio-Med Big Data Center, CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, 200031 Shanghai, China

⁷CAS Key Laboratory of Tissue Microenvironment and Tumor, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, 200031 Shanghai, China

*Correspondence: touhuangtao@126.com (Tao Huang); cai_yud@126.com (Yu-Dong Cai)

†These authors contributed equally.

Academic Editor: Alike K. Maunakea

Submitted: 17 March 2022 Revised: 26 May 2022 Accepted: 26 May 2022 Published: 27 June 2022

Abstract

Background: COVID-19 displays an increased mortality rate and higher risk of severe symptoms with increasing age, which is thought to be a result of the compromised immunity of elderly patients. However, the underlying mechanisms of aging-associated immunodeficiency against *Severe acute respiratory syndrome coronavirus 2* (SARS-CoV-2) remains unclear. Epigenetic modifications show considerable changes with age, causing altered gene regulations and cell functions during the aging process. The DNA methylation patterns among patients with coronavirus 2019 disease (COVID-19) who had different ages were compared to explore the effect of aging-associated methylation modifications in SARS-CoV-2 infection. **Methods:** Patients with COVID-19 were divided into three groups according to age. Boruta was used on the DNA methylation profiles of the patients to remove irrelevant features and retain essential signature sites to identify substantial aging-associated DNA methylation changes in COVID-19. Next, these features were ranked using the minimum redundancy maximum relevance (mRMR) method, and the feature list generated by mRMR was processed into the incremental feature selection method with decision tree (DT), random forest, *k*-nearest neighbor, and support vector machine to obtain the key methylation sites, optimal classifier, and decision rules. **Results:** Several key methylation sites that showed distinct patterns among the patients with COVID-19 who had different ages were identified, and these methylation modifications may play crucial roles in regulating immune cell functions. An optimal classifier was built based on selected methylation signatures, which can be useful to predict the aging-associated disease risk of COVID-19. **Conclusions:** Existing works and our predictions suggest that the methylation modifications of genes, such as *NHLH2*, *ZEB2*, *NWD1*, *ELOVL2*, *FGGY*, and *FHL2*, are closely associated with age in patients with COVID-19, and the 39 decision rules extracted with the optimal DT classifier provides quantitative context to the methylation modifications in elderly patients with COVID-19. Our findings contribute to the understanding of the epigenetic regulations of aging-associated COVID-19 symptoms and provide the potential methylation targets for intervention strategies in elderly patients.

Keywords: SARS-CoV-2; DNA methylation signature; age; feature selection; classification algorithm

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which belongs to subfamily *Coronavirinae*, is a highly pathogenic coronavirus that triggered the coronavirus disease 2019 (COVID-19) pandemic. The SARS-CoV-2 genome is approximately 29.9 kb and contains four structural proteins and 16 non-structural proteins. Certain non-structural proteins can inhibit the host's viral defense function by influencing the mRNA splicing, translation, and transport of secreted proteins. Molecular structure mutations have different kinds, and the fast mutation

speed causes the virus to spread rapidly, causing great challenges to epidemic prevention and control.

At present, in addition to layered prevention strategies, including wearing masks, vaccination is the most effective way to fight infection. Various countries and regions have adopted an active vaccination strategy, and the COVID-19 pneumonia has been controlled to a certain extent. However, some problems persist. On the one hand, vaccine coverage in some parts of the world is low and is not enough to form an immune barrier. On the other hand, virus variants, such as the alpha, beta, and B.1.617.2 (delta) variants, have



appeared in the population. In comparison with the original SARS-CoV-2, the delta variant spreads faster and can cause more infections. The results of two studies in Canada and Scotland showed that the delta variant may cause more serious diseases than the previous variants. Moreover, an increasing number of people infected with the new coronavirus has experienced various symptoms, such as huge tongue and wrong sense of smell. Aside from causing pneumonia, SARS-CoV-2 can also cause thrombosis and other neurological diseases. After the acute SARS-CoV-2 infection of patients with COVID-19, cell metabolism becomes disordered and may eventually develop into diabetes [1,2]. In addition, SARS-CoV-2 targets pancreatic islet endocrine cells [3] and destroys its structure, resulting in abnormal insulin levels in the body.

Viral mutations make existing vaccines incapable of protecting the human body efficiently, especially people who only received one shot of the vaccine [4]. An increasing number of patients has experienced various sequelae. These undesirable factors are closely related to host genetic factors. Researchers analyzed the DNA methylation data of lung tissue, screened the methylation variants of the *ACE2* gene, and found that changes in the human epigenome may be related to COVID-19 risk [5]. In addition, scientists found that the severity and poor prognosis of COVID-19 are affected by old age [6–10]. Elderly patients with COVID-19 can produce a high level of lymphocyte activating factor. Patients with severe COVID-19 have high levels of inflammatory cytokines, such as interleukin (IL)-6 and IL-1 β [11–15], and the concentration of inflammatory cytokines increases with age [16]. IL-10 is correlated with COVID-19 severity [17]. High IL-10 and PD-L1 concentrations have been detected in critically ill patients and are therefore considered biomarkers of immune failure [18]. IL-10 was reported to play an important role in the COVID-19 process [19]. Circulating hepatocyte growth factor can regulate cell proliferation and migration [20], angiogenesis, tissue regeneration, and other cellular processes. It has a high concentration in critically ill patients with COVID-19 [21] and increases with age. These findings may help explain the high incidence of COVID-19 among the elderly and patients with underlying diseases [22]. In addition to advanced age, risk factors, such as underlying diseases [9,23–29] and complications [30–33], are related to COVID-19 severity.

Statistical analysis of high-throughput data can provide an important reference for revealing the pathophysiology and pathogenic mechanism of COVID-19. Our team has long been committed to using machine learning analysis methods to screen disease-related signatures and explain their pathogenic mechanisms. Accordingly, we aimed to further explore the pathogenic mechanism of COVID-19 and determined the molecular information related to disease complications based on the blood methylation profiles of 407 patients with mild or se-

vere COVID-19 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168739>). We used appropriate calculation methods to select important epigenetic information, which may help in the effective understanding of severe COVID-19 with age. Boruta [34,35] and minimal redundancy maximum relevance (mRMR) [36] were adopted to select the methylation sites with the highest correlation with the age of patients with COVID-19, and the selected sites were ranked to obtain a list of methylation sites. This list was fed into the incremental feature selection (IFS) method [37], which integrated four classification algorithms, to identify aging-associated methylation sites and construct decision rules, which provide a quantitative description of the correlation between methylation changes and aging in patients with COVID-19. Furthermore, efficient classifiers were built with essential aging-associated methylation sites. Some of the important methylation sites obtained from the analysis are supported through literature review. Our study revealed the aging-associated methylation changes that may impact antiviral immune responses in SARS-CoV-2 infection. Our findings contribute to the understanding of epigenetic regulations in aging-associated COVID-19 symptoms and provide the potential methylation targets for intervention strategies in elderly patients.

2. Materials and Methods

In the present study, feature selection methods and machine learning algorithms were used to filter out key methylation site features and decision rules. The detailed analysis flow is shown in Fig. 1, and its detailed description is presented below.

2.1 Dataset

The COVID-19 blood methylation profiles analyzed in this study were collected from the Gene Expression Omnibus (GEO) database with access ID GSE168739 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168739>) [38]. Among the 407 patients with COVID-19, 111 were below 35 years old, 132 were 35–45 years old, and 164 were over 45 years old. Age groups were considered the classification labels for this study. This dataset contains 865,149 DNA methylation sites. The detailed information of the samples can be found at the GEO website <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE168739> and from the descriptions of Castro de Moura *et al.*'s study [38].

2.2 Boruta Feature Selection

The investigated COVID-19 methylation profiles involved a large number of methylation sites. Evidently, not all sites are related to the age of patients with COVID-19. Irrelevant sites were excluded. As elaborated in [35], Boruta [34] is an excellent feature selection method and always provides best performance among random forest (RF)-based feature selection methods. Thus, in this study, it was employed to analyze the methylation profiles.

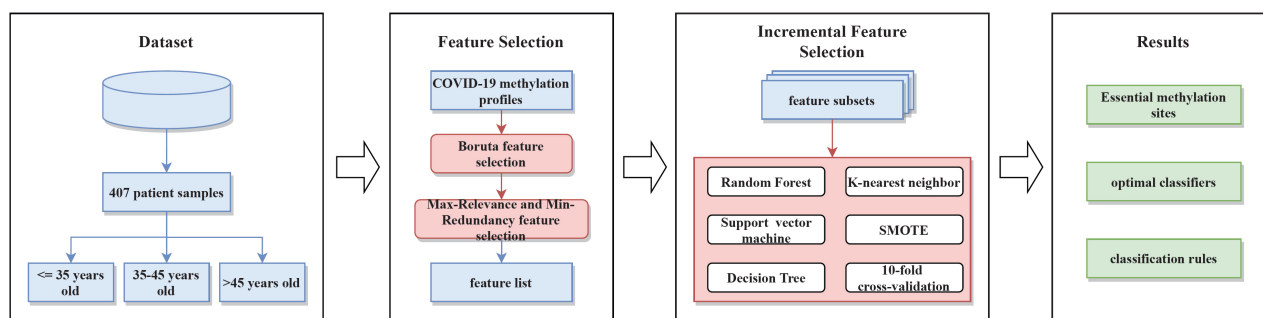


Fig. 1. Flow chart for classifying samples from three age groups of patients with COVID-19. The dataset was analyzed by Boruta and mRMR individually to obtain a feature list. The list was fed into the incremental feature selection method to extract essential methylation sites, classification rules and build optimal classifiers.

Boruta is a wrapper approach that uses a RF [39] classification algorithm. The importance of features is evaluated by comparing them with shadow features. For a dataset with n features, Boruta first creates a shadow feature for each real feature, whose values are produced by shuffling values under the real feature. Accordingly, a new dataset with $2n$ features is generated and fed into a RF to compute the importance of all features. Feature importance is defined as the accuracy loss of RF caused by the permutation of values under this feature. The Z score of each shadow feature is defined as the average loss divided by standard deviation of the accuracy loss. The maximum Z score among shadow features, called MZSA, was found. The real features with importance significantly lower than the MZSA was tagged as “unimportant”, whereas those with importance significantly higher than the MZSA was tagged as “important”. The “unimportant” features and shadow features were removed. The procedure was repeated until all remaining features were tagged as “important” or RF ran the predefined times. The “important” features were output as the outcome of Boruta.

In this study, the Boruta program retrieved from https://github.com/scikit-learn-contrib/boruta_py was applied on the methylation profiles mentioned in Section 2.1, and default parameters were set to execute such program. The methylation features selected by Boruta were further investigated in the following analysis.

2.3 Minimum Redundancy Maximum Relevance

The features selected by Boruta were further analyzed by mRMR. mRMR is a feature selection method proposed by Peng *et al.* [36] and is extensively utilized in bioinformatics and biomedical research.

The mRMR method evaluates the importance of features based on maximum relevance and minimum redundancy, which are quantified based on mutual information (MI). The outcome of the mRMR method is a feature list, named the mRMR feature list. The list is generated with the following procedures. The mRMR method first creates

an empty mRMR feature list and gradually adds one feature that has the most relevance to the classification labels and the least redundancies to features already in the list. The procedure stops when all features are in the list. Evidently, features with high ranks in the list can comprise an optimal feature subspace for a classification algorithm. These features are essential for the classification problem.

The mRMR program used in this article was accessed at <http://home.penglab.com/proj/mRMR/> and processed using default parameters. An mRMR feature list was obtained, which was denoted by F in this study, by analyzing the methylation profiles with features selected by Boruta.

2.4 Incremental Feature Selection

Although the mRMR method evaluates the importance by outputting an mRMR feature list, selecting features from this list to optimally represent samples is still a problem. In view of this problem, the IFS method was employed in this study. The IFS method is an excellent feature selection algorithm that can determine optimal features [37] from a feature list for a given classification algorithm. IFS first produces a series of feature subsets from a feature list (mRMR feature list in this study) according to a given step interval. For example, if the step interval is set to 1, the first feature subset includes the top feature in the list, the second feature subset contains the top two features, and so on. Next, for each feature subset, a training dataset, where samples are represented by features in the subset, is fed into one classification algorithm to build the classifier. This classifier is assessed by 10-fold cross-validation [40]. The classifier with the best performance can be found after all the classifiers have been tested. This classifier is called the optimal classifier, and features used in the classifier are termed as optimal features in this study.

2.5 Synthetic Minority Oversampling Technique

The performances of all classifiers, which can be influenced by several factors, were evaluated by IFS. Sample distribution (patients with COVID-19 in this study) is one

of the major factors. As mentioned in Section 2.1, different groups contained different numbers of samples. Therefore, the classifier directly built on such dataset will not be efficient enough. Accordingly, synthetic minority oversampling technique (SMOTE) [41] was adopted in this study.

SMOTE aims to balance the number of samples in the training process. This task is completed by using the k -nearest neighbor (kNN) technique [42] to generate new samples for minor classes, resulting in equal sample size for each class in the dataset. In detail, a sample, denoted by x , is randomly selected from one minor class. Its kNNs in this class are discovered, and one neighbor, denoted by y , is randomly selected. A new sample is constructed based on x and y , which is defined as their linear combination with randomly produced combination coefficients. As this new sample has strong associations with x and y , it is placed in the same minor class. This procedure is executed several times until the minor class has same number of samples in the major class.

In this analysis, the Python version of the SMOTE program was downloaded from <https://github.com/scikit-learn-contrib/imbalanced-learn>, and the parameters were set to default. The samples produced by SMOTE were only used in IFS and were not involved in Boruta and mRMR.

2.6 Classification Algorithm

A classification algorithm is necessary to perform the IFS method. In this study, four different classification algorithms were selected, namely, RF [39], kNN [42], support vector machine (SVM) [43], and decision tree (DT) [44]. We compared the performances of the classifiers generated based on these algorithms in IFS. These algorithms are widely used to tackle various biological and medical problems [45–55]. They are described in this section and were programmed in Python via the Scikit-learn module with default parameters.

2.6.1 Random Forest

RF is an ensemble learning method based on a bagging algorithm, which consists of many decision trees. In the training procedure, samples and features are randomly selected to construct a DT for several times. These DTs comprise the RF. For a query sample, each DT in the forest makes a prediction. RF integrates the predictions with majority voting. Considering that RF uses an ensemble algorithm, it has a superior performance over DT.

2.6.2 k-nearest Neighbor

kNN is a supervised learning algorithm that performs classification by measuring the distances between a query sample and training samples. Training samples are ranked with increasing order of their distances to the query sample. Then, kNN selects the top k training samples and estimates the frequency of occurrence of these k sample labels. Finally, the class label with the highest frequency is regarded

as the label for the query sample.

2.6.3 Support Vector Machine

SVM is a well-known machine learning algorithm that has attracted extensive attention. The original SVM finds out the optimal hyperplane that can separate samples in two class with maximum margins. However, this hyperplane does not exist in many cases. In this case, SVM was used to map the training samples into a high-dimensional feature space through one kernel function. Subsequently, the optimal hyperplane can be constructed in the high-dimensional space. The query sample can also be mapped into the high-dimensional feature space. Its class is determined by the side of the hyperplane where it is located. For multiclassification problems, SVM adopts the “one against the rest” strategy, which regards each class as positive and all other samples as negative to construct a binary SVM classifier.

2.6.4 Decision Tree

Many classification algorithms, including RF, kNN, and SVM, are complete black-box algorithms. Their classification principle learned from a training dataset is very difficult to understand. Although efficient classifiers can be set up based on these algorithms, few insights can be extracted from these classifiers. For the problem investigated in this study, we cannot obtain useful clues to uncover methylation patterns on patients with COVID-19 under different age groups. Accordingly, DT, a white-box algorithm, was also adopted in this study. The classification procedure of DT is completely open, providing opportunities for us to understand its classification principle. A tree can be built by applying DT to a training dataset. From this tree, several IF–THEN rules can be obtained. The conditions in the IF clause indicate a special pattern for the result in the THEN clause. A further investigation of the obtained rules is helpful to uncover different patterns among patients with COVID-19 with different ages.

2.7 Performance Measurement

In the present study, overall accuracy and Mathews correlation coefficient (MCC) [56] were applied to evaluate the performance of classifiers. Overall accuracy is one of the most important measurements for multiclass classification. It is defined as the proportion of correctly predicted samples among all samples. However, this measurement is not perfect when the sizes of classes have great differences. In this case, the MCC is much more accurate. Its formula is as follows:

$$\text{MCC} = \frac{\text{cov}(X, Y)}{\sqrt{\text{cov}(X, X) \cdot \text{cov}(Y, Y)}}, \quad (1)$$

where X stands for the matrix of the predicted labels, Y represents the matrix of real labels, and $\text{cov}(*, *)$ is used to calculate the correlation coefficient of the two matrices. MCC

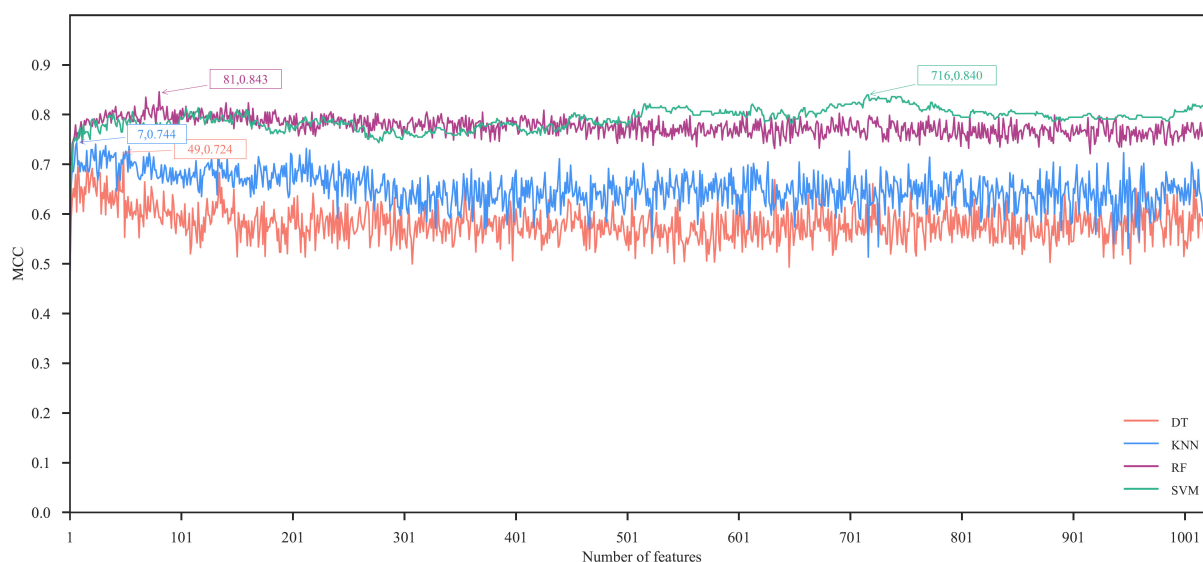


Fig. 2. IFS curves of different classifiers on the different number of methylation site features. DT, KNN, RF, and SVM provided the highest MCC values of 0.724, 0.744, 0.843, and 0.840, respectively, using the top 49, 7, 81, and 716 features, respectively. The RF classifier with the top 81 features provided the best performance.

ranges from -1 to $+1$. A classifier with a MCC closer to 1 has a stronger performance. This study adopted MCC as the key measurement.

In addition, the accuracy on each class was also computed to evaluate the performance of classifiers on each class. It is defined as the proportion of correctly predicted samples in one class among all samples in the same class.

3. Results

3.1 Results of Feature Selection

The COVID-19 methylation profiles were first processed by Boruta to exclude insubstantial features. A total of 1027 important sites were screened from 865,149 methylation sites as presented in **Supplementary Table 1**. The Boruta algorithm obtains a considerably good pre-processing result, which also reduces the computation time required to run the following analysis. Next, mRMR was used to evaluate the importance of the retained features and sort them in the mRMR feature list. The list is also shown in **Supplementary Table 1**.

3.2 Results of IFS with Classification Algorithms

The mRMR feature list was imported into IFS and combined with several classification algorithms to classify samples into different age groups. The step interval of IFS was set to 1, inducing a total of 1027 feature subsets. On each feature subset, one classifier was built based on one of the four classification algorithms. All classifiers were evaluated by 10-fold cross-validation. Overall accuracy, MCC, and accuracy on each class were calculated as evaluation metrics for each classifier, and the results are shown in **Supplementary Table 2**. For an accurate presentation, we plotted the IFS curve with the number of features in the x-axis

and the MCC metric in the y-axis for each classification algorithm as presented in Fig. 2. According to Fig. 2, MCC reached the maximum values of 0.724, 0.744, 0.843, and 0.840 for DT, kNN, RF, and SVM, respectively, using the top 49, 7, 81, and 716 features, respectively. The optimal classifiers were also obtained for the four algorithms using the above features. The overall accuracies of DT, kNN, RF, and SVM were 0.818, 0.830, 0.897, and 0.894, respectively (Table 1). Evidently, the optimal RF and SVM classifiers were superior to the other two optimal classifiers. A bar chart was plotted to show their performance on three classes to fully display the performance of four optimal classifiers (Fig. 3). The optimal RF and SVM classifiers had much better performances on the patients aged 35–45 and >45 years, and their superiority was not clear on the patients aged ≤ 35 years old. Although the classification performances of the optimal DT and kNN classifiers were worse than the other two algorithms, they resulted in a decent performance, indicating the effectiveness of our analysis.

Table 1. Performance of the optimal classifiers based on different classification algorithms.

Classification algorithm	Number of features	Overall accuracy	MCC
Decision tree	49	0.818	0.724
k-nearest neighbor	7	0.830	0.744
Random forest	81	0.897	0.843
Support vector machine	716	0.894	0.840

Although the optimal RF and SVM classifiers provided almost equal performance, their efficiencies were not the same. The optimal RF classifier used 81 top features,

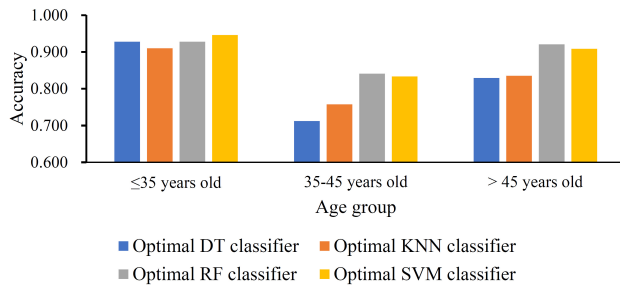


Fig. 3. Performance of four optimal classifiers on three age groups. The optimal RF and SVM classifiers provided evident better performance on two age groups than the other two optimal classifiers.

whereas the optimal SVM classifier adopted the top 716 features. Furthermore, RF is much faster than SVM. Thus, the optimal RF classifier was more proper than the optimal SVM classifier as a tool to predict the aging-associated risk of COVID-19.

3.3 Results of Decision Rules

As mentioned in Section 2.6.4, DT can help uncover methylation patterns in patients with COVID-19 under different age groups. As 49 features were used in the optimal DT classifier, we applied DT to all samples represented by these 49 features. A large tree was obtained, from which a total of 39 rules were extracted. These rules are listed in **Supplementary Table 3**. Among these 39 rules, 11 rules are predictions for the category of patients below 35 years of age, 18 rules are predictions for the category of patients between 35 and 45 years of age, and 10 rules are predictions for the category of patients over 45 years of age. A detailed description of these rules is presented in Section 4.2.

4. Discussion

Feature selection methods, such as Boruta, mRMR, and IFS, and classification algorithms, such as DT, RF, kNN, and SVM, were performed to analyze the methylation profiles of patients with COVID-19 and obtain a list of important candidate methylation sites. However, the use of classifiers alone to screen for methylation modification sites is not enough. Accordingly, we further applied the learning algorithm, DT, to clarify the expression rules in patients. We obtained 39 classification rules through rigorous analysis. The top-ranked features in the mRMR result can indicate important methylation modification information related to aging-induced risk of COVID-19. The classification rules present the methylation level of identified sites for classifying different age groups. We focused on valuable features and decision rules, because they can identify important DNA methylation sites and suggest their essential role as epigenetic susceptibility sites in patients infected with SARS-CoV-2. We collected the scientific findings of other researchers and initially summarized the experimen-

tal evidence of the above-mentioned features and rules to prove the accuracy of the prediction.

4.1 Analysis of Top Features Identified via mRMR

We initially identified 49 important features with the most relevance in the classification. We found that the methylation levels of some CpGs are highly associated with the age factor in patients with COVID-19, which was consistent with previous research findings [57–59]. Some methylation sites are located in the annotated coding region of genes. Next, we briefly summarized the experimental evidence of important features to provide more references for the exploration of age-related COVID-19 mechanisms.

The first methylation site in the feature ranking is cg21863110, and its methylation modification is closely related to aging phenotype. Aging is one of the important risk factors for cancer [60]. An increasing number of elderly patients are at higher risk of death owing to the COVID-19 outbreak. Studies showed that DNA damage has a time-dependent characteristic [61] and that aging induces specific DNA methylation [62–66]. Cg21863110 is located in the *NHLH2* gene (ENSG00000177551) interval, and functional enrichment analysis showed that *NHLH2* was associated with signaling pathways, including primary immunodeficiency signaling pathway, leukocyte extravasation signaling pathway, pluripotent stem cell hematopoiesis, and antigen presentation pathway. The above research evidence suggests that *NHLH2* in elderly patients with COVID-19 exhibits abnormal hypermethylation [67], causing primary immunodeficiency or abnormal antigen presentation, which ultimately leads to more severe COVID-19 symptoms in elderly patients.

The methylation site, cg00573770, is located at chr2:145278485 and belongs to the shelf region of the CpG island. Several published epigenome-wide association studies (EWASs) have reported the role of cg00573770 in aging [59,68,69]. Cg00573770 is associated with *ZEB2* (ENSG00000169554) and is located in the promoter region of the *ZEB2* gene. *ZEB2* is a zinc finger transcription factor that can bind DNA molecules in tandem with *ZEB1* and directly compete for E-protein binding sites. *ZEB2* expression can be induced by the transcription factor, T-bet, which triggers the differentiation of cytotoxic T lymphocytes to the terminal state [70]. *ZEB2* deletion can cause the loss of antigen-specific CD8⁺ T cells in patients with primary and secondary infections. In addition, transcription factor *ZEB2* can promote epithelial–mesenchymal transition, regulate NK cell maturation, and remarkably reduce memory cell production [71], as well as maintain the tissue specificity of macrophages [72]. The low methylation level of *ZEB2* in the elderly population leads to its abnormal expression, which may impair the specificity of macrophages and the maturation of NK cells, ultimately causing severe reactions in elderlies infected with SARS-CoV-2.

Cg19344626 is located in chr19:16830749 and the

open sea of the CpG island. The top trait in the EWAS atlas associated with this methylation site is aging [68,73–76]. Cg19344626 is located in the promoter of *NWD1* (ENSG00000188039). *NWD1* is a member of the innate immune protein subfamily, has a gene size of approximately 98 kb, and contains 19 exons and 1,358 amino acids. Substantial evidence suggests that *NWD1* methylation is associated with aging [76] and age-related diseases [75] and can regulate circulating cytokines and leukocyte function [77]. The levels of cytokines, such as TNF [78], IL-6 [79], and IL-10 [80], are increased in the elderly. These factors can reflect the chronic inflammatory state of the body and increase with age, gradually causing immune aging and death [81]. Another study reported that DNA methylation levels at several loci are remarkably associated with changes in the levels of the inflammatory marker, C-reactive protein [82]. The findings suggest that the age-related loss of DNA methylation increases the levels of immune-related factors by increasing transcriptional activity [83,84]. We screened the *NWD1*-related methylation site, cg19344626, from samples from patients with COVID-19 who had different ages. Combined with existing studies, we speculated that *NWD1* methylation may be an age-related risk factor for COVID-19. Our findings provide a reference for the molecular mechanisms of differential COVID-19 symptoms.

4.2 Analysis of Decision Rules Identified by the DT Method

We constructed 38 decision rules from the 49 features of 407 patients using the DT algorithm. Based on these quantitative rules, methylation signatures can be used to predict the elderly patients with COVID-19 who had a high risk of severe disease. We next briefly describe the relevant research evidence based on the biological significance of these rules.

The cg16867657 site is located at chr6:11044877, which belongs to the island area of the CpG island. Our analysis found that the methylation levels of cg16867657 were different in the different age groups. The methylation level of cg16867657 was significantly ($p = 4.65 \times 10^{-37}$) associated with age [59], which was consistent with the results of another study [76]. The genes associated with cg16867657 include *ELOVL2* (ENSG00000197977) and *RPI-62D2.3* (ENSG00000230314). The protein-coding gene *ELOVL2* encodes a type of transmembrane protein, which is involved in the formation of polyunsaturated fatty acids. The expression products of *ELOVL2* gene has transferase activity, partial acyl transfer, and fatty acid elongase activity. Methylation studies in elderly twins revealed the strongest age-related CpG methylation pattern in the promoter region of the *ELOVL2* gene [74]. The methylation modification of *ELOVL2* may play an important role in age-related diseases. Aging can induce the gut microbiota changes and lead to the dysregulation of *ELOVL2* expression [85]. Based on the whole-blood DNA analysis

of 64 subjects, the researchers identified the CpG islands of the *ELOVL2*, *FHL2*, and *PENK* genes, whose methylation levels were strongly correlated with age. Among them, *ELOVL2* methylation levels gradually increased with age (Spearman correlation coefficient = 0.92) and appeared to be a very promising biomarker of aging [86]. Another study based on monocytes and T cell lines found a remarkable age-related methylation site on the *ELOVL2* gene at chromosome 6 [87]. The lncRNA of *ELOVL2* is associated with immune cell infiltration in breast cancer and affects the prognosis of patients [88]. The methylation level of *ELOVL2* in peripheral blood cells was higher than that in EBV-transformed lymphoblastoid cell line, and the correlation between chronological age and aging-related methylation sites in *ELOVL2* was strong [89]. Our decision rules showed that the hypermethylation of cg16867657 is indicative of the older group of patients with COVID-19. The hypermethylation of *ELOVL2* leads to the downregulation of its expression level, which has adverse effects on the body's immunity. Moreover, this effect has become more severe with age in patients infected with SARS-CoV-2. Aging or senescence is an important risk factor for the severity of illness and death of patients with COVID-19 [90]. The immune function decreases with age and may make the elderly vulnerable to COVID-19. The patients with COVID-19 who had high methylation level of *ELOVL2* may have a higher disease risk. These findings help in explaining the severe symptoms of COVID-19 and provide important references for the development of a therapeutic vaccine.

Cg11425788 lies in chr1:60136100 and is located in the open sea region of the CpG island. Traits related to this methylation site include aging, prostate cancer, and mortality according to EWAS atlas [68,91]. Cg11425788 is located in the gene body of *FGGY*. The product encoded by *FGGY* has kinase and phosphotransferase activities, and the receptor is an alcohol group, which can phosphorylate carbohydrates, such as ribulose and ribitol. Multiple studies reported the effect of *FGGY* in aging phenotype [68,73,92]. In addition, a Chinese lung squamous cell carcinoma (LUSC) cohort study found that hypomethylation-induced *FGGY* retrotransposition (LINE-1-*FGGY*) can overcome the local immune evasion of LUSC and promote LUSC progression [93]. In the present study, the methylation level of cg11425788 was much lower in middle-aged and elderly patients with COVID-19, which implies that the hypomethylation-induced transcriptional regulation of *FGGY* may contribute to viral immune escape, which in turn aggravates the infection symptoms of patients.

Another important methylation site is cg22454769, which is located at chr2:106015767 and belongs to the island region of the CpG island. The most important traits in EWAS atlas related to cg22454769 is aging [89]. Cg22454769 is related to the *FHL2* gene (ENSG00000115641) and is located in the promoter re-

gion of *FHL2*. The protein coding gene, *FHL2*, with four half-LIM domain 2, belongs to the four half-Lim-only protein family members and is involved in the assembly of cell outer membranes [94,95]. *FHL2* can interact with cell surface receptors, kinases, cytoplasmic junctions, nuclear transcription factors, and other molecules and participate in peroxisome proliferator-activated receptor alpha and lipid metabolism regulation pathways. The study found that 97% of 1214 age-related DNA methylation sites in breast tissue were elevated in methylation levels, particularly in autosomal CpG islands and non-enhancers. Among them, the methylation level of 15 transcription regulators, such as *FHL2*, increases the risk of breast cancer with age [96]. The results of studies based on genome-wide DNA methylation in pancreatic islets have shown that aging is associated with increased levels of DNA methylation at many loci. The methylation level of *FHL2* in pancreatic islets is related to age and insulin secretion. Abnormal methylation levels in aging populations may be associated with impaired islet function and increased diabetes risk [97]. In addition, life processes, such as epithelial–mesenchymal transition, tissue repair, cell proliferation, inflammation regulation, apoptosis process, cell adhesion, and migration, are associated with the role of *FHL2* [94,98,99]. After the human body is infected with a virus, the *FHL2* protein is transferred to the nucleus to support type I interferon transcription. This defense mechanism can inhibit the proliferation of pathogens and reduce tissue damage. *FHL2* can also help the transcription of pro-inflammatory cytokines, such as IL-6 and IL-8 [98]. In the present study, the methylation level of cg22454769 in *FHL2* was increased in elderly patients with a value exceeding the threshold of 0.514. Combined with existing research evidence, we speculated that changes in the specific methylation sites of *FHL2* may increase the risk of disease in elderly patients by affecting the body's immunity. It may be a biomarker for monitoring age-related COVID-19 risk and may help to explore the pathogenic mechanisms of COVID-19.

5. Conclusions

This study aimed to determine the methylation signatures associated with aging in patients with COVID-19 through computational analysis. Based on our computational studies, we carefully selected key methylation features using machine learning methods, including Boruta, mRMR, IFS, and four classification algorithms. The results showed that the identified key features are consistent with published academic studies. We established a computational framework to detect key methylation site information, and the results were accurate and credible. However, the correlation between the methylation candidate molecules and COVID-19 requires in-depth cellular and molecular biological verification. The important age-related methylation signatures obtained in this study are helpful to understand the pathogenesis of COVID-19 risk

and provide references for the targeted intervention of the disease.

Abbreviations

SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; COVID-19, coronavirus disease 2019; mRMR, minimum redundancy maximum relevance; MI, mutual information; IFS, incremental feature selection; SMOTE, synthetic minority oversampling technique; RF, random forest; kNN, k-nearest neighbor; SVM, support vector machine; DT, decision tree; MCC, Mathews correlation coefficient.

Author Contributions

TH and YDC designed the research study. LC, GHH, and SJD performed the research. HPL and WG analyzed the data. LC, HPL and GHH wrote the manuscript. All authors contributed to editorial changes in the manuscript. All authors read and approved the final manuscript.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

Not applicable.

Funding

This research was funded by the Strategic Priority Research Program of Chinese Academy of Sciences [XDB38050200, XDA26040304], National Key R&D Program of China [2018YFC0910403], the Fund of the Key Laboratory of Tissue Microenvironment and Tumor of Chinese Academy of Sciences [202002].

Conflict of Interest

The authors declare no conflict of interest. YDC is serving as the editorial board member of this journal. We declare that YDC had no involvement in the peer review of this article and has no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to AKM.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at <https://doi.org/10.31083/j.fbl2707204>.

References

- [1] Tang X, Uhl S, Zhang T, Xue D, Li B, Vandana JJ, *et al.* SARS-CoV-2 infection induces beta cell transdifferentiation. *Cell Metabolism*. 2021; 33: 1577–1591.e7.
- [2] Wu CT, Lidsky PV, Xiao Y, Lee IT, Cheng R, Nakayama T, *et al.* SARS-CoV-2 infects human pancreatic β cells and elicits β cell impairment. *Cell Metabolism*. 2021; 33: 1565–1576.e5.

- [3] Yang L, Han Y, Nilsson-Payant BE, Gupta V, Wang P, Duan X, *et al.* A Human Pluripotent Stem Cell-based Platform to Study SARS-CoV-2 Tropism and Model Virus Infection in Human Cells and Organoids. *Cell Stem Cell.* 2020; 27: 125–136.e7.
- [4] Lopez Bernal J, Andrews N, Gower C, Gallagher E, Simmons R, Thelwall S, *et al.* Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. *New England Journal of Medicine.* 2021; 385: 585–594.
- [5] Muus C, Luecken MD, Eraslan G, Waghray A, Heimberg G, Sikkema L, *et al.* Integrated analyses of single-cell atlases reveal age, gender, and smoking status associations with cell type-specific expression of mediators of SARS-CoV-2 viral entry and highlights inflammatory programs in putative target cells. *bioRxiv.* 2020.
- [6] Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet.* 2020; 395: 1054–1062.
- [7] Zhang J, Cao Y, Tan G, Dong X, Wang B, Lin J, *et al.* Clinical, radiological, and laboratory characteristics and risk factors for severity and mortality of 289 hospitalized COVID-19 patients. *Allergy.* 2021; 76: 533–550.
- [8] Ou M, Zhu J, Ji P, Li H, Zhong Z, Li B, *et al.* Risk factors of severe cases with COVID-19: a meta-analysis. *Epidemiology and Infection.* 2020; 148: e175.
- [9] Wolff D, Nee S, Hickey NS, Marschollek M. Risk factors for Covid-19 severity and fatality: a structured literature review. *Infection.* 2021; 49: 15–28.
- [10] Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, *et al.* Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. *The Journal of the American Medical Association.* 2020; 323: 1061–1069.
- [11] Schulte-Schrepping J RN, Paclik D, Bassler K, Schlickeiser S, Zhang B, *et al.* Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment. *Cell.* 2020; 182: 1419–1440.e23.
- [12] Chen G, Wu D, Guo W, Cao Y, Huang D, Wang H, *et al.* Clinical and immunological features of severe and moderate coronavirus disease 2019. *Journal of Clinical Investigation.* 2020; 130: 2620–2629.
- [13] Giamarellos-Bourboulis EJ, Netea MG, Rovina N, Akinosoglou K, Antoniadou A, Antonakos N, *et al.* Complex Immune Dysregulation in COVID-19 Patients with Severe Respiratory Failure. *Cell Host and Microbe.* 2020; 27: 992–1000.e3.
- [14] Lucas C, Wong P, Klein J, Castro TBR, Silva J, Sundaram M, *et al.* Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature.* 2020; 584: 463–469.
- [15] Ong EZ, Chan YFZ, Leong WY, Lee NMY, Kalimuddin S, Haja Mohideen SM, *et al.* A Dynamic Immune Response Shapes COVID-19 Progression. *Cell Host and Microbe.* 2020; 27: 879–882.e2.
- [16] Kilic G, Bulut O, Jaeger M, Ter Horst R, Koeken V, Moorlag S, *et al.* The Immunological Factors Predisposing to Severe Covid-19 Are Already Present in Healthy Elderly and Men. *Frontiers in Immunology.* 2021; 12: 720090.
- [17] Zhao Y, Qin L, Zhang P, Li K, Liang L, Sun J, *et al.* Longitudinal COVID-19 profiling associates IL-1RA and IL-10 with disease severity and RANTES with mild disease. *JCI Insight.* 2020; 5: e139834.
- [18] Blackburn SD, Wherry EJ. IL-10, T Cell Exhaustion and Viral Persistence. *Trends in Microbiology.* 2007; 15: 143–146.
- [19] Lu L, Zhang H, Dauphars DJ, He Y. A Potential Role of Interleukin 10 in COVID-19 Pathogenesis. *Trends in Immunology.* 2021; 42: 3–5.
- [20] Komarowska I, Coe D, Wang G, Haas R, Mauro C, Kishore M, *et al.* Hepatocyte Growth Factor Receptor c-Met Instructs T Cell Cardiotropism and Promotes T Cell Migration to the Heart via Autocrine Chemokine Release. *Immunity.* 2015; 42: 1087–1099.
- [21] Yang Y, Shen C, Li J, Yuan J, Wei J, Huang F, *et al.* Plasma IP-10 and MCP-3 levels are highly associated with disease severity and predict the progression of COVID-19. *Journal of Allergy and Clinical Immunology.* 2020; 146: 119–127.e4.
- [22] Corley MJ, Ndhlovu LC. DNA Methylation Analysis of the COVID-19 Host Cell Receptor, Angiotensin I Converting Enzyme 2 Gene (ACE2) in the Respiratory System Reveal Age and Gender Differences. *Preprints.* 2020; 2020030295.
- [23] Du H, Dong X, Zhang J, Cao Y, Akdis M, Huang P, *et al.* Clinical characteristics of 182 pediatric COVID-19 patients with different severities and allergic status. *Allergy.* 2021; 76: 510–532.
- [24] Li R, Tian J, Yang F, Lv L, Yu J, Sun G, *et al.* Clinical characteristics of 225 patients with COVID-19 in a tertiary Hospital near Wuhan, China. *Journal of Clinical Virology.* 2020; 127: 104363.
- [25] Huang S, Wang J, Liu F, Liu J, Cao G, Yang C, *et al.* COVID-19 patients with hypertension have more severe disease: a multicenter retrospective observational study. *Hypertension Research.* 2020; 43: 824–831.
- [26] Guan WJ, Liang WH, Zhao Y, Liang HR, Chen ZS, Li YM, *et al.* Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis. *The European Respiratory Journal.* 2020; 55: 2000547.
- [27] Lighter J, Phillips M, Hochman S, Sterling S, Johnson D, Francois F, *et al.* Obesity in Patients Younger than 60 Years is a Risk Factor for COVID-19 Hospital Admission. *Clinical Infectious Diseases.* 2020; 71: 896–897.
- [28] Ng JH, Hirsch JS, Wanchoo R, Sachdeva M, Sakhiya V, Hong S, *et al.* Outcomes of patients with end-stage kidney disease hospitalized with COVID-19. *Kidney International.* 2020; 98: 1530–1539.
- [29] Tian J, Yuan X, Xiao J, Zhong Q, Yang C, Liu B, *et al.* Clinical characteristics and risk factors associated with COVID-19 disease severity in patients with cancer in Wuhan, China: a multicentre, retrospective, cohort study. *The Lancet Oncology.* 2020; 21: 893–903.
- [30] Hirsch JS, Ng JH, Ross DW, Sharma P, Shah HH, Barnett RL, *et al.* Acute kidney injury in patients hospitalized with COVID-19. *Kidney International.* 2020; 98: 209–218.
- [31] Ackermann M, Verleden SE, Kuehnel M, Haverich A, Welte T, Laenger F, *et al.* Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19. *New England Journal of Medicine.* 2020; 383: 120–128.
- [32] Nadkarni GN, Lala A, Bagiella E, Chang HL, Moreno PR, Pujadas E, *et al.* Anticoagulation, Bleeding, Mortality, and Pathology in Hospitalized Patients with COVID-19. *Journal of the American College of Cardiology.* 2020; 76: 1815–1826.
- [33] Bompard F, Monnier H, Saab I, Tordjman M, Abdoul H, Fournier L, *et al.* Pulmonary embolism in patients with COVID-19 pneumonia. *European Respiratory Journal.* 2020; 56: 2001365.
- [34] Kursa MB, Rudnicki WR. Feature selection with the Boruta package. *Journal of Statistical Software.* 2010; 36: 1–13.
- [35] Kursa MB. Robustness of Random Forest-based gene selection methods. *BMC Bioinformatics.* 2014; 15: 8.
- [36] Hanchuan Peng, Fuhui Long, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 2005; 27: 1226–1238.
- [37] Liu H, Setiono R. Incremental Feature Selection. *Applied Intelligence.* 1998; 9: 217–230.
- [38] Castro de Moura M, Davalos V, Planas-Serra L, Alvarez-Errico D, Arribas C, Ruiz M, *et al.* Epigenome-wide association study of COVID-19 severity with respiratory failure. *EBioMedicine.* 2021; 66: 103339.

- [39] Breiman L. Random forests. *Machine Learning*. 2001; 45: 5–32.
- [40] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2* (pp. 1137–1143). Morgan Kaufmann Publishers Inc.: Montreal, Quebec, Canada. 1995.
- [41] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*. 2002; 16: 321–357.
- [42] Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*. 1967; 13: 21–27.
- [43] Cortes C, Vapnik V. Support-vector networks. *Machine Learning*. 1995; 20: 273–297.
- [44] Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. *IEEE Transactions on Systems, Man, and Cybernetics*. 1991; 21: 660–674.
- [45] Chen L, Li Z, Zhang S, Zhang Y, Huang T, Cai Y. Predicting RNA 5-Methylcytosine Sites by Using Essential Sequence Features and Distributions. *BioMed Research International*. 2022; 2022: 4035462.
- [46] Zhang Y, Zeng T, Chen L, Huang T, Cai Y. Determining protein–protein functional associations by functional rules based on gene ontology and KEGG pathway. *Biochimica Et Biophysica Acta (BBA) - Proteins and Proteomics*. 2021; 1869: 140621.
- [47] Ding S, Wang D, Zhou X, Chen L, Feng K, Xu X, *et al.* Predicting Heart Cell Types by Using Transcriptome Profiles and a Machine Learning Method. *Life*. 2022; 12: 228.
- [48] Li Z, Wang D, Liao H, Zhang S, Guo W, Chen L, *et al.* Exploring the genomic patterns in human and mouse cerebellums via single-cell sequencing and machine learning method. *Frontiers in Genetics*. 2022; 13: 857851.
- [49] Yang Y, Chen L. Identification of drug-disease associations by using multiple drug and disease networks. *Current Bioinformatics*. 2022; 17: 48–59.
- [50] Li X, Lu L, Chen L. Identification of protein functions in mouse with a label space partition method. *Mathematical Biosciences and Engineering*. 2022; 19: 3820–3842.
- [51] Zhou J, Chen L, Wang T, Liu M. iATC-FRAKEL: a simple multi-label web server for recognizing anatomical therapeutic chemical classes of drugs with their fingerprints only. *Bioinformatics*. 2020; 36: 3568–3569.
- [52] Liu H, Hu B, Chen L, Lu L. Identifying Protein Subcellular Location with Embedding Features Learned from Networks. *Current Proteomics*. 2021; 18: 646–660.
- [53] Chen W, Chen L, Dai Q. iMPT-FDNPL: Identification of Membrane Protein Types with Functional Domains and a Natural Language Processing Approach. *Computational and Mathematical Methods in Medicine*. 2021; 2021: 7681497.
- [54] Chen L, Wang S, Zhang Y, Li J, Xing Z, Yang J, *et al.* Identify Key Sequence Features to Improve CRISPR sgRNA Efficacy. *IEEE Access*. 2017; 5: 26582–26590.
- [55] Zhao X, Chen L, Lu J. A similarity-based method for prediction of drug side effects with heterogeneous information. *Mathematical Biosciences*. 2018; 306: 136–144.
- [56] Matthews BW. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica Et Biophysica Acta (BBA) - Protein Structure*. 1975; 405: 442–451.
- [57] Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, *et al.* Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genetics*. 2012; 8: e1002629.
- [58] Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MP, van Eijk K, *et al.* Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biology*. 2012; 13: R97.
- [59] Zaghlool SB, Al-Shafai M, Al Muftah WA, Kumar P, Falchi M, Suhre K. Association of DNA methylation with age, gender, and smoking in an Arab population. *Clinical Epigenetics*. 2015; 7: 6.
- [60] Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, *et al.* Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Research*. 2010; 20: 440–446.
- [61] Hoeijmakers JHJ. DNA Damage, Aging, and Cancer. *New England Journal of Medicine*. 2009; 361: 1475–1485.
- [62] Issa JJ, Ottaviano YL, Celano P, Hamilton SR, Davidson NE, Baylin SB. Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nature Genetics*. 1994; 7: 536–540.
- [63] Issa JP, Vertino PM, Boehm CD, Newsham IF, Baylin SB. Switch from monoallelic to biallelic human IGF2 promoter methylation during aging and carcinogenesis. *Proceedings of the National Academy of Sciences*. 1996; 93: 11757–11762.
- [64] Ahuja N, Li Q, Mohan AL, Baylin SB, Issa JP. Aging and DNA methylation in colorectal mucosa and cancer. *Cancer research*. 1998; 58: 5489–5494.
- [65] Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, *et al.* Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genetics*. 2009; 5: e1000602.
- [66] Bjornsson HT, Sigurdsson MI, Fallin MD, Irizarry RA, Aspelund T, Cui H, *et al.* Intra-individual Change over Time in DNA Methylation with Familial Clustering. *The Journal of the American Medical Association*. 2008; 299: 2877–2883.
- [67] Rakyan VK, Down TA, Maslau S, Andrew T, Yang T, Beyan H, *et al.* Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Research*. 2010; 20: 434–439.
- [68] Chitralla KN, Hernandez DG, Nalls MA, Mode NA, Zonderman AB, Ezike N, *et al.* Race-specific alterations in DNA methylation among middle-aged African Americans and Whites with metabolic syndrome. *Epigenetics*. 2020; 15: 462–482.
- [69] Li C, Gao W, Gao Y, Yu C, Lv J, Lv R, *et al.* Age prediction of children and adolescents aged 6–17 years: an epigenome-wide analysis of DNA methylation. *Aging*. 2018; 10: 1015–1026.
- [70] Dominguez CX, Amezcua RA, Guan T, Marshall HD, Joshi NS, Kleinstein SH, *et al.* The transcription factors ZEB2 and T-bet cooperate to program cytotoxic T cell terminal differentiation in response to LCMV viral infection. *Journal of Experimental Medicine*. 2015; 212: 2041–2056.
- [71] van Helden MJ, Goossens S, Daussy C, Mathieu A, Faure F, Marçais A, *et al.* Terminal NK cell maturation is controlled by concerted actions of T-bet and Zeb2 and is essential for melanoma rejection. *Journal of Experimental Medicine*. 2015; 212: 2015–2025.
- [72] Scott CL, T’Jonck W, Martens L, Todorov H, Sichien D, Soen B, *et al.* The Transcription Factor ZEB2 is Required to Maintain the Tissue-Specific Identities of Macrophages. *Immunity*. 2018; 49: 312–325.e5.
- [73] Pérez RF, Santamarina P, Tejedor JR, Urdinguio RG, Álvarez-Pitti J, Redon P, *et al.* Longitudinal genome-wide DNA methylation analysis uncovers persistent early-life DNA methylation changes. *Journal of Translational Medicine*. 2019; 17: 15.
- [74] Wang Y, Karlsson R, Lampa E, Zhang Q, Hedman ÅK, Almgren M, *et al.* Epigenetic influences on aging: a longitudinal genome-wide methylation study in old Swedish twins. *Epigenetics*. 2018; 13: 975–987.
- [75] Verschuur CP, McEwen LM, Kohli V, Wolfson C, Bowdish DM, Raina P, *et al.* The relation between DNA methylation patterns and serum cytokine levels in community-dwelling adults: a preliminary study. *BMC Genetics*. 2017; 18: 57.
- [76] Florath I, Butterbach K, Muller H, Bewerunge-Hudler M, Brenner H. Cross-sectional and longitudinal changes in DNA methylation with age: an epigenome-wide analysis revealing over 60

- novel age-associated CpG sites. *Human Molecular Genetics*. 2014; 23: 1186–1201.
- [77] McEwen LM, Goodman SJ, Kobor MS, Jones MJ. The DNA Methylome: an Interface between the Environment, Immunity, and Ageing. *The Ageing Immune System and Health* (pp. 35–52). Springer: Cham. 2017.
- [78] Verschoor CP, Johnstone J, Millar J, Parsons R, Lelic A, Loeb M, *et al.* Alterations to the frequency and function of peripheral blood monocytes and associations with chronic disease in the advanced-age, frail elderly. *PLoS ONE*. 2014; 9: e104522.
- [79] Roubenoff R, Harris TB, Abad LW, Wilson PWF, Dallal GE, Dinarello CA. Monocyte Cytokine Production in an Elderly Population: Effect of Age and Inflammation. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*. 1998; 53: M20–M26.
- [80] Mohanty S, Joshi SR, Ueda I, Wilson J, Blevins TP, Siconolfi B, *et al.* Prolonged Proinflammatory Cytokine Production in Monocytes Modulated by Interleukin 10 after Influenza Vaccination in Older Adults. *Journal of Infectious Diseases*. 2015; 211: 1174–1184.
- [81] Franceschi C, Bonafè M. Centenarians as a model for healthy aging. *Biochemical Society Transactions*. 2003; 31: 457–461.
- [82] Ligthart S, Marzi C, Aslibekyan S, Mendelson MM, Conneely KN, Tanaka T, *et al.* DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. *Genome Biology*. 2016; 17: 255.
- [83] Hermsdorff HH, Mansego ML, Campion J, Milagro FI, Zulet MA, Martinez JA. TNF-alpha promoter methylation in peripheral white blood cells: relationship with circulating TNFalpha, truncal fat and n-6 PUFA intake in young women. *Cytokine*. 2013; 64: 265–271.
- [84] Gowers IR, Walters K, Kiss-Toth E, Read RC, Duff GW, Wilson AG. Age-related loss of CpG methylation in the tumour necrosis factor promoter. *Cytokine*. 2011; 56: 792–797.
- [85] Albouery M, Buteau B, Gregoire S, Cherbuy C, Pais de Barros JP, Martine L, *et al.* Age-Related Changes in the Gut Microbiota Modify Brain Lipid Composition. *Frontiers in Cellular and Infection Microbiology*. 2019; 9: 444.
- [86] Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, *et al.* Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell*. 2012; 11: 1132–1134.
- [87] Reynolds LM, Taylor JR, Ding J, Lohman K, Johnson C, Siscovick D, *et al.* Age-related variations in the methylome associated with gene expression in human monocytes and T cells. *Nature Communications*. 2014; 5: 5366.
- [88] Xiao R, Yang M, Tan Y, Ding R, Li D. Identification of Five Immune-Related lncRNAs Predicting Survival and Tumor Microenvironment Characteristics in Breast Cancer. *Computational and Mathematical Methods in Medicine*. 2021; 2021: 6676692.
- [89] Taniguchi I, Iwaya C, Ohnaka K, Shibata H, Yamamoto K. Genome-wide DNA methylation analysis reveals hypomethylation in the low-CpG promoter regions in lymphoblastoid cell lines. *Human Genomics*. 2017; 11: 8.
- [90] Chen Y, Klein SL, Garibaldi BT, Li H, Wu C, Osevala NM, *et al.* Aging in COVID-19: Vulnerability, immunity and intervention. *Ageing Research Reviews*. 2021; 65: 101205.
- [91] Reynolds CA, Tan Q, Munoz E, Jylhävä J, Hjelmborg J, Christiansen L, *et al.* A decade of epigenetic change in aging twins: Genetic and environmental contributions to longitudinal DNA methylation. *Aging Cell*. 2020; 19: e13197.
- [92] Sharp GC, Ho K, Davies A, Stergiakouli E, Humphries K, McArdle W, *et al.* Distinct DNA methylation profiles in subtypes of orofacial cleft. *Clinical Epigenetics*. 2017; 9: 63.
- [93] Zhang R, Zhang F, Sun Z, Liu P, Zhang X, Ye Y, *et al.* LINE-1 Retrotransposition Promotes the Development and Progression of Lung Squamous Cell Carcinoma by Disrupting the Tumor-Suppressor Gene FGGY. *Cancer Research*. 2019; 79: 4453–4465.
- [94] Johannessen M, Møller S, Hansen T, Moens U, Ghelue MV. The multifunctional roles of the four-and-a-half-LIM only protein FHL2. *Cellular and Molecular Life Sciences CMLS*. 2006; 63: 268–284.
- [95] Kleiber K, Strebhardt K, Martin BT. The biological relevance of FHL2 in tumour cells and its role as a putative cancer target. *Anticancer Research*. 2007; 27: 55–61.
- [96] Song M, Brasky TM, Weng DY, McElroy JP, Marian C, Higgins MJ, *et al.* Landscape of genome-wide age-related DNA methylation in breast tissue. *Oncotarget*. 2017; 8: 114648–114662.
- [97] Bacos K, Gillberg L, Volkov P, Olsson AH, Hansen T, Pedersen O, *et al.* Blood-based biomarkers of age-associated epigenetic changes in human islets associate with insulin secretion and diabetes. *Nature Communications*. 2016; 7: 11089.
- [98] Wixler V. The role of FHL2 in wound healing and inflammation. *The FASEB Journal*. 2019; 33: 7799–7809.
- [99] Tran MK, Kurakula K, Koenis DS, de Vries CJ. Protein-protein interactions of the LIM-only protein FHL2 and functional implication of the interactions relevant in cardiovascular disease. *Biochimica et Biophysica Acta*. 2016; 1863: 219–228.