Identification of Genes Associated with the Impairment of Olfactory and Gustatory Functions in COVID-19 via Machine-Learning Methods

Jingxin Ren [1,†], Yuhang Zhang [2,†] (ID), Wei Guo [3], Kaiyan Feng [4], Ye Yuan [5], Tao Huang [6,7,*] (ID) and Yu-Dong Cai [1,*] (ID)

1   School of Life Sciences, Shanghai University, Shanghai 200444, China
2   Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA
3   Key Laboratory of Stem Cell Biology, Shanghai Jiao Tong University School of Medicine (SJTUSM) & Shanghai Institutes for Biological Sciences (SIBS), Chinese Academy of Sciences (CAS), Shanghai 200030, China
4   Department of Computer Science, Guangdong AIB Polytechnic College, Guangzhou 510507, China
5   Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, and Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China
6   Bio-Med Big Data Center, CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China
7   CAS Key Laboratory of Tissue Microenvironment and Tumor, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China
*   Correspondence: huangtao@sibs.ac.cn (T.H.); caiyudong@staff.shu.edu.cn (Y.-D.C.);
     Tel.: +86-21-54923269 (T.H.); +86-21-66136132 (Y.D.C.)
†   These authors contributed equally to this work.

Abstract: The coronavirus disease 2019 (COVID-19), as a severe respiratory disease, affects many parts of the body, and approximately 20–85% of patients exhibit functional impairment of the senses of smell and taste, some of whom even experience the permanent loss of these senses. These symptoms are not life-threatening but severely affect patients' quality of life and increase the risk of depression and anxiety. The pathological mechanisms of these symptoms have not been fully identified. In the current study, we aimed to identify the important biomarkers at the expression level associated with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection-mediated loss of taste or olfactory ability, and we have suggested the potential pathogenetic mechanisms of COVID-19 complications. We designed a machine-learning-based approach to analyze the transcriptome of 577 COVID-19 patient samples, including 84 COVID-19 samples with a decreased ability to taste or smell and 493 COVID-19 samples without impairment. Each sample was represented by 58,929 gene expression levels. The features were analyzed and sorted by three feature selection methods (least absolute shrinkage and selection operator, light gradient boosting machine, and Monte Carlo feature selection). The optimal feature sets were obtained through incremental feature selection using two classification algorithms: decision tree (DT) and random forest (RF). The top genes identified by these multiple methods (H3-5, NUDT5, and AOC1) are involved in olfactory and gustatory impairments. Meanwhile, a high-performance RF classifier was developed in this study, and three sets of quantitative rules that describe the impairment of olfactory and gustatory functions were obtained based on the optimal DT classifiers. In summary, this study provides a new computation analysis and suggests the latent biomarkers (genes and rules) for predicting olfactory and gustatory impairment caused by COVID-19 complications.

Keywords: COVID-19; olfactory function; gustatory function; gene identification; machine learning

1. Introduction

The coronavirus disease 2019, which is also known as COVID-19, is a respiratory infectious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus and has spread worldwide [1]. With more than 500 million confirmed cases

and 6 million deaths, COVID-19 has been one of the most severe and life-threatening infectious diseases [1,2]. In general, the specific symptoms of COVID-19 may appear 2–14 days after infection by the virus, including fever, chills, cough, fatigue, and loss of taste or smell [3,4]. One of the most typical symptoms and the long-term sequela of COVID-19 is a reduced sense of smell or taste (also known as hypogeusia/ageusia or hyposmia/anosmia) [5,6], which has been widely observed in patients with COVID-19 worldwide.

Early in May 2020, researchers from the Guy's and St. Thomas' Hospitals observed the renewed onset of smell and taste disturbances in patients with COVID-19 [5]; of the 382 patients, 11.5% suffered from severe loss of smell after treatment; after 1 week, more than 82% recovered, but 18% reported persistent and complete loss of smell or taste [5]. A follow-up study with systematic analyses conducted in July 2020 on JAMA Otolaryngology Head Neck Surg reported that 12 out of 202 patients had unchanged or worsening loss of taste or smell 4 weeks after their infection [6]. Therefore, a decreased sense of taste or smell in patients with COVID-19 during the course of the disease or even after their recovery is a common complication.

As for the pathological mechanisms underlying the loss of taste or smell in COVID-19, a review conducted by Brazilian researchers summarized the potential interactions between the inflammatory reactions triggered by a viral infection and the olfactory system [7]. According to their analyses, SARS-CoV-2 may target horizontal basal cells, Bowmans' glands, and nonneuronal supporting cells but not olfactory sensory neurons [7]. Given the long recovery time for the ability to taste or smell, transmission from olfactory sensory neurons to the central nervous system may be disrupted by infection. Another review summary of COVID-19 showed reduced smell and taste and speculated that the loss of the dendrite layer of the olfactory sensory neurons (OSN: neurons that sense the odorant information and transmit it to the central neural system) might directly cause temporary or permanent loss of smell (anosmia) in patients with COVID-19 [8]. As for the loss of taste (ageusia), the taste buds, which are the fundamental units for taste, have high expression levels of ACE2, which is the major target for SARS-CoV-2 infection recognition [8]. Therefore, it has been speculated that SARS-CoV-2 may act as a probable source of taste dysfunctions [9].

Approximately 10–20% of patients with COVID-19 suffer from permanent taste or smell loss as a major complication. Some studies have summarized and speculated the potential pathological mechanisms for taste or smell loss. However, analyses were performed at the cell level rather than at the gene level in most reviews. In the present study, according to the transcriptomic profiles of 577 patients with COVID-19, we attempted to identify the key pathogenic regulatory factors of taste or smell loss at the gene level by using multiple machine learning algorithms, including three feature selection methods: least absolute shrinkage and selection operator (LASSO) [10], Monte Carlo feature selection (MCFS) [11], and light gradient boosting machine (LightGBM) [12], and two classification algorithms: decision tree (DT) [13] and random forest (RF) [14]. The feature selection methods were used to deeply analyze the transcriptomic profiles, yielding three feature lists. Then, each list was fed into an incremental feature selection (IFS) [15] method to screen out the essential biomarker genes, construct efficient classifiers and classification rules. Some genes (e.g., H3-5, NUDT5, and AOC1) and rules can be confirmed to be associated with a decreased sense of taste and smell during SARS-CoV-2 infection. This study demonstrated a novel potential pathological mechanism for viral infection complications.

2. Materials and Methods

2.1. Data

Serum transcriptome data from 577 COVID-19-positive patients were obtained from the Gene Expression Omnibus (GEO) database under registration number GSE198449 [16,17]. The patients were divided into two groups according to whether the patients had impaired olfactory and gustatory function. A total of 84 patients showed a reduced sense of taste or smell. Each sample patient had 58,929 gene expression levels.

2.2. Feature Selection Methods

In recent years, several feature selection methods have been proposed to analyze the complicated dataset. Their purpose was to screen out essential features that played important roles in classifying samples. However, each method has advantages and disadvantages. Given a dataset, it is impossible to extract all essential features by using one single feature selection method. To mine essential features as completely as possible, the usage of multiple feature selection methods is a feasible scheme. Thus, we employed three feature selection methods: LASSO [10], MCFS [11], and LightGBM [12] to analyze the investigated transcriptome data mentioned in Section 2.1. These methods have different principles, which were helpful to overview the data from different points of view, thereby extracting more essential features hidden in different ways.

LASSO. Based on the Nonnegative Garrote proposed by Leobreiman [18], the LASSO algorithm was first proposed by Robert Tibshirani [10] in 1996 to obtain a statistical regression model and construct a first-order penalty function. Overfitting can be effectively reduced by regularizing the coefficients of some variables to zero, and features that contribute less to prediction are ignored. The absolute values of the coefficients are proportional to the importance of the features. The ranking of relevant features can be obtained based on such values. This study adopted the LASSO program in SCIKIT-LEARN [19], which was performed with default parameters.

LightGBM. The LightGBM [12] introduces one-sided gradient sampling, exclusive feature bundling, and histogram algorithms compared with the traditional gradient-boosting DT framework. The samples are segmented according to the outputs of an ensemble of classifiers, and the importance of each feature is estimated according to the number of times it is involved in the building of all DTs. LightGBM has a fast training speed and small memory footprint and is suitable for handling large-scale data while ensuring high accuracy. Features can be ranked in a list with the decreasing order of the above times. The LightGBM program sourced from https://lightgbm.readthedocs.io/en/latest/ accessed on 10 May 2020 was used in this study. It was also executed with default parameters.

MCFS. The MCFS was originally developed by Draminski et al. [11]. It selects some features randomly and repeatedly many times to form the p subsets of features. For each subset of features, t trees are constructed by randomly splitting the dataset into a training set and a test set t times. Therefore, $p \times t$ trees are constructed, and their performance is assessed. The importance of each feature is determined according to its involvement in these DT classifiers. A feature is considered important if it is involved in the splitting of tree nodes. This importance is defined as the relative importance (RI) score, which is calculated as follows:

$$RI_g = \sum_{\tau=1}^{p \times t} (\omega A_{CC})^u \sum_{ng(\tau)} IG(ng(\tau)) \left(\frac{no.in\ ng(\tau)}{no.in\ \tau} \right)^v \tag{1}$$

In the formula, ωA_{CC} is the weighted accuracy of the tree τ, $ng(\tau)$ is a node in tree τ, its information gain is denoted as $IG(ng(\tau))$, and $no.in\ ng(\tau)/no.in\ \tau$ denotes the sample size of $ng(\tau)/\tau$; u and v are two positive reals weighting the ωA_{CC} and the ratio $no.in\ ng(\tau)/no.in\ \tau$, respectively. Accordingly, we ranked features in a list according to the decreasing order of their RI scores. Here, we used the MCFS program retrieved from http://www.ipipan.eu/staff/m.draminski/mcfs.html accessed on 4 June 2019. Likewise, it was also executed with default parameters.

Three feature selection methods were applied to the transcriptome data. Each can generate a feature list. For easy descriptions, these lists were called LASSO, LightGBM, and MCFS feature lists.

2.3. Incremental Feature Selection

Although the above three feature selection methods can sort features in lists according to their importance, it is still not adequate to determine the optimal feature subspace for classification. In the present study, IFS [15] was used to complete the task of extracting

optimal subspace. From each feature list, a series of feature subsets were constructed by setting the step size at 5; that is, each feature subset has five more features than the previous subset. For each feature subset, samples were represented by features in this subset, on which a classifier was built with a given classification algorithm. All classifiers were evaluated through cross-validation [20]. By comparing the performance of all classifiers, the classifier with the best performance can be obtained, which is called the optimal classifier. The feature subset of such a classifier was termed the optimal feature subset.

2.4. Synthetic Minority Oversampling Technique

The number of samples in two classes in the dataset differed significantly, which may produce preferences for major classes if the classifier was directly based on such a dataset. To solve this problem, the synthetic minority oversampling technique (SMOTE) method [21] was employed in this study. It randomly selects a sample in the minority class, and Euclidean distances with other samples in the same class are calculated for the determination of k-nearest neighbors. A point lying on the line between the selected sample and one of its k-nearest neighbors is randomly selected as a new sample, which is put into the minority class. This process is repeated for the generation of new samples until a dataset balanced in number is obtained. We used the SMOTE package obtained from https://github.com/scikit-learn-contrib/imbalanced-learn accessed on 24 March 2020. Default parameters were used.

2.5. Classification Algorithm

To execute the IFS method, one classification algorithm is necessary. In this study, we attempted two classification algorithms: DT [13] and RF [14].

Decision tree. The DT algorithm constructs a tree-like structure in which each internal node holds a test on a feature, branches hold the conclusion of the test, and each leaf node holds a class label. A tree is grown according to selected features and conditions used in splitting based on information gain. During classification, starting from the root node, a sample is allocated to a sub-node or child node along the branch that satisfies the test of a node until it reaches a leaf node where a class label is assigned to the sample [13]. A DT can produce a group of classification rules that are easy to interpret and provide insights into biological mechanisms. Each rule represents a path from the root to one leaf. In this study, we used the CART algorithm with node ranking by the Gini coefficient. The program was taken from the SCIKIT-LEARN [19] package, and it was executed with default parameters.

Random forest. The RF algorithm is an ensemble learning based on DT algorithms [14,22–26] and creates a number of independent DT classifiers that do not interfere with one another. These classifiers were constructed by randomly taking samples from the training set and features. By combining the prediction results of all the DTs, the final classification decision is the class label that receives the most votes. As an integrated algorithm, RF tends to have higher accuracy than DT and can effectively prevent overfitting. The RF program in the SCIKIT-LEARN [19] package was used in this study and executed with default parameters.

2.6. Performance Evaluation

In the IFS method, lots of classifiers were set up. These classifiers were evaluated by 10-fold cross-validation. To evaluate the prediction quality, several measurements were adopted in this study. Generally, the F1-measure is a widely used measurement in binary classification [27–30]. The calculation procedure is as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

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

where TP is the true positive, FP is the false positive, and FN is the false negative. Classifier performance increases with F1-measure.

In addition, we further employed two other measurements: prediction accuracy (ACC) and Matthew correlation coefficient (MCC) [31]. ACC is defined as the proportion of correctly predicted samples, and MCC can be computed by

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \tag{5}$$

where TN represents the true negative.

3. Results

The whole workflow of the computational analysis is shown in Figure 1. We screened and extracted the key features to distinguish the COVID-19 samples with olfactory and gustatory impairment from the other COVID-19 samples. The quantitative classification rules were obtained, and a high-performance RF classifier was built. The results of each stage are summarized in this section.
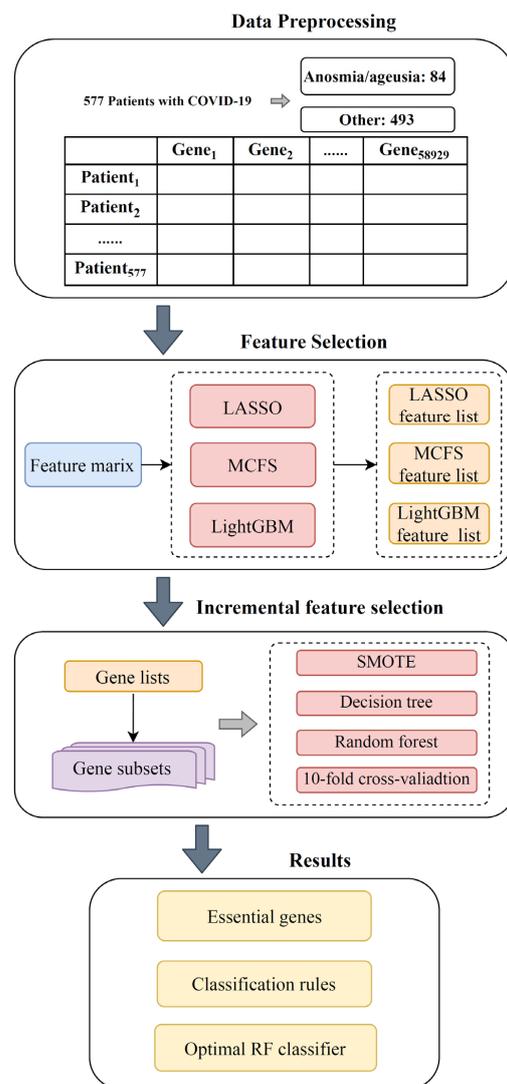

Figure 1. Flow chart of the entire computational analysis. A total of 58,929 gene features from 577 COVID-19 samples were ranked by feature importance using three feature selection methods: LASSO,

LightGBM, and MCFS. The three feature lists were fed into the IFS computational framework, containing two classification algorithms. Finally, on the basis of the IFS results and the white-box algorithm DT, the optimal feature subset, classification rules, and optimal RF classifier were obtained.

3.1. Feature Ranking Results

A total of 577 samples were used in this study, each of which was represented by 58,929 gene expression levels. The full sets of gene features were ranked using three feature selection methods, yielding three feature lists (the LASSO, LightGBM, and MCFS feature lists), which are shown in Table S1. Table 1 shows the top 10 gene features in the three feature lists. These were considered to be the most essential genes.

Table 1. Top 10 gene features in the three feature lists.

Rank	LASSO Feature List	LightGBM Feature List	MCFS Feature List
1	ENSG00000254624.1 (OR4R3P)	ENSG00000225611.1 (CCDC13-AS2)	ENSG00000171360.3 (KRT38)
2	ENSG00000223379.1 (Lnc-SPATA31A7-1)	ENSG00000259515.1 (Antisense to CDC42BPB)	ENSG00000149798.5 (CDC42EP2)
3	ENSG00000168528.12 (SERINC2)	ENSG00000104973.18 (MED25)	ENSG00000124549.14 (BTN2A3P)
4	ENSG00000278931.1 (ANKRD20A3 Pseudogene)	ENSG00000010361.13 (FUZ)	ENSG00000230526.1 (Lnc-CDH23-3)
5	ENSG00000203581.7 (OR1F2P)	ENSG00000163032.11 (VSNL1)	ENSG00000225611.1 (CCDC13-AS2)
6	ENSG00000234134.1 (EDRF1-AS1)	ENSG00000236496.2 (GPS2P1)	ENSG00000010361.13 (FUZ)
7	ENSG00000229596.3 (MYL12BP3)	ENSG00000149798.5 (CDC42EP2)	ENSG00000092054.13 (MYH7)
8	ENSG00000261707.1 (Lnc-ADAMTS18-3)	ENSG00000101544.9 (ADNP2)	ENSG00000131370.16 (SH3BP5)
9	ENSG00000229190.1 (Lnc-HACD1-2)	ENSG00000259674.1 (RPL7AP75)	ENSG00000158710.14 (TAGLN2)
10	ENSG00000183562.3 (Lnc-SLC22A18-2)	ENSG00000165406.16 (MARCHF8)	ENSG00000264204.2 (AGAP7P)

3.2. Results of Incremental Feature Selection

According to the three lists, a number of feature subsets were constructed using the IFS method, and the step size was 5. In order to save time, we used the first 10,000 genes from each list to construct the feature subsets. After balancing the training set with the SMOTE, the RF and DT classifiers were built on each feature subset. The performance of all classifiers was evaluated through 10-fold cross-validation, and the F1-measure was selected as the major measurement. The detailed evaluation results are shown in Table S2. The IFS curves were plotted for visualization, as shown in Figure 2, where the number of features was used as the horizontal co-ordinate, and the F1-measure was used as the vertical co-ordinate.

For the LASSO feature list, the IFS curves of DT and RF are illustrated in Figure 2A. When the top 3510 features in this list were used, the RF yielded the highest F1-measure of 0.916. Accordingly, the optimal RF classifier can be built with these features. The ACC and MCC of such a classifier are listed in Table 2. For DT, its best performance was obtained by using the top 800 features in the list. Then, the optimal DT classifier was built using these features. The F1-measure of this classifier was 0.551. Clearly, such a classifier was far inferior to the optimal RF classifier.
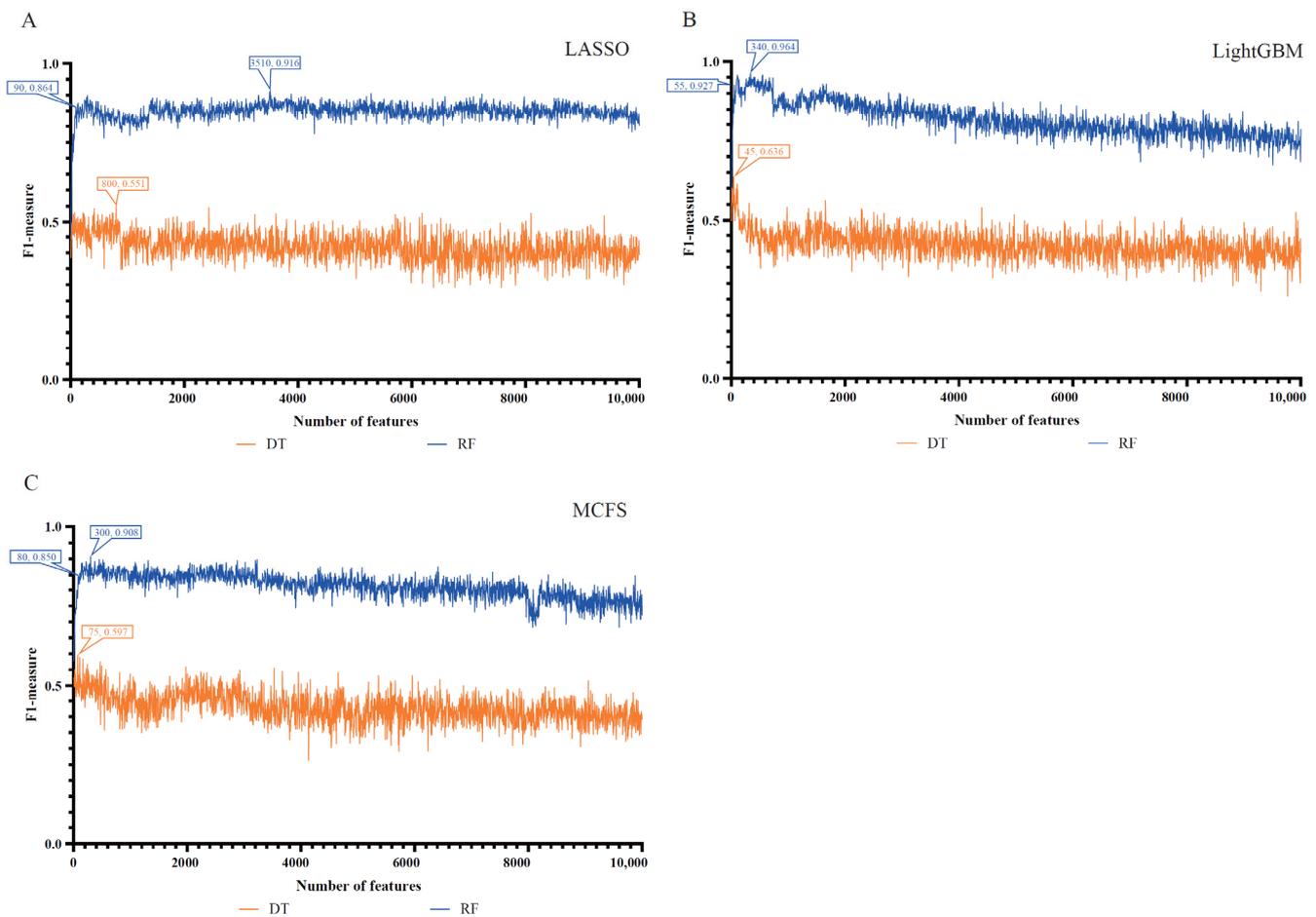
A

B

C

Figure 2. IFS curves of decision tree (DT) and random forest (RF) from different feature lists. (**A**) IFS curves based on the LASSO feature list. (**B**) IFS curves based on the LightGBM feature list. (**C**) IFS curves based on the MCFS feature list. RF classifiers always provide better performance than the DT classifiers.

Table 2. Performance of the optimal and feasible random forest classifiers on the three feature lists.

Feature Lists	Number of Features	F1 Measure	MCC [a]	ACC [b]
LASSO feature	90	0.864	0.841	0.960
list	3510	0.916	0.902	0.976
LightGBM	55	0.927	0.915	0.979
feature list	340	0.964	0.958	0.990
MCFS feature	80	0.850	0.825	0.957
list	300	0.908	0.893	0.974

[a]: Mathew's correlation coefficient; [b]: Prediction accuracy.

For the LightGBM feature list, the two IFS curves are shown in Figure 2B. By using the same operation, the optimal RF and DT classifiers were built using the top 340 and 45 features in the list. They generated an F1-measure of 0.964 and 0.636, respectively. Evidently, the optimal RF classifier was much better than the optimal DT classifier. The detailed performance of the optimal RF classifier is listed in Table 2. Clearly, it yielded better performance when compared to the optimal RF classifier from the LASSO feature list.

For the MCFS feature list, the IFS curves of DT and RF are shown in Figure 2C. RF generated the highest F1-measure (0.908) when the top 300 features were adopted, whereas the highest F1-measure for DT was 0.597 when the top 75 features were used. Thus, the optimal DT and RF classifiers can be set up with corresponding optimal features. Again, the optimal RF classifier was far superior to the optimal DT classifier. The detailed performance

of the optimal RF classifier is listed in Table 2. When comparing the performance of the optimal RF classifiers with the other two feature lists, the optimal classifier from the MCFS feature list was almost equal to that of the LASSO feature list and slightly weaker than that of the LightGBM feature list. Thus, the optimal RF classifiers from the LightGBM feature list can be a latent useful tool to identify COVID-19 samples with olfactory and gustatory impairment from other COVID-19 samples.

3.3. Intersection of Essential Features on Different Feature Lists

In Section 3.2, three optimal RF classifiers were built based on different feature lists. However, the numbers of optimal features used in these classifiers were generally large. Thus, to improve the analytical process, the essential features among these optimal features should be extracted. By checking the IFS results with RF on each feature list, we can find out an RF classifier that used much fewer features and provided a slightly lower performance. From the LASSO feature list, such an RF classifier adopted the top 90 features, which yielded an F1-measure of 0.864. The RF classifiers on the other two feature lists adopted the top 55 (LightGBM feature list) and 80 (MCFS feature list) features. The detailed performance of the above three RF classifiers is listed in Table 2. Evidently, much fewer features were involved in these classifiers. However, their performance was a little lower than that of the optimal RF classifiers. These results indicated that these features were essential among the optimal features. The above RF classifiers were called feasible RF classifiers for convenience. Three feature subsets were constructed, which consisted of those features used by the three feasible RF classifiers. In order to show the relationship between these feature subsets, a Venn diagram was plotted, as shown in Figure 3. The detailed results of the intersection set are provided in Table S3. It can be observed that one gene feature was included in all three of the feature subsets, and 19 gene features belonged to two exact subsets, indicating that these genes were identified to be essential by multiple feature selection methods. Some genes can be confirmed to be associated with the impairment of gustatory and olfactory functions in patients with COVID-19, which is discussed in Section 4.
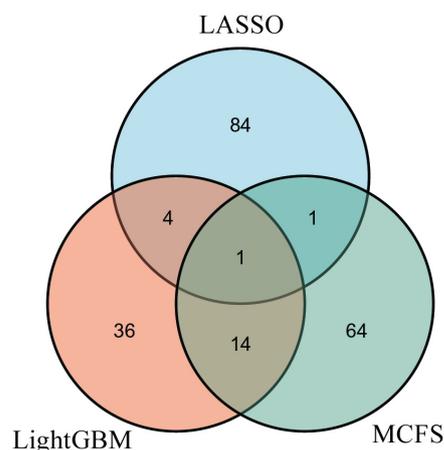

Figure 3. Venn diagram of the feature subsets used in the three feasible random forest classifiers from the three feature lists that were obtained by LASSO, LightGBM, and MCFS, respectively. The overlapping circles indicate the number of genes that were deemed to be essential by multiple feature selection methods.

3.4. Classification Rules

Although DT gave a lower performance than RF in the IFS method, it has a great advantage that RF does not own. As DT is a white-box algorithm, it can provide clues that are easy to interpret and analyze, thereby providing useful insights into understanding the essential expression differences between the COVID-19 samples with a reduced sense of taste or smell and those without impairment. According to the IFS results using DT

on the three feature lists, the optimal DT classifiers adopted the top 800 features in the LASSO list, the top 45 features in the LightGBM list, and the top 75 features in the MCFS list. All samples were represented by these features, and three trees learned from these representations. Then, three rule groups were obtained, which are provided in Table S4. A total of 48, 43, and 51 classification rules were included in the three rule groups. The number of rules for the two classes in each group is shown in Figure 4. Each rule contains the tests of the expression levels of several genes (quantitatively), and the impairment of the olfactory and gustatory functions of the samples can be predicted according to the division of these rules. Some gene rules that contribute significantly to the prediction will be discussed in detail in Section 4.
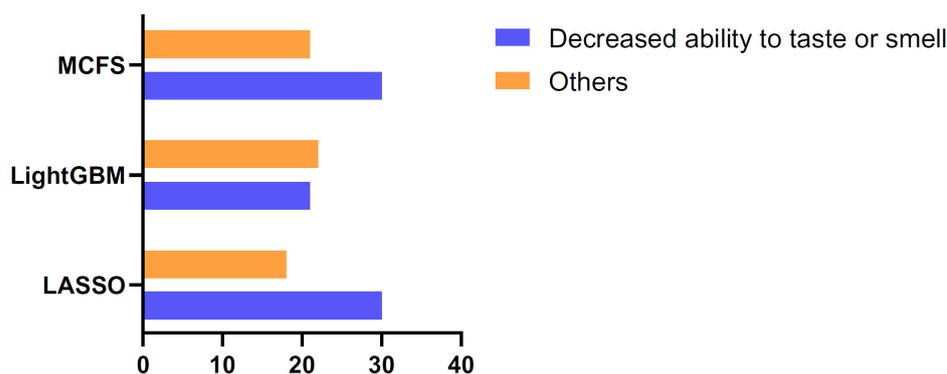
Figure 4. Bar chart showing the number of rules for the two classes of COVID-19 samples. Three optimal DT classifiers were constructed on the LASSO, LightGBM, and MCFS feature lists, and different numbers of classification rules were obtained by these classifiers. The number of rules for the COVID-19 samples with a reduced sense of taste or smell and other COVID-19 samples is plotted in the figure.

4. Discussion

We identified the various biomarkers associated with the sequelae (loss of smell and taste) in COVID-19 infection. LASSO [10], MCFS [11], and LightGBM [12] enabled us not only to identify some of the potential biomarkers associated with COVID-19-induced taste or smell loss at the circulating transcriptomics level but also helped us to establish the quantitative rules for patient clustering on the basis of the optimal DT classifiers. A detailed discussion of some of the top gene features and quantitative rules is provided below.

4.1. Relationships of Top Features in Different Lists

Three feature selection methods (LASSO, LightGBM, and MCFS) were used to analyze the expression data. As these methods have different principles, they can theoretically screen out different essential features. In order to confirm this, we selected the top 10, 50, and 100 features in the three lists and investigated their relationship. Three Venn diagrams were plotted for the top 10, 50, and 100 features, as shown in Figure 5. It can be observed that only a few features were deemed to be important by the multiple feature selection methods. There was only zero or one feature identified by all three methods when different top genes were selected in the three lists. The number of features identified by the two methods was also low. This result proved that LASSO, LightGBM, and MCFS could find the different essential gene features, thereby increasing the probability of extracting all the latent essential genes. Furthermore, the common features identified by LightGBM and MCFS were evidently more than those identified by LASSO and LightGBM or LASSO and MCFS. As LightGBM and MCFS are both DT-based methods, the features identified by them were more similar to those identified by LASSO.
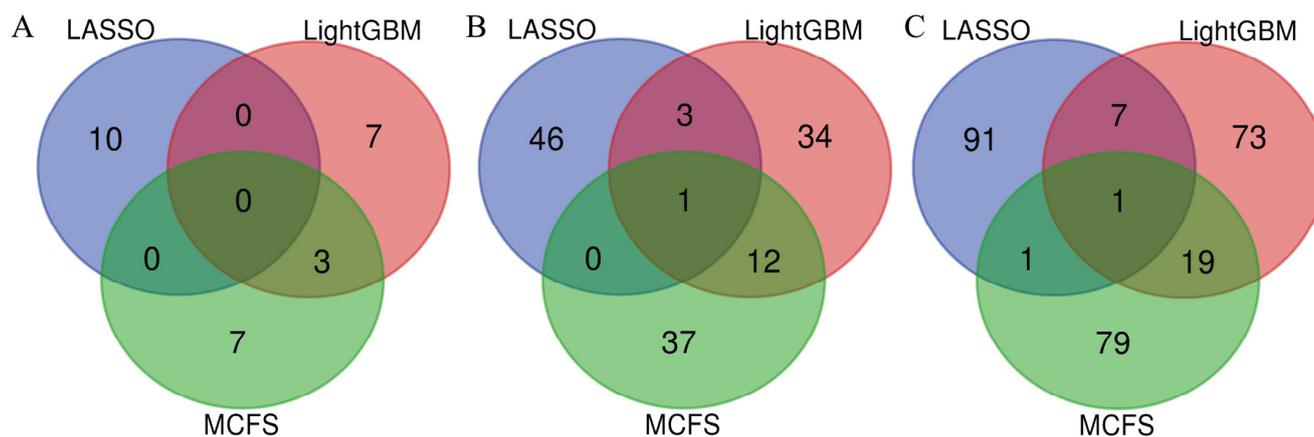
Figure 5. Venn diagram to show the relationship between the top features in the three feature lists. (**A**) Venn diagram for the top ten features in the three lists. (**B**) Venn diagram for the top 50 features in the three lists. (**C**) Venn diagram for the top 100 features in the three lists. The top features identified by different methods were quite different.

4.2. Biomarkers Predicted by One or More Feature Selection Methods

The biomarkers that were predicted by all three feature selection methods. Only one biomarker (**ENSG00000234134.1**) was recognized by all three methods. This gene is a novel gene without a functional description in the protein products from the Ensemble dataset [32]. However, at the transcriptomics level, the gene is associated with an effective lncRNA and regulates the functional gene EDRF1, which is a specific gene for erythroid cell differentiation, thus contributing to alpha and gamma-globin regulation [33,34]. Erythroid membrane antigens are associated with Type II congenital smell loss [35]. In 2022, the down-regulation of the odor pathway has been reported to be associated with COVID-19-induced anosmia [36]. The dysfunction of non-neuron-related SARS-CoV-2 entry genes may help explain the smell or taste loss in patients with COVID-19 [37]. These publications indicate that the taste or smell loss in COVID-19 patients is related to in situ odor dysfunction, which the gene EDRF1 also participated in through the GATA-1-mediated pathway [38].

The biomarkers that were predicted by the two feature selection methods. Few biomarkers were predicted by the two machine learning models. **KRT38** (**ENSG00000171360.3**) was recognized by LightGBM and MCFS. The keratin gene family has tissue-specific patterns, especially during inflammation and particularly the regional inflammation caused by SARS-CoV-2 infection [39]. A recently published abstract described the association between alveolar regeneration and the Keratin gene family during COVID-19 pathogenic progression [40]. Early in 2021, researchers recognized that, as the main target of SARS-CoV-2, alveolar cells might transmit the effects of viral infection to nearly all cells and may eventually initiate the dysfunction of olfactory cells [41,42]. **H3-5** (**ENSG00000188375.5**) was predicted by LightGBM and LASSO. Histone modification and DNA methylation initiate the impairment of the sense of smell during SARS-CoV-2 infection [43]. Therefore, we speculated that H3-5 should be one of the key biomarkers for predicting smell or taste loss in patients with COVID-19. **NUDT5 (ENSG00000165609.13)** was predicted by LightGBM and LASSO. It is reported to participate in the viral infection associated with signal transmission [44,45]. As for its specific role in association with COVID-19 compliments, the low expression level of the gene has been observed in severely infected cells [46]. The gene is involved in signal transmission in the respiratory system in COVID-19, indicating its potential olfactory regulatory effects [47]. These biomarkers, according to publications, may be associated with pathological decreases in smell or taste in COVID-19.

The biomarkers that were predicted by only one feature selection method. Some biomarkers were recognized by a single machine-learning model. These biomarkers may be associated with taste or smell loss during COVID-19 pathogenesis. The first predicted

gene is **BTN2A3P** (ENSG00000124549.14, predicted by MCFS), which is reported to be associated with immune recognition and signal transduction [48]. Given that the loss of taste or smell is highly correlated with abnormal immune responses in the olfactory regulatory networks [49], the gene could participate in the regulation of olfactory impairment through these responses. **AOC1** (ENSG00000002726.20, predicted by LASSO) and another epidemiological factor, NOS2, are the two major factors reported to be associated with the neurological disorders causing taste and smell loss [50], validating the efficacy and accuracy of our prediction. **CDC42BPB** (ENSG00000259515.1, predicted by LightGBM) could be related to a viral infection that decreases the sense of smell or taste. CDC42BPB is a binding protein kinase of the functional protein CDC42, which interacts with Rac1 and PAK1 to regulate hormone olfactory neuroblasts [51] and plays a specific role in smelling capacity.

4.3. Quantitative Rules for COVID-19 Patient Clustering

By using three optimal DT classifiers, we established a series of quantitative rules that contributed to the distinction of patients with or without smell or taste dysfunction.

First, for the rule group in the LASSO feature list, the top rules involved the specific gene **BRCA1 (ENSG00000012048.22)**. According to the rule, the low expression of the gene may indicate a loss of smell or taste. BRCA1 is reported to be a specific proliferative regulator for taste bud cells in the taste epithelium [52]. Therefore, the gene is a quantitative parameter for COVID-19 complement prediction. Another rule-associated gene is **KATNIP (ENSG00000047578.13)**, which is said to be associated with microtubule function and lysosome delivery and has a low expression level in patients with a decreased sense of taste or smell [53]. Lysosome delivery is associated with decreased smell and taste capacities [54], and the low expression of KATNIP may contribute to patient clustering. Similarly, a low **IL23R (ENSG00000162594.15)** expression level can indicate patients without dysfunctions, corresponding to an enhanced inflammation-induced decrease in the sense of smell or taste in COVID-19 [7,55,56].

Moreover, we built another rule group for the MCFS feature list for the accurate clustering of patients with COVID-19. **CDC42EP2 (ENSG00000149798.5)**, which is reported to be functionally associated with CDC42, which is similar to CDC42BPB, is highly expressed in patients with taste or smell dysfunction but has low expression in patients without these complications. CDC42 may interact with Rac1 and PAK1 and thereby regulate hormone olfactory neuroblasts [51]. When highly expressed, it may inhibit the function of hormone olfactory neuroblasts. This feature validates our rules. **PBXIP1 (ENSG00000163346.17)** and **DHRS9 (ENSG00000073737.16)** are both reported to be quantitative biomarkers with high expression levels in patients with smell or taste dysfunction. Both genes are negatively associated with taste bud functioning [57,58].

The rules established for the LightGBM feature list are similar to those established for the MCFS feature list. The top prediction gene is **CDC42EP2 (ENSG00000149798.5)**. The expression of **ADNP2 (ENSG00000101544.9)** and **CLPP (ENSG00000125656.10)** might contribute to the prediction of patients with smell or taste dysfunctions. The associations between the two genes, the regulatory directions of the genes, and their olfactory functions have been validated [59,60]. Therefore, both rules can facilitate the identification of patients with a decreased sense of taste or smell.

The above biomarkers, especially those predicted by more than one algorithm, are functionally related to a decreased sense of taste or smell during viral infections, especially SARS-CoV-2 infection. By using multiple machine-learning algorithms, we were able to identify the potential regulatory factors as completely as possible. As for the quantitative rules built for the effective and accurate clustering of patients, the top rule-associated genes at the transcriptomics level were functionally associated with COVID-19 and taste or smell regulation, validating the efficacy and accuracy of our analysis.

4.4. Limitation of This Study

In this study, we designed a computational analysis to investigate the serum transcriptome data of COVID-19 samples associated with a reduced ability to taste or smell, trying to discover the latent biomarkers for such patients. Some discovered biomarkers were confirmed to have associations with a decreased sense of taste or smell during viral infections. However, some limitations exist in this study. First, the definition of smell or taste loss was based on the original questionnaire from previous papers [16,17], which has not been widely accepted. Second, the discussion on the discovered genes was only based on a literature review. This study did not provide solid evidence (through wet experiments) to link the identified genes with a decreased sense of taste or smell. Third, three feature selection methods were adopted to analyze the serum transcriptome data. They can investigate the data from different points of view. However, it was not clear whether these methods could cover all points of view. Adding other feature methods might produce additional biomarker genes.

5. Conclusions

We applied a set of advanced machine-learning methods to the analysis of serum transcriptomic data from patients with COVID-19 to reveal the genes associated with olfactory and gustatory impairments in COVID-19. First, we ranked all genes according to their importance by using three methods. The top genes in the ranked list might increase our understanding of the underlying mechanisms of olfactory and gustatory impairments. Subsequently, by using two classification algorithms, we identified the best-performing classifier that helped us screen patients with COVID-19 and olfactory and gustatory impairments. The quantitative rules obtained by using the DT classifiers facilitated the determination of the functional impairment profile.

References

1. Hu, B.; Guo, H.; Zhou, P.; Shi, Z.-L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **2021**, *19*, 141–154. [CrossRef] [PubMed]
2. Baud, D.; Qi, X.; Nielsen-Saines, K.; Musso, D.; Pomar, L.; Favre, G. Real estimates of mortality following COVID-19 infection. *Lancet Infect. Dis.* **2020**, *20*, 773. [CrossRef] [PubMed]

3. Struyf, T.; Deeks, J.J.; Dinnes, J.; Takwoingi, Y.; Davenport, C.; Leeflang, M.M.; Spijker, R.; Hooft, L.; Emperador, D.; Domen, J. Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19. *Cochrane Database Syst. Rev.* **2022**, *5*, CD013665. [PubMed]

4. Larsen, J.R.; Martin, M.R.; Martin, J.D.; Kuhn, P.; Hicks, J.B. Modeling the onset of symptoms of COVID-19. *Front. Public Health* **2020**, *8*, 473. [CrossRef] [PubMed]

5. Hopkins, C.; Surda, P.; Whitehead, E.; Kumar, B.N. Early recovery following new onset anosmia during the COVID-19 pandemic—An observational cohort study. *J. Otolaryngol.-Head Neck Surg.* **2020**, *49*, 26. [CrossRef] [PubMed]

6. Boscolo-Rizzo, P.; Borsetto, D.; Fabbris, C.; Spinato, G.; Frezza, D.; Menegaldo, A.; Mularoni, F.; Gaudioso, P.; Cazzador, D.; Marciani, S.; et al. Evolution of altered sense of smell or taste in patients with mildly symptomatic COVID-19. *JAMA Otolaryngol.–Head Neck Surg.* **2020**, *146*, 729–732. [CrossRef]

7. Glezer, I.; Bruni-Cardoso, A.; Schechtman, D.; Malnic, B. Viral infection and smell loss: The case of COVID-19. *J. Neurochem.* **2021**, *157*, 930–943. [CrossRef]

8. Meunier, N.; Briand, L.; Jacquin-Piques, A.; Brondel, L.; Pénicaud, L. COVID 19-induced smell and taste impairments: Putative impact on physiology. *Front. Physiol.* **2021**, *11*, 625110. [CrossRef]

9. Srinivasan, M. Taste dysfunction and long COVID-19. *Front. Cell. Infect. Microbiol.* **2021**, *11*, 716563. [CrossRef]

10. Tibshirani, R. Regression shrinkage and selection via the lasso: A retrospective. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2011**, *73*, 273–282. [CrossRef]

11. Dramiński, M.; Rada-Iglesias, A.; Enroth, S.; Wadelius, C.; Koronacki, J.; Komorowski, J. Monte carlo feature selection for supervised classification. *Bioinformatics* **2007**, *24*, 110–117. [CrossRef] [PubMed]

12. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3146–3154.

13. Safavian, S.R.; Landgrebe, D. A survey of decision tree classifier methodology. *IEEE Trans. Syst. Man Cybern.* **1991**, *21*, 660–674. [CrossRef]

14. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

15. Liu, H.A.; Setiono, R. Incremental feature selection. *Appl. Intell.* **1998**, *9*, 217–230. [CrossRef]

16. Soares-Schanoski, A.; Sauerwald, N.; Goforth, C.W.; Periasamy, S.; Weir, D.L.; Lizewski, S.; Lizewski, R.; Ge, Y.; Kuzmina, N.A.; Nair, V.D.; et al. Asymptomatic SARS-CoV-2 infection is associated with higher levels of serum il-17c, matrix metalloproteinase 10 and fibroblast growth factors than mild symptomatic COVID-19. *Front. Immunol.* **2022**, *13*, 821730. [CrossRef]

17. Sauerwald, N.; Zhang, Z.; Ramos, I.; Nair, V.D.; Soares-Schanoski, A.; Ge, Y.; Mao, W.; Alshammary, H.; Gonzalez-Reiche, A.S.; van de Guchte, A.; et al. Pre-infection antiviral innate immunity contributes to sex differences in SARS-CoV-2 infection. *Cell Syst.* **2022**, *13*, 924–931.e924. [CrossRef]

18. Breiman, L. Better subset regression using the nonnegative garrote. *Technometrics* **1995**, *37*, 373–384. [CrossRef]

19. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. ; Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

20. Kohavi, R. 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*; Morgan Kaufmann Publishers Inc.: Montreal, QC, Canada, 1995; pp. 1137–1143.

21. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. Smote: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]

22. Ren, J.; Zhou, X.; Guo, W.; Feng, K.; Huang, T.; VCai, Y.-D. Identification of methylation signatures and rules for sarcoma subtypes by machine learning methods. *BioMed Res. Int.* **2022**, *2022*, 5297235. [CrossRef] [PubMed]

23. Huang, F.; Ma, Q.; Ren, J.; Li, J.; Wang, F.; Huang, T.; Cai, Y.-D. Identification of smoking associated transcriptome aberration in blood with machine learning methods. *BioMed Res. Int.* **2023**, *2023*, 5333361. [CrossRef] [PubMed]

24. Zhou, X.; Ding, S.; Wang, D.; Chen, L.; Feng, K.; Huang, T.; Li, Z.; Cai, Y.-D. Identification of cell markers and their expression patterns in skin based on single-cell rna-sequencing profiles. *Life* **2022**, *12*, 550. [CrossRef]

25. Huang, F.; Fu, M.; Li, J.; Chen, L.; Feng, K.; Huang, T.; Cai, Y.-D. Analysis and prediction of protein stability based on interaction network, gene ontology, and kegg pathway enrichment scores. *BBA—Proteins Proteom.* **2023**, *1871*, 140889. [CrossRef] [PubMed]

26. Wang, H.; Chen, L. Pmptce-hnea: Predicting metabolic pathway types of chemicals and enzymes with a heterogeneous network embedding algorithm. *Curr. Bioinform.* **2023**. [CrossRef]

27. Wu, C.; Chen, L. A model with deep analysis on a large drug network for drug classification. *Math. Biosci. Eng.* **2023**, *20*, 383–401. [CrossRef] [PubMed]

28. Ran, B.; Chen, L.; Li, M.; Han, Y.; Dai, Q. Drug-drug interactions prediction using fingerprint only. *Comput. Math. Methods Med.* **2022**, *2022*, 7818480. [CrossRef]

29. Powers, D. Evaluation: From precision, recall and f-measure to roc., informedness, markedness & correlation. *J. Mach. Learn. Technol.* **2011**, *2*, 37–63.

30. Tang, S.; Chen, L. Iatc-nfmlp: Identifying classes of anatomical therapeutic chemicals based on drug networks, fingerprints and multilayer perceptron. *Curr. Bioinform.* **2022**, *17*, 814–824.

31. Matthews, B. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochim. Et Biophys. Acta (BBA)-Protein Struct.* **1975**, *405*, 442–451. [CrossRef]

32. Cunningham, F.; Achuthan, P.; Akanni, W.; Allen, J.; Amode, M.R.; Armean, I.M.; Bennett, R.; Bhai, J.; Billis, K.; Boddu, S. Ensembl 2019. *Nucleic Acids Res.* **2019**, *47*, D745–D751. [CrossRef] [PubMed]

33. Wang, D.; Li, Y.; Shen, B. A novel erythroid differentiation related gene edrf1 upregulating globin gene expression in hel cells. *Chin. Med. J.* **2002**, *115*, 1701–1705. [PubMed]

34. Wang, D.; Li, Y.; Shen, B. Initial function analysis of a novel erythroid differentiation related geneedrf1. *Sci. China Ser. C Life Sci.* **2001**, *44*, 489–496. [CrossRef] [PubMed]

35. Stateman, W.A.; Henkin, R.I.; Knöppel, A.B.; Flegel, W.A. Erythrocyte membrane antigen frequencies in patients with type ii congenital smell loss. *Am. J. Otolaryngol.* **2015**, *36*, 146–152. [CrossRef] [PubMed]

36. Zazhytska, M.; Kodra, A.; Hoagland, D.A.; Frere, J.; Fullard, J.F.; Shayya, H.; McArthur, N.G.; Moeller, R.; Uhl, S.; Omer, A.D.; et al. ; Non-cell-autonomous disruption of nuclear architecture as a potential cause of COVID-19-induced anosmia. *Cell* **2022**, *185*, 1052–1064.e1012. [CrossRef]

37. Brann, D.H.; Tsukahara, T.; Weinreb, C.; Lipovsek, M.; Van den Berge, K.; Gong, B.; Chance, R.; Macaulay, I.C.; Chou, H.-J.; Fletcher, R.B.; et al. Non-neuronal expression of SARS-CoV-2 entry genes in the olfactory system suggests mechanisms underlying COVID-19-associated anosmia. *Sci. Adv.* **2020**, *6*, eabc5801. [CrossRef]

38. Duncheng, W.; Shuhua, M.; Shengqi, W.; Beifen, S. Antisense edrf1 gene inhibited gata-1 transcription factor DNA-binding activity in k562 cells. *Sci. China Ser. C Life Sci.* **2002**, *45*, 289–297.

39. Ho, M.; Thompson, B.; Fisk, J.N.; Nebert, D.W.; Bruford, E.A.; Vasiliou, V.; Bunick, C.G. Update of the keratin gene family: Evolution, tissue-specific expression patterns, and relevance to clinical disorders. *Hum. Genom.* **2022**, *16*, 1. [CrossRef]

40. Ting, C.; Aspal, M.; Vaishampayan, N.; Huang, S.; Wang, F.; Farver, C.; Zemans, R. Ineffectual alveolar regeneration in COVID-19 ards: Persistence of keratin-8 positive transitional cells. In *Tp105. Tp105 Basic Mechanisms of Lung Infections: From SARS-CoV-2 to Influenza*; American Thoracic Society: New York, NY, USA, 2021; p. A4163.

41. Hawkes, C.H. Smell, taste and COVID-19: Testing is essential. *QJM Int. J. Med.* **2021**, *114*, 83–91. [CrossRef]

42. Bhatta, S.; Sharma, D.; Sharma, S.; Maharjan, L.; Bhattachan, S.; Shah, M.K.; Singhal, A.; Ghanpur, A.D.; Ganesuni, D.; Saindani, S.J. Smell and taste disturbance in COVID-19 patients: A prospective multicenteric review. *Indian J. Otolaryngol. Head Neck Surg.* **2021**, *74*, 2978–2984. [CrossRef]

43. Shirvaliloo, M. Epigenomics in COVID-19; the link between DNA methylation, histone modifications and SARS-CoV-2 infection. *Epigenomics* **2021**, *13*, 745–750. [CrossRef] [PubMed]

44. Kronsbein, H.C.; Jastorff, A.M.; Maccarrone, G.; Stalla, G.; Wurst, W.; Holsboer, F.; Turck, C.W.; Deussing, J.M. Crhr1-dependent effects on protein expression and posttranslational modification in att-20 cells. *Mol. Cell. Endocrinol.* **2008**, *292*, 1–10. [CrossRef]

45. Özkan, H.; Öztürk, D.G.; Korkmaz, G. Transcriptional factor repertoire of breast cancer in 3d cell culture models. *Cancers* **2022**, *14*, 1023. [CrossRef] [PubMed]

46. Schonfeld, E. Single Cell Genetic Analysis of COVID-19: A Deep Learning Approach. 2022. Available online: https://www.researchsquare.com/article/rs-1754568/latest.pdf (accessed on 4 October 2022).

47. Jha Bhaskar, A.; Fadda, A.; Merce, C.; Mugo, E.; Droll, D.; Clayton, C. Depletion of the trypanosome pumilio domain protein puf2 or of some other essential proteins causes transcriptome changes related to coding region length. *Eukaryot. Cell* **2014**, *13*, 664–674.

48. de Groot, N.; Groen, R.; Orie, V.; Bruijnesteijn, J.; de Groot, N.G.; Doxiadis, G.G.; Bontrop, R.E. Analysis of macaque btn3a genes and transcripts in the extended mhc: Conserved orthologs of human γδ t cell modulators. *Immunogenetics* **2019**, *71*, 545–559. [CrossRef] [PubMed]

49. Rank, A.; Tzortzini, A.; Kling, E.; Schmid, C.; Claus, R.; Löll, E.; Burger, R.; Römmele, C.; Dhillon, C.; Müller, K. One year after mild COVID-19: The majority of patients maintain specific immunity, but one in four still suffer from long-term symptoms. *J. Clin. Med.* **2021**, *10*, 3305. [CrossRef]

50. Macías, Á.M. The intestinal perspective of COVID-19: Nos2 and aoc1 genes as epidemiological factors, and a homeopathic approach to their functional improvement. *Homœopathic Links* **2020**, *33*, 196–221. [CrossRef]

51. González-Martínez, D.; Kim, S.-H.; Hu, Y.; Guimond, S.; Schofield, J.; Winyard, P.; Vannelli, G.B.; Turnbull, J.; Bouloux, P.-M. Anosmin-1 modulates fibroblast growth factor receptor 1 signaling in human gonadotropin-releasing hormone olfactory neuroblasts through a heparan sulfate-dependent mechanism. *J. Neurosci.* **2004**, *24*, 10384–10392. [CrossRef]

52. Cohn, Z.J.; Kim, A.; Huang, L.; Brand, J.; Wang, H. Lipopolysaccharide-induced inflammation attenuates taste progenitor cell proliferation and shortens the life span of taste bud cells. *BMC Neurosci.* **2010**, *11*, 72. [CrossRef]

53. Starling, G.P.; Philips, B.A.; Ganesh, S.; King, J.S. Katnip is required to maintain microtubule function and lysosomal delivery to autophagosomes and phagosomes. *Mol. Biol. Cell* **2023**, *34*, ar12. [CrossRef]

54. Gahl, W.A.; Reed, G.F.; Thoene, J.G.; Schulman, J.D.; Rizzo, W.B.; Jonas, A.J.; Denman, D.W.; Schlesselman, J.J.; Corden, B.J.; Schneider, J.A. Cysteamine therapy for children with nephropathic cystinosis. *N. Engl. J. Med.* **1987**, *316*, 971–977. [CrossRef] [PubMed]

55. Khani, E.; Khiali, S.; Beheshtirouy, S.; Entezari-Maleki, T. Potential pharmacologic treatments for COVID-19 smell and taste loss: A comprehensive review. *Eur. J. Pharmacol.* **2021**, *912*, 174582. [CrossRef] [PubMed]

56. de Melo, G.D.; Lazarini, F.; Levallois, S.; Hautefort, C.; Michel, V.; Larrous, F.; Verillaud, B.; Aparicio, C.; Wagner, S.; Gheusi, G. COVID-19–related anosmia is associated with viral persistence and inflammation in human olfactory epithelium and brain infection in hamsters. *Sci. Transl. Med.* **2021**, *13*, eabf8396. [CrossRef] [PubMed]

57. Chen, W.; Chen, G. The roles of vitamin a in the regulation of carbohydrate, lipid, and protein metabolism. *J. Clin. Med.* **2014**, *3*, 453–479. [CrossRef] [PubMed]

58. Diego, V.P.; Rainwater, D.L.; Wang, X.-L.; Cole, S.A.; Curran, J.E.; Johnson, M.P.; Jowett, J.B.; Dyer, T.D.; Williams, J.T.; Moses, E.K. Genotype× adiposity interaction linkage analyses reveal a locus on chromosome 1 for lipoprotein-associated phospholipase a2, a marker of inflammation and oxidative stress. *Am. J. Hum. Genet.* **2007**, *80*, 168–177. [CrossRef]

59. Malishkevich, A.; Leyk, J.; Goldbaum, O.; Richter-Landsberg, C.; Gozes, I. Adnp/adnp2 expression in oligodendrocytes: Implication for myelin-related neurodevelopment. *J. Mol. Neurosci.* **2015**, *57*, 304–313. [CrossRef]

60. Wang, Y.T.; Lim, Y.; McCall, M.N.; Huang, K.-T.; Haynes, C.M.; Nehrke, K.; Brookes, P.S. Cardioprotection by the mitochondrial unfolded protein response requires atf5. *Am. J. Physiol.-Heart Circ. Physiol.* **2019**, *317*, H472–H478. [CrossRef]