



OPEN ACCESS

EDITED BY

Quan Zou,
University of Electronic Science and
Technology of China, China

REVIEWED BY

Bo Zhou,
Shanghai University of Medicine and
Health Sciences, China
Jing Yang,
ShanghaiTech University, China

*CORRESPONDENCE

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

[†]These authors have contributed equally
to this work

SPECIALTY SECTION

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

RECEIVED 02 February 2023

ACCEPTED 07 March 2023

PUBLISHED 17 March 2023

CITATION

Li H, Ma Q, Ren J, Guo W, Feng K, Li Z,
Huang T and Cai Y-D (2023), Immune
responses of different COVID-19
vaccination strategies by analyzing
single-cell RNA sequencing data from
multiple tissues using machine
learning methods.
Front. Genet. 14:1157305.
doi: 10.3389/fgene.2023.1157305

COPYRIGHT

© 2023 Li, Ma, Ren, Guo, Feng, Li, Huang
and Cai. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original author(s)
and the copyright owner(s) are credited
and that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Immune responses of different COVID-19 vaccination strategies by analyzing single-cell RNA sequencing data from multiple tissues using machine learning methods

Hao Li^{1†}, Qinglan Ma^{2†}, Jingxin Ren^{2†}, Wei Guo³, Kaiyan Feng⁴,
Zhandong Li¹, Tao Huang^{5,6*} and Yu-Dong Cai^{2*}

¹College of Food Engineering, Jilin Engineering Normal University, Changchun, China, ²School of Life Sciences, Shanghai University, Shanghai, China, ³Key Laboratory of Stem Cell Biology, Shanghai Institutes for Biological Sciences (SIBS), Shanghai Jiao Tong University School of Medicine (SJTUSM), Chinese Academy of Sciences (CAS), Shanghai, China, ⁴Department of Computer Science, Guangdong AIB Polytechnic College, Guangzhou, China, ⁵Bio-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, ⁶CAS 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

Multiple types of COVID-19 vaccines have been shown to be highly effective in preventing SARS-CoV-2 infection and in reducing post-infection symptoms. Almost all of these vaccines induce systemic immune responses, but differences in immune responses induced by different vaccination regimens are evident. This study aimed to reveal the differences in immune gene expression levels of different target cells under different vaccine strategies after SARS-CoV-2 infection in hamsters. A machine learning based process was designed to analyze single-cell transcriptomic data of different cell types from the blood, lung, and nasal mucosa of hamsters infected with SARS-CoV-2, including B and T cells from the blood and nasal cavity, macrophages from the lung and nasal cavity, alveolar epithelial and lung endothelial cells. The cohort was divided into five groups: non-vaccinated (control), 2*adenovirus (two doses of adenovirus vaccine), 2*attenuated (two doses of attenuated virus vaccine), 2*mRNA (two doses of mRNA vaccine), and mRNA/attenuated (primed by mRNA vaccine, boosted by attenuated vaccine). All genes were ranked using five signature ranking methods (LASSO, LightGBM, Monte Carlo feature selection, mRMR, and permutation feature importance). Some key genes that contributed to the analysis of immune changes, such as *RPS23*, *DDX5*, *PFN1* in immune cells, and *IRF9* and *MX1* in tissue cells, were screened. Afterward, the five feature sorting lists were fed into the feature incremental selection framework, which contained two classification algorithms (decision tree [DT] and random forest [RF]), to construct optimal classifiers and generate quantitative rules. Results showed that random forest classifiers could provide relative higher performance than decision tree classifiers, whereas the DT classifiers provided quantitative rules that indicated special gene expression levels under different vaccine strategies. These findings may help us to develop better protective vaccination programs and new vaccines.

KEYWORDS

immune response, COVID-19 vaccination, SARS-CoV-2 infection, machine learning method, classification rule

1 Introduction

Since the outbreak of a novel coronavirus, known as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), in late 2019, there has been an unprecedented global impact. In particular, as of 28 September 2022, more than 616 million cases have been diagnosed, and more than 6.5 million deaths have been reported worldwide (Fabbri et al., 2014). The World Health Organization named the disease caused by SARS-CoV-2 as coronavirus disease 2019 (COVID-19). Fever, sore throat, dry cough, and symptoms of pneumonia are common clinical manifestations of COVID-19, and severe COVID-19 can even lead to death (Guan et al., 2020; Parasher, 2021). Licensed vaccines have proven highly effective in preventing symptomatic and asymptomatic SARS-CoV-2 infections and reducing COVID-19-related hospitalizations and deaths (Haas et al., 2021; Castro Dopico et al., 2022), and they have given the world hope to defeat SARS-CoV-2.

A variety of COVID-19 vaccines have been marketed in response to the massive spread of SARS-CoV-2, such as mRNA vaccines, inactivated/attenuated whole virus vaccines, adenovirus vector vaccines, and recombinant protein vaccines. mRNA vaccines such as the widely used BNT16b2, which contains mRNA that can encode the SARS-CoV-2 spike protein (Mabrouk et al., 2022), have been reported effective against infection, with effectivity accounting for 89.5%–99.2% against alpha variants, 75%–96.4% against beta, and 42%–84.4% against delta (Fiolet et al., 2022). Attenuated vaccines have been used against measles virus, rubella virus, and influenza virus (Okamura and Ebina, 2021). Viruses with slow rates of proliferation in the human body were mostly attenuated through adaptation to cold culture conditions. (Makino et al., 1974; Parks et al., 2001). Live-attenuated vaccines can induce immune responses against multiple antigens and activate higher mucosal immune responses compared with other current COVID-19 vaccines (Han et al., 2021; Okamura and Ebina, 2021), which have a better and long-lasting immune effect. In general, adenoviral vector vaccines modify replication-deficient adenoviruses to express SARS-CoV-2 S protein or its epitopes (Feng et al., 2020). Viral vector vaccines can combine the safety benefits of inactivated vaccines with the immunological benefits of attenuated vaccines (Baron et al., 2018). For example, ChAdOx1 has been reported to have 74.5% protection against alpha and 67.0% protection against delta (Lopez Bernal et al., 2021).

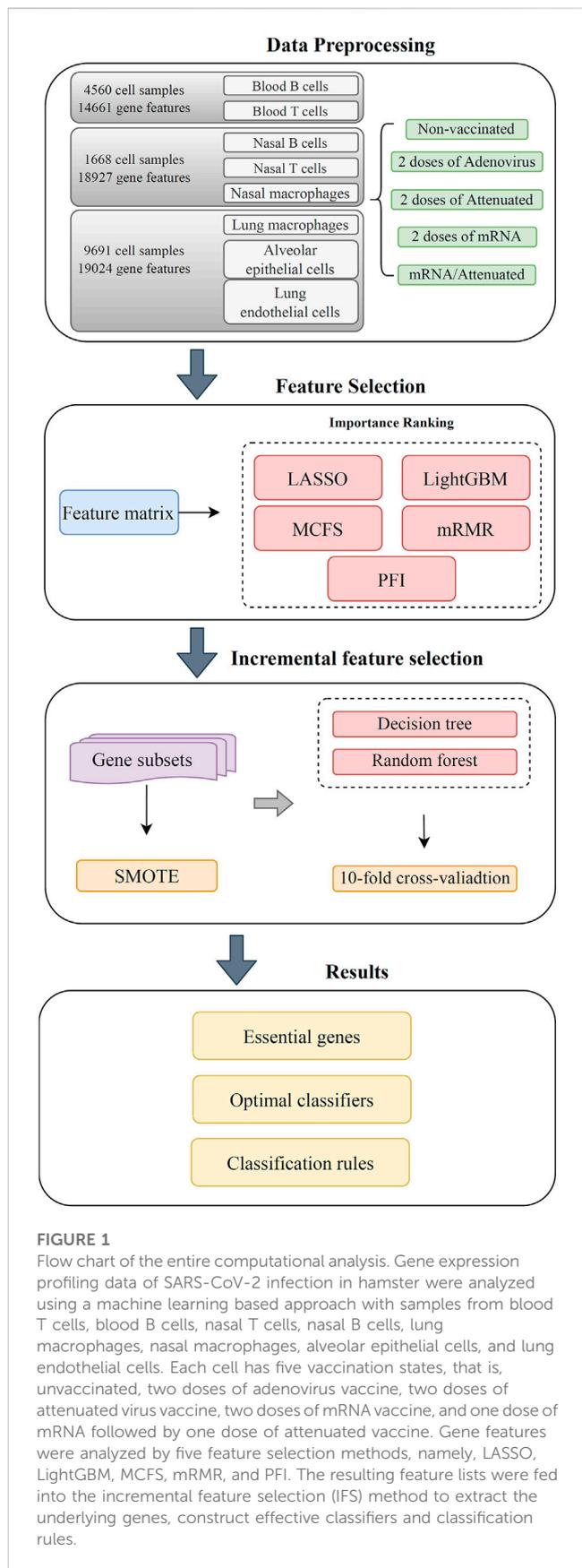
Almost all approved COVID-19 vaccines are effective in inducing protective systemic immunity, including the induction of T-cell responses (cellular immunity) (Oberhardt et al., 2021; Wherry and Barouch, 2022) and B-cell responses (antibody immunity) (Turner et al., 2021), along with the production of long-lived memory T cells and memory B cells (Sette and Crotty, 2022). Vaccine composition and dose can potentially affect the development of different immune responses. “Homologous prime-boost” vaccination is when subjects are given the same type of vaccine in a second dose as the first (Mahase, 2021), whereas “heterologous prime-boost” vaccination is when different vaccine strategies are combined in the primary and booster phases (He et al., 2021). The majority of studies have concluded that “heterologous prime-boost” vaccination has a

protective immunological advantage over “homologous prime-boost” vaccination (Benning et al., 2021; Fabricius et al., 2021; Gao et al., 2022), whereas “heterologous prime-boost” immunization may induce severe side effects (Liu X. et al., 2021; Hillus et al., 2021). However, few writers have been able to draw on any systematic comparison of “homologous prime-boost” vaccination and “heterologous prime-boost” vaccination.

This study was designed to compare the protective capacity of different vaccination strategies, including mRNA vaccine, adenoviral vector vaccine, and modified live-attenuated vaccine. The mRNA vaccine BNT16b2 and the adenovirus vaccine ChAdOx1 have received the majority of attention in recent studies, whereas comparison studies on attenuated vaccine are limited. Cell samples in eight cell types from the blood, lungs, and nasal mucosa of Syrian hamsters were divided into five groups: non-vaccinated (control), 2*adenovirus (two doses of adenovirus vaccine), 2*attenuated (two doses of attenuated virus vaccine), 2*mRNA (two doses of mRNA vaccine), and mRNA/attenuated (primed by mRNA vaccine, boosted by attenuated vaccine). Based on single-cell data on gene expression in Syrian hamsters infected with SARS-CoV-2 by nasal drip 21 days after two doses of vaccine, machine learning based analysis was designed to explore differences in immune memory protection induced by different prime-boost vaccination strategies and target cell immune status after SARS-CoV-2 infection. Five feature ranking algorithms: least absolute shrinkage and selection operator (LASSO) (Ranstan and Cook, 2018), light gradient boosting machine (LightGBM) (Ke et al., 2017), Monte Carlo feature selection (MCFS) (Dramiński and Koronacki, 2018), max-relevance and min-redundancy (mRMR) (Peng et al., 2005), and permutation feature importance (PFI) (Fisher et al., 2019) were applied to the single-cell data on each cell type, yielding five feature lists. These lists were fed into incremental feature selection (IFS) (Liu and Setiono, 1998), which incorporated decision tree (DT) (Safavian and Landgrebe, 1991) and random forest (RF) (Breiman, 2001), to extract important features, build effective classifiers and classification rules. The classifier and rules can be used to monitor the level of immunity and disease risk in SARS-CoV-2-infected patients following different vaccine combination. The features (e.g., *RPS23*, *DDX5*, *PFN1* in immune cells, and *IRF9* and *MX1* in tissue cells) and rules identified in this study could be helpful in the research for prime-boost vaccination methods, providing improved protection and duration.

2 Materials and methods

The entire workflow used in this study is shown in Figure 1. After grouping the obtained expression profile data on each cell type, the genes were ranked using several feature ranking algorithms, and a number of ranked lists were generated. Then, each list was fed into the IFS method with DT or RF. Two optimal classifiers were constructed. The methods involved are described in detail in this section.



2.1 Data

Expression profiling data for different cell types from Syrian hamsters were obtained from the GEO database under accession number GSE200596 (Nouailles et al., 2022). These data describe the cellular response to SARS-CoV-2 infection in hamsters vaccinated with mRNA vaccine, adenovirus vaccine, and attenuated virus vaccine for 21 days. Data were obtained from immune cells and tissue cells, including blood T cells, blood B cells, nasal T cells, nasal B cells, lung macrophages, nasal macrophages, alveolar epithelial and lung endothelial cells. Cell samples in each cell type were divided into five groups based on vaccination status: non-vaccine group (control group), 2*adenovirus group (two doses of adenovirus vaccine), 2*attenuated group (two doses of attenuated virus vaccine sCPD9), 2*mRNA group (two doses of mRNA vaccine), and mRNA/attenuated group (primed by mRNA vaccine and boosted by attenuated vaccine sCPD9). Table 1 demonstrates the number of cells in each group for eight cell types. Each sample from the blood, nasal cavity, and lungs contained 14661, 18927, and 19024 genes, respectively. Using genes as features and five groups as sample labels, they were entered into a machine learning framework for the analysis of the classification problem.

2.2 Feature ranking algorithms

Each sample was represented by a large number of features. It is necessary to understand which of these genes are associated with COVID-19 vaccination and SARS-CoV-2 infection. The genes involved in each cell type were analyzed using five ranking algorithms and sorted by their importance. These algorithms included LASSO (Ranstam and Cook, 2018), LightGBM (Ke et al., 2017), MCFS (Dramiński and Koronacki, 2018), mRMR (Peng et al., 2005), and PFI (Fisher et al., 2019). These methods have been widely practiced in solving life science problems (Zhao et al., 2018; Li et al., 2022a; Li et al., 2022b; Li Z. et al., 2022; Lu et al., 2022; Huang et al., 2023a; Huang et al., 2023b).

2.2.1 Least absolute shrinkage and selection operator

LASSO is a regression analysis method that can accomplish feature selection. It inputs the feature matrix into a first-order penalty function that treats the features as independent variables. This penalty function contains L1-type regularization terms. After optimization, features that tend to contribute more greatly affect the outcome of the function, a process is executed to adjust the coefficients of the independent variable. Consequently, the coefficients of some features decrease to zero, which are considered as redundant features by the algorithm and eliminated. The magnitude of the absolute value of the coefficients of the independent variables is picked up to determine the importance of the corresponding features. Accordingly, features can be ranked in a list. To execute LASSO, the package collected in Scikit-learn (Pedregosa et al., 2011) was used in this study. Default parameters were adopted.

TABLE 1 Sample size of five vaccination strategies from eight cell types.

Cell type	COVID-19 vaccination strategy	Sample size
Blood B cells	Non-vaccinated	316
	Adenovirus vaccine + Adenovirus vaccine	524
	Attenuated vaccine + Attenuated vaccine	1,018
	mRNA vaccine + mRNA vaccine	547
	mRNA vaccine + Attenuated vaccine	722
Blood T cells	Non-vaccinated	90
	Adenovirus vaccine + Adenovirus vaccine	242
	Attenuated vaccine + Attenuated vaccine	523
	mRNA vaccine + mRNA vaccine	272
	mRNA vaccine + Attenuated vaccine	306
Nasal B cells	Non-vaccinated	88
	Adenovirus vaccine + Adenovirus vaccine	38
	Attenuated vaccine + Attenuated vaccine	20
	mRNA vaccine + mRNA vaccine	61
	mRNA vaccine + Attenuated vaccine	96
Nasal T cells	Non-vaccinated	54
	Adenovirus vaccine + Adenovirus vaccine	43
	Attenuated vaccine + Attenuated vaccine	12
	mRNA vaccine + mRNA vaccine	45
	mRNA vaccine + Attenuated vaccine	61
Nasal macrophages	Non-vaccinated	417
	Adenovirus vaccine + Adenovirus vaccine	303
	Attenuated vaccine + Attenuated vaccine	45
	mRNA vaccine + mRNA vaccine	180
	mRNA vaccine + Attenuated vaccine	205
Lung macrophages	Non-vaccinated	1,437
	Adenovirus vaccine + Adenovirus vaccine	804
	Attenuated vaccine + Attenuated vaccine	753
	mRNA vaccine + mRNA vaccine	1,030
	mRNA vaccine + Attenuated vaccine	645
Alveolar epithelial cells	Non-vaccinated	614
	Adenovirus vaccine + Adenovirus vaccine	481
	Attenuated vaccine + Attenuated vaccine	481
	mRNA vaccine + mRNA vaccine	356
	mRNA vaccine + Attenuated vaccine	286
Lung endothelial cells	Non-vaccinated	869
	Adenovirus vaccine + Adenovirus vaccine	527
	Attenuated vaccine + Attenuated vaccine	724
	mRNA vaccine + mRNA vaccine	362
	mRNA vaccine + Attenuated vaccine	322

2.2.2 Light gradient boosting machine

The LightGBM method is derived from the gradient boosting DT, which is a tree structure. It is suitable for handling high-dimensional data because it can bundle mutually exclusive features during computation. A leaf-wise growth strategy was used to determine the attributes of the instances, and only the branches with high efficiency were extended. Therefore, the higher the degree of participation in the construction of the tree, the higher the degree of feature contribution it represents. Thus, features can be ranked in accordance with the degree of involvement. The present study adopted the LightGBM program obtained from <https://lightgbm.readthedocs.io/en/latest/>. For convenience, it was executed with default parameters.

2.2.3 Monte Carlo feature selection

The MCFS method is executed by constructing a number of independent DTs. The features and training samples used to build these trees are randomly selected. The random selection yields p subsets of features, and for each feature subset, t datasets are built by randomly splitting training and test samples. A DT is built on each dataset. Thus, $p \times t$ classification trees can be constructed. The importance of each feature is expressed using the relative importance (RI) score, which can be computed by

$$RI_g = \sum_{\tau=1}^{p \times t} (\omega_{ACC})^u \sum_{ng(\tau)} IG(ng(\tau)) \left(\frac{no.inng(\tau)}{no.in\tau} \right)^v, \quad (1)$$

In the formula, ω_{ACC} is the weighted accuracy of the tree τ ; $ng(\tau)$ is a node in the tree τ whose information gain is denoted as $IG(ng(\tau))$, and $no.inng(\tau)$ denotes the sample size of $ng(\tau)$, $no.in\tau$ denotes the sample size in the root of τ . In addition, u and v are two positive numbers weighting ω_{ACC} and $no.inng(\tau)/no.in\tau$, respectively. The higher the RI score of a feature, the more important it is. Features can be sorted in a list with the decreasing order of their RI values. The MCFS program was retrieved from <http://www.ipipan.eu/staff/m.draminski/mcfs.html>, which was performed with default parameters.

2.2.4 Max-relevance and min-redundancy

mRMR aims to select features that are least correlated with other features but have maximum correlation with the target variable. The correlation between the features and target variable and the redundancy between features are all measured by mutual information (MI). It first creates an empty list of features and selects one feature in each round. Generally, the feature with the highest correlation to target variable and lowest redundancy to features already in the list is selected and appended to the list. The process is repeated until all features are in the list. The mRMR package adopted in this study was obtained from <http://home.penglab.com/proj/mRMR/>. It was run using default parameters.

2.2.5 Permutation feature importance

RF is a powerful classification algorithm. It can also be used to evaluate the importance of features. Its logic is simple. If the values of a feature are permuted randomly in such a way that it causes a larger prediction error, then the feature is more important. Conversely, if it does not cause a change in the prediction result, then the feature is considered unimportant. Features are ranked in a list in terms of the change of prediction error. Here, the PFI program

was downloaded from scikit-learn (Pedregosa et al., 2011). It was performed with default parameters.

Above feature ranking algorithms were applied to the expression profiling data on each cell type. For easy descriptions, the lists generated by these five algorithms were called LASSO, LightGBM, MCFS, mRMR and PFI feature lists.

2.3 Incremental feature selection

Above five algorithms only sorted features in five lists, which did not tell us which features can be picked up for setting up classifiers. However, these lists had a common trait, that is, features with high ranks were more important than others. This indicated that some top features in the list can be used to build a classifier with good performance. In view of this, the IFS method (Liu and Setiono, 1998) was employed in this study, which can determine the features that achieve the best classification performance for one classification algorithm. It transforms the feature list into a series of feature subsets, where the features in each subset are taken from the top ones of the list, but each subset contains a different number of features. The number of features in each subset is incremented by a step compared with the previous subset. For example, if the step is 10, the first subset contains the first 10 features of the list, the second subset contains the first 20 features of the list, and so on. Then, these subsets are fed into one classification algorithm to construct classifiers, and their performance is evaluated using 10-fold cross-validation (Kohavi, 1995). The performance of these classifiers is observed, and the optimal classifier is selected, at which point the feature subset is the optimal feature subset.

2.4 Synthetic minority oversampling technique

The sample sizes were not consistent across inoculation strategies, for example, in the nasal macrophage dataset, the sample size of the non-vaccine group was 9.3 times larger than that of the two* attenuated group. These unbalanced data sets lead to preferences in the results of the classifier. The synthetic minority oversampling technique (SMOTE) method (Chawla et al., 2002) was used to tackle such problem in this study. It adds new samples to minority classes for enlarging its size. In detail, SMOTE randomly selects a sample from a minority class and then determines the k nearest samples of the selected sample in the same class using the Euclidean distance as a metric. On the line segment between one randomly selected nearest neighbor and the current sample, a random point is selected and treated as a newly generated sample. This process is repeated until the data set is balanced. Here, we used the program downloaded from <https://github.com/scikit-learn-contrib/imbalanced-learn> to implement SMOTE. The default parameters were adopted.

2.5 Classification algorithm

As previously described, IFS must be coupled with a classification algorithm. In this study, DT (Safavian and Landgrebe, 1991; Zhang et al., 2021a; Zhang et al., 2021b) and RF (Breiman, 2001; Chen et al., 2021; Ran et al., 2022; Yang and

Chen, 2022; Wang and Chen 2023) were used to construct the classifiers. Their brief introduction is as below.

2.5.1 Random forest

RF is one of the most classic classification algorithms in machine learning. In fact, it is an ensemble algorithm, which contains several DTs. Each tree is constructed by randomly selecting samples and features and the selected samples are as many as the training samples but can be same for some samples. For a test sample, each tree provides its decision. The result of RF is determined in accordance with the majority rule on all decisions. To implement RF, the corresponding package in scikit-learn (Pedregosa et al., 2011) was employed. For convenience, it was performed with default parameters.

2.5.2 Decision tree

Although RF is a powerful classification algorithm, the underlying classification principle is difficult to capture as it is a black-box algorithm. In this case, few medical insights can be obtained. DT is a classic white-box algorithm as the classification procedures are completely open, which provides more opportunities to understand the classification principle. It can be represented by a tree, where each internal node represents a feature with a threshold and each leaf node indicates the predicted result (class label). In addition to the tree representation, DT can also be represented by a group of rules. Each rule is generated by a path from the root to one leaf node. These rules imply the essential clues hidden in the investigated dataset. Similar to RF, the DT package in scikit-learn (Pedregosa et al., 2011) was employed to construct DT classifiers in IFS method.

2.6 Performance evaluation

The F1-measure is often used in machine learning to evaluate the performance of classifiers (Powers, 2011; Liang et al., 2020; Tang and Chen, 2022; Wu and Chen, 2022; Li et al., 2023; Wu and Chen, 2023). For multi-classification problems, F1-measure is defined for each class, which can be computed by

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i} \quad (2)$$

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i} \quad (3)$$

$$\text{F1-measure}_i = \frac{2 \times \text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i} \quad (4)$$

where TP_i , FP_i and FN_i represent true positive, false positive, and false negative for the i th class, i is the index of one class. To evaluate the overall performance of the classifiers, the F1-measure values on all classes can be integrated, including two measurements: macro F1 and weighted F1. Macro F1 is the direct average of all F1-measure values, whereas weighted F1 further considered the weights of F1-measure values on different classes. The weighted F1 can be expressed by

$$\text{Weighted F1} = \sum_i^L w_i \times \text{F1-measure}_i \quad (5)$$

where L represents the number of classes and w_i represents the proportion of samples in the i th class to overall sample.

In addition, prediction accuracy (ACC) and Matthews correlation coefficients (Matthews, 1975; Gorodkin, 2004; Wang

and Chen, 2022) were also used for evaluation. ACC is one of the most widely used measurements, which is defined as the proportion of correctly predicted samples. However, such measurement is not very accurate when the dataset is imbalanced. For such dataset, MCC is a more objective measurement. It can be computed by

$$\text{MCC} = \frac{\text{cov}(X, Y)}{\sqrt{\text{cov}(X, X)\text{cov}(Y, Y)}} \quad (6)$$

where X and Y are two matrices, indicating the true and predicted classes of all samples, $\text{cov}(X, Y)$ stands for the correlation coefficient of two matrices.

3 Results

3.1 Feature ranking results

The expression profiling data on each cell type was analyzed by five feature ranking algorithms. Each algorithm yielded one feature list. Totally, five feature lists (LASSO, LightGBM, MCFS, mRMR and PFI feature lists) were obtained for each cell type. All these lists on eight cell types are provided in [Supplementary Table S1](#).

3.2 Results of incremental feature selection

For each cell type, five feature lists were obtained, as listed in [Supplementary Table S1](#). Each list was fed into IFS workflow one by one. Although huge number of features were included in each list, only a few features may be highly related to indicate the differences on immune responses of different vaccination status. Thus, it was not necessary to consider all features in the list. Here, we focused on the top 2000 features in each list and adopted step 10 to construct feature subsets in IFS method. Accordingly, 200 feature subsets were constructed, on each of which one DT classifier and one RF classifier were set up. SMOTE was employed to tackle imbalanced problem when building each classifier. All classifiers were evaluated by 10-fold cross-validation. Detailed evaluation results are shown in the [Supplementary Table S2](#). Weighted F1 was selected as the major measurement. Several IFS curves were plotted to show the performance of DT and RF under different numbers of top features in each list, as shown in [Figures 2–9](#).

3.2.1 IFS results of immune cells

For blood B cells, the IFS curves of DT and RF are illustrated in [Figures 2A, B](#), respectively. It can be observed from [Figure 2A](#) that DT classifier with the top 70 features in the MCFS feature list can generate the highest weighted F1 of 0.712. As for RF, the best RF classifier adopted the top 200 features in the LightGBM feature list ([Figure 2B](#)). The detailed performance of above two classifiers is listed in [Table 2](#). Clearly, the best RF classifier was superior to the best DT classifier. Furthermore, IFS results with RF were generally better than those with DT.

For blood T cells, [Figures 3A, B](#) show the IFS curves of DT and RF on five feature lists. From [Figure 3A](#), DT classifier with top 1,060 features in the mRMR feature list can generate perfect

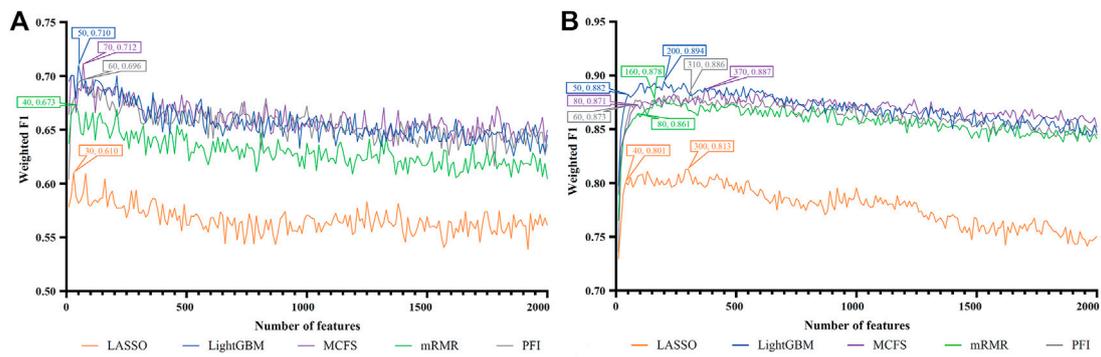


FIGURE 2
IFS curves of two classification algorithms on five feature lists for blood B cells. (A) IFS curves of the decision tree (DT). (B) IFS curves of the random forest (RF). The best DT/RF classifier used top 70/200 features in the MCFS/LightGBM feature list.

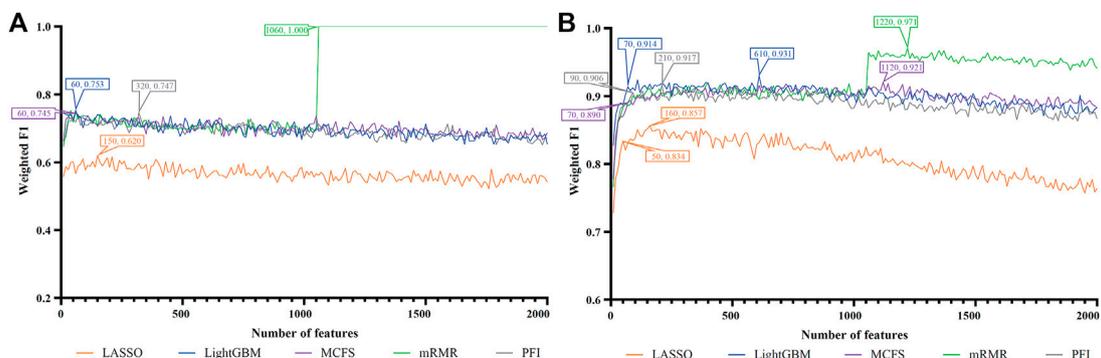


FIGURE 3
IFS curves of two classification algorithms on five feature lists for blood T cells. (A) IFS curves of the decision tree (DT). (B) IFS curves of the random forest (RF). The best DT/RF classifier used top 1,060/1,220 features in the mRMR/mRMR feature list.

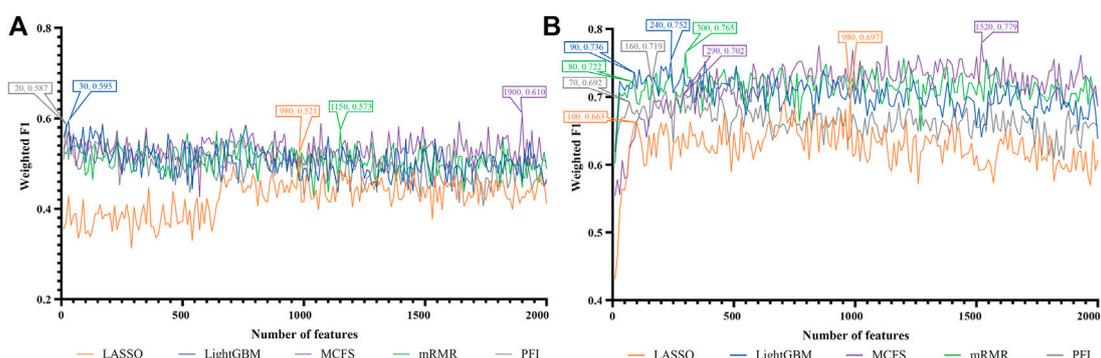


FIGURE 4
IFS curves of two classification algorithms on five feature lists for nasal B cells. (A) IFS curves of the decision tree (DT). (B) IFS curves of the random forest (RF). The best DT/RF classifier used top 1900/1,520 features in the MCFS/MCFS feature list.

performance with weighted F1 = 1. For RF, the best performance with weighted F1 = 0.971 was obtained using top 1,220 features in the mRMR feature list (Figure 3B). The detailed performance

of these two classifiers is provided in Table 2. It is amazing that this DT classifier provided better performance than the RF classifier.

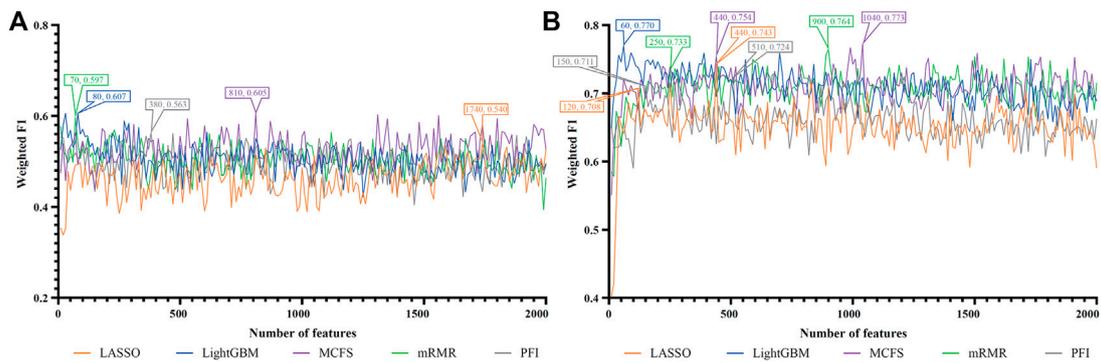


FIGURE 5 IFS curves of two classification algorithms on five feature lists for nasal T cells. **(A)** IFS curves of the decision tree (DT). **(B)** IFS curves of the random forest (RF). The best DT/RF classifier used top 80/1,040 features in the LightGBM/MCFS feature list.

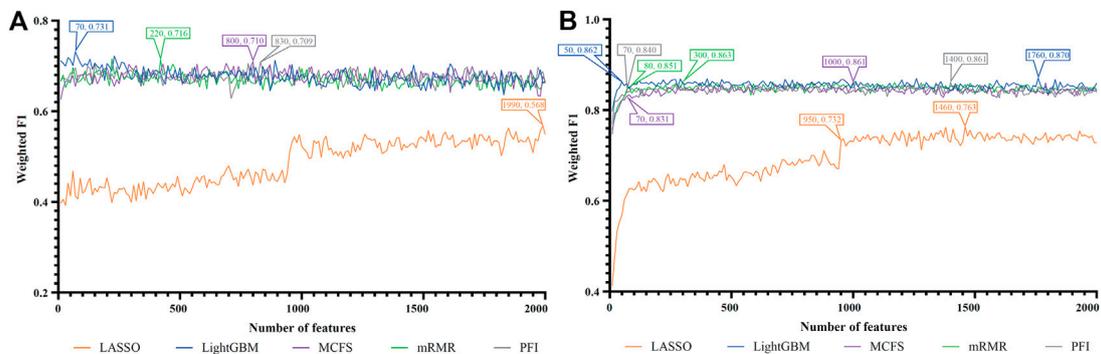


FIGURE 6 IFS curves of two classification algorithms on five feature lists for nasal macrophages. **(A)** IFS curves of the decision tree (DT). **(B)** IFS curves of the random forest (RF). The best DT/RF classifier used top 70/1760 features in the LightGBM/LightGBM feature list.

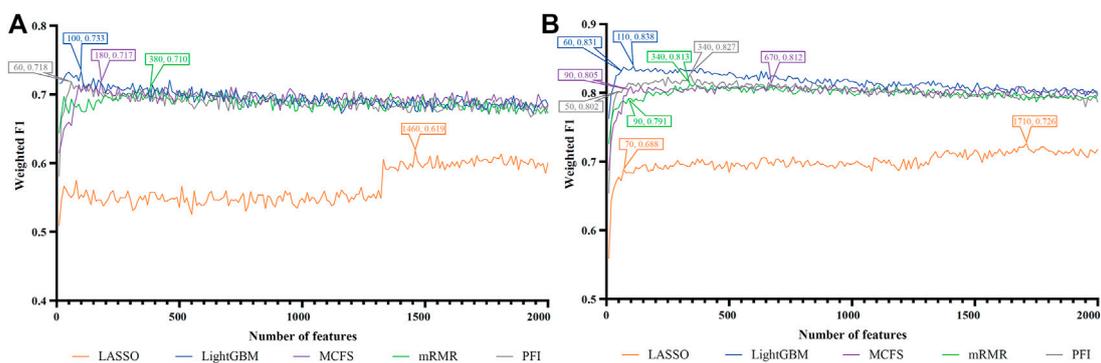


FIGURE 7 IFS curves of two classification algorithms on five feature lists for lung macrophages. **(A)** IFS curves of the decision tree (DT). **(B)** IFS curves of the random forest (RF). The best DT/RF classifier used top 100/110 features in the LightGBM/LightGBM feature list.

For Nasal B cells, the IFS curves of DT and RF on five feature lists are shown in Figures 4A, B, respectively. When using DT as the classification algorithm, its best performance was obtained by using

top 1900 features in the MCFS feature list (Figure 4A). In this case, DT yielded the weighted F1 of 0.610. As for the other classification algorithm, RF, it can be observed from Figure 4B that the top

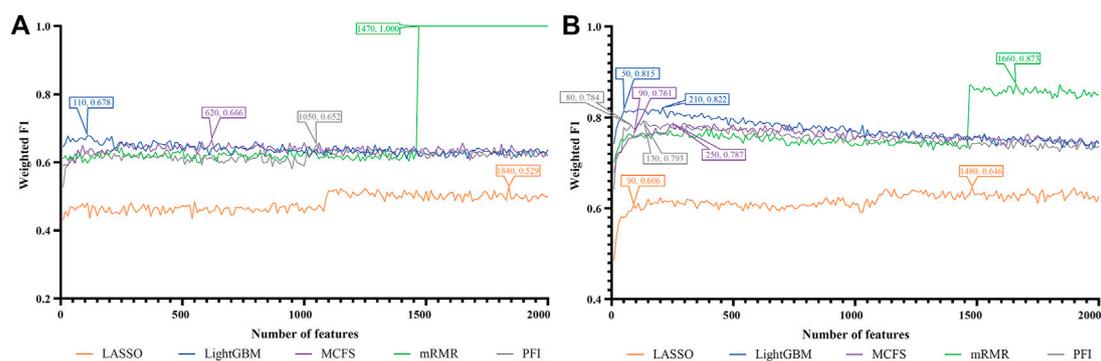


FIGURE 8

IFS curves of two classification algorithms on five feature lists for lung alveolar epithelial cells. (A) IFS curves of the decision tree (DT). (B) IFS curves of the random forest (RF). The best DT/RF classifier used top 1,470/1,660 features in the mRMR/mRMR feature list.

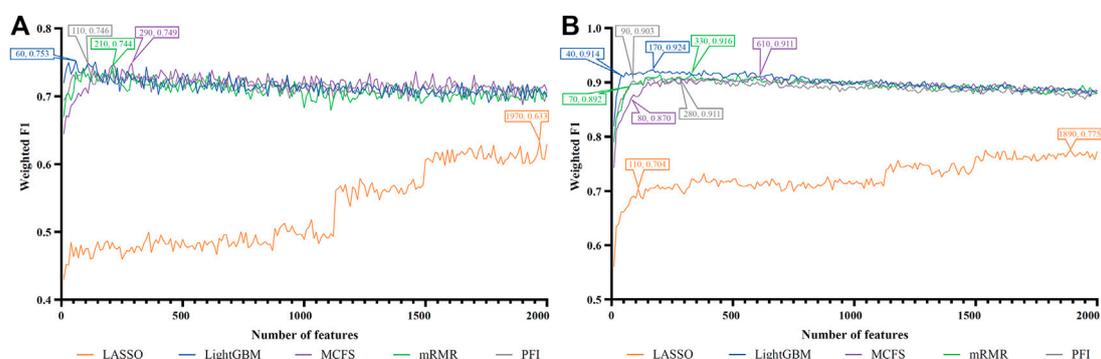


FIGURE 9

IFS curves of two classification algorithms on five feature lists for lung endothelial cells. (A) IFS curves of the decision tree (DT). (B) IFS curves of the random forest (RF). The best DT/RF classifier used top 60/170 features in the LightGBM/LightGBM feature list.

1,520 features in the MCFS features can support it in producing the best weighted F1 of 0.779. The detailed performance of above DT and RF classifiers is listed in Table 2. Generally, RF classifiers in this cell type on different feature lists were better than DT classifiers.

For Nasal T cells, the IFS curves of DT and RF on five feature lists are provided in Figures 5A, B, respectively. By observing Figure 5A, DT yielded the highest weighted F1 of 0.607 when top 80 features in the LightGBM feature list were adopted. For RF, its highest performance with weighted F1 of 0.773 was accessed when top 1,040 features in the MCFS feature list were used (Figure 5B). Table 2 also shows the detailed performance of above DT and RF classifiers. Evidently, RF classifiers on different lists were superior to DT classifiers according to the IFS results on this cell type.

For Nasal macrophages, the IFS curves of DT and RF on five feature lists are shown in Figures 6A, B. By observing the five IFS curves of DT, as shown in Figure 6A, the highest weighted F1 was 0.731, which was obtained by using top 70 features in the LightGBM feature list. With the same operation, the highest weighted F1 of RF was 0.870 when top 1760 features in the LightGBM feature list were employed. The detailed performance of above DT and RF classifiers

is also listed in Table 2. Again, the RF classifiers on different lists provided the better performance than DT classifiers.

For Lung macrophages, IFS curves of DT and RF are illustrated in Figures 7A, B, respectively. With the same arguments, DT and RF yielded the highest performance when top 100 and 110, respectively, features in the LightGBM feature list were used. They yielded the weighted F1 of 0.733 and 0.838, respectively. Detailed performance of such two classifiers is listed in Table 2. RF classifiers on different lists also generated better performance than DT classifiers.

3.2.2 IFS results of tissue cells

For alveolar epithelial cells, the IFS curves of DT and RF on five feature lists are provided in Figures 8A, B, respectively. For DT, it can yield the highest weighted F1 of 1.000 (i.e., the perfect performance) when top 1,470 features in the mRMR feature list were used, which can be observed from Figure 8A. As for RF, its best performance was obtained by using top 1,660 features in the mRMR feature list, which produced the weighted F1 of 0.873 (Figure 8B). The detailed performance of above two classifiers is listed in Table 2. Although above DT classifier was better than

TABLE 2 Performance of the best classifiers for eight cell types based on two classification algorithms.

Cell type	Classification algorithm (Feature list)	Number of features	Weight F1	MCC	ACC
Blood B cells	DT (MCFS feature list)	70	0.712	0.627	0.71
	RF (LightGBM feature list)	200	0.894	0.863	0.894
Blood T cells	DT (mRMR feature list)	1,060	1.000	1.000	1.000
	RF (mRMR feature list)	1,220	0.971	0.961	0.971
Nasal B cells	DT (MCFS feature list)	1900	0.610	0.483	0.611
	RF (MCFS feature list)	1,520	0.779	0.706	0.779
Nasal T cells	DT (LightGBM feature list)	80	0.607	0.489	0.605
	RF (MCFS feature list)	1,040	0.773	0.715	0.777
Nasal macrophages	DT (LightGBM feature list)	70	0.731	0.64	0.727
	RF (LightGBM feature list)	1760	0.870	0.825	0.870
Lung macrophages	DT (LightGBM feature list)	100	0.733	0.660	0.733
	RF (LightGBM feature list)	110	0.838	0.795	0.839
Alveolar epithelial cells	DT (mRMR feature list)	1,470	1.000	1.000	1.000
	RF (mRMR feature list)	1,660	0.873	0.838	0.873
Lung endothelial cells	DT (LightGBM feature list)	60	0.753	0.678	0.750
	RF (LightGBM feature list)	170	0.924	0.901	0.924

above RF classifier, the optimal DT classifiers on other four feature lists were generally weaker than the optimal RF classifiers on the same feature list.

For lung endothelial cells, [Figure 9A](#) shows the IFS curves of DT on five feature lists. It can be observed that DT yielded the best performance with weighted F1 of 0.753 when top 60 features in the LightGBM feature list were adopted. As for RF, its IFS curve is provided in [Figure 9B](#), from which we can see that the highest weighted F1 was 0.924. Such performance was obtained by using top 170 features in the LightGBM feature list. The detailed performance of above DT and RF classifiers is listed in [Table 2](#). Clearly, the RF classifier was superior to DT classifier. Furthermore, from [Figure 9](#), DT classifiers were evidently weaker than RF classifiers on the same feature list.

3.2.3 Intersection of different feature lists

According to [Figures 2–9](#), several optimal classifiers employed lots of top features in the corresponding lists. In this case, their efficiencies were not very high. For each of such classifiers, we want to find out another classifier which adopted much less features, whereas its performance was a little lower than the optimal classifier. These classifiers were called feasible classifiers for convenience. The difference on the performance of feasible and optimal classifiers on different feature lists for eight cell types is provided in [Table 3](#) (if exist). It can be observed that the weighted F1 of one feasible classifier was very close to that of the optimal classifier. The proportions were higher than 90%. However, the features used in feasible classifiers were much less than those used in the optimal classifiers. Most proportions were lower than 40%. Such results further indicated that features used in feasible classifiers were most important, which can capture the essential differences on immune responses between different vaccination strategies.

For each cell type, different features were used in the feasible classifiers on different feature lists. Some features may be adopted in multiple feasible classifiers, which can be deemed as more important than others. To show the relationship between five feature subsets used in five feasible classifiers (if feasible classifier was not available, optimal classifier was used), a Venn diagram was plotted for each cell type, as shown in [Figure 10](#). The intersection results for eight cell types are presented in [Supplementary Table S3](#). Some gene features occurred in multiple feature subsets would be analyzed in [Section 4.1](#).

3.3 Classification rules

Based on the IFS curves shown in [Figures 2–9](#), the performance of DT classifiers is generally lower than that of RF classifiers. However, as mentioned in the introduction of DT ([Section 2.5.2](#)), the interpretability of DT classifiers for prediction can help us analyze their biological significance, which cannot be obtained from RF classifiers. Based on the optimal DT classifiers on different feature lists for each cell type, we extracted the number of optimal features for these DT classifiers. These features were used to represent each sample and a large tree was learned from such representation of all samples. A group of quantitative classification rules can be extracted from such tree. [Supplementary Tables S4–11](#) provide the rule groups yielded by DT on different feature lists for eight cell types. Each rule contained several conditions and one result, describing the expression levels of genes under the corresponding vaccination strategies.

TABLE 3 Difference between feasible and optimal classifiers on five feature lists for eight cell types.

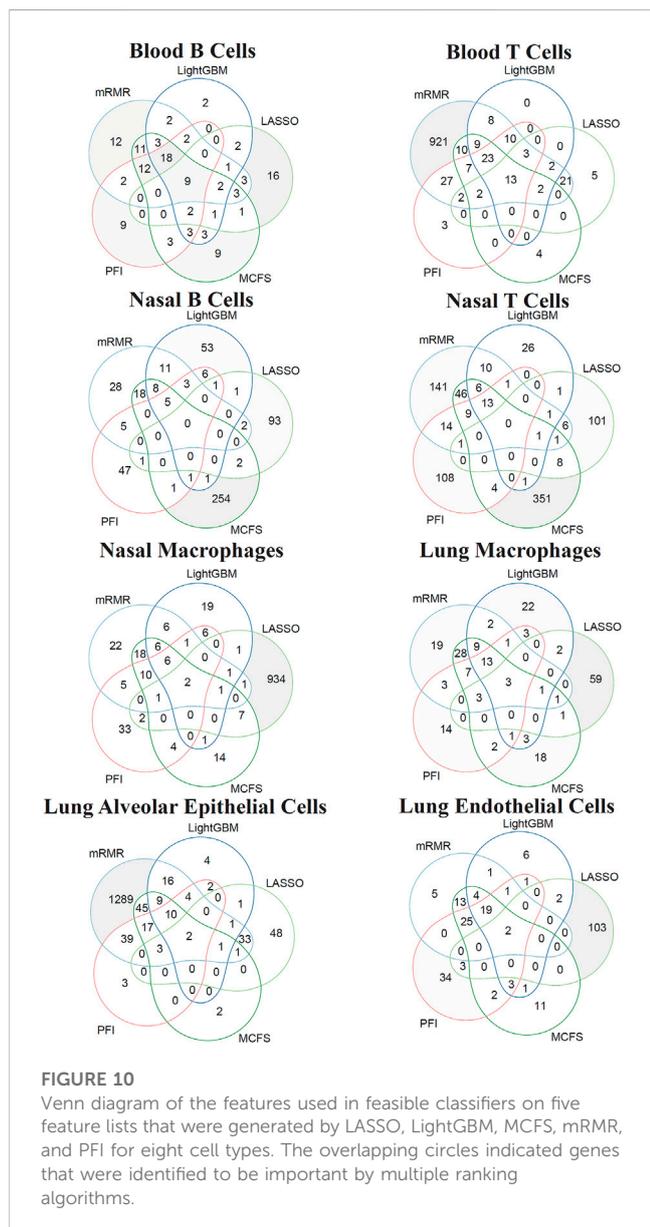
Cell Type	Feature list	Classification algorithm	Proportion to the optimal classifier	
			Number of features	Weighted F1
Blood B cells	LASSO feature list	RF	13.33%	98.52%
	LightGBM feature list	RF	25.00%	98.66%
	MCFS feature list	RF	21.62%	98.20%
	mRMR feature list	RF	50.00%	98.06%
	PFI feature list	RF	19.35%	98.53%
Blood T cells	LASSO feature list	RF	31.25%	97.32%
	LightGBM feature list	RF	11.48%	98.17%
	MCFS feature list	RF	6.25%	96.63%
	mRMR feature list ^a	-	-	-
	PFI feature list	RF	42.86%	98.80%
Nasal B cells	LASSO feature list	RF	10.20%	95.12%
	LightGBM feature list	RF	37.50%	97.87%
	MCFS feature list	RF	19.08%	90.12%
	mRMR feature list	RF	26.67%	94.38%
	PFI feature list	RF	43.75%	96.24%
Nasal T cells	LASSO feature list	RF	27.27%	95.29%
	LightGBM feature list ^a	-	-	-
	MCFS feature list	RF	42.31%	97.54%
	mRMR feature list	RF	27.78%	95.94%
	PFI feature list	RF	29.41%	98.20%
Nasal macrophages	LASSO feature list	RF	65.07%	95.94%
	LightGBM feature list	RF	2.84%	99.08%
	MCFS feature list	RF	7.00%	96.52%
	mRMR feature list	RF	26.67%	98.61%
	PFI feature list	RF	5.00%	97.56%
Lung macrophages	LASSO feature list	RF	4.09%	94.77%
	LightGBM feature list	RF	54.55%	99.16%
	MCFS feature list	RF	13.43%	99.14%
	mRMR feature list	RF	26.47%	97.29%
	PFI feature list	RF	14.71%	96.98%
Alveolar epithelial cells	LASSO feature list	RF	6.08%	93.81%
	LightGBM feature list	RF	23.81%	99.15%
	MCFS feature list	RF	36.00%	96.70%
	mRMR feature list ^a	-	-	-
	PFI feature list	RF	61.54%	98.87%
Lung endothelial cells	LASSO feature list	RF	5.82%	90.84%
	LightGBM feature list	RF	23.53%	98.92%

(Continued on following page)

TABLE 3 (Continued) Difference between feasible and optimal classifiers on five feature lists for eight cell types.

Cell Type	Feature list	Classification algorithm	Proportion to the optimal classifier	
			Number of features	Weighted F1
	MCFS feature list	RF	13.11%	95.50%
	mRMR feature list	RF	21.21%	97.38%
	PFI feature list	RF	32.14%	99.12%

*Feasible classifier on the corresponding feature list was not identified.



4 Discussion

In this study, we integrated multiple machine learning approaches to perform in-depth analysis of single-cell transcriptome data under different COVID-19 vaccine strategies

using hamsters as experimental subjects. The effectiveness of various COVID-19 vaccination techniques to provide protection is closely correlated with the gene expression patterns of certain immune and tissue cells. Several optimal classifiers were constructed, which can be used to predict vaccination strategies for two doses of adenovirus vaccine, two doses of attenuated virus vaccine, two doses of mRNA vaccine, one dose of mRNA and one dose of attenuated vaccine. The tissue cells included alveolar epithelial and endothelial cells from the lungs, whereas the immune cells included B cells, T cells, and macrophages from the blood, nasal, and lungs, respectively. Some essential gene features identified by the computational analysis might be crucial and the classification rules can imply the expression levels of key genes in different vaccine strategies after SARS-CoV-2 infection. Thus, the features and rules identified in this study may provide evidence for the immune memory capacity of different vaccination strategies and help advance more effective vaccination methods to combat SAR-CoV-2 infection. Based on the newly released publications, some essential gene features and quantitative rules can be confirmed to play crucial roles in anti-viral responses.

4.1 Analysis of top features in SARS-CoV-2-infected hamsters for distinguishing different vaccination strategies

Based on our computational analysis, we identified a set of essential genes differentially expressed in immune cells and lung tissue cells to identify vaccine recipients with different prime-boost vaccination after SARS-CoV-2 infection. Recent studies have demonstrated the mechanism of some genes in the antiviral process. One or two genes were selected for detailed analysis for each cell type, which are listed in Table 4.

4.1.1 Top features in immune cells

In blood B cells, *RPS23* (ENSG00000186468) is a 40S ribosomal protein (Barrado-Gil et al., 2020) that plays a role in ribosome assembly and protein translation, which may be related to antibody production by B cells. Moreover, *RPS23* plays an important role in physiological and pathological processes such as tumorigenesis, immune signaling, and development (Zhou et al., 2015). *RPS23* has also been reported to be a new antimicrobial peptide that can recognize and kill potential pathogens (Ma et al., 2020). Furthermore, the expression level of *RPS23* may be related to the immune response induced after vaccination. Two recent studies have found that *RPS23* expression was changed after inactivated

TABLE 4 Top features identified by the computational analysis in immune and lung cells.

Classification	Cell type	Gene symbol	Description
Immune cells	Blood B cells	RPS23	ribosomal protein S23
		TPT1	tumor protein, translationally-controlled 1
	Nasal B cells	IFIT3	interferon induced protein with tetratricopeptide repeats 3
	Blood T cells	EEF1A1	eukaryotic translation elongation factor 1 alpha 1
		UBA52	ubiquitin A-52 residue ribosomal protein fusion product 1
	Nasal T cells	DDX5	DEAD-box helicase 5
		DEF6	DEF6 guanine nucleotide exchange factor
	Lung macrophages	PFN1	profilin 1
		RPSA	ribosomal protein SA
	Nasal macrophages	ISG15	ISG15 ubiquitin like modifier
Lung cells	Alveolar epithelial cells	IRF9	interferon regulatory factor 9
	Lung endothelial Cells	MX1	MX dynamin like GTPase 1
		MX2	MX dynamin like GTPase 2

vaccination (Pisano et al., 2021), indicating its potential role in immune response. *TPT1* (ENSG00000133112) is involved in the regulation of apoptosis (Bruneel et al., 2005), and it is also related to the regulation of protein synthesis in immune cells (Arowolo et al., 2021). Moreover, *TPT1* is involved in the viral response (Leong and Chow, 2006). Based on recent publications, *TPT1* plays an important role in the development of COVID-19 (Hasankhani et al., 2021), and it can be used to predict COVID-19 (Akbulut et al., 2022). Therefore, *TPT1* may be involved in the antiviral response induced by SARS-CoV-2 infection, thereby promoting the exploration of the immune memory capacity induced by different vaccines.

In nasal B cells, *IFIT3* (ENSG00000119917) belongs to the interferon-stimulated gene (ISG) family, and it is involved in immune processes, including innate immunity, inflammatory response, and antiviral immunity (de Veer et al., 1998; Fleith et al., 2018). In addition, *IFIT3* is differentially expressed in B cells and monocytes in patients with autoimmune diseases (Fang et al., 2021), indicating that the *IFIT3* gene may be involved in B cell-mediated humoral immunity. With regard to the relationship between *IFIT3* and viral infection, *IFIT3* was found to be differentially expressed in response to infection with RNA viruses (Zhou et al., 2013; Feng et al., 2018) and was considered to have predictive potential for COVID-19 because the expression level can be affected by SARS-CoV-2 infection (Shaath et al., 2020; Gao et al., 2021).

In blood T cells, *EEF1A1* (ENSG00000156508) encodes the same type of alpha subunit of a complex, namely, elongation factor-1, which is responsible for aminoacyl tRNAase delivery to the ribosome; promotes cell growth and proliferation; and inhibits apoptosis (Mills and Gago, 2021). Huang et al. found that the expression of *EEF1A1* was positively correlated with the number of initial CD4⁺ T cells (Huang and Zhou, 2022), indicating that *EEF1A1* may be associated with cellular immunity. In addition, *EEF1A1* could inhibit viral growth (Zhang et al., 2015), and it is

associated with inflammatory responses (Maruyama et al., 2007). The *EEF1A1* protein has been reported to play a key role in several viral infections by interacting with viral proteins (Sikora et al., 2009; Zhang et al., 2015). Based on a recent study, SARS-CoV-2 infection affects *EEF1A1* expression, and it may be associated with the suppression of viral RNA replication. Ubiquitin A-52 residue ribosomal protein fusion product 1, *UBA52* (ENSG00000221983), is a ubiquitin-encoding gene encoding ubiquitin fusion proteins (Kobayashi et al., 2016). *UBA52* participates in H5N1 viral replication (Wang et al., 2018), which is linked to viral infection. *UBA52* deficiency may cause cell cycle arrest and inhibit protein synthesis (Mao et al., 2018), revealing its potential role in T cells performing antiviral functions. In addition, *UBA52*, as a ubiquitin-encoding gene, might be associated with antigen processing and MHC II antigen presentation, which is consistent with the role of *UBA52* in the proteasomal degradation of CD4⁺ T cells after SARS-CoV-2 infection identified by Tiwari et al. (Tiwari et al., 2022).

In nasal T cells, *DDX5* (ENSG00000108654), also known as p68, is a typical member of the dead box ATP-dependent RNA unwinding enzyme family (Lane and Hoeffler, 1980). *DDX5* gene encodes a protein that plays an important role in RNA metabolism (Zonta et al., 2013; Dardenne et al., 2014). A recent study has focused on the function of *DDX5* in regulating cellular life cycles, cancer and development, and spermatogenesis (Hashemi et al., 2019; Legrand et al., 2019; Hu et al., 2022). Notably, *DDX5* has been associated with multiple viral infections. For example, *DDX5* could inhibit RNA transcription of hepatitis B virus (Zhang et al., 2016) and enhance RNA transcription of hepatitis C virus (Goh et al., 2004), and *DDX5* may promote SARS-CoV replication (Chen J. Y. et al., 2009). In addition, a recent study has found that *DDX5* is involved in the regulation of SARS-CoV-2 replication (Ariumi, 2022), thereby identifying the ability of the immune memory of COVID-19 vaccine. *DEF6* (ENSG0000023892), also known as IRF4-binding protein or SWAP-70-like bridging protein (SLAT) of T cells, is a specific guanine nucleotide exchange factor for Rho GTPase

Cdc42 and Rac1 (Deng et al., 2020). *DEF6* is expressed in myeloid cells, and it controls innate immunity (Chen Q. et al., 2009). Thus, it is strongly related to immunity. Moreover, mutations or deletions of *DEF6* can lead to immune dysregulation diseases (Fournier et al., 2021). *DEF6*, as a feature gene, is highly expressed in T cells, and it plays an important role in T cell proliferation, Th1/Th2 lineage differentiation, and function. It is also involved in T cell receptor signaling regulation (Izawa et al., 2017; Deng et al., 2020). Some researchers have also found that *DEF6* deficiency adversely affects the function of memory T cells (Rossi et al., 2011).

In lung macrophages, *PFN1* (ENSG00000108518) is a key actin regulatory protein that is involved in the regulation of actin filament assembly (Mouneimne et al., 2012), which may be related to the migration of macrophages to the site of infection. *PFN1* may also be crucial for viral transcriptional activation and airway hyperresponsiveness (Leng et al., 2021). As *PFN1* expression is altered by SARS-CoV-2 infection (Shen et al., 2020), it can be identified as a biomarker to detect COVID-19. *RPSA* (ENSG00000168028) is an important component of the small ribosomal subunit with a wide range of physiological functions, including RNA processing, cell migration, and angiogenesis (Bernard et al., 2009; O'Donohue et al., 2010; Rea et al., 2012). *RPSA* also plays a role in regulating the mitogen-activated protein kinase (MAPK) signaling pathway (Givant-Horwitz et al., 2004), and many viral infections have been associated with deviations from well-balanced control of the MAPK signaling cascade, such as Ebola virus (Strong et al., 2008) and influenza A virus (Mizumura et al., 2003). *RPSA* has been found to be expressed in a variety of immune cells, including neutrophils, monocytes, and T cells (Sun et al., 2020), to participate in the immune process. In macrophages, *RPSA* expression levels were altered after infection with *Mycoplasma pleuropneumoniae* and porcine circovirus type 2 (Liu M. et al., 2021) or after BCG vaccination (Liu et al., 2022).

In nasal macrophages, an abundantly induced ISG, *ISG15* (ENSG00000187608), is crucial for viral infection (Morales and Lenschow, 2013). In the beginning of the innate response to viral infection, *ISG15* has been shown to be substantially increased as an effector and signaling molecule (Freitas et al., 2020). In addition, *ISG15* can prevent viral replication by interfering with the exocytosis and endogenous translation machinery that viruses rely on to grow (Okumura et al., 2007). Following SARS-CoV-2 infection, a study found that the secretion of *ISG15* exacerbated the inflammatory response (Cao, 2021), indicating the immunological role of *ISG15* in COVID-19. In macrophages, the expression of *ISG15* can promote macrophage polarization toward a pro-inflammatory and antiviral M1 phenotype to produce more antiviral factors (Freitas et al., 2020). Furthermore, macrophages can display increased autophagy and mitophagy of infected cells under *ISG15* stimulation (Swaim et al., 2017).

4.1.2 Top features in lung tissue cells

In lung alveolar epithelial cells, *IRF9* (ENSG00000213928) is a key component of the type I and type III interferon signaling pathways, which controls the antiviral response of cells to type I and type III interferons (Stark and Darnell, 2012; Lazear et al., 2019). The antiviral ability of *IRF9* against common viruses such as respiratory viruses has been well demonstrated (Hernandez et al., 2018; Bravo Garcia-Morato et al., 2019). A recent study revealed that the high expression level of

IRF9 in SARS-CoV-2-infected cells controls the ISGF-3-dependent response to type I and type III interferons, thereby accelerating the initiation of the immune response (Ahmed, 2020). Therefore, the expression level of the *IRF9* gene is related to the degree of SARS-CoV-2 infection of alveolar epithelial cells.

In lung endothelial cells, *MX1* (ENSG00000157601) and *MX2* (ENSG00000183486) encode two different guanosine triphosphate (GTP)-metabolizing proteins that differ remarkably in viral specificity and mechanism of action. *MX1* has a wide antiviral activity against RNA and DNA viruses, whereas *MX2* is only effective against certain viruses, such as HIV (Jung et al., 2019). *MX1* is involved in the antiviral innate response, and it regulates neutrophil activity and brings neutrophils into the tissues for immune functions (Henarejos-Castillo et al., 2020). *MX1* can be induced by SARS-CoV-2 infection (Senapati et al., 2020; Halfmann et al., 2022). Based on a study conducted in 2020, SARS-CoV-2 can induce strong expression of *MX1* in the lungs of infected hamsters (Halfmann et al., 2022). Thus, the expression of *MX1* and *MX2* could be used to determine the degree of lung infection.

4.2. Analysis of Classification Rules in SARS-CoV-2-infected Hamsters for Distinguishing Different Vaccination Strategies

Besides essential genes, quantitative rules were another main output of the computational analysis, which are provided in Supplementary Tables S4–11. Each rule contained several gene features and thresholds. It is quite difficult to confirm the underlying expression patterns of each rule. Here, we extracted some important conditions for detailed analysis. For each cell type, we focused on one important gene such that different results (class labels) can be outputted with different thresholds and tendencies. The conditions for each cell type are listed in Table 5.

4.2.1 Classification rules in immune cells

In blood B cells, *PAX5* (ENSG00000196092) is upregulated in samples with two doses of attenuated vaccination and mRNA/attenuated vaccination but downregulated in unvaccinated samples. *PAX5* is a crucial gene, which is known as a key factor for B cell proliferation and differentiation (Mullighan et al., 2007). Harris et al. found that *PAX5* binds to Fbxo7 transcription in pre-B cells (Harris et al., 2021). *FBX O 7* is known for its important role in lymphocyte development and differentiation (Ballesteros Reviriego et al., 2019). Thus, *PAX5* might be involved in the positive regulation of B cell proliferation and differentiation. The expression of *PAX5* is essential for memory B cell development after antigen encounter (Johnson et al., 2005; Nutt and Tarlinton, 2011). In addition, *PAX5* expression declines as plasma cells differentiate (Urbánek et al., 1994; Cobaleda et al., 2007), which may partially reflect immunological memory activation. Based on our classification rules, the expression of *PAX5* in B cells may indicate that specific vaccine combinations induce better B cell memory.

In nasal B cells, *IFIT3* (ENSG00000119917) was identified by our computational method, which was shown to be upregulated in unvaccinated and heterologous vaccinated samples. However, the upregulation of *IFIT3* expression was remarkable in mRNA/

TABLE 5 Representative conditions for different cell types.

Cell Type	Rules	Parameters	Predicted class
Blood B cells	Condition 0	ENSG0000019609(PAX5) \leq 2.1383	Non-vaccinated
	Condition 1	ENSG0000019609(PAX5) $>$ 0.3466	2*Attenuated
	Condition 2	ENSG0000019609(PAX5) $>$ 0.8959	mRNA/Attenuated
Nasal B cells	Condition 3	ENSG00000119917(IFIT3) $>$ 0.3466	Non-vaccinated
	Condition 4	ENSG00000119917(IFIT3) $>$ 0.8959	mRNA/Attenuated
Blood T cells	Condition 5	ENSG00000221983(UBA52) \leq 2.4414	Non-vaccinated
	Condition 6	ENSG00000221983(UBA52) $>$ 2.4414	2*Adenovirus
	Condition 7	ENSG00000221983(UBA52) $>$ 2.1910	2*Attenuated
	Condition 8	ENSG00000221983(UBA52) $>$ 2.7403	mRNA/Attenuated
Nasal T cells	Condition 9	ENSG00000233927(RPS28) \leq 4.0687	Non-vaccinated
	Condition 10	ENSG00000233927(RPS28) $>$ 4.0687	2*Adenovirus
	Condition 11	ENSG00000233927(RPS28) $>$ 4.0687	2*Attenuated
Lung macrophages	Condition 12	ENSG00000163563(MNDA) $>$ 3.5972	Non-vaccinated
	Condition 13	ENSG00000163563(MNDA) \leq 3.2385	2*Adenovirus
	Condition 14	ENSG00000163563(MNDA) \leq 3.5972	2*mRNA
	Condition 15	ENSG00000163563(MNDA) \leq 2.8029	mRNA/Attenuated
Nasal macrophages	Condition 16	ENSG00000160932(LY6E) $>$ 2.5249	Non-vaccinated
	Condition 17	ENSG00000160932(LY6E) \leq 2.5249	2*Adenovirus
	Condition 18	ENSG00000160932(LY6E) \leq 2.5249	2*Attenuated
	Condition 19	ENSG00000160932(LY6E) \leq 2.5249	2*mRNA
	Condition 20	ENSG00000160932(LY6E) \leq 2.5249	mRNA/Attenuated
Lung alveolar epithelial cells	Condition 21	ENSG00000187608(ISG15) $>$ 0.8959	Non-vaccinated
	Condition 22	ENSG00000187608(ISG15) \leq 0.8959	2*Adenovirus
	Condition 23	ENSG00000187608(ISG15) \leq 0.8959	2*Attenuated
Lung endothelial cells	Condition 24	ENSG00000204264(PSMB8) $>$ 0.3466	Non-vaccinated
	Condition 25	ENSG00000204264(PSMB8) \leq 1.9945	2*Attenuated

attenuated vaccination samples. *IFIT3* was found to be involved in viral responses (Metz et al., 2013). *IFIT3* is an IFN-inducible protein whose expression is increased by viral infection and IFN treatment (Pidugu et al., 2019). Although no direct evidence is found for the role of *IFIT3* expression in B cells, altered *IFIT3* expression induced by SARS-CoV-2 infection has been widely demonstrated. *IFIT3* could be related to immune response to SARS-CoV-2 infection based on the findings of several studies, demonstrating that *IFIT3* is strongly expressed in the pulmonary inflammatory cells of patients with COVID-19 (Shaath et al., 2020; Vishnubalaji et al., 2020). Moreover, *IFIT3* was found to play an important role in limiting the replication of RNA viruses, including SARS-CoV-2 (Metz et al., 2013; Pfaender et al., 2020; Martin-Sancho et al., 2021). Collectively, the expression level of *IFIT3* may indicate the immune response to viral infection in B cells, which can be used to compare the

immunological memory induced by various vaccination strategies.

In blood T cells, *UBA52* (ENSG00000221983) was identified as a rule gene. *UBA52* expression in T cells was shown to be upregulated in recipients with two doses of adenovirus vaccination, two doses of attenuated vaccination, and mRNA/attenuated vaccination. As previously discussed, *UBA52* was considered as a signature gene in blood T cells. As a ubiquitin-encoding gene (Kobayashi et al., 2016), *UBA52* was found to be closely associated with proteasomal degradation in CD4⁺ T cells (V⁺Kovski et al., 2019). Picciotto et al. indicated that *UBA52* is rapidly upregulated after T-cell activation (de Picciotto et al., 2022), and it may be involved in effector T-cell activation. In addition, *UBA52* was found to be highly expressed in patients with COVID-19 (Jiang et al., 2022), which may be related to COVID-19 pathogenesis. Thus, the differential expression of

UBA52 in blood T cells helps to distinguish different prime-boost vaccination strategies.

In nasal T cells, ribosomal gene *RPS28* (ENSG00000233927) was identified, whose basic function is to participate in protein synthesis, folding, and assembly (Kim et al., 2019). Based on our rules, *RPS28* was upregulated in samples with two doses of adenovirus vaccine and two doses of attenuated vaccine. *RPS28* has been reported to control the generation of MHC class I peptides by regulating non-canonical translation (Wei et al., 2019), leading to differential antigen presentation in cells. A study on melanoma found that mutations in ribosomal proteins resulting in the deletion of *RPS28* caused greater killing of melanoma cells by CD8⁺ T cells (Dersh et al., 2021), indicating the association of *RPS28* with CD8⁺ T cells. Therefore, the expression of *RPS28* in nasal T cells may help to distinguish different vaccine combinations and predict immune memory activation.

In lung macrophages, *MNDA* (ENSG00000163563) was identified, whose function is thought to be related to immune cells (Metcalf et al., 2014). *MNDA* was downregulated in samples receiving COVID-19 vaccines, with the greatest downregulation in mRNA/attenuated vaccine recipients. *MNDA* is an interferon-inducible gene, whose protein contains a pyridine structural domain that plays a role in programmed cell death and inflammation-related signaling (Bottardi et al., 2020). *MNDA* was strongly expressed in activated macrophages linked to inflammation but not in normal tissue cells (Miranda et al., 1999), indicating the relationship between *MNDA* expression and tissue inflammation. In monocytes, *MNDA* was found to be remarkably upregulated after IFN α exposure (Briggs et al., 1994), and it could be a major regulator of monocyte and granulocyte lineage (Milot et al., 2012). Thus, the downregulation of the immune-related gene *MNDA* in lung macrophages may be due to the good protective capacity of the vaccine to keep the lungs free from viral infection.

In nasal macrophages, *LY6E* (ENSG00000160932) was identified as a rule gene, whose expression was shown to be downregulated in all vaccination strategies except for controls. *LY6E* encodes an interferon-inducible protein, which has been shown to regulate viral infection in a cell type-dependent manner (Godfrey et al., 1992). *LY6E* is involved in the regulation of infection by a variety of viruses, and it was found to promote HIV-1 (Yu et al., 2017), yellow fever virus (Schoggins et al., 2011), and influenza A virus (Mar et al., 2018) infection. Therefore, the reduced expression of the *LY6E* gene in nasal macrophages of samples with COVID-19 vaccination may be due to the fact that COVID-19 vaccination helped to avoid SARS-CoV-2 attack on the nasal cavity.

4.2.2 Classification rules in lung tissue cells

In lung alveolar epithelial cells, *ISG15* (ENSG00000187608) was identified as a rule gene in lung alveolar epithelial cells, which is an IFN α -stimulated gene that plays an important role in the antiviral response (Swaim et al., 2020). *ISG15* was downregulated in samples with two doses of adenovirus or attenuated vaccination and upregulated in controls, reflecting the protective ability of COVID-19 vaccination on target cells. It is hypothesized that *ISG15* can prevent viral assembly by tagging newly translated viral proteins (Shin et al., 2020). *ISG15* has also

been found to drive antiviral immune functions by modifying viral proteins, inhibiting viral replication, and regulating host signaling pathways associated with viral infection (Perng and Lenschow, 2018). In addition, *ISG15* expression exacerbates the inflammatory response of COVID-19 (Cao, 2021), partially indicating the tissue damage caused by SARS-CoV-2 infection. Thus, the expression of *ISG15* on alveolar epithelial cells may reflect virus-induced damage, helping to compare the protective capacity of COVID-19 vaccines.

In lung endothelial cells, the *PSMB8* (ENSG00000204264) gene was found to be upregulated in controls and downregulated in samples receiving two doses of attenuated COVID-19 vaccines based on our rule. *PSMB8* encodes the proteasome 20S subunit Beta 8, and it is involved in the positive regulation of apoptosis (Yang et al., 2009; Jean-Baptiste et al., 2017). More et al. found that *PSPM8* is involved in mediating viral infection and synthesis in target cells, indicating the potential role of *PSPM8* in viral infection (More et al., 2019). In addition, *PSMB8* is involved in regulating cytokine secretion during viral infection (Servaas et al., 2021). Furthermore, in patients with mild COVID-19, the high expression of *PSMB8* could promote M1 macrophage polarization (Desterke et al., 2021). The extensive involvement of *PSPM8* in viral infection may help us to identify lung damage caused by SARS-CoV-2 infection.

5 Conclusion

In investigating the differences in immune changes induced by SARS-CoV-2 infection under different vaccination strategies, this study designed a machine learning based framework to analyze expression profile datasets from lung tissue cells (endothelial cells and alveolar epithelial cells) and immune cells from different sites (B cells, T cells, and macrophages). Five feature ranking methods and two classification algorithms were used to obtain key genes and easily understand quantitative classification rules associated with COVID-19 vaccination and SARS-CoV-2 infection. These results revealed the pathways of action of different vaccination regimens in COVID-19, which could lead to the development of safe and long-lasting vaccination regimens.

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=GSE200596>.

Author contributions

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

Funding

This research was funded by the National Key R&D Program of China [2022YFF1203202], Strategic Priority Research Program of Chinese Academy of Sciences [XDA26040304 and XDB38050200], the Fund of the Key Laboratory of Tissue Microenvironment and Tumor of Chinese Academy of Sciences [202002], Shandong Provincial Natural Science Foundation [ZR2022MC072].

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.

References

- Ahmed, F. (2020). A network-based analysis reveals the mechanism underlying vitamin D in suppressing cytokine storm and virus in SARS-CoV-2 infection. *Front. Immunol.* 11, 590459. doi:10.3389/fimmu.2020.590459
- Akbulut, S., Yağın, F. H., and Olak, C. (2022). Prediction of COVID-19 based on genomic biomarkers of metagenomic next-generation sequencing data using artificial intelligence Technology. *Erciyes Med. J.* 44, 544–548. doi:10.14744/etd.2022.00868
- Ariumi, Y. (2022). Host cellular RNA helicases regulate SARS-CoV-2 infection. *J. Virol.* 96, e0000222. doi:10.1128/jvi.00002-22
- Arowolo, O., Pobezinsky, L., and Suvorov, A. (2021). Chemical exposures affect innate immune response to SARS-CoV-2. *Int. J. Mol. Sci.* 22, 12474. doi:10.3390/ijms222212474
- Ballesteros Reviriego, C., Clare, S., Arends, M. J., Cambridge, E. L., Swiatkowska, A., Caetano, S., et al. (2019). FBXO7 sensitivity of phenotypic traits elucidated by a hypomorphic allele. *PLoS One* 14, e0212481. doi:10.1371/journal.pone.0212481
- Baron, M. D., Iqbal, M., and Nair, V. (2018). Recent advances in viral vectors in veterinary vaccinology. *Curr. Opin. Virol.* 29, 1–7. doi:10.1016/j.coviro.2018.02.002
- Barrado-Gil, L., Del Puerto, A., Muñoz-Moreno, R., Galindo, I., Cuesta-Gejjo, M., Urquiza, J., et al. (2020). African swine fever virus ubiquitin-conjugating enzyme interacts with host translation machinery to regulate the host protein synthesis. *Front. Microbiol.* 11, 622907. doi:10.3389/fmicb.2020.622907
- Benning, L., Töllner, M., Hidmark, A., Schaefer, M., Nusshag, C., Kälble, F., et al. (2021). Heterologous ChAdOx1 nCoV-19/BNT162b2 prime-boost vaccination induces strong humoral responses among health care workers. *Vaccines* 9, 857. doi:10.3390/vaccines9080857
- Bernard, A., Gao-Li, J., Franco, C. A., Bouceba, T., Huet, A., and Li, Z. (2009). Laminin receptor involvement in the anti-angiogenic activity of pigment epithelium-derived factor. *J. Biol. Chem.* 284, 10480–10490. doi:10.1074/jbc.M809259200
- Bottardi, S., Guieze, R., Bourgoin, V., Fotouhi-Ardakani, N., Dougé, A., Darracq, A., et al. (2020). MND4 controls the expression of MCL-1 and BCL-2 in chronic lymphocytic leukemia cells. *Exp. Hematol.* 88, 68–82. doi:10.1016/j.exphem.2020.07.004
- Bravo García-Morato, M., Calvo Apalategi, A., Bravo-Gallego, L. Y., Blázquez Moreno, A., Simón-Fuentes, M., Garmendia, J. V., et al. (2019). Impaired control of multiple viral infections in a family with complete IRF9 deficiency. *J. Allergy Clin. Immunol.* 144, 309–312. doi:10.1016/j.jaci.2019.02.019
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi:10.1023/a:1010933404324
- Briggs, R. C., Briggs, J. A., Ozer, J., Sealy, L., Dworkin, L. L., Kingsmore, S. F., et al. (1994). The human myeloid cell nuclear differentiation antigen gene is one of at least two related interferon-inducible genes located on chromosome 1q that are expressed specifically in hematopoietic cells. *Blood* 83, 2153–2162. doi:10.1182/blood.v83.8.2153.bloodjournal8382153
- Bruneel, A., Labas, V., Mailloux, A., Sharma, S., Royer, N., Vinh, J., et al. (2005). Proteomics of human umbilical vein endothelial cells applied to etoposide-induced apoptosis. *Proteomics* 5, 3876–3884. doi:10.1002/pmic.200401239
- Cao, X. (2021). ISG15 secretion exacerbates inflammation in SARS-CoV-2 infection. *Nat. Immunol.* 22, 1360–1362. doi:10.1038/s41590-021-01056-3
- Castro Dopico, X., Ols, S., Loré, K., and Karlsson Hedestam, G. B. (2022). Immunity to SARS-CoV-2 induced by infection or vaccination. *J. Intern. Med.* 291, 32–50. doi:10.1111/joim.13372
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* 16, 321–357. doi:10.1613/jair.953

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/fgene.2023.1157305/full#supplementary-material>

Chen, J. Y., Chen, W. N., Poon, K. M., Zheng, B. J., Lin, X., Wang, Y. X., et al. (2009a). Interaction between SARS-CoV helicase and a multifunctional cellular protein (Ddx5) revealed by yeast and mammalian cell two-hybrid systems. *Arch. Virol.* 154, 507–512. doi:10.1007/s00705-009-0323-y

Chen, Q., Gupta, S., and Pernis, A. B. (2009b). Regulation of TLR4-mediated signaling by IBP/Def6, a novel activator of Rho GTPases. *J. Leukoc. Biol.* 85, 539–543. doi:10.1189/jlb.0308219

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

Cobaleda, C., Schebesta, A., Delogu, A., and Busslinger, M. (2007). Pax5: The guardian of B cell identity and function. *Nat. Immunol.* 8, 463–470. doi:10.1038/ni1454

Dardenne, E., Polay Espinoza, M., Fattet, L., Germann, S., Lambert, M. P., Neil, H., et al. (2014). RNA helicases DDX5 and DDX17 dynamically orchestrate transcription, miRNA, and splicing programs in cell differentiation. *Cell Rep.* 7, 1900–1913. doi:10.1016/j.celrep.2014.05.010

De Picciotto, S., Devita, N., Hsiao, C. J., Honan, C., Tse, S. W., Nguyen, M., et al. (2022). Selective activation and expansion of regulatory T cells using lipid encapsulated mRNA encoding a long-acting IL-2 mutein. *Nat. Commun.* 13, 3866. doi:10.1038/s41467-022-31130-9

De Veer, M. J., Sim, H., Whisstock, J. C., Devenish, R. J., and Ralph, S. J. (1998). IFI60/ISG60/IFIT4, a new member of the human IFI54/IFIT2 family of interferon-stimulated genes. *Genomics* 54, 267–277. doi:10.1006/geno.1998.5555

Deng, Z., Ng, C., Inoue, K., Chen, Z., Xia, Y., Hu, X., et al. (2020). Def6 regulates endogenous type-I interferon responses in osteoblasts and suppresses osteogenesis. *Elife* 9, e59659. doi:10.7554/eLife.59659

Dersh, D., Holly, J., and Yewdell, J. W. (2021). A few good peptides: MHC class I-based cancer immunosurveillance and immunoevasion. *Nat. Rev. Immunol.* 21, 116–128. doi:10.1038/s41577-020-0390-6

Desterke, C., Turhan, A. G., Bennaceur-Griscelli, A., and Griscelli, F. (2021). HLA-dependent heterogeneity and macrophage immunoproteasome activation during lung COVID-19 disease. *J. Transl. Med.* 19, 290. doi:10.1186/s12967-021-02965-5

Dramiński, M., and Koronacki, J. (2018). rmcfs: An R package for Monte Carlo feature selection and interdependency discovery. *J. Stat. Softw.* 85, 1–28. doi:10.18637/jss.v085.i12

Fabbri, E., Borgatti, M., Montagner, G., Bianchi, N., Finotti, A., Lampronti, I., et al. (2014). Expression of microRNA-93 and Interleukin-8 during *Pseudomonas aeruginosa*-mediated induction of proinflammatory responses. *Am. J. Respir. Cell Mol. Biol.* 50, 1144–1155. doi:10.1165/rcmb.2013-0160OC

Fabricius, D., Ludwig, C., Scholz, J., Rode, I., Tsamadou, C., Jacobsen, E. M., et al. (2021). mRNA vaccines enhance neutralizing immunity against SARS-CoV-2 variants in convalescent and ChAdOx1-primed subjects. *Vaccines (Basel)* 9, 918. doi:10.3390/vaccines9080918

Fang, Q., Li, T., Chen, P., Wu, Y., Wang, T., Mo, L., et al. (2021). Comparative analysis on abnormal methylation of differentially expressed genes and disease pathways in the immune cells of RA and SLE. *Front. Immunol.* 12, 668007. doi:10.3389/fimmu.2021.668007

Feng, B., Zhang, Q., Wang, J., Dong, H., Mu, X., Hu, G., et al. (2018). IFIT1 expression patterns induced by H9N2 virus and inactivated viral particle in human umbilical vein endothelial cells and bronchus epithelial cells. *Mol. Cells* 41, 271–281. doi:10.14348/molcells.2018.2091

- Feng, L., Wang, Q., Shan, C., Yang, C., Feng, Y., Wu, J., et al. (2020). An adenovirus-vectored COVID-19 vaccine confers protection from SARS-CoV-2 challenge in rhesus macaques. *Nat. Commun.* 11, 4207. doi:10.1038/s41467-020-18077-5
- Fiolet, T., Kherabi, Y., Macdonald, C. J., Ghosn, J., and Peiffer-Smadja, N. (2022). Comparing COVID-19 vaccines for their characteristics, efficacy and effectiveness against SARS-CoV-2 and variants of concern: A narrative review. *Clin. Microbiol. Infect.* 28, 202–221. doi:10.1016/j.cmi.2021.10.005
- Fisher, A., Rudin, C., and Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *J. Mach. Learn. Res.* 20, 177–181.
- Fleith, R. C., Mears, H. V., Leong, X. Y., Sanford, T. J., Emmott, E., Graham, S. C., et al. (2018). IFIT3 and IFIT2/3 promote IFIT1-mediated translation inhibition by enhancing binding to non-self RNA. *Nucleic Acids Res.* 46, 5269–5285. doi:10.1093/nar/gky191
- Fournier, B., Tusseau, M., Villard, M., Malcus, C., Chopin, E., Martin, E., et al. (2021). DEF6 deficiency, a mendelian susceptibility to EBV infection, lymphoma, and autoimmunity. *J. Allergy Clin. Immunol.* 147, 740–743.e9. doi:10.1016/j.jaci.2020.05.052
- Freitas, B. T., Scholte, F. E. M., Bergeron, É., and Pegan, S. D. (2020). How ISG15 combats viral infection. *Virus Res.* 286, 198036. doi:10.1016/j.virusres.2020.198036
- Gao, X., Liu, Y., Zou, S., Liu, P., Zhao, J., Yang, C., et al. (2021). Genome-wide screening of SARS-CoV-2 infection-related genes based on the blood leukocytes sequencing data set of patients with COVID-19. *J. Med. Virol.* 93, 5544–5554. doi:10.1002/jmv.27093
- Gao, Y., Cai, C., Wullimann, D., Niessl, J., Rivera-Ballesteros, O., Chen, P., et al. (2022). Immunodeficiency syndromes differentially impact the functional profile of SARS-CoV-2-specific T cells elicited by mRNA vaccination. *Immunity* 55, 1732–1746.e5. doi:10.1016/j.immuni.2022.07.005
- Givant-Horwitz, V., Davidson, B., and Reich, R. (2004). Laminin-induced signaling in tumor cells: The role of the M(r) 67,000 laminin receptor. *Cancer Res.* 64, 3572–3579. doi:10.1158/0008-5472.CAN-03-3424
- Godfrey, D. I., Masantonio, M., Tucek, C. L., Malin, M. A., Boyd, R. L., and Hugo, P. (1992). Thymic shared antigen-1. A novel thymocyte marker discriminating immature from mature thymocyte subsets. *J. Immunol.* 148, 2006–2011. doi:10.4049/jimmunol.148.7.2006
- Goh, P. Y., Tan, Y. J., Lim, S. P., Tan, Y. H., Lim, S. G., Fuller-Pace, F., et al. (2004). Cellular RNA helicase p68 relocalization and interaction with the hepatitis C virus (HCV) NS5B protein and the potential role of p68 in HCV RNA replication. *J. Virol.* 78, 5288–5298. doi:10.1128/jvi.78.10.5288-5298.2004
- Gorodkin, J. (2004). Comparing two K-category assignments by a K-category correlation coefficient. *Comput. Biol. Chem.* 28, 367–374. doi:10.1016/j.compbiolchem.2004.09.006
- Guan, W. J., Ni, Z. Y., Hu, Y., Liang, W. H., Ou, C. Q., He, J. X., et al. (2020). Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* 382, 1708–1720. doi:10.1056/NEJMoa2002032
- Haas, E. J., Angulo, F. J., McLaughlin, J. M., Anis, E., Singer, S. R., Khan, F., et al. (2021). Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalizations, and deaths following a nationwide vaccination campaign in Israel: An observational study using national surveillance data. *Lancet* 397, 1819–1829. doi:10.1016/S0140-6736(21)00947-8
- Halfmann, P. J., Nakajima, N., Sato, Y., Takahashi, K., Accola, M., Chiba, S., et al. (2022). SARS-CoV-2 interference of influenza virus replication in Syrian hamsters. *J. Infect. Dis.* 225, 282–286. doi:10.1093/infdis/jiab587
- Han, X., Xu, P., and Ye, Q. (2021). Analysis of COVID-19 vaccines: Types, thoughts, and application. *J. Clin. Lab. Anal.* 35, e23937. doi:10.1002/jcla.23937
- Harris, R., Randle, S., and Laman, H. (2021). Analysis of the FBXO7 promoter reveals overlapping Pax5 and c-Myb binding sites functioning in B cells. *Biochem. Biophys. Res. Commun.* 554, 41–48. doi:10.1016/j.bbrc.2021.03.052
- Hasankhani, A., Bahrami, A., Sheybani, N., Aria, B., Hemati, B., Fatehi, F., et al. (2021). Differential Co-expression network analysis reveals key hub-high traffic genes as potential therapeutic targets for COVID-19 pandemic. *Front. Immunol.* 12, 789317. doi:10.3389/fimmu.2021.789317
- Hashemi, V., Masjedi, A., Hazhir-Karzar, B., Tanomand, A., Shotorbani, S. S., Hojjat-Farsangi, M., et al. (2019). The role of DEAD-box RNA helicase p68 (DDX5) in the development and treatment of breast cancer. *J. Cell Physiol.* 234, 5478–5487. doi:10.1002/jcp.26912
- He, Q., Mao, Q., An, C., Zhang, J., Gao, F., Bian, L., et al. (2021). Heterologous prime-boost: Breaking the protective immune response bottleneck of COVID-19 vaccine candidates. *Emerg. Microbes Infect.* 10, 629–637. doi:10.1080/22221751.2021.1902245
- Henarejos-Castillo, I., Sebastian-Leon, P., Devesa-Peiro, A., Pellicer, A., and Diaz-Gimeno, P. (2020). SARS-CoV-2 infection risk assessment in the endometrium: Viral infection-related gene expression across the menstrual cycle. *Fertil. Steril.* 114, 223–232. doi:10.1016/j.fertnstert.2020.06.026
- Hernandez, N., Melki, I., Jing, H., Habib, T., Huang, S. S. Y., Danielson, J., et al. (2018). Life-threatening influenza pneumonitis in a child with inherited IRF9 deficiency. *J. Exp. Med.* 215, 2567–2585. doi:10.1084/jem.20180628
- Hillus, D., Schwarz, T., Tober-Lau, P., Vanshylla, K., Hastor, H., Thibeault, C., et al. (2021). Safety, reactogenicity, and immunogenicity of homologous and heterologous prime-boost immunisation with ChAdOx1 nCoV-19 and BNT162b2: A prospective cohort study. *Lancet Respir. Med.* 9, 1255–1265. doi:10.1016/S2213-2600(21)00357-X
- Hu, M., Zheng, H., Wu, J., Sun, Y., Wang, T., and Chen, S. (2022). DDX5: An expectable treater for viral infection—a literature review. *Ann. Transl. Med.* 10, 712. doi:10.21037/atm-22-2375
- Huang, F., Fu, M., Li, J., Chen, L., Feng, K., Huang, T., et al. (2023a). Analysis and prediction of protein stability based on interaction network, gene ontology, and KEGG pathway enrichment scores. *BBA - Proteins Proteomics* 1871, 140889. doi:10.1016/j.bbapap.2023.140889
- Huang, F., Ma, Q., Ren, J., Li, J., Wang, F., Huang, T., et al. (2023b). Identification of smoking associated transcriptome aberration in blood with machine learning methods. *BioMed Res. Int.* 2023, 5333361. doi:10.1155/2023/5333361
- Huang, J., and Zhou, Q. (2022). Identification of the relationship between hub genes and immune cell infiltration in vascular endothelial cells of proliferative diabetic retinopathy using bioinformatics methods. *Dis. Markers* 2022, 7231046. doi:10.1155/2022/7231046
- Izawa, K., Martin, E., Soudais, C., Bruneau, J., Boutboul, D., Rodriguez, R., et al. (2017). Inherited CD70 deficiency in humans reveals a critical role for the CD70-CD27 pathway in immunity to Epstein-Barr virus infection. *J. Exp. Med.* 214, 73–89. doi:10.1084/jem.20160784
- Jean-Baptiste, V. S. E., Xia, C. Q., Clare-Salzler, M. J., and Horwitz, M. S. (2017). Type 1 diabetes and type 1 interferonopathies: Localization of a type 1 common thread of virus infection in the pancreas. *EBioMedicine* 22, 10–17. doi:10.1016/j.ebiom.2017.06.014
- Jiang, Y., Yan, Q., Liu, C. X., Peng, C. W., Zheng, W. J., Zhuang, H. F., et al. (2022). Insights into potential mechanisms of asthma patients with COVID-19: A study based on the gene expression profiling of bronchoalveolar lavage fluid. *Comput. Biol. Med.* 146, 105601. doi:10.1016/j.compbiomed.2022.105601
- Johnson, K., Shapiro-Shelef, M., Tunyaplin, C., and Calame, K. (2005). Regulatory events in early and late B-cell differentiation. *Mol. Immunol.* 42, 749–761. doi:10.1016/j.molimm.2004.06.039
- Jung, H. E., Oh, J. E., and Lee, H. K. (2019). Cell-penetrating Mx1 enhances anti-viral resistance against mucosal influenza viral infection. *Viruses* 11, 109. doi:10.3390/v11020109
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Adv. neural Inf. Process. Syst.* 30, 3146–3154.
- Kim, H. K., Xu, J., Chu, K., Park, H., Jang, H., Li, P., et al. (2019). A tRNA-derived small RNA regulates ribosomal protein S28 protein levels after translation initiation in humans and mice. *Cell Rep.* 29, 3816–3824. doi:10.1016/j.celrep.2019.11.062
- Kobayashi, M., Oshima, S., Maeyashiki, C., Nibe, Y., Otsubo, K., Matsuzawa, Y., et al. (2016). The ubiquitin hybrid gene UBA52 regulates ubiquitination of ribosome and sustains embryonic development. *Sci. Rep.* 6, 36780. doi:10.1038/srep36780
- Kohavi, R. (1995). "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the 14th international joint conference on Artificial intelligence - volume 2* (Montreal, Quebec, Canada: Morgan Kaufmann Publishers Inc.).
- Lane, D. P., and Hoeffler, W. K. (1980). SV40 large T shares an antigenic determinant with a cellular protein of molecular weight 68,000. *Nature* 288, 167–170. doi:10.1038/288167a0
- Lazear, H. M., Schoggins, J. W., and Diamond, M. S. (2019). Shared and distinct functions of type I and type III interferons. *Immunity* 50, 907–923. doi:10.1016/j.immuni.2019.03.025
- Legrand, J. M. D., Chan, A. L., La, H. M., Rossello, F. J., nkö, M. L., Fuller-Pace, F. V., et al. (2019). DDX5 plays essential transcriptional and post-transcriptional roles in the maintenance and function of spermatogonia. *Nat. Commun.* 10, 2278. doi:10.1038/s41467-019-09972-7
- Leng, L., Li, M., Li, W., Mou, D., Liu, G., Ma, J., et al. (2021). Sera proteomic features of active and recovered COVID-19 patients: Potential diagnostic and prognostic biomarkers. *Signal Transduct. Target Ther.* 6, 216. doi:10.1038/s41392-021-00612-5
- Leong, W. F., and Chow, V. T. (2006). Transcriptomic and proteomic analyses of rhabdomyosarcoma cells reveal differential cellular gene expression in response to enterovirus 71 infection. *Cell Microbiol.* 8, 565–580. doi:10.1111/j.1462-5822.2005.00644.x
- Li, H., Huang, F., Liao, H., Li, Z., Feng, K., Huang, T., et al. (2022a). Identification of COVID-19-specific immune markers using a machine learning method. *Front. Mol. Biosci.* 9, 952626. doi:10.3389/fmolb.2022.952626
- Li, H., Zhang, S., Chen, L., Pan, X., Li, Z., Huang, T., et al. (2022b). Identifying functions of proteins in mice with functional embedding features. *Front. Genet.* 13, 909040. doi:10.3389/fgene.2022.909040
- Li, J., Huang, F., Ma, Q., Guo, W., Feng, K., Huang, T., et al. (2023). Identification of genes related to immune enhancement caused by heterologous ChAdOx1-BNT162b2 vaccines in lymphocytes at single-cell resolution with machine learning methods. *Front. Immunol.* 14. doi:10.3389/fimmu.2023.1131051

- Li, Z., Mei, Z., Ding, S., Chen, L., Li, H., Feng, K., et al. (2022c). Identifying methylation signatures and rules for COVID-19 with machine learning methods. *Front. Mol. Biosci.* 9, 908080. doi:10.3389/fmolb.2022.908080
- Liang, H., Chen, L., Zhao, X., and Zhang, X. (2020). Prediction of drug side effects with a refined negative sample selection strategy. *Comput. Math. Methods Med.* 2020, 1573543. doi:10.1155/2020/1573543
- Liu, H. A., and Setiono, R. (1998). Incremental feature selection. *Appl. Intell.* 9, 217–230. doi:10.1023/a:1008363719778
- Liu, H., Su, L., Zhu, T., Zhu, X., Zhu, Y., Peng, Y., et al. (2022). Comparative analysis on proteomics profiles of intracellular and extracellular M.tb and BCG from infected human macrophages. *Front. Genet.* 13, 847838. doi:10.3389/fgene.2022.847838
- Liu, M., Li, N., Guo, W., Jia, L., Jiang, H., Li, Z., et al. (2021a). RPSA distribution and expression in tissues and immune cells of pathogen-infected mice. *Microb. Pathog.* 152, 104609. doi:10.1016/j.micpath.2020.104609
- Liu, X., Shaw, R. H., Stuart, A. S. V., Greenland, M., Aley, P. K., Andrews, N. J., et al. (2021b). Safety and immunogenicity of heterologous versus homologous prime-boost schedules with an adenoviral vectored and mRNA COVID-19 vaccine (Com-COV): A single-blind, randomised, non-inferiority trial. *Lancet* 398, 856–869. doi:10.1016/S0140-6736(21)01694-9
- Lopez Bernal, J., Andrews, N., Gower, C., Gallagher, E., Simmons, R., Thelwall, S., et al. (2021). Effectiveness of covid-19 vaccines against the B.1.617.2 (delta) variant. *N. Engl. J. Med.* 385, 585–594. doi:10.1056/NEJMoa2108891
- Lu, J., Li, J., Ren, J., Ding, S., Zeng, Z., Huang, T., et al. (2022). Functional and embedding feature analysis for pan-cancer classification. *Front. Oncol.* 12, 979336. doi:10.3389/fonc.2022.979336
- Ma, Z., Qu, B., Yao, L., Gao, Z., and Zhang, S. (2020). Identification and functional characterization of ribosomal protein S23 as a new member of antimicrobial protein. *Dev. Comp. Immunol.* 110, 103730. doi:10.1016/j.dci.2020.103730
- Mabrouk, M. T., Huang, W. C., Martinez-Sobrido, L., and Lovell, J. F. (2022). Advanced materials for SARS-CoV-2 vaccines. *Adv. Mater.* 34, e2107781. doi:10.1002/adma.202107781
- Mahase, E. (2021). Covid-19: Vaccine brands can be mixed in "extremely rare occasions," says Public Health England. *says Public Health Engl. Bmj* 372, n12. doi:10.1136/bmj.n12
- Makino, S., Sasaki, K., Nakamura, N., Nakagawa, M., and Nakajima, S. (1974). Studies on the modification of the live AIK measles vaccine. II. Development and evaluation of the live AIK-C measles vaccine. *Kitasato Arch. Exp. Med.* 47, 13–21.
- Mao, J., O'gorman, C., Sutovsky, M., Zigo, M., Wells, K. D., and Sutovsky, P. (2018). Ubiquitin A-52 residue ribosomal protein fusion product 1 (Uba52) is essential for preimplantation embryo development. *Biol. Open* 7, bio035717. doi:10.1242/bio.035717
- Mar, K. B., Rinkenberger, N. R., Boys, I. N., Eitson, J. L., McDougall, M. B., Richardson, R. B., et al. (2018). LY6E mediates an evolutionarily conserved enhancement of virus infection by targeting a late entry step. *Nat. Commun.* 9, 3603. doi:10.1038/s41467-018-06000-y
- Martin-Sancho, L., Lewinski, M. K., Pache, L., Stoneham, C. A., Yin, X., Becker, M. E., et al. (2021). Functional landscape of SARS-CoV-2 cellular restriction. *Mol. Cell* 81, 2656–2668.e8. doi:10.1016/j.molcel.2021.04.008
- Maruyama, T., Nara, K., Yoshikawa, H., and Suzuki, N. (2007). Txk, a member of the non-receptor tyrosine kinase of the Tec family, forms a complex with poly(ADP-ribose) polymerase 1 and elongation factor 1alpha and regulates interferon-gamma gene transcription in Th1 cells. *Clin. Exp. Immunol.* 147, 164–175. doi:10.1111/j.1365-2249.2006.03249.x
- Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta* 405, 442–451. doi:10.1016/0005-2795(75)90109-9
- Metcalf, R. A., Monabati, A., Vyas, M., Roncador, G., Gualco, G., Bacchi, C. E., et al. (2014). Myeloid cell nuclear differentiation antigen is expressed in a subset of marginal zone lymphomas and is useful in the differential diagnosis with follicular lymphoma. *Hum. Pathol.* 45, 1730–1736. doi:10.1016/j.humpath.2014.04.004
- Metz, P., Reuter, A., Bender, S., and Bartenschlager, R. (2013). Interferon-stimulated genes and their role in controlling hepatitis C virus. *J. Hepatol.* 59, 1331–1341. doi:10.1016/j.jhep.2013.07.033
- Mills, A., and Gago, F. (2021). On the need to tell apart fraternal twins eEF1A1 and eEF1A2, and their respective outfits. *Int. J. Mol. Sci.* 22, 6973. doi:10.3390/ijms22136973
- Milot, E., Fotouhi-Ardakani, N., and Filep, J. G. (2012). Myeloid nuclear differentiation antigen, neutrophil apoptosis and sepsis. *Front. Immunol.* 3, 397. doi:10.3389/fimmu.2012.00397
- Miranda, R. N., Briggs, R. C., Shults, K., Kinney, M. C., Jensen, R. A., and Cousar, J. B. (1999). Immunocytochemical analysis of MND1A in tissue sections and sorted normal bone marrow cells documents expression only in maturing normal and neoplastic myelomonocytic cells and a subset of normal and neoplastic B lymphocytes. *Hum. Pathol.* 30, 1040–1049. doi:10.1016/s0046-8177(99)90221-6
- Mizumura, K., Hashimoto, S., Maruoka, S., Gon, Y., Kitamura, N., Matsumoto, K., et al. (2003). Role of mitogen-activated protein kinases in influenza virus induction of prostaglandin E2 from arachidonic acid in bronchial epithelial cells. *Clin. Exp. Allergy* 33, 1244–1251. doi:10.1046/j.1365-2222.2003.01750.x
- Morales, D. J., and Lenschow, D. J. (2013). The antiviral activities of ISG15. *J. Mol. Biol.* 425, 4995–5008. doi:10.1016/j.jmb.2013.09.041
- More, S., Zhu, Z., Lin, K., Huang, C., Pushparaj, S., Liang, Y., et al. (2019). Long non-coding RNA PSM88-AS1 regulates influenza virus replication. *RNA Biol.* 16, 340–353. doi:10.1080/15476286.2019.1572448
- Mouneimne, G., Hansen, S. D., Selfors, L. M., Petrak, L., Hickey, M. M., Gallegos, L. L., et al. (2012). Differential remodeling of actin cytoskeleton architecture by profilin isoforms leads to distinct effects on cell migration and invasion. *Cancer Cell* 22, 615–630. doi:10.1016/j.ccr.2012.09.027
- Mullighan, C. G., Goorha, S., Radtke, I., Miller, C. B., Coustan-Smith, E., Dalton, J. D., et al. (2007). Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 446, 758–764. doi:10.1038/nature05690
- Nouailles, G., Adler, J. M., Pennitz, P., Peidli, S., Alves, G. T., Baumgart, M., et al. (2022). A live attenuated vaccine confers superior mucosal and systemic immunity to SARS-CoV-2 variants. *bioRxiv* 2005, 492138.
- Nutt, S. L., and Tarlinton, D. M. (2011). Germinal center B and follicular helper T cells: Siblings, cousins or just good friends? *Nat. Immunol.* 12, 472–477. doi:10.1038/ni.2019
- O'donohue, M. F., Choesmel, V., Faubladier, M., Fichant, G., and Gleizes, P. E. (2010). Functional dichotomy of ribosomal proteins during the synthesis of mammalian 40S ribosomal subunits. *J. Cell Biol.* 190, 853–866. doi:10.1083/jcb.201005117
- Oberhardt, V., Luxemburger, H., Kemming, J., Schulien, I., Ciminski, K., Giese, S., et al. (2021). Rapid and stable mobilization of CD8(+) T cells by SARS-CoV-2 mRNA vaccine. *Nature* 597, 268–273. doi:10.1038/s41586-021-03841-4
- Okamura, S., and Ebina, H. (2021). Could live attenuated vaccines better control COVID-19? *Vaccine* 39, 5719–5726. doi:10.1016/j.vaccine.2021.08.018
- Okumura, F., Zou, W., and Zhang, D. E. (2007). ISG15 modification of the eIF4E cognate 4EHP enhances cap structure-binding activity of 4EHP. *Genes Dev.* 21, 255–260. doi:10.1101/gad.1521607
- Parasher, A. (2021). COVID-19: Current understanding of its pathophysiology, clinical presentation and treatment. *Postgrad. Med. J.* 97, 312–320. doi:10.1136/postgradmedj-2020-138577
- Parks, C. L., Lerch, R. A., Walpita, P., Wang, H. P., Sidhu, M. S., and Udem, S. A. (2001). Comparison of predicted amino acid sequences of measles virus strains in the Edmonston vaccine lineage. *J. Virol.* 75, 910–920. doi:10.1128/JVI.75.2.910-920.2001
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Analysis Mach. Intell.* 27, 1226–1238. doi:10.1109/TPAMI.2005.159
- Perng, Y. C., and Lenschow, D. J. (2018). ISG15 in antiviral immunity and beyond. *Nat. Rev. Microbiol.* 16, 423–439. doi:10.1038/s41579-018-0020-5
- Pfaender, S., Mar, K. B., Michailidis, E., Kratzel, A., Boys, I. N., V'kovski, P., et al. (2020). LY6E impairs coronavirus fusion and confers immune control of viral disease. *Nat. Microbiol.* 5, 1330–1339. doi:10.1038/s41564-020-0769-y
- Pidugu, V. K., Pidugu, H. B., Wu, M. M., Liu, C. J., and Lee, T. C. (2019). Emerging functions of human IFIT proteins in cancer. *Front. Mol. Biosci.* 6, 148. doi:10.3389/fmolb.2019.00148
- Pisano, M. P., Grandi, N., and Tramontano, E. (2021). Human endogenous retroviruses (HERVs) and mammalian apparent LTRs retrotransposons (MaLRs) are dynamically modulated in different stages of immunity. *Biol. (Basel)* 10, 405. doi:10.3390/biology10050405
- Powers, D. (2011). Evaluation: From precision, recall and f-measure to roc, informedness, markedness & correlation. *J. Mach. Learn. Technol.* 2, 37–63.
- Ran, B., Chen, L., Li, M., Han, Y., and Dai, Q. (2022). Drug-Drug interactions prediction using fingerprint only. *Comput. Math. Methods Med.* 2022, 7818480. doi:10.1155/2022/7818480
- Ranstrom, J., and Cook, J. (2018). LASSO regression. *J. Br. Surg.* 105, 1348. doi:10.1002/bjs.10895
- Rea, V. E., Rossi, F. W., De Paulis, A., Ragno, P., Selleri, C., and Montuori, N. (2012). 67 kDa laminin receptor: Structure, function and role in cancer and infection. *Infez. Med.* 20 (2), 8–12.
- Rossi, D., Deaglio, S., Dominguez-Sola, D., Rasi, S., Vaisitti, T., Agostinelli, C., et al. (2011). Alteration of BIRC3 and multiple other NF- κ B pathway genes in splenic marginal zone lymphoma. *Blood* 118, 4930–4934. doi:10.1182/blood-2011-06-359166
- Safavian, S. R., and Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man, Cybern.* 21, 660–674. doi:10.1109/21.97458
- Schoggins, J. W., Wilson, S. J., Panis, M., Murphy, M. Y., Jones, C. T., Bieniasz, P., et al. (2011). A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* 472, 481–485. doi:10.1038/nature09907

- Senapati, S., Kumar, S., Singh, A. K., Banerjee, P., and Bhagavatula, S. (2020). Assessment of risk conferred by coding and regulatory variations of TMPRSS2 and CD26 in susceptibility to SARS-CoV-2 infection in human. *J. Genet.* 99, 53. doi:10.1007/s12041-020-01217-7
- Servaas, N. H., Mariotti, B., Van Der Kroef, M., Wichers, C. G. K., Pandit, A., Bazzoni, F., et al. (2021). Characterization of long non-coding RNAs in systemic sclerosis monocytes: A potential role for PSMB8-AS1 in altered cytokine secretion. *Int. J. Mol. Sci.* 22, 4365. doi:10.3390/ijms22094365
- Sette, A., and Crotty, S. (2022). Immunological memory to SARS-CoV-2 infection and COVID-19 vaccines. *Immunol. Rev.* 310, 27–46. doi:10.1111/imr.13089
- Shaath, H., Vishnubalaji, R., Elkord, E., and Alajez, N. M. (2020). Single-cell transcriptome analysis highlights a role for neutrophils and inflammatory macrophages in the pathogenesis of severe COVID-19. *Cells* 9, 2374. doi:10.3390/cells9112374
- Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., et al. (2020). Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell* 182, 59–72. doi:10.1016/j.cell.2020.05.032
- Shin, D., Mukherjee, R., Grewe, D., Bojkova, D., Baek, K., Bhattacharya, A., et al. (2020). Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* 587, 657–662. doi:10.1038/s41586-020-2601-5
- Sikora, D., Greco-Stewart, V. S., Miron, P., and Pelchat, M. (2009). The hepatitis delta virus RNA genome interacts with eEF1A1, p54(nrb), hnRNP-L, GAPDH and ASF/SF2. *Virology* 390, 71–78. doi:10.1016/j.virol.2009.04.022
- Stark, G. R., and Darnell, J. E., Jr. (2012). The JAK-STAT pathway at twenty. *Immunity* 36, 503–514. doi:10.1016/j.immuni.2012.03.013
- Strong, J. E., Wong, G., Jones, S. E., Grolla, A., Theriault, S., Kobinger, G. P., et al. (2008). Stimulation of Ebola virus production from persistent infection through activation of the Ras/MAPK pathway. *Proc. Natl. Acad. Sci. U. S. A.* 105, 17982–17987. doi:10.1073/pnas.0809698105
- Sun, Q., Li, N., Jia, L., Guo, W., Jiang, H., Liu, B., et al. (2020). Ribosomal protein SA-positive neutrophil elicits stronger phagocytosis and neutrophil extracellular trap formation and subdues pro-inflammatory cytokine secretion against Streptococcus suis serotype 2 infection. *Front. Immunol.* 11, 585399. doi:10.3389/fimmu.2020.585399
- Swaim, C. D., Canadeo, L. A., Monte, K. J., Khanna, S., Lenschow, D. J., and Huibregtse, J. M. (2020). Modulation of extracellular ISG15 signaling by pathogens and viral effector proteins. *Cell Rep.* 31, 107772. doi:10.1016/j.celrep.2020.107772
- Swaim, C. D., Scott, A. F., Canadeo, L. A., and Huibregtse, J. M. (2017). Extracellular ISG15 signals cytokine secretion through the LFA-1 integrin receptor. *Mol. Cell* 68, 581–590. doi:10.1016/j.molcel.2017.10.003
- Tang, S., and Chen, L. (2022). iATC-NFMLP: Identifying classes of anatomical therapeutic chemicals based on drug networks, fingerprints and multilayer perceptron. *Curr. Bioinforma.* 17, 814–824. doi:10.2174/1574893617666220318093000
- Tiwari, R., Mishra, A. R., Gupta, A., and Nayak, D. (2022). Structural similarity-based prediction of host factors associated with SARS-CoV-2 infection and pathogenesis. *J. Biomol. Struct. Dyn.* 40, 5868–5879. doi:10.1080/07391102.2021.1874532
- Turner, J. S., O'halloran, J. A., Kalaidina, E., Kim, W., Schmitz, A. J., Zhou, J. Q., et al. (2021). SARS-CoV-2 mRNA vaccines induce persistent human germinal centre responses. *Nature* 596, 109–113. doi:10.1038/s41586-021-03738-2
- Urbánek, P., Wang, Z. Q., Fetka, I., Wagner, E. F., and Busslinger, M. (1994). Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. *Cell* 79, 901–912. doi:10.1016/0092-8674(94)90079-5
- V'kovski, P., Gerber, M., Kelly, J., Pfaender, S., Ebert, N., Braga Lagache, S., et al. (2019). Determination of host proteins composing the microenvironment of coronavirus replicase complexes by proximity-labeling. *Elife* 8, e42037. doi:10.7554/eLife.42037
- Vishnubalaji, R., Shaath, H., and Alajez, N. M. (2020). Protein coding and long noncoding RNA (lncRNA) transcriptional landscape in SARS-CoV-2 infected bronchial epithelial cells highlight a role for interferon and inflammatory response. *Genes (Basel)* 11, 760. doi:10.3390/genes11070760
- Wang, H., and Chen, L. (2023). PMPTCE-HNEA: Predicting metabolic pathway types of chemicals and enzymes with a heterogeneous network embedding algorithm. *Curr. Bioinforma.* 18. doi:10.2174/1574893618666230224121633
- Wang, Q., Li, Q., Liu, T., Chang, G., Sun, Z., Gao, Z., et al. (2018). Host interaction analysis of PA-N155 and PA-N182 in chicken cells reveals an essential role of UBA52 for replication of H5N1 avian influenza virus. *Front. Microbiol.* 9, 936. doi:10.3389/fmicb.2018.00936
- Wang, R., and Chen, L. (2022). Identification of human protein subcellular location with multiple networks. *Curr. Proteomics* 19, 344–356. doi:10.2174/1570164619666220531113704
- Wei, J., Kishton, R. J., Angel, M., Conn, C. S., Dalla-Venezia, N., Marcel, V., et al. (2019). Ribosomal proteins regulate MHC class I peptide generation for immunosurveillance. *Mol. Cell* 73, 1162–1173. doi:10.1016/j.molcel.2018.12.020
- Wherry, E. J., and Barouch, D. H. (2022). T cell immunity to COVID-19 vaccines. *Science* 377, 821–822. doi:10.1126/science.add2897
- Wu, C., and Chen, L. (2023). A model with deep analysis on a large drug network for drug classification. *Math. Biosci. Eng.* 20, 383–401. doi:10.3934/mbe.2023018
- Wu, Z., and Chen, L. (2022). Similarity-based method with multiple-feature sampling for predicting drug side effects. *Comput. Math. Methods Med.* 2022, 9547317. doi:10.1155/2022/9547317
- Yang, Y., and Chen, L. (2022). Identification of drug–disease associations by using multiple drug and disease networks. *Curr. Bioinforma.* 17, 48–59. doi:10.2174/1574893616666210825115406
- Yang, Z., Gagarin, D., St Laurent, G., 3rd, Hammell, N., Toma, I., Hu, C. A., et al. (2009). Cardiovascular inflammation and lesion cell apoptosis: A novel connection via the interferon-inducible immunoproteasome. *Arterioscler. Thromb. Vasc. Biol.* 29, 1213–1219. doi:10.1161/ATVBAHA.109.189407
- Yu, J., Liang, C., and Liu, S. L. (2017). Interferon-inducible LY6E protein promotes HIV-1 infection. *J. Biol. Chem.* 292, 4674–4685. doi:10.1074/jbc.M116.755819
- Zhang, H., Xing, Z., Mani, S. K., Bancel, B., Durantel, D., Zoulim, F., et al. (2016). RNA helicase DEAD box protein 5 regulates Polycomb repressive complex 2/Hox transcript antisense intergenic RNA function in Hepatitis B virus infection and hepatocarcinogenesis. *Hepatology* 64, 1033–1048. doi:10.1002/hep.28698
- Zhang, Y.-H., Li, H., Zeng, T., Chen, L., Li, Z., Huang, T., et al. (2021a). Identifying transcriptomic signatures and rules for SARS-CoV-2 infection. *Front. Cell Dev. Biol.* 8, 627302. doi:10.3389/fcell.2020.627302
- Zhang, Y.-H., Zeng, T., Chen, L., Huang, T., and Cai, Y.-D. (2021b). Determining protein–protein functional associations by functional rules based on gene ontology and KEGG pathway. *Biochimica Biophysica Acta (BBA) - Proteins Proteomics* 1869, 140621. doi:10.1016/j.bbapap.2021.140621
- Zhang, Z., Lin, W., Li, X., Cao, H., Wang, Y., and Zheng, S. J. (2015). Critical role of eukaryotic elongation factor 1 alpha 1 (EEF1A1) in avian reovirus sigma-C-induced apoptosis and inhibition of viral growth. *Arch. Virol.* 160, 1449–1461. doi:10.1007/s00705-015-2403-5
- Zhao, X., Chen, L., and Lu, J. (2018). A similarity-based method for prediction of drug side effects with heterogeneous information. *Math. Biosci.* 306, 136–144. doi:10.1016/j.mbs.2018.09.010
- Zhou, X., Liao, W. J., Liao, J. M., Liao, P., and Lu, H. (2015). Ribosomal proteins: Functions beyond the ribosome. *J. Mol. Cell Biol.* 7, 92–104. doi:10.1093/jmcb/mjv014
- Zhou, X., Michal, J. J., Zhang, L., Ding, B., Lunney, J. K., Liu, B., et al. (2013). Interferon induced IFIT family genes in host antiviral defense. *Int. J. Biol. Sci.* 9, 200–208. doi:10.7150/ijbs.5613
- Zonta, E., Bittencourt, D., Samaan, S., Germann, S., Dutertre, M., and Auboeuf, D. (2013). The RNA helicase DDX5/p68 is a key factor promoting c-fos expression at different levels from transcription to mRNA export. *Nucleic Acids Res.* 41, 554–564. doi:10.1093/nar/gks1046