Check for updates

Identification of methylation signatures associated with CAR T cell in B-cell acute lymphoblastic leukemia and non-hodgkin's lymphoma

Jiwei Song[1†], FeiMing Huang[2†], Lei Chen[3†], KaiYan Feng[4], Fangfang Jian[5], Tao Huang[6,7]* and Yu-Dong Cai[2]*

[1]College of Life Science, Changchun Sci-Tech University, Shuangyang, China, [2]School of Life Sciences, Shanghai University, Shanghai, China, [3]College of Information Engineering, Shanghai Maritime University, Shanghai, China, [4]Department of Computer Science, Guangdong AIB Polytechnic College, Guangzhou, China, [5]Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [6]Bio-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, Shanghai, China, [7]CAS Key Laboratory of Tissue Microenvironment and Tumor, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China

CD19-targeted CAR T cell immunotherapy has exceptional efficacy for the treatment of B-cell malignancies. B-cell acute lymphocytic leukemia and non-Hodgkin's lymphoma are two common B-cell malignancies with high recurrence rate and are refractory to cure. Although CAR T-cell immunotherapy overcomes the limitations of conventional treatments for such malignances, failure of treatment and tumor recurrence remain common. In this study, we searched for important methylation signatures to differentiate CAR-transduced and untransduced T cells from patients with acute lymphoblastic leukemia and non-Hodgkin's lymphoma. First, we used three feature ranking methods, namely, Monte Carlo feature selection, light gradient boosting machine, and least absolute shrinkage and selection operator, to rank all methylation features in order of their importance. Then, the incremental feature selection method was adopted to construct efficient classifiers and filter the optimal feature subsets. Some important methylated genes, namely, *SERPINB6*, *ANK1*, *PDCD5*, *DAPK2*, and *DNAJB6*, were identified. Furthermore, the classification rules for distinguishing different classes were established, which can precisely describe the role of methylation features in the classification. Overall, we applied advanced machine learning approaches to the high-throughput data, investigating the mechanism of CAR T cells to establish the theoretical foundation for modifying CAR T cells.

Introduction

The chimeric antigen receptor (CAR) T cell immunotherapy is a type of pericyte therapy in which T cells are genetically modified to express chimeric antigen receptors that detect and kill tumor cells in patients (1). In the USA, over 70,000 people are diagnosed with non-Hodgkin's lymphoma (NHL) annually, with a 5-year survival rate of roughly 70% (2). Acute lymphoblastic leukemia (ALL) is the most common pediatric cancer, accounting for approximately 25% of all pediatric cancer cases, and has a high recurrence rate (3). CD19-targeted CAR T-cell immunotherapy has a high response rate in B-cell ALL and B-cell NHL, especially in ALL, with a treatment effectiveness of 90% (4). However, CAR-T therapy is not effective for all tumor patients, and drug-resistant relapse occurs in approximately 50% of patients treated with CD19-targeted CAR-T (5). Some CAR T cells become exhausted, resulting in an increase in inhibitory receptors and a loss of effector function (6, 7). Creating a long-lasting therapeutic response is an essential problem that demands a better knowledge of the cellular and molecular processes that drive CAR T cell proliferation, contraction, and persistence in patients. Studying the specific functions of CAR-transduced T-cells at the molecular level, such as epigenetic level, can help in the understanding of the deeper mechanisms of CAR-T cell immunotherapy and clinical identification of potential targets for effective cancer treatment.

CAR consists of an antigen recognition domain, a co-stimulatory region, and a T cell activation region (8–10). Through multiple signaling cascades, the costimulatory region and T cell activation region activate the CAR T cells, which exhibit proliferative and cytotoxic properties (11). Activated CAR T cells have different gene expression patterns compared with regular T cells, which are influenced by epigenetic modifications (12). DNA may be modified in various ways, the most frequent of which is direct nucleotide methylation. Methylation of promoters results in a decrease in gene expression and suppression of transcription. High-expression genes have high levels of methylation at introns but low levels of methylation at the promoter or regulatory areas (13, 14). Epigenetic imprinting is emerging as a unifying subject in the study of immunological memory and the correlation of long-lasting antitumor responses.

Modifications in DNA methylation shape the overall immune response by altering the phenotype and function of CAR T cells. Zebley et al. showed that alterations in DNA methylation are linked to the proliferation and contraction of CAR T cells and that CD19-targeted CAR T cells acquire DNA methylation features over time. These results suggested that these cells are developing into a progenitor subset of exhausted T cells (15). Meanwhile, Wang et al. discovered that CAR T cells treated with low doses of the demethylating drug decitabine had stronger antitumor, proliferation, and cytokine release abilities.

This result indicates the presence of methylation in CAR T cells that inhibit their oncogenic functions (16). Among the large number of methylation sites, traditional biological experiments cannot meet the requirement of searching for methylation sites that affect the proliferation, failure, and oncogenic functions of CAR T cells. Therefore, this study was focused on how to combine advanced computational methods, such as machine learning, to mine CAR T cell-specific methylation sites to find potential sites for the sustained activation of CAR T cells.

Herein, we devised a process to rapidly screen CAR T cells for specific methylation sites. First, the methylation sites were analyzed and sorted by three feature ranking methods, namely, least absolute shrinkage and selection operator (LASSO) (17), light gradient boosting machine (LightGBM) (18), and Monte Carlo feature selection (MCFS) (19). Then, the incremental feature selection (IFS) (20) method was used to estimate the importance of feature subsets, which were constructed from three ranked methylation site lists, by evaluating the performance of classifiers on these subsets. One optimal feature subset was obtained from each list generated by one feature ranking method. The intersection of all obtained optimal feature subsets was investigated. The methylation sites that recurred multiple times were considered to be highly correlated with the specific functions of the CAR T cell, because the three feature ranking methods used different and independent concepts. Moreover, we also used decision trees (DTs) (21) to create quantitative classification rules that can accurately describe the composition of features for distinguishing each class. All in all, we identified the methylation sites associated with specific functions of the CAR T cells on a large scale using an efficient machine learning based framework and provided a functional description of highly ranked methylation sites in conjunction with the literature.

Materials and methods

Data and preprocessing

The T-cell methylation profiles of 157 patients with B-cell malignancies, including ALL and NHL, were downloaded from the GEO database under the accession number GSE179414 (22). The dataset comprised 77 ALL and 37 NHL cases, who were treated with CART19 cells. These two groups of patients were injected with CAR-transduced T cells and were referred to as ALL transduced and NHL transduced samples, respectively. Meanwhile, 13 ALL and 30 NHL cases were also included in the dataset, but they were not given CART19 cells. These patients were injected with CAR-untransduced T cells and were referred to as ALL untransduced and NHL untransduced

samples. Each group was deemed as a class in this study. We investigated their essential differences by studying the classification problem on these classes. Furthermore, each sample in the dataset was represented by 865,859 methylation sites. These sites were termed as features in this investigation.

Feature ranking methods

A large number of methylation sites were involved in the investigated methylation profiles, which were deemed as features in this study. Evidently, a small proportion of features were highly related to distinguish the CAR-untransduced and -transduced T cells. The powerful feature analysis method in machine learning was necessary. Here, three such methods were employed, including MCFS (19), LightGBM (18), and LASSO (17).

Monte Carlo feature selection

The MCFS algorithm is a DT-based method for determining the relevance of features. This method was first proposed by Micha et al. and has been widely used in tackling various complex medical and biological problems, showing promise in solving such problems (19, 23, 24).

The procedures of MCFS can be summarized as the following steps: (1) s feature subsets are randomly constructed from all features; (2) For each feature subset, t DTs are constructed by randomly selecting training and test samples from the original datasets; (3) After t×s DTs have been built, each feature g is evaluated by the relative importance (RI), which can be computed as follows:

$$RI_g = \sum_{\tau=1}^{st} (wAcc)^u \sum_{n_g(\tau)} IG(n_g(\tau)) \left(\frac{no\,.\,in\ n_g(\tau)}{no\,.\,in\ \tau} \right)^v, \quad (1)$$

where wAcc is the weighted accuracy; IG(ng(τ)) stands for the information gain (IG) of ng(τ) (a DT node with the feature g); no.in ng(τ) stands for the number of samples in ng(τ); no.in τ stands for the sample sizes in the tree root; and u and v are two settled positive integers. According to the RI values of all features, they are ranked in a feature list by the decreasing order of their RI values.

In this study, we adopted the MCFS program downloaded from http://www.ipipan.eu/staff/m.draminski/mcfs.html. Default parameters were used to execute such program, where u and v were set to one.

Light gradient boosting machine

LightGBM is an iterative boosting tree classifier proposed by Microsoft and is a modified version of the gradient boosting DT (18). LightGBM uses the total number of times (i.e., T_Split) that each feature is involved in the trees iteratively created and the gain (i.e., T_Gain) that a feature is utilized for splitting in all DTs

as measurements of feature relevance for the prediction. They are defined as

$$T\_Split = \sum_{t=1}^{K} Split_t, \quad (2)$$

$$T\_Gain = \sum_{t=1}^{K} Gain_t, \quad (3)$$

where K is the K DTs generated by K iterations. Here, we used T_Split as a metric to measure the importance of features, i.e., features were sorted in the decreasing order of their T_Split values.

This study adopted the LightGBM program retrieved from https://lightgbm.readthedocs.io/en/latest/. It was performed with its default parameters.

Least absolute shrinkage and selection operator

The LASSO algorithm is a feature selection method based on linear regression models that selects and compresses variables to prevent overfitting (17). This method uses the L1 paradigm to create a penalty function that selectively removes lower-correlation variables by imposing a bigger penalty on the larger value of the feature variables. This process results in a model with fewer feature variables and effectively avoiding overfitting. If the coefficients of the input features did not contribute positively to the training of the machine learning model, they were scaled down. As a result, the features could be ranked according to their coefficients.

Here, the LASSO package integrated in Scikit-learn (25) was used and its default parameters were adopted.

Incremental feature selection

By three feature ranking methods, all features were ranked in three lists. Evidently, top features in each list were important. However, determining the number of top features was still a problem. Thus, the IFS method (20, 26–28) was employed, which can determine the suitable number of top features. The procedures of IFS method can be divided as follows: (1) Several feature subsets are constructed based on the ranked feature list, which consists of some top features in the list; (2) A classifier is constructed on samples represented by features in each subset and its performance is evaluated by ten-fold cross-validation (29); (3) The classifier with the best performance can be found and the feature subset used in this classifier is picked up as the optimal feature subset. As three ranked feature list was produced in this study, IFS method was executed on each list. Three optimal feature subsets were obtained. We drew Venn diagrams for these three feature sets to display and analyze their intersection results.

Synthetic minority over sampling technique

As described in Section Data and preprocessing, the size of the class with the most samples (77) was about six times as large as that of the class with the least samples (13). Given the imbalance in sample size, when building and evaluating the classifiers, the predicted results would be biased toward the classes with a larger sample size, reducing the generalization ability for the model. In view of this, synthetic minority oversampling technique (SMOTE) algorithm was used in this study to effectively achieve data balance by enlarging the size of each minority class (30, 31). It generates new samples for each minority class by the linear combination of two samples in the same minority class, which are near enough. Finally, all classes have the same number of samples. This study adopted the SMOTE program obtained from https://github.com/scikitlearn-contrib/imbalanced-learn. Likewise, its default parameters were used.

Classification algorithm

In the IFS method, classifiers were built to evaluate the importance of constructed feature subsets. A certain classification algorithm was necessary to execute the IFS method. In this study, we applied four classification algorithms, namely, K-nearest neighbor (KNN) (32), support vector machine (SVM) (33), random forest (RF) (34), and DT (21). The purpose of employing these algorithms was to fully test the importance of each constructed feature subset and select the best one.

The KNN algorithm is one of the most classic classification algorithms. Its principle is quite simple. However, its performance is still acceptable in some cases. Given a test sample, KNN finds its k nearest neighbors in the training dataset. According to the labels of these neighbors, the label of the test sample can be determined.

The SVM is a classification algorithm based on statistical learning theory. It generally maps samples into a high-dimensional space by using a kernel function and linearly separates them by finding the maximum margin separating hyperplane. For a test sample, it is also mapped into the same high-dimensional space and its class is determined by the side of hyperplane that the test sample is located.

The RF is also a classic classification algorithm, which is quite different from SVM. In fact, it is an ensemble algorithm, which contains several DTs. Each DT is built by randomly selecting features and samples. For a test sample, each DT gives its prediction. RF integrates these predictions using majority voting.

For the above three classification algorithms, their classification principles are quite difficult for us to understand. Thus, few insights can be extracted from them. DT has its merits in this regard. It is a white-box algorithm, whose classification procedures are completely open, giving opportunities for us to uncover its principle learned from the given dataset, thereby access more knowledge from the dataset. A DT is a tree structure consisting of a series of nodes and branches that use logical operations. Two types of nodes are contained in a DT, they are branch and leaf nodes. The branch node is always related to one feature. According the threshold, samples in a branch node are classified into two groups. The leaf node stands for one class. Samples that reach such node are assigned the corresponding class label. During predictions, it starts at the root node and sorts the test samples down the tree according to the thresholds defined at each branch node. Furthermore, a DT can be represented by lots of if-then rules. Each rule is constructed by a path from the root node to one leaf node. From these rules, a clearer picture on each class can be uncovered.

Above algorithms have been applied to construct various classifiers in dealing with biological and medical problems (35–40). In this study, we used the corresponding Python Scikit-learn packages (25) of above four classification algorithms to implement them.

Performance evaluation

To evaluate the performance of all classifiers constructed in the IFS method, several measurements were employed. First, as multi-class classifiers, the overall accuracy (ACC) was adopted, which is the most accepted measurements. It is defined as the proportion of correctly predicted samples among all samples. However, such measurement is not perfect if the sizes of classes are quite different. Thus, we also employed the Matthew correlation coefficients (MCC) (41), which is deemed as a balanced measurement. As four classes were involved, the MCC in multi-class was adopted, which is defined as

$$MCC = \frac{cov(X, Y)}{\sqrt{cov(X, X)cov(Y, Y)}}, \qquad (4)$$

where X and Y are two binary matrices, indicating the true and predicted class of each sample.

In addition, we computed the precision, recall and F1-score for each class, which is defined as

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

$$Recall = \frac{TP}{TP + FN}, \qquad (6)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \qquad (7)$$

where *TP* stands for the number of samples in such class which are correctly predicted, *FP* is the number of samples in other classes which are classified into this class, *FN* is the number of samples in such class which are wrongly predicted. According to F1-score on each class, the macro F1 and weighted F1 are further computed to give a whole evaluation on classifiers. For macro F1, it is defined as the mean of all F1-score values on all classes, whereas weighted F1 integrates all F1-score values by further considering the class sizes, that is, it is the weighted mean of F1-score values.

In this study, weighted F1 was picked up as the key measurement. Other measurements were provided as reference.

Functional enrichment analysis

By analyzing the T-cell methylation profiles downloaded from the GEO with several machine learning algorithms, the optimal feature subsets, containing several methylation probes, were obtained. After taking the union operation on these subsets and mapping features in the union set onto the genes, ClusterProfiler in R was used to calculate the enrichment of these genes on GO terms and KEGG pathways (42). The p-value was corrected with FDR, and 0.05 was chosen as the cutoff value. Only the GO terms and KEGG pathways with FDR<0.05 were considered statistically significant.

Results

We built a machine learning based framework for analyzing CAR-transduced and untransduced T cells in different B-cell malignancies and further constructed efficient classifiers to discriminate CAR-transduced and untransduced T cells. The entire procedures are illustrated in Figure 1. The detailed results were listed in this section.

Results of feature selecting methods

Each sample was represented by a large number of features (methylation sites). They were deeply analyzed by three feature ranking methods (MCFS, LightGBM, and LASSO). Each method produced one feature list, which is provided in Table S1. It was necessary to pointed out that only features with evaluation score (RI for MCFS, T_Split for LightGBM and coefficient for LASSO) larger than zero were provided in Table S1. The top-ranked features are considered to be important because of their participation in the classification. Their biological significance and the reasons why they are important as core classification features would be discussed in Section Discussion.

Results of IFS method

The three ordered feature lists created by three feature ranking methods were fed into the IFS method one by one and four classification algorithms (DT, KNN, RF and SVM) were used in the IFS method. To save time, we only considered the top 1000 features in each list. For each list, 1000 possible feature subsets were constructed, on which 1000 classifiers with one give classification algorithm were built and evaluated by ten-fold cross-validation. The evaluated results, including measurements listed in Section Performance evaluation, are available in Table S2.

For the feature list yielded by MCFS method, we plotted an IFS curve for each classification algorithm to illustrate
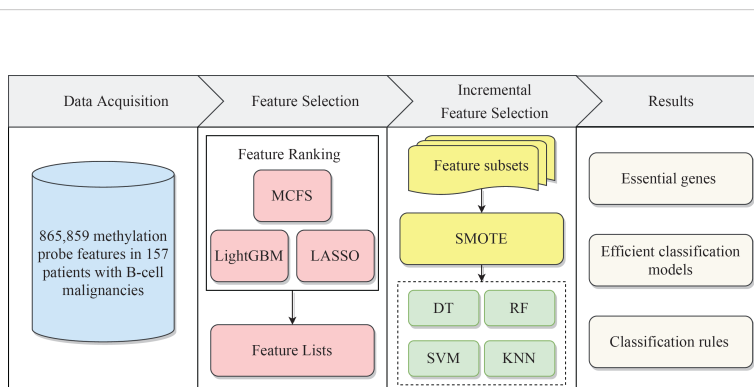
FIGURE 1
Flow chart of the entire analytical process. The 865,859 methylation probe signals on patients with B-cell malignancies were ranked according to feature importance by using three feature ranking algorithms, namely, MCFS, LightGBM, and LASSO. Then, three ordered feature lists were fed into the incremental feature selection (IFS) method, which incorporates four classification algorithms. Finally, based on the IFS results, the essential genes, efficient classification models and classification rules were extracted.

FIGURE 2
IFS curves for displaying the performance of four classification algorithms on the feature list yielded by MCFS method. The best classifiers on different algorithms yield the weight F1 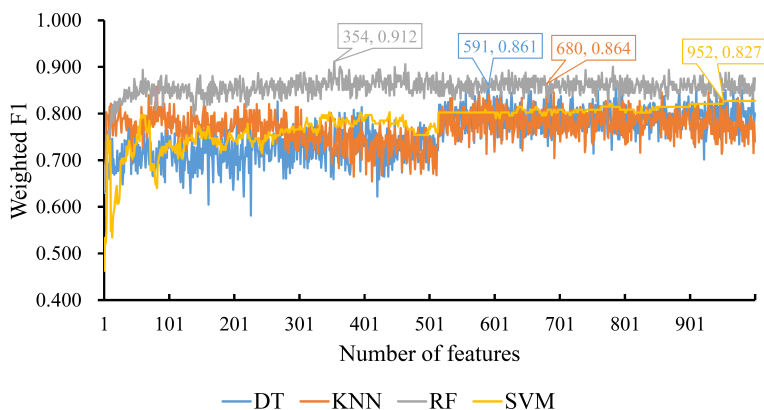values of 0.861, 0.864, 0.912 and 0.827, respectively, which use top 591, 680, 354 and 952, respectively, features in the list.

TABLE 1 Performance of the best classifiers using different classification algorithms and feature ranking methods.
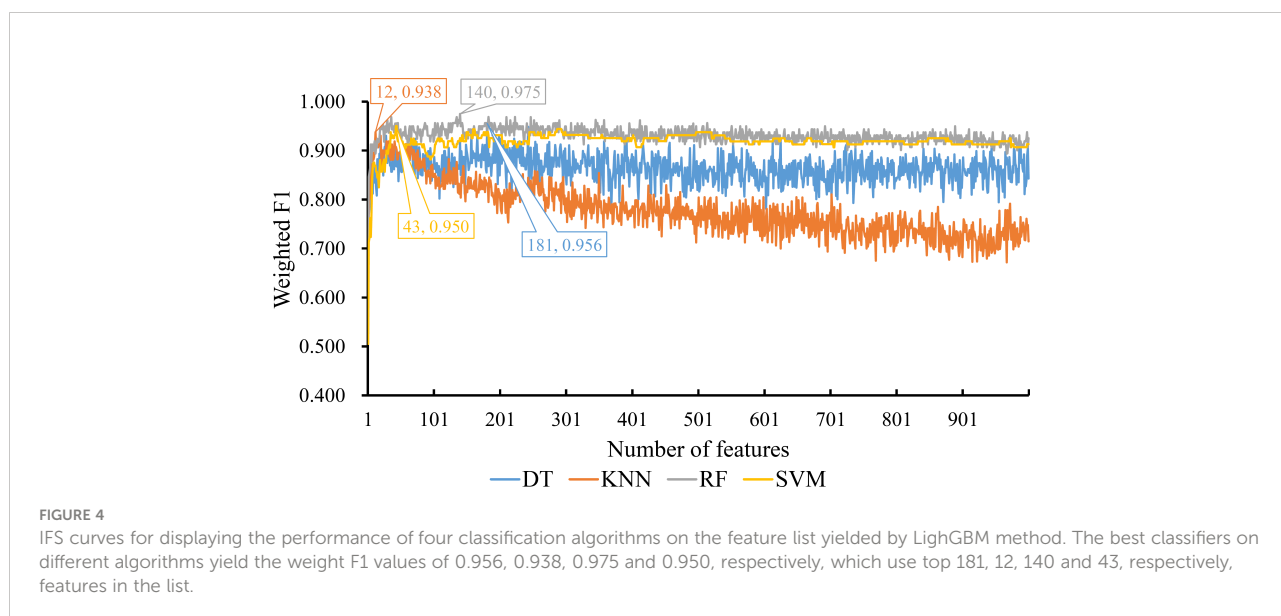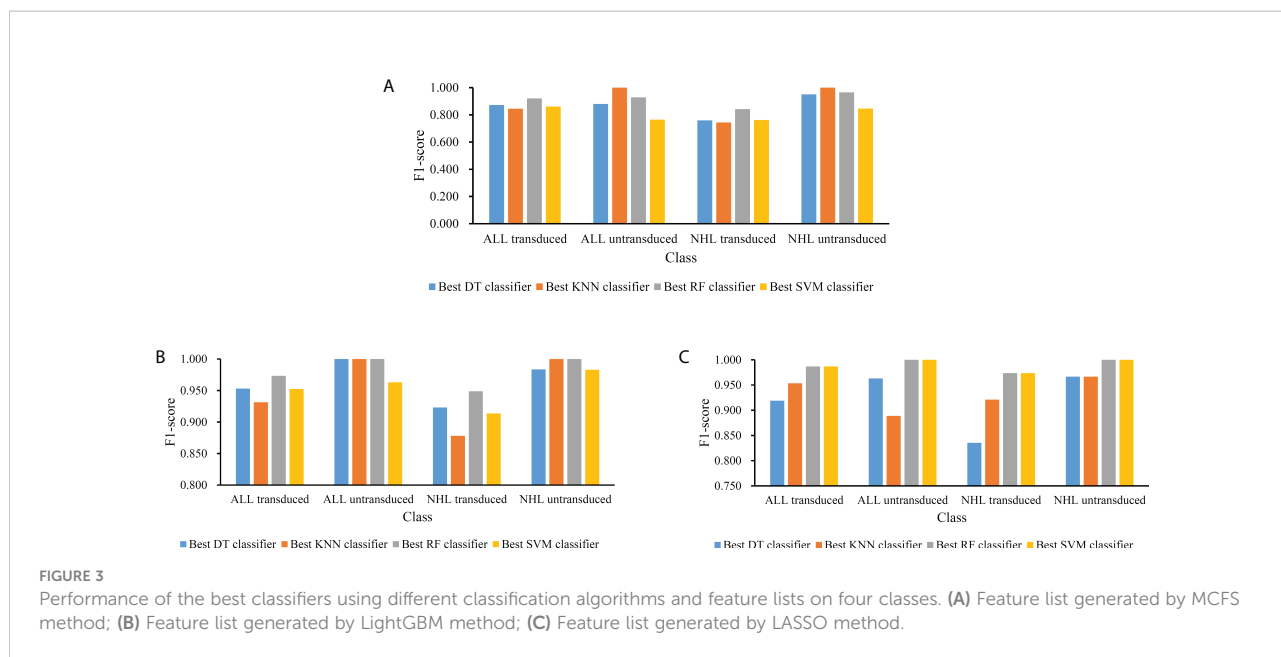
Feature ranking method	Classification algorithm	Weighted F1	Macro F1	ACC	MCC
MCFS	DT	0.861	0.866	0.860	0.792
	KNN	0.864	0.897	0.860	0.801
	RF	0.912	0.914	0.911	0.866
	SVM	0.827	0.808	0.822	0.746
LightGBM	DT	0.956	0.965	0.955	0.935
	KNN	0.938	0.952	0.936	0.909
	RF	0.975	0.981	0.975	0.963
	SVM	0.950	0.953	0.949	0.927
LASSO	DT	0.912	0.921	0.911	0.869
	KNN	0.943	0.933	0.943	0.914
	RF	0.987	0.990	0.987	0.981
	SVM	0.987	0.990	0.987	0.981

its performance on different feature subsets, which is shown in Figure 2. It can be observed that DT, KNN, RF and SVM can yielded the highest weighted F1 values of 0.861, 0.864, 0.912 and 0.827, respectively. These values were obtained by using top 591, 680, 354 and 952, respectively, features in the list, which comprised the optimal feature subsets for these four classification algorithms. With the optimal feature subsets, we can build the best DT, KNN, RF and SVM classifiers. The values of macro F1, ACC and MCC of these classifiers are listed in Table 1. Furthermore, their performance on four classes is illustrated in Figure 3. Evidently, the best RF classifier was superior to other three best classifiers and the best DT classifier was only better than the best SVM classifier.

For the feature list generated by LightGBM method, four IFS curves were also drawn, which are shown in Figure 4. When top 181, 12, 140 and 43 features in the list were adopted, four classification algorithms produced the highest weighted F1 values of 0.956, 0.938, 0.975 and 0.950, respectively. These features constituted the optimal feature subset for each classification algorithm. Furthermore, the best DT/KNN/RF/SVM classifier was built with its corresponding optimal feature subset. The detailed performance of these best classifiers is listed in Table 1 and shown in Figure 3. Likewise, the best RF classifier still provided the highest performance. As for the best DT classifier, it was a little better than the best KNN and SVM classifiers.

For the last feature list generated by LASSO method, IFS curves were also plotted, as shown in Figure 5. The highest weighted F1 for DT, KNN, RF and SVM were 0.912, 0.943, 0.987 and 0.987, respectively. To reach such performance, top 9, 12, 28 and 111, respectively, features were used. These features comprised the optimal feature subset for each classification algorithm. Similarly, the best DT, KNN, RF and SVM classifiers were constructed with the corresponding optimal feature subsets. Table 1 and Figure 3 provide their detailed performance. The best RF classifier provided equal performance to the best SVM classifier. However, the best RF classifier adopted much less features than the best SVM classifier. Thus, this classifier was still deemed to be better than other three classifiers. On the other hand, the best DT classifier gave the lowest performance among all four best classifiers.

Based on the above arguments, the best RF classifier always provided better performance than other three best classifiers on each feature list. Among three RF classifiers built on three feature lists, the RF classifier on the list generated by LASSO provided the highest performance. Such classifier can be a useful tool to

FIGURE 3
Performance of the best classifiers using different classification algorithms and feature lists on four classes. **(A)** Feature list generated by MCFS method; **(B)** Feature list generated by LightGBM method; **(C)** Feature list generated by LASSO method.

FIGURE 4
IFS curves for displaying the performance of four classification algorithms on the feature list yielded by LighGBM method. The best classifiers on different algorithms yield the weight F1 values of 0.956, 0.938, 0.975 and 0.950, respectively, which use top 181, 12, 140 and 43, respectively, features in the list.

discriminate CAR-transduced and untransduced T cells. On the other hand, the DT classifiers generally gave the low performance. However, they can provide more clues to uncover the differences between CAR-transduced and untransduced T cells.

Feature intersection

As mentioned above, on each feature list yielded by one feature ranking method, the best RF classifier was always better than other three best classifiers. Thus, its optimal feature subset was picked up as the optimal feature subset for one feature ranking method. In

detail, the optimal feature subsets for MCFS, LightGBM and Lasso consisted of the top 354, 140, and 28 features in the lists generated by MCFS, LightGBM, and LASSO, respectively.

For each above-mentioned optimal feature subset, features in such subset were mapped onto their related genes, which comprised the optimal gene subset. Concretely, the optimal gene set for MCFS, LightGBM and LASSO contained 231, 97 and 16 genes, respectively. Detailed genes in these sets are provided in Table S3. The intersection of these three gene sets was investigated and plotted in one Venn diagram, as shown in Figure 6. It can be observed that no genes were contained in all three optimal gene sets, three genes (*SERPINB6*, *ANK1*, *OST4*)
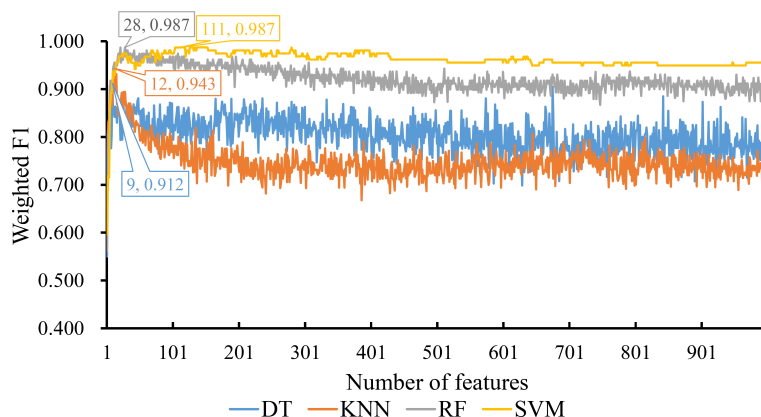
FIGURE 5
IFS curves for displaying the performance of four classification algorithms on the feature list yielded by LASSO method. The best classifiers on different algorithms yield the weight F1 values of 0.912, 0.943, 0.987 and 0.987, respectively, which use top 9, 12, 28 and 111, respectively, features in the list.
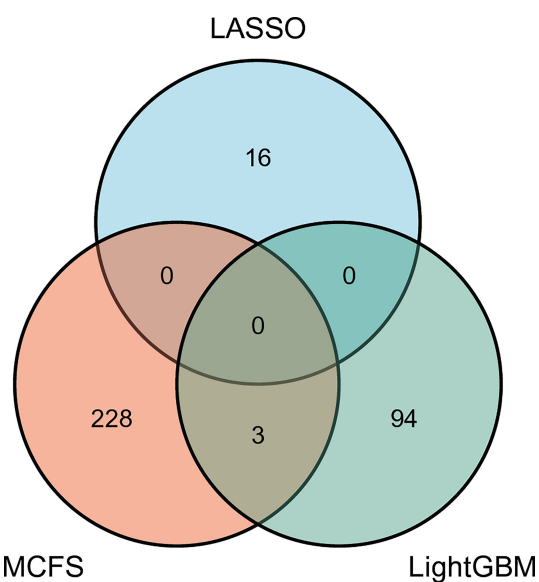

FIGURE 6
Venn diagram to show the intersection of the optimal gene sets for MCFS, LightGBM, and LASSO. Three genes are contained in two optimal gene sets, indicating their importance.

were in two optimal gene sets. Overlapped genes would be discussed in Section Discussion.

Classification rules

Although the DT classifiers were generally weaker than RF classifiers, they can provide clues hidden in the investigated
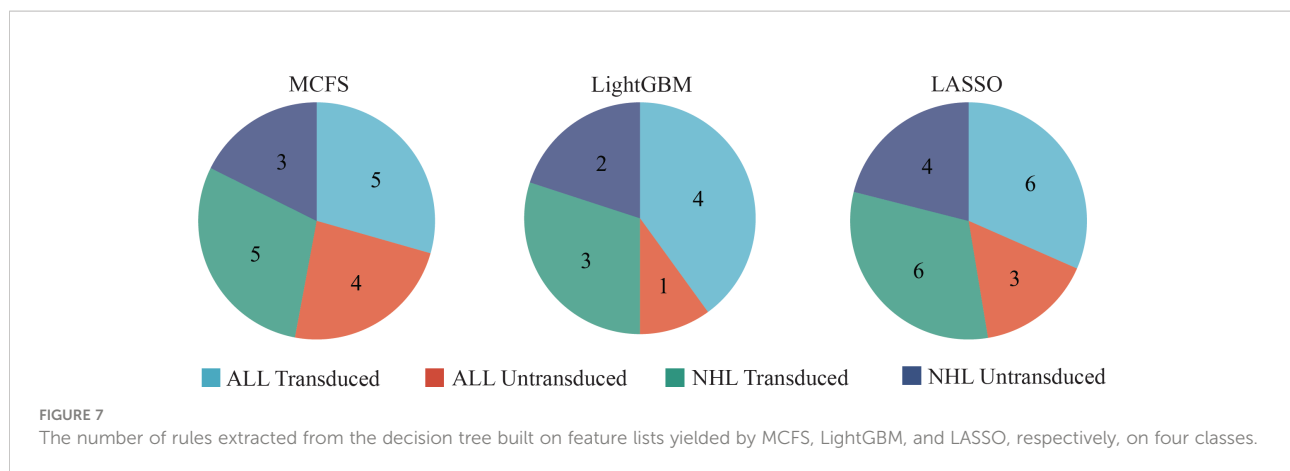
methylation profiles, which cannot be extracted by other classifiers. According to the IFS results on three feature lists, the best DT classifiers used top 591, 181 and 9 features in three lists, respectively. With these features, three DTs were learned on all samples. Each DT induced a rule set, which contained 17, 10 and 19 rules, respectively. Detailed rules are listed in Table S4. In each rule set, each class was assigned at least one rule, as shown in Figure 7. Following the rules in each rule set, we can determine the class of a test sample. Furthermore, their most contributions were the clear descriptions on different methylation patterns on CAR-transduced and untransduced T cells. This would be discussed in Section Discussion.

Results of enrichment analysis

The optimal feature subsets for three feature ranking methods were determined by the IFS method. We mapped the methylation probes in three optimal feature subsets to genes, yielding a total of 341 genes. Then, the functional enrichment analysis was performed on these genes. The enrichment results are provided in Table S5. Two GO terms were enriched by 341 genes, whereas no KEGG pathways were enriched by these genes with FDR<0.05. GO enrichment result indicated that 12 of these genes were involved in the splicing process, suggesting that the transcripts of these genes may be involved in regulating CAR T-cell processes.

Discussion

In this study, we applied several advanced machine learning algorithms to deeply mine the T-cell methylation profiles of

FIGURE 7

The number of rules extracted from the decision tree built on feature lists yielded by MCFS, LightGBM, and LASSO, respectively, on four classes.

patients with B-cell malignancies. Latent important genes were obtained and interesting classification rules were constructed. This section gave extensive analysis on these genes and rules.

Top ranked genes in multiple algorithms

The first gene was *SERPINB6* (targeted by probes cg27001747 and cg04181408), which encoded a member of the serpin superfamily and ovalbumin-serpin subfamily (43, 44). *SERPINB6* appeared in both LightGBM and MCFS in the subset of optimal features. Both methylation probes were linked to the promoter and found in the 5′-UTR region of *SERPINB6*, suggesting that they may affect the transcriptional regulation of this gene. Serpinb9, a homolog of SERPINB6, has been shown to protect T cells from Granzyme-B leaked from granules and also participates in T cell homeostasis (45, 46). Although the function of SERPINB6 in T cells has yet to be established, this protein is important to other immune cells. In neutrophils and monocytes, SERPINB6 inhibits Cathepsin G, thereby preventing programmed necrosis (47). Thus, SERPINB6 may play a role in the normal functioning of CAR-T cells. However, more research is needed to confirm this concept. Furthermore, SERPINB6 methylation has been linked to the risk of CLL pathogenicity (48). This result demonstrates the precision by which our method can identify CAR-T cell-specific genes and differential genes in B-cell malignancies.

The next probes identified were cg09405790 and cg02172579, which both targeted the gene body of *ANK1*. *ANK1* was found in the subset of optimum features in LightGBM and MCFS. ANK1 is a modular adaptor protein that mediates the connection of integral membrane proteins to the spectrin cytoskeleton (49). *ANK1* methylation has been shown to regulate the expression of microRNA-486-5p, which inhibits Interleukin-22 production by helper T cells *via* the Dock1/NF-B/Snail signaling pathway. Such process results in cancer suppression (50, 51).

The cg18756060 and cg04001935 probes were designed to detect the DNA methylation status in a specific intergenic region on chromosome 2 (chr2:27294139-27294915) according to GRCh37. Such region has been shown to be the coding region of gene *OST4*, and the protein encoded by *OST4* is an important subunit of oligosaccharyltransferase (OST). Similar to *SERPINB6* and *ANK1*, OST4 is an intersection feature of the optimal feature subsets of LightGBM and MCFS. Eukaryotic OSTs catalyze the N-glycosylation of nascent polypeptides in the lumen of the endoplasmic reticulum, a conserved biosynthetic process that diversifies the structure and function of proteins (52). Kumar et al. found that N-glycosylation activity remained elevated during the activation and expansion of human T cells, and lymphocytes in a resting state had lower N-glycosylation activity (53). These results suggest that OST was involved in T cell activation in transduced CARs, and that OST activity was influenced by methylation of *OST4*.

PDCD5 (also known as *TFAR19*), which is targeted by the optimal features cg13563193, has been generally reported to participate in immunoregulation. *PDCD5* is at the top of the list of feature rankings obtained with the LASSO method. PDCD5 interacted with FOXP3 to promote FOXP3 acetylation, hence reducing effector cytokine production (54). Meanwhile, the methylation signal of PDCD5 was primarily found in the promoter region, which negatively regulated the PDCD5 expression and thus relaxed the immunosuppressive effect of Treg. This activity could explain the mechanism by which the CAR-T cells were activated and therefore appeared in our list. In addition, in hepatocellular carcinoma, the PDCD5 overexpression stimulates the promoter activity of KLF9, and the upregulation of KLF9 inhibits cell migration and proliferation (55). This phenomenon also suggests that the cg13563193 methylation signature may suppress the expression level of PDCD5. Yuan et al. have discovered that PDCD5 inhibits the production of proinflammatory mediators and promotes the secretion of anti-inflammatory cytokines by modifying the T-lymphocyte homeostasis (56). The two

hallmark clinical toxicities associated with CAR-T cell therapy are cytokine release syndrome (CRS) and neurotoxicity (57, 58). The characteristics of CRS produce massive inflammation, suggesting a possible involvement of PDCD5 in this process.

The probe cg07632860 was developed to detect the methylation status of the transcription start site of the *DAPK2*. *DAPK2* is at the top of the list of feature rankings obtained with LASSO. *DAPK2* encodes a member of the serine/threonine protein kinase family, which functions as a tumor suppressor and regulates autophagic and apoptotic processes in various cell types (59, 60). When T lymphocytes are activated, they secrete inflammatory cytokines, such as TNF- and IL-6. During this process, DAPK2 is activated by T cell receptor, which inhibits T-cell activation (61, 62). We discovered that cg07632860 targeted the regulatory region of DAPK2, implying that it may limit the expression level of the protein. Meanwhile, DAPK2 expression has been found to be downregulated in ALL and NHL (63). Low levels of DAPK lead to T-cell activation, which implies the CAR-T cell activation mode. Furthermore, the inflammatory cytokines IL-17 and IL-32 have been demonstrated to use DAPK2 as a signaling mediator (64). Whether the production of cytokine storm, one of the side effects of CAR-T immunotherapy, is linked to DAPK2 is worthy of investigation.

The next predicted gene, *DNAJB6*, targeted by cg18753341, encodes a member of the DNAJ protein family, which is one of two key groups of molecular chaperones involved in biological activities, such as protein folding and oligomeric protein complex assembly. Strict control of the cell cycle process is essential for the proper functioning of T lymphocytes. Slfn1 has been shown to play an important role in the establishment and maintenance of T lymphocyte quiescence (65). Overexpression of DnaJB6 increases Slfn1 nuclear accumulation and causes cell-cycle arrest, whereas Slfn1 is mostly sequestered in the cytoplasm, and no cell-cycle arrest has been detected in *DnaJB6* knock-down cells (66). Furthermore, transgenic expression of DNAJB6 in T cells blocks Slfn1 degradation, enhances its nuclear import, and results in T cell proliferation suppression when T cell receptors are activated (66). In addition, DNAJB6 is neurotoxic when overexpressed in primary neurons, suggesting that it may be a potential locus for CAR-T treatment to eliminate side effects (67).

Analysis of classification rules

In addition to the functional analysis of the top-ranked features, we also mined the specific rules used to distinguish each class based on the classification tree structure of the DTs. The rules of each class consisted of methylation probes and their signal intensities, and each methylation probe was linked to a gene to describe its function in greater depth.

The first rule was aimed to distinguish T cells derived from patients with ALL that have been transduced with CAR. *MYCN*,

which is targeted by cg13799853, was an important site with low methylation, according to the classification rule based on LASSO results. In our classification rules, *MYCN* exhibited lower methylation levels. *MYCN* has been demonstrated to have lower methylation levels in relapsed children with B-cell acute lymphoblastic leukemia (B-ALL), which was consistent with the usage of *MYCN* in this study as a key feature to differentiate B-ALL (68). MYCN also downregulates DKK3 expression and activates the Wnt/β-catenin signaling pathway at the transcriptional level, boosting the development of B-ALL (69). Meanwhile, MYCN apparently decreases the interferon signaling, promoting a non-inflamed and T-cell infiltration-poor ("cool") tumor microenvironment (70). In the classification rule based on the MCFS results, *HDGF* targeted by cg18593717 was an important locus, which exhibited a lower methylation level. HDGF has been demonstrated to cause Foxp3 + Treg differentiation and that Tregs decrease CD8+ cytotoxic T cell activity (71). This phenomenon suggests that *HDGF* may act as a potential gene driving the activation of CAR-T cell.

The second rule was used to distinguish the T cells derived from patients with ALL without transduced CARs. After constructing the DT by using the optimal subset obtained after MCFS, the classification rules were established. Among them, hypomethylation of the *ZBTB7A*, also known as LRF, was an important quantitative rule. Many studies have shown that ZBTB7A is closely associated with B and T cell differentiation and plays an important role in their fate decisions (72, 73). Meanwhile, dysregulation in B-cell maturation can lead to the development of autoimmune syndromes and B-cell malignancies (73). *ZBTB7* was described in the rules in our study, because it plays an important role in both immune processes and cancer development.

The next two rules were used to distinguish between CAR-transduced and untransduced T cells derived from patients with NHL. *TP73* targeted by cg10654015 appeared in our rules and exhibited a higher methylation status. *TP73* has been demonstrated to be frequently methylated in NHLs (74). This result is consistent with the highly methylated results of *TP73* found in our study, indicating the accuracy of our method. Furthermore, *TP73* deletion has been shown to impact lymphoma formation by several mechanisms, such as altered gene expression patterns, defective early T-cell growth, impaired apoptosis, and chromosomal abnormality accumulation (75). This phenomenon suggests that *TP73* may be a potential target for the modification of CAR-T cells.

Conclusion

We applied a powerful computational strategy based on DNA methylation probe data to uncover the features of CAR T cells across diverse B-cell malignancies. The outcomes can be summarized in the three key components. First, a series of

methylation signatures and genes were extracted, which can be used to distinguish cells from four different origins. The findings provided a theoretical foundation to precisely modify CAR T cells and treat B-cell malignancies. Second, efficient multi-class classifiers were built to aid in a more accurate delineation of T cells prior to treatment. The delineation of T cells facilitated the screening for T cells that could efficiently suppress cancer *in vivo* and further improve those that were not successfully transduced. Finally, some classification rules were built to specifically distinguish a particular class of cells. These rules aided to better understand the specific functions of CAR T cells by describing the degree of gene methylation.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE179414.

Author contributions

TH and YD-C designed the study. LC and KYF performed the experiments. JS, FMH and FJ analyzed the results. JS, FMH and LC wrote the manuscript. All authors contributed to the research and reviewed the manuscript.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.976262/full#supplementary-material

SUPPLEMENTARY TABLE 1
Feature lists obtained by MCFS, LightGBM and LASSO.

SUPPLEMENTARY TABLE 2
Performance of IFS with different classification algorithms and feature lists.

SUPPLEMENTARY TABLE 3
Optimal gene sets based on MCFS, LightGBM, and LASSO methods, and their intersection results.

SUPPLEMENTARY TABLE 4
Classification rules generated by decision tree.

SUPPLEMENTARY TABLE 5
GO and KEGG enrichment results on the optimal gene sets of three feature ranking methods.

References

1. Singh AK, Mcguirk JP. CAR T cells: continuation in a revolution of immunotherapy. *Lancet Oncol* (2020) 21:e168–78. doi: 10.1016/S1470-2045(19)30823-X

2. Bethesda M. *SEER cancer stat facts: Non-Hodgkin lymphoma* . Available at: https://seer.cancer.gov/statfacts/html/nhl.html (Accessed Mar 2022).

3. Ward E, Desantis C, Robbins A, Kohler B, Jemal A. Childhood and adolescent cancer statistic. *CA Cancer J Clin* (2014) 64:83–103. doi: 10.3322/caac.21219

4. Maude SL, Frey N, Shaw PA, Aplenc R, Barrett DM, Bunin NJ, et al. Chimeric antigen receptor T cells for sustained remissions in leukemia. *N Engl J Med* (2014) 371:1507–17. doi: 10.1056/NEJMoa1407222

5. Shah NN, Fry TJ. Mechanisms of resistance to CAR T cell therapy. *Nat Rev Clin Oncol* (2019) 16:372–85. doi: 10.1038/s41571-019-0184-6

6. Wherry EJ, Kurachi M. Molecular and cellular insights into T cell exhaustion. *Nat Rev Immunol* (2015) 15:486–99. doi: 10.1038/nri3862

7. Schietinger A, Philip M, Krisnawan VE, Chiu EY, Delrow JJ, Basom RS, et al. Tumor-specific T cell dysfunction is a dynamic antigen-driven differentiation program initiated early during tumorigenesis. *Immunity* (2016) 45:389–401. doi: 10.1016/j.immuni.2016.07.011

8. Kalos M, Levine BL, Porter DL, Katz S, Grupp SA, Bagg A, et al. T Cells with chimeric antigen receptors have potent antitumor effects and can establish memory in patients with advanced leukemia. *Sci Transl Med* (2011) 3:95ra73. doi: 10.1126/scitranslmed.3002842

9. Brentjens RJ, Davila ML, Riviere I, Park J, Wang X, Cowell LG, et al. CD19-targeted T cells rapidly induce molecular remissions in adults with chemotherapy-refractory acute lymphoblastic leukemia. *Sci Transl Med* (2013) 5:177ra138. doi: 10.1126/scitranslmed.3005930

10. Brudno JN, Lam N, Vanasse D, Shen YW, Rose JJ, Rossi J, et al. Safety and feasibility of anti-CD19 CAR T cells with fully human binding domains in patients with b-cell lymphoma. *Nat Med* (2020) 26:270–80. doi: 10.1038/s41591-019-0737-3

11. Larson RC, Maus MV. Recent advances and discoveries in the mechanisms and functions of CAR T cells. *Nat Rev Cancer* (2021) 21:145–61. doi: 10.1038/s41568-020-00323-z

12. Akbari B, Ghahri-Saremi N, Soltantoyeh T, Hadjati J, Ghassemi S, Mirzaei HR. Epigenetic strategies to boost CAR T cell therapy. *Mol Ther* (2021) 29:2640–59. doi: 10.1016/j.ymthe.2021.08.003

13. Ball MP, Li JB, Gao Y, Lee JH, Leproust EM, Park IH, et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* (2009) 27:361–8. doi: 10.1038/nbt.1533

14. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* (2012) 13:484–92. doi: 10.1038/nrg3230

15. Zebley CC, Brown C, Mi T, Fan Y, Alli S, Boi S, et al. CD19-CAR T cells undergo exhaustion DNA methylation programming in patients with acute lymphoblastic leukemia. *Cell Rep* (2021) 37:110079. doi: 10.1016/j.celrep.2021.110079

16. Wang Y, Tong C, Dai H, Wu Z, Han X, Guo Y, et al. Low-dose decitabine priming endows CAR T cells with enhanced and persistent antitumour potential *via* epigenetic reprogramming. *Nat Commun* (2021) 12:409. doi: 10.1038/s41467-020-20696-x

17. Tibshirani RJ. Regression shrinkage and selection *via* the LASSO. journal of the royal statistical society. *Ser B: Methodological* (1996) 73:273–82.

18. Ke G, Meng Q, Finely T, Wang T, Chen W, Ma W, et al. LightGBM: A highly efficient gradient boosting decision tree; *Adv Neural Inf Process Syst* 30 (NIP2017) (2017).

19. Micha D, Rada-Iglesias A, Enroth S, Wadelius C, Koronacki J, Komorowski J. Monte Carlo Feature selection for supervised classification. *Bioinformatics* (2008) 24:110–7. doi: 10.1093/bioinformatics/btm486

20. Liu HA, Setiono R. Incremental feature selection. *Appl Intell* (1998) 9:217–30. doi: 10.1023/A:1008363719778

21. Safavian SR, Landgrebe D. A survey of decision tree classifier methodology. *IEEE Trans systems man cybernetics* (1991) 21:660–74. doi: 10.1109/21.97458

22. Garcia-Prieto CA, Villanueva L, Bueno-Costa A, Davalos V, González-Navarro EA, Urbano-Ispizua JM, et al. Epigenetic profiling and response to CD19 chimeric antigen receptor T-cell therapy in b-cell malignancies. *J Natl Cancer Inst* (2021) 114:436–45. doi: 10.1093/jnci/djab194

23. Chen L, Li J, Zhang YH, Feng K, Wang S, Zhang Y, et al. Identification of gene expression signatures across different types of neural stem cells with the Monte-Carlo feature selection method. *J Cell Biochem* (2018) 119:3394–403. doi: 10.1002/jcb.26507

24. Chen X, Jin Y, Feng Y. Evaluation of plasma extracellular vesicle MicroRNA signatures for lung adenocarcinoma and granuloma with Monte-Carlo feature selection method. *Front Genet* (2019) 10:367. doi: 10.3389/fgene.2019.00367

25. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res* (2011) 12:2825–30.

26. Chen L, Li Z, Zhang S, Zhang Y-H, Huang T, Cai Y-D. Predicting RNA 5-methylcytosine sites by using essential sequence features and distributions. *BioMed Res Int* (2022) 2022:4035462. doi: 10.1155/2022/4035462

27. 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. doi: 10.3390/life12020228

28. Zhou X, Ding S, Wang D, Chen L, Feng K, Huang T, et al. Identification of cell markers and their expression patterns in skin based on single-cell RNA-sequencing profiles. *Life* (2022) 12:550. doi: 10.3390/life12040550

29. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *International joint conference on artificial intelligence*. San Francisco, CA, United States: Morgan Kaufmann Publishers Inc. (1995). p. 1137–45.

30. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. *J Artif Intell Res* (2002) 16:321–57. doi: 10.1613/jair.953

31. Pan X, Chen L, Liu I, Niu Z, Huang T, Cai YD. Identifying protein subcellular locations with embeddings-based node2loc. *IEEE/ACM Trans Comput Biol Bioinform* (2022) 19:666–75. doi: 10.1109/TCBB.2021.3080386

32. Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inf Theory* (1967) 13:21–7. doi: 10.1109/TIT.1967.1053964

33. Cortes C, Vapnik V. Support-vector networks. *Mach Learn* (1995) 20:273–97. doi: 10.1007/BF00994018

34. Breiman L. Random forests. *Mach Learn* (2001) 45:5–32. doi: 10.1023/A:1010933404324

35. Chen W, Chen L, Dai Q. iMPT-FDNPL: identification of membrane protein types with functional domains and a natural language processing approach. *Comput Math Methods Med* (2021) 2021:7681497. doi: 10.1155/2021/7681497

36. Onesime M, Yang Z, Dai Q. Genomic island prediction *via* chi-square test and random forest algorithm. *Comput Math Methods Med* (2021) 2021:9969751. doi: 10.1155/2021/9969751

37. Wang Y, Xu Y, Yang Z, Liu X, Dai Q. Using recursive feature selection with random forest to improve protein structural class prediction for low-similarity sequences. *Comput Math Methods Med* (2021) 2021:5529389. doi: 10.1155/2021/5529389

38. Li X, Lu L, Chen L. Identification of protein functions in mouse with a label space partition method. *Math Biosci Eng* (2022) 19:3820–42. doi: 10.3934/mbe.2022176

39. Wu Z, Chen L. Similarity-based method with multiple-feature sampling for predicting drug side effects. *Comput Math Methods Med* (2022) 2022:9547317. doi: 10.1155/2022/9547317

40. Yang Y, Chen L. Identification of drug–disease associations by using multiple drug and disease networks. *Curr Bioinf* (2022) 17:48–59. doi: 10.2174/1574893616666210825115406

41. Matthews B. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta (BBA)-Protein Structure* (1975) 405:442–51. doi: 10.1016/0005-2795(75)90109-9

42. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* (2021) 2:100141. doi: 10.1016/j.xinn.2021.100141

43. Gettins PG. Serpin structure, mechanism, and function. *Chem Rev* (2002) 102:4751–804. doi: 10.1021/cr010170+

44. Lin HY, Muller YA, Hammond GL. Molecular and structural basis of steroid hormone binding and release from corticosteroid-binding globulin. *Mol Cell Endocrinol* (2010) 316:3–12. doi: 10.1016/j.mce.2009.06.015

45. Hirst CE, Buzza MS, Bird CH, Warren HS, Cameron PU, Zhang M, et al. The intracellular granzyme b inhibitor, proteinase inhibitor 9, is up-regulated during accessory cell maturation and effector cell degranulation, and its overexpression enhances CTL potency. *J Immunol* (2003) 170:805–15. doi: 10.4049/jimmunol.170.2.805

46. Azzi J, Skartsis N, Mounayar M, Magee CN, Batal I, Ting C, et al. Serine protease inhibitor 6 plays a critical role in protecting murine granzyme b-producing regulatory T cells. *J Immunol* (2013) 191:2319–27. doi: 10.4049/jimmunol.1300851

47. Burgener SS, Leborgne NGF, Snipas SJ, Salvesen GS, Bird PI, Benarafa C. Cathepsin G inhibition by Serpinb1 and Serpinb6 prevents programmed necrosis in neutrophils and monocytes and reduces GSDMD-driven inflammation. *Cell Rep* (2019) 27:3646–3656.e3645. doi: 10.1016/j.celrep.2019.05.065

48. Berndt SI, Camp NJ, Skibola CF, Vijai J, Wang Z, Gu J, et al. Meta-analysis of genome-wide association studies discovers multiple loci for chronic lymphocytic leukemia. *Nat Commun* (2016) 7:10933. doi: 10.1038/ncomms10933

49. Cunha SR, Mohler PJ. Ankyrin protein networks in membrane formation and stabilization. *J Of Cell And Mol Med* (2009) 13:4364–76. doi: 10.1111/j.1582-4934.2009.00943.x

50. Tessema M, Yingling CM, Picchi MA, Wu G, Ryba T, Lin Y, et al. ANK1 methylation regulates expression of MicroRNA-486-5p and discriminates lung tumors by histology and smoking status. *Cancer Lett* (2017) 410:191–200. doi: 10.1016/j.canlet.2017.09.038

51. Li H, Mou Q, Li P, Yang Z, Wang Z, Niu J, et al. MiR-486-5p inhibits IL-22-induced epithelial-mesenchymal transition of breast cancer cell by repressing Dock1. *J Cancer* (2019) 10:4695–706. doi: 10.7150/jca.30596

52. Dumax-Vorzet A, Roboti P, High S. OST4 is a subunit of the mammalian oligosaccharyltransferase required for efficient n-glycosylation. *J Cell Sci* (2013) 126:2595–606. doi: 10.1242/jcs.115410

53. Kumar V, Heinemann FS, Ozols J. Interleukin-2 induces n-glycosylation in T-cells: characterization of human lymphocyte oligosaccharyltransferase. *Biochem Biophys Res Commun* (1998) 247:524–9. doi: 10.1006/bbrc.1998.8780

54. Xiao J, Liu C, Li G, Peng S, Hu J, Qu L, et al. PDCD5 negatively regulates autoimmunity by upregulating FOXP3(+) regulatory T cells and suppressing Th17 and Th1 responses. *J Autoimmun* (2013) 47:34–44. doi: 10.1016/j.jaut.2013.08.002

55. Fu DZ, Cheng Y, He H, Liu HY, Liu YF. The fate of krüppel-like factor 9-positive hepatic carcinoma cells may be determined by the programmed cell death protein 5. *Int J Oncol* (2014) 44:153–60. doi: 10.3892/ijo.2013.2147

56. Yuan F, Wang J, Zhang K, Li Z, Guan Z. Programmed cell death 5 transgenic mice attenuates adjuvant induced arthritis by 2 modifying the T lymphocytes balance. *Biol Res* (2017) 50:40. doi: 10.1186/s40659-017-0145-4

57. Hay KA, Hanafi LA, Li D, Gust J, Liles WC, Wurfel MM, et al. Kinetics and biomarkers of severe cytokine release syndrome after CD19 chimeric antigen receptor-modified T-cell therapy. *Blood* (2017) 130:2295–306. doi: 10.1182/blood-2017-06-793141

58. Neelapu SS, Tummala S, Kebriaei P, Wierda W, Gutierrez C, Locke FL, et al. Chimeric antigen receptor T-cell therapy - assessment and management of toxicities. *Nat Rev Clin Oncol* (2018) 15:47–62. doi: 10.1038/nrclinonc.2017.148

59. Bialik S, Kimchi A. The death-associated protein kinases: structure, function, and beyond. *Annu Rev Biochem* (2006) 75:189–210. doi: 10.1146/annurev.biochem.75.103004.142615

60. Michie AM, Mccaig AM, Nakagawa R, Vukovic M. Death-associated protein kinase (DAPK) and signal transduction: regulation in cancer. *FEBS J* (2010) 277:74–80. doi: 10.1111/j.1742-4658.2009.07414.x

61. Chuang YT, Fang LW, Lin-Feng MH, Chen RH, Lai MZ. The tumor suppressor death-associated protein kinase targets to TCR-stimulated NF-kappa b activation. *J Immunol* (2008) 180:3238–49. doi: 10.4049/jimmunol.180.5.3238

62. Lai MZ, Chen RH. Regulation of inflammation by DAPK. *Apoptosis* (2014) 19:357–63. doi: 10.1007/s10495-013-0933-4

63. Tur MK, Daramola AK, Gattenlöhner S, Herling M, Chetty S, Barth S. Restoration of DAP kinase tumor suppressor function: A therapeutic strategy to selectively induce apoptosis in cancer cells using immunokinase fusion proteins. *Biomedicines* (2017) 5:59. doi: 10.3390/biomedicines5040059

64. Turner-Brannen E, Choi KY, Arsenault R, El-Gabalawy H, Napper S, Mookherjee N. Inflammatory cytokines IL-32 and IL-17 have common signaling intermediates despite differential dependence on TNF-receptor 1. *J Immunol* (2011) 186:7127–35. doi: 10.4049/jimmunol.1002306

65. Schwarz DA, Katayama CD, Hedrick SM. Schlafen, a new family of growth regulatory genes that affect thymocyte development. *Immunity* (1998) 9:657–68. doi: 10.1016/S1074-7613(00)80663-9

66. Zhang Y, Yang Z, Cao Y, Zhang S, Li H, Huang Y, et al. The Hsp40 family chaperone protein DnaJB6 enhances Schlafen1 nuclear localization which is critical for promotion of cell-cycle arrest in T-cells. *Biochem J* (2008) 413:239–50. doi: 10.1042/BJ20071510

67. Smith C, D'mello SR. Cell and context-dependent effects of the heat shock protein DNAJB6 on neuronal survival. *Mol Neurobiol* (2016) 53:5628–39. doi: 10.1007/s12035-015-9452-3

68. Bhatia P, Singh M, Singh A, Sharma P, Trehan A, Varma N. Epigenetic analysis reveals significant differential expression of miR-378C and miR-128-2-5p in a cohort of relapsed pediatric b-acute lymphoblastic leukemia cases. *Int J Lab Hematol* (2021) 43:1016–23. doi: 10.1111/ijlh.13477

69. Kong D, Zhao L, Sun L, Fan S, Li H, Zhao Y, et al. MYCN is a novel oncogenic target in adult b-ALL that activates the wnt/β-catenin pathway by suppressing DKK3. *J Cell Mol Med* (2018) 22:3627–37. doi: 10.1111/jcmm.13644

70. Seier JA, Reinhardt J, Saraf K, Ng SS, Layer JP, Corvino D, et al. Druggable epigenetic suppression of interferon-induced chemokine expression linked to MYCN amplification in neuroblastoma. *J Immunother Cancer* (2021) 9:e001335. doi: 10.1136/jitc-2020-001335

71. Sun AM, Li CG, Zhang YQ, Lin SM, Niu HR, Shi YS. Hepatocarcinoma cell-derived hepatoma-derived growth factor (HDGF) induces regulatory T cells. *Cytokine* (2015) 72:31–5. doi: 10.1016/j.cyto.2014.12.001

72. Maeda T, Merghoub T, Hobbs RM, Dong L, Maeda M, Zakrzewski J, et al. Regulation of b versus T lymphoid lineage fate decision by the proto-oncogene LRF. *Science* (2007) 316:860–6. doi: 10.1126/science.1140881

73. Sakurai N, Maeda M, Lee SU, Ishikawa Y, Li M, Williams JC, et al. The LRF transcription factor regulates mature b cell development and the germinal center response in mice. *J Clin Invest* (2011) 121:2583–98. doi: 10.1172/JCI45682

74. Martinez-Delgado B, Melendez B, Cuadros M, Garcia MJ, Nomdedeu J, Rivas C, et al. Frequent inactivation of the p73 gene by abnormal methylation or LOH in non-hodgkin's lymphomas. *Int J Cancer* (2002) 102:15–9. doi: 10.1002/ijc.10618

75. Nemajerova A, Palacios G, Nowak NJ, Matsui S, Petrenko O. Targeted deletion of p73 in mice reveals its role in T cell development and lymphomagenesis. *PloS One* (2009) 4:e7784. doi: 10.1371/journal.pone.0007784