

## Article

# Integrating Expression Data-Based Deep Neural Network Models with Biological Networks to Identify Regulatory Modules for Lung Adenocarcinoma

Lei Fu <sup>1,†</sup>, Kai Luo <sup>1,†</sup>, Junjie Lv <sup>1</sup>, Xinyan Wang <sup>2</sup>, Shimei Qin <sup>1</sup>, Zihan Zhang <sup>1</sup>, Shibin Sun <sup>1</sup>, Xu Wang <sup>1</sup>, Bei Yun <sup>1</sup>, Yuehan He <sup>1</sup>, Weiming He <sup>3</sup>, Wan Li <sup>1,\*</sup>  and Lina Chen <sup>1,\*</sup>

<sup>1</sup> College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150000, China  
<sup>2</sup> Department of Respiratory, Second Affiliated Hospital of Harbin Medical University, Harbin 150000, China  
<sup>3</sup> Institute of Opto-Electronics, Harbin Institute of Technology, Harbin 150000, China  
\* Correspondence: liwan@hrbmu.edu.cn (W.L.); chenlina@ems.hrbmu.edu.cn (L.C.)  
† These authors contributed equally to this work.

**Simple Summary:** The growing evidence suggested that competing endogenous RNAs (ceRNAs) have significant associations with tumor occurrence and progression, yet the regulatory mechanism of them in lung adenocarcinoma remains unclear. Identification of the regulatory modules for lung adenocarcinoma is a critical and fundamental step towards understanding the regulatory mechanisms during carcinogenesis. Deep neural network (DNN) models have become a powerful tool to intelligently recognize the sophisticated relationships of ceRNAs appropriately. In this paper, multiple deep neural network models were constructed using the expression data to identify regulatory modules for lung adenocarcinoma in biological networks. Three identified regulatory modules association with lung adenocarcinoma were validated from three aspects, i.e., literature review, functional enrichment analysis, and an independent dataset. The regulatory relationships between RNAs were validated in various datasets, including CPTAC, TCGA and an expression profile from the GEO database. Our study will contribute to improving the understanding of regulatory mechanisms in the carcinogenesis of lung adenocarcinoma and provide schemes for identifying novel regulatory modules of other cancers.

**Abstract:** Lung adenocarcinoma is the most common type of primary lung cancer, but the regulatory mechanisms during carcinogenesis remain unclear. The identification of regulatory modules for lung adenocarcinoma has become one of the hotspots of bioinformatics. In this paper, multiple deep neural network (DNN) models were constructed using the expression data to identify regulatory modules for lung adenocarcinoma in biological networks. First, the mRNAs, lncRNAs and miRNAs with significant differences in the expression levels between tumor and non-tumor tissues were obtained. mRNA DNN models were established and optimized to mine candidate mRNAs that significantly contributed to the DNN models and were in the center of an interaction network. Another DNN model was then constructed and potential ceRNAs were screened out based on the contribution of each RNA to the model. Finally, three modules comprised of miRNAs and their regulated mRNAs and lncRNAs with the same regulation direction were identified as regulatory modules that regulated the initiation of lung adenocarcinoma through ceRNAs relationships. They were validated by literature and functional enrichment analysis. The effectiveness of these regulatory modules was evaluated in an independent lung adenocarcinoma dataset. Regulatory modules for lung adenocarcinoma identified in this study provided a reference for regulatory mechanisms during carcinogenesis.

**Keywords:** regulatory module; deep neural network; competing endogenous RNA; lung adenocarcinoma; biological network



**Citation:** Fu, L.; Luo, K.; Lv, J.; Wang, X.; Qin, S.; Zhang, Z.; Sun, S.; Wang, X.; Yun, B.; He, Y.; et al. Integrating Expression Data-Based Deep Neural Network Models with Biological Networks to Identify Regulatory Modules for Lung Adenocarcinoma. *Biology* **2022**, *11*, 1291. <https://doi.org/10.3390/biology11091291>

Academic Editors: Shibiao Wan, Chunjie Jiang and Shengli Li

Received: 10 August 2022

Accepted: 29 August 2022

Published: 30 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

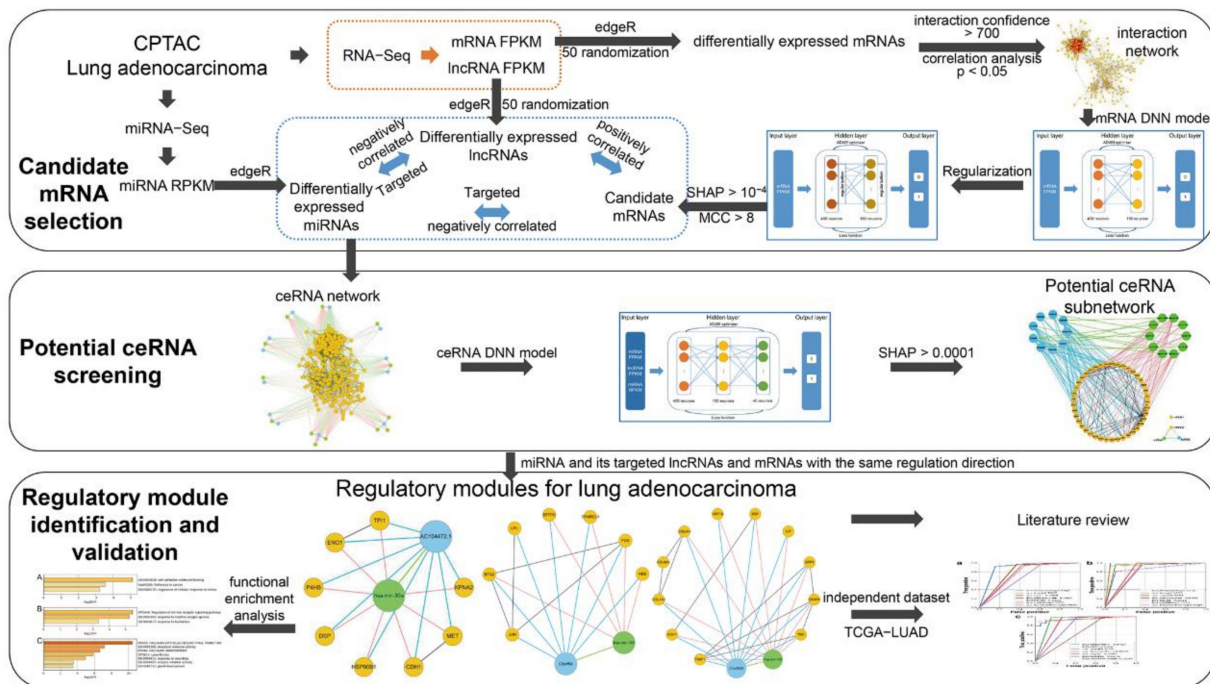
Lung cancer is one of the most common causes of cancer deaths in the world [1], of which the most common type is lung adenocarcinoma that comprises about 40% of all lung cancer cases. It remains one of the most aggressive and lethal tumor types [2], despite the understanding of the pathogenesis and new treatments for lung adenocarcinoma having improved [3]. Many experiments have confirmed that regulations of RNA molecules are closely related to the occurrence and development of lung adenocarcinoma [4]. Therefore, regulatory modules of lung adenocarcinoma identified based on biological networks and expression data are beneficial to understand the carcinogenesis.

Most proteins activate and function through their interactions. Thus, protein interactions and their networks are very important in most biological functions and processes [5]. Protein interaction networks can help us better understand the disease process, which is of great significance in the identification of disease proteins/gene [6,7]. Roudi et al. identified differentially expressed genes (DEGs) at each stage of lung adenocarcinoma in four datasets from the Gene Expression Omnibus (GEO) database. Co-expression clusters and biological pathways were identified for common and unique DEGs, respectively. Five hub genes crucial for lung adenocarcinoma were observed from a protein interaction network of common DEGs among all stages, and confirmed using an independent dataset collected from The Cancer Genome Atlas (TCGA) [8]. Better understanding of diseases could be obtained through disease proteins/genes identified from protein interaction networks.

A competing endogenous RNA (ceRNA) network links the function of protein-coding mRNAs with the function of non-coding RNAs (ncRNAs, including microRNAs (miRNAs) and long non-coding RNAs (lncRNAs)) [9], which better explains the respective roles of different RNAs in biological processes. Dysregulation of their expression has been implicated in various diseases, including cancer. For example, Jafarinejad-Farsangi et al. predicted top miRNAs targeting SARS-CoV-2 genome and differentially expressed genes (DEGs) in the lungs of patients infected with SARS-CoV-2 [10]. Hsa-mir-130a has been proven by many clinical trials and bioinformatics studies to be a marker gene that widely participated in various types of tumors by down-regulating a variety of key proto-oncogenes [11]. Li et al. identified 11 gastric cancer-specific lncRNAs, 9 miRNAs, and 41 mRNAs from a gastric cancer ceRNA network generated from miRcode and miRTarBase based on bioinformatics [12]. These ncRNAs have been extracted through ceRNA regulatory relationships and served as determined or potential tumor suppressors or therapeutic targets [13]. A better understanding of the underlying mechanisms of these regulations and their roles in cancer initiation is essential for the development of more robust clinical diagnostic tools.

In the field of biology, deep learning has greatly improved the reliability of biological big data analysis, when compared to traditional machine learning with the advantages of self-learning and high generalization ability. The evaluation criteria obtained by deep learning using a large volume of patient data not only have a wide range of confidence, but also greatly reduce the cost of clinical diagnosis [14]. Deep neural network (DNN) models have become a powerful tool of deep learning and artificial intelligence. Good progress has been made in the application of DNNs in various biological branches, and the diagnosis applying DNNs can even reach the level of experienced clinical experts [15–17].

In this paper, regulatory modules for lung adenocarcinoma were identified using multiple DNN models in interaction and ceRNA networks (Figure 1). This would not only improve the ability in the identification of regulatory modules for lung adenocarcinoma, but also provide a basis for the understanding of regulatory mechanisms during carcinogenesis.



**Figure 1.** The workflow of this study. Firstly, expression data of mRNAs, miRNAs and lncRNAs were extracted from CPTAC, and differential analysis was performed to obtain significant differentially expressed mRNAs, miRNAs and lncRNAs. Protein interaction and co-expression analysis were performed on differential mRNAs to construct the interaction network, and DNN models were established to screen out candidate mRNAs with significant contribution to the DNN models. In addition, multiple experimentally validated miRNA target databases were combined for these candidate mRNAs, differential miRNAs and lncRNAs to form a ceRNA network. Then a ceRNA DNN model was performed to identify potential ceRNAs. Finally, modules comprised of miRNAs and their regulated mRNAs and lncRNAs with the same regulation direction were identified as regulatory modules for lung adenocarcinoma from the potential ceRNA subnetwork. These valuable tumor regulatory modules were validated by literature review, functional enrichment analysis and an independent lung adenocarcinoma dataset from The Cancer Genome Atlas (TCGA-LUAD).

## 2. Materials and Methods

### 2.1. Data

In this study, RNA-Seq data of 204 samples (including 102 tumor tissues and matched paracancerous non-tumor tissues) and miRNA-Seq data of 202 samples (including 101 tumor tissues and matched paracancerous non-tumor tissues) for lung adenocarcinoma were acquired from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) project. A total of 100 pairs of tumor tissues and matched paracancerous non-tumor tissues) shared by RNA-Seq and miRNA-Seq data were selected for data analysis. The pathological characteristics of the lung adenocarcinoma patients are presented in Table 1.

**Table 1.** The pathological characteristics of the lung adenocarcinoma patients.

	CPTAC
<b>Patient (n)</b>	100
<b>Age, years</b>	
median	63.5
range	35–81
<b>Sex (%)</b>	
male	63 (63%)
female	37 (37%)

Table 1. Cont.

	CPTAC
<b>Tumor_grade (%)</b>	
G1	7 (7%)
G2	55 (55%)
G3	37 (37%)
GX	1 (1%)
<b>Ajcc_pathologic_stage (%)</b>	
Stage I	54 (54%)
Stage II	29 (29%)
Stage III	17 (17%)

For RNA-Seq data, the Ensemble database was used to annotate symbol and gene\_biotype attributes. According to the gene\_biotype attribute, 14,560 lncRNAs and 19,450 mRNAs were extracted. Fragments per kilobase of exon per million reads mapped (FPKM) values were chosen as the representative measure of mRNA or lncRNA expression. Read counts for 1886 miRNAs were first obtained from the miRNA-Seq data. Then Reads Per Kilobase per Million mapped reads (RPKM) values were calculated for each sample using the Reads per million mapped reads (RPM) for each miRNA.

The data have undergone quality control and normalization using the edgeR TMM method after removing RNAs with more than 1/3 of the missing values for RNA-Seq data or 2 missing values for miRNA-Seq data. Then edgeR was used to detect significant differentially expressed mRNAs, lncRNAs and miRNAs between lung adenocarcinoma and corresponding paracancerous non-tumor tissues based on their read counts.  $|\log_2(\text{fold-change})| > 1$  and FDR adjusted  $p$  value  $< 0.05$  were used as thresholds.

In order to ensure the stability and reliability of the obtained significant differentially expressed mRNAs and lncRNAs, due to their large size, 4/5 of the samples were randomly selected each time to detect significant differentially expressed mRNAs and lncRNAs. The intersection of the differentially expressed mRNAs and lncRNAs obtained after 50 times of randomization was taken as the significant differentially expressed mRNAs and lncRNAs. Because of the small number of miRNAs and sparse expression values, the differential expression analysis was performed one time to obtain significant differentially expressed miRNAs [18]. A total of 4888 significant differentially expressed RNAs, including 3399 mRNAs (1463 up-regulated and 1936 down-regulated), 1098 lncRNAs (519 up-regulated and 579 down-regulated) and 391 miRNAs (330 up-regulated and 61 down-regulated) were obtained.

## 2.2. DNN Model

### 2.2.1. DNN Model Feature

The interaction network for significant differentially expressed mRNAs was constructed according to interactions between proteins they encoded and their expression correlation. Interactions with confidence  $> 700$  from the STRING database and significantly co-expressed ( $p$ -value  $< 0.05$ ) were screened out. An interaction network containing 1122 mRNAs and 13455 pairs of interactions was obtained. The FPKM values of these 1122 mRNAs were the features for the mRNA DNN models.

Based on experimentally validated miRNA targets (mRNAs and lncRNAs) from TarBase, mirTarBase, miRecords, targetScan and starBase, the ceRNA triplets (mRNA-miRNA-lncRNA) were constructed if miRNAs were negatively correlated with lncRNAs and mRNAs, respectively, and mRNAs positively correlated with lncRNAs ( $Z$ -Score  $> 0$ ,  $p$ -value  $< 0.05$ ). The ceRNA network was constructed accordingly. The FPKM/RPKM values of these mRNAs, miRNAs and lncRNAs were employed as the features for the ceRNA DNN models.

### 2.2.2. DNN Model Construction

Multiple fully connected DNN models were built with Google TensorFlow 2.0 architecture. The architecture of each fully connected DNN model is composed of an input layer, multiple hidden layers, and an output layer. The output layer with 1/0 label indicates the sample to be a cancer one or not. The Adaptive Moment Estimation (ADAM) optimizer with default parameters supplied by Keras was chosen due to the small sample size [19]. Here, the loss function of binary cross-entropy was used.

The performance of a DNN model is associated with many model training related parameters, including batch size, number of epochs and the learning rate in the training process. The model training requires the accumulation of multiple learning rounds. In each round, a batch of training sets will be randomly selected according to the learning rate. The more samples each batch has, the faster the convergence of the model will be, and the weaker the generalization ability will be. Since the purpose of this analysis was to obtain regulatory modules for lung adenocarcinoma, and the sample size and characteristic number were relatively small, the batch size = 16 and epoch  $\geq$  1000 was set. The learning rate should be high to prevent under-fitting if the batch number is large and be low to prevent overfitting when the batch number is small. Therefore, the learning rate was set to 0.0001 for more learning epochs.

In order to prevent overfitting in the learning process, DNNs could be further optimized from two aspects: regularization layer and Drop layer. In the regularization layer, each multidimensional feature is regularized after the neuron layer to make the gap smaller and reduce the dependence of high-weight features; while in the Drop layer, some eigenvalues are discarded randomly. In order to preserve the integrity of features, regularization was used to prevent overfitting in this paper.

Features that contributed significantly to the DNN models were more biologically significant. SHapley Additive exPlanations (SHAP) is a game-theoretic approach to explain the output of a machine learning model. The SHAP value represents the contribution of a feature to the machine learning model. In this paper, DeepExplainer of the python SHAP module was used to approximate SHAP values for the DNN models [20]. The impact of each feature on each sample was obtained using force\_plot. Then the arithmetic average of absolute values for the impact representing the importance of the feature to all samples, denoted as the SHAP value, was calculated by summary\_plot. The larger the SHAP value is, the greater the contribution of the feature to the corresponding DNN model. The identification of candidate and potential RNAs in this paper was based on the SHAP values of RNAs in DNN models.

### 2.2.3. DNN Model Evaluation

The samples were split randomly into a 70% training set and 30% validation set. The randomization was repeated 50 times. The model was then evaluated with the closeness between the expected output and the actual output by two measures: accuracy and loss function. Accuracy indicates the precision of deep learning. The closer the accuracy approaches to 1, the more accurate the prediction will be. The accuracy of training sets is generally higher than that of validation sets. The loss represents the degree of deviation between the training or validation results and the real results. The closer to 0, the better the fitting with the real results will be. The loss of the validation sets is generally not less than that of the training sets. DNN models were evaluated via the accuracy and loss curves for the training and validation sets.

## 2.3. Regulatory Modules for Lung Adenocarcinoma

### 2.3.1. Candidate mRNA Selection

In order to select candidate mRNAs that not only contributed significantly to the mRNA DNN models, but also were in the center of the interaction network, their contribution to mRNA DNN models and centrality values were jointly analyzed by Jointplot of the

python Seaborn module. Jointplot is a Seaborn function that plots a scatter graph for two variables with distinct histograms at the plot's upper edge and right sides.

On the one hand, the contribution of each mRNA to the mRNA DNN models was evaluated by SHAP values. On the other hand, central mRNAs were identified by the plug-in cytoHubba of Cytoscape using Maximal Clique Centrality (MCC) algorithm since MCC had a better performance in predicting PPI network central nodes among all the centrality measures [21]. Therefore, Candidate mRNAs were selected as those with high values of both measures.

### 2.3.2. Potential ceRNA Screening

In order to reveal the ceRNA regulatory mechanism in lung adenocarcinoma, a ceRNA DNN model was constructed. SHAP values were also used to evaluate the contribution of each RNA (mRNA, miRNA or lncRNA) to the ceRNA DNN model. RNAs with SHAP value > 0.0001 were screened out as potential ceRNAs and used to reconstruct a potential ceRNA subnetwork with ceRNA regulatory and mRNA interaction relationships.

### 2.3.3. Regulatory Module Identification and Validation

Modules comprising one miRNA and its regulated lncRNAs and mRNAs with the same up/down regulation direction were selected from the potential ceRNA subnetwork as regulatory modules for lung adenocarcinoma.

Literature review was conducted by searching in the PubMed database for all articles published in English Language on the topics of the identified regulatory modules and lung adenocarcinoma.

The metascape platform was used to conduct functional enrichment analysis based on GO, KEGG, wikiPathways and Hallmark databases for mRNAs in regulatory modules for lung adenocarcinoma. Categories with the minimum overlap number 3 and the hypergeometric test Benjamini–Hochberg adjusted  $p$ -value < 0.05 were selected.

In order to reflect the effectiveness of identified regulatory modules for lung adenocarcinoma, 551 unpaired samples from the lung adenocarcinoma dataset from The Cancer Genome Atlas (TCGA-LUAD) (including 497 tumor and 54 paracancerous non-tumor samples, completely different from the training dataset) were used as an independent dataset. Traditional machine learning methods, including K-nearest neighbor (KNN), Support Vector Machine (SVM), decision tree, Multi-feature Bayesian, Logistic regression, and random forest were also applied to sample classification using identified regulatory modules. Their performance was compared to that of the DNN model by area under receiver operating characteristic curves (AUC) values of the area under the receiver operating characteristic (ROC) curves.

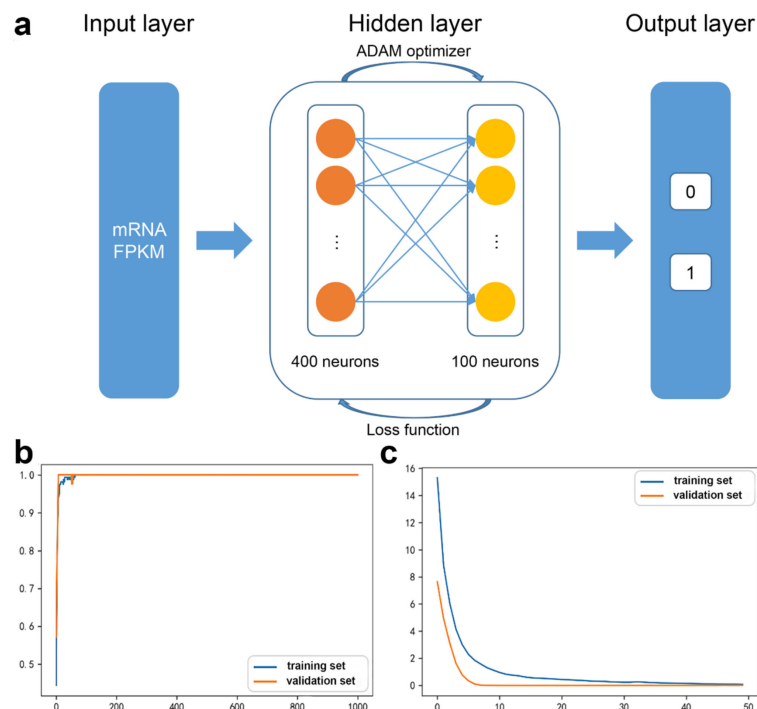
## 3. Results

### 3.1. Candidate mRNAs

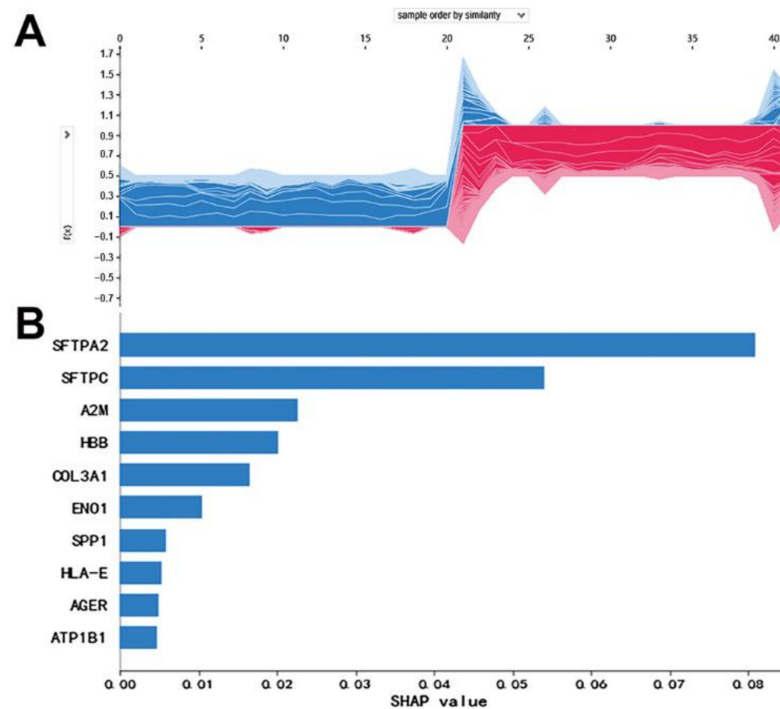
For the mRNA DNN models, the initial input layer was set to the FPKM values of 1122 mRNAs from the interaction network, and two hidden layers of 400 and 100 neurons were established, with the 0/1 label as the output layer (Figure 2a).

The accuracy curve and the loss curve were both accordant with the general law of deep learning (Figure 2b,c). The loss curve for the training set and the validation set decreased with the increase in iteration number. However, when the training times were between 5 and 10 the loss for the training set decreased, while for the validation set it was stable, indicating the occurrence of overfitting.

In order to alleviate the imbalance of each feature and eliminate the phenomenon of overfitting, regularization layers were added after each hidden layer. The python SHAP module was applied to interpret the contribution of each mRNA to the DNN model (Figure 3).

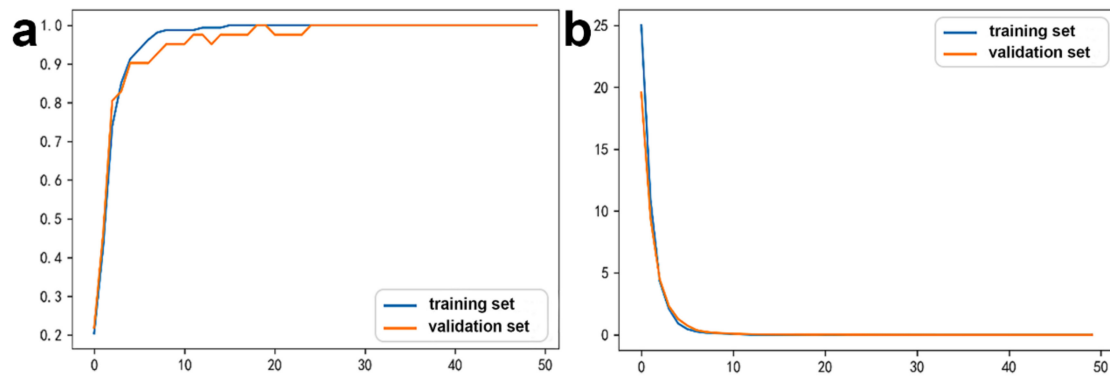


**Figure 2.** The mRNA DNN model: (a) The DNN model structure, (b) accuracy curve, and (c) loss curve.



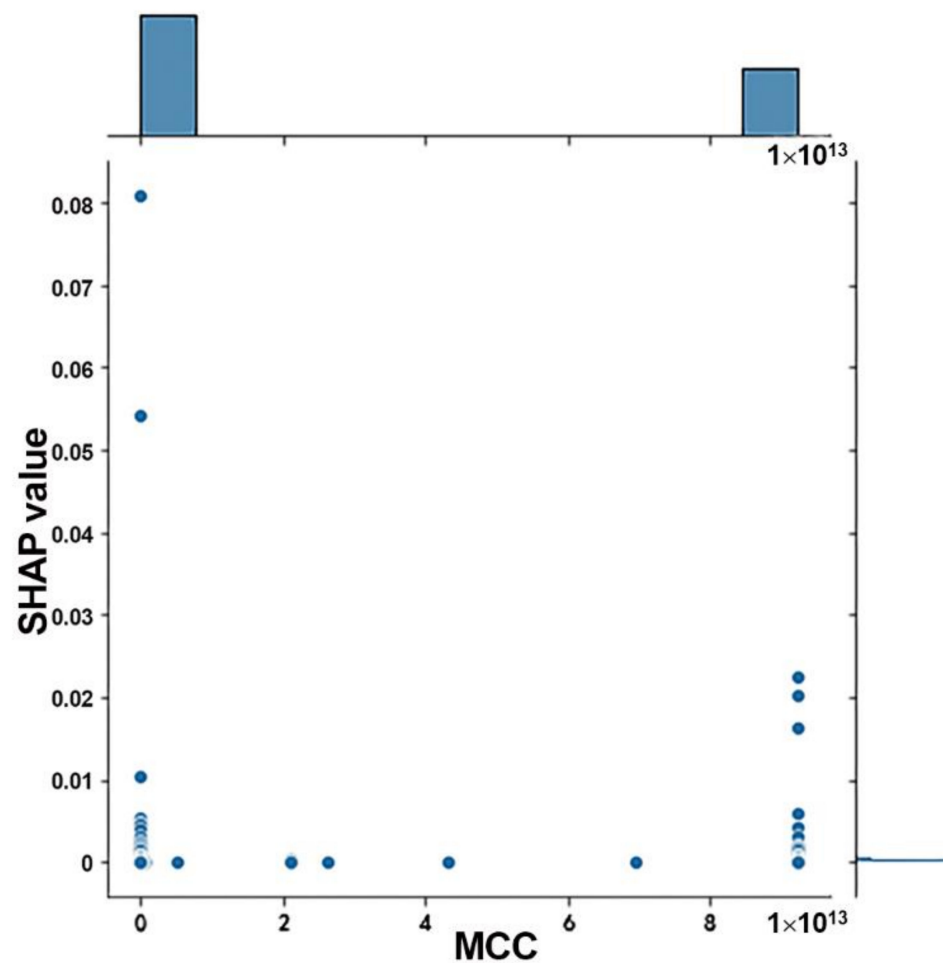
**Figure 3.** The SHAP values of the mRNAs: (A) Contribution of each mRNA to partial individual sample. Red represents the positive influence, and blue represents the negative influence. The abscissa represents the samples, and the ordinate represents the SHAP values. (B) The SHAP values of top 10 mRNAs based on their contribution to all samples.

The 18 mRNAs with SHAP value = 0 were removed and the DNN model was relearned (Figure 4). The accuracy was 97.04%. The validation loss was slightly greater than the training loss, and the two curves tended to converge. Hence, the regularization optimization was effective.



**Figure 4.** The accuracy and loss of the optimized mRNA DNN model after removing mRNAs with SHAP value = 0. (a) Accuracy curve and (b) loss curve.

In order to select candidate mRNAs that had not only high contribution on DNN models, but also central properties in the interaction network, the SHAP values of DNN models and the MCC values of network nodes were jointly analyzed. It was demonstrated that the mRNAs were mainly clustered in two categories (Figure 5). Therefore, a total of 699 mRNAs with  $MCC > 8$  and SHAP value  $> 10^{-4}$  were selected as candidate mRNAs.



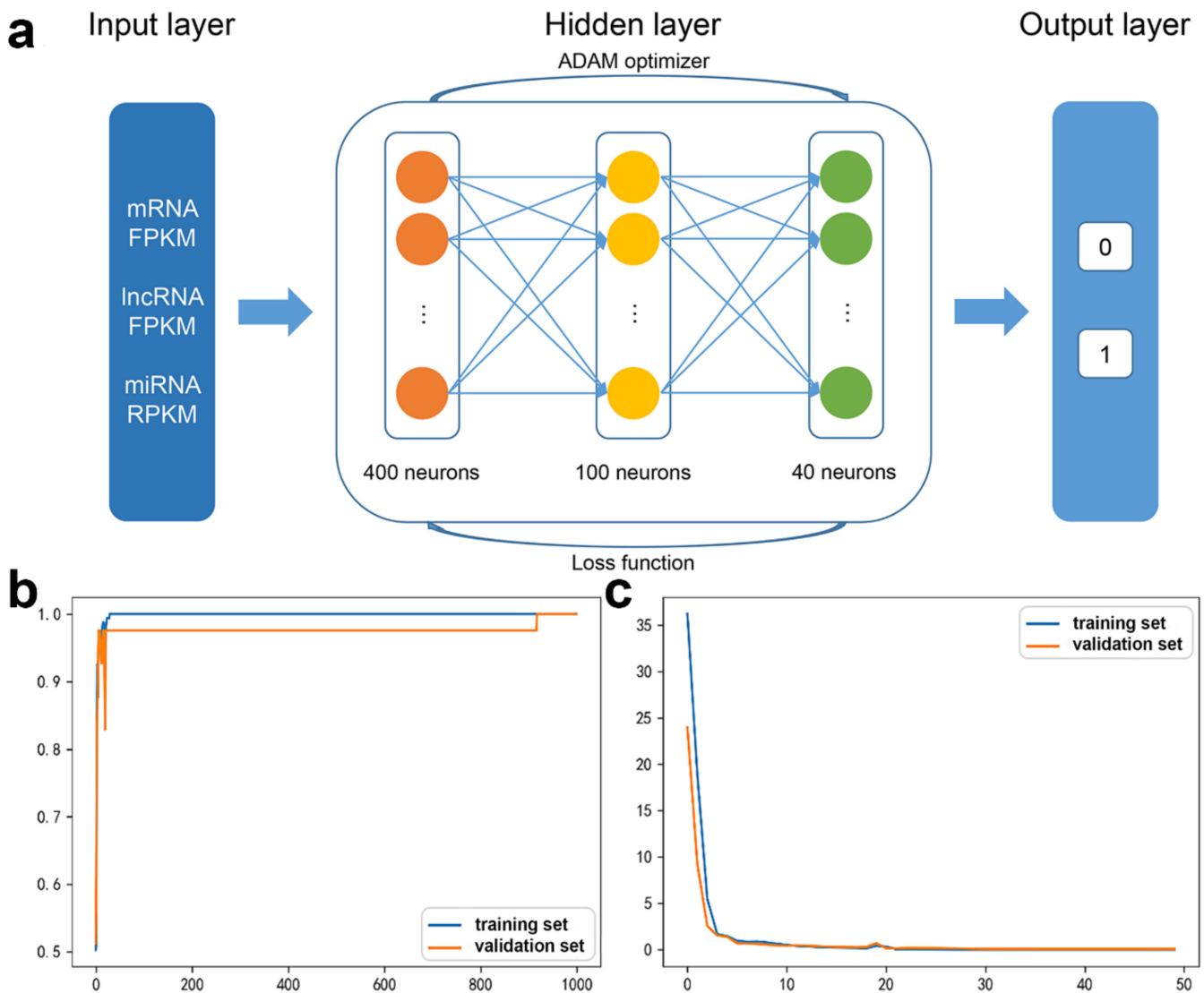
**Figure 5.** The joint distribution of SHAP and MCC values. MRNAs are mainly clustered in two categories.



### 3.2. Potential ceRNAs

Based on miRNA target information, candidate mRNAs obtained by previous DNN models and significant differentially expressed lncRNAs targeted by significant differentially expressed miRNAs were extracted. Pearson correlation coefficients of miRNA-lncRNA, miRNA-mRNA and mRNA-lncRNA were calculated to screen ceRNA triplets ( $p$ -value < 0.05). Finally, 518 ceRNAs triplets were screened out, containing 309 mRNAs, 13 miRNAs and 12 lncRNAs. Combining the interaction network and screened ceRNA triplets, a ceRNA network comprised of 270 mRNAs, 13 miRNAs and 12 lncRNAs was constructed after removing isolated nodes.

To achieve a better performance, the regularization layers in the previous optimized mRNA DNN model were removed and a layer of 40 neurons was added to expand the capacity of the neural network. FPKM values for 13 miRNAs and 12 lncRNAs were used as features for the ceRNA DNN model (Figure 6a). The accuracy curve finally converged, and the loss curve uniformly converged to 0 (Figure 6b,c). Therefore, the ceRNA DNN model was feasible and appropriate.



**Figure 6.** The ceRNA DNN model: (a) The DNN model structure, (b) accuracy curve, and (c) loss curve.

SHAP values were used to evaluate the contribution of each RNA (mRNA, miRNA or lncRNA) to the ceRNA DNN model. RNAs with SHAP value > 0.0001 (including 40 mRNAs, 10 miRNAs and 9 lncRNAs) were screened out as potential ceRNAs (Figure 7) and used to

reconstruct the potential ceRNA subnetwork with ceRNA regulatory and mRNA interaction relationships (Figure 8).

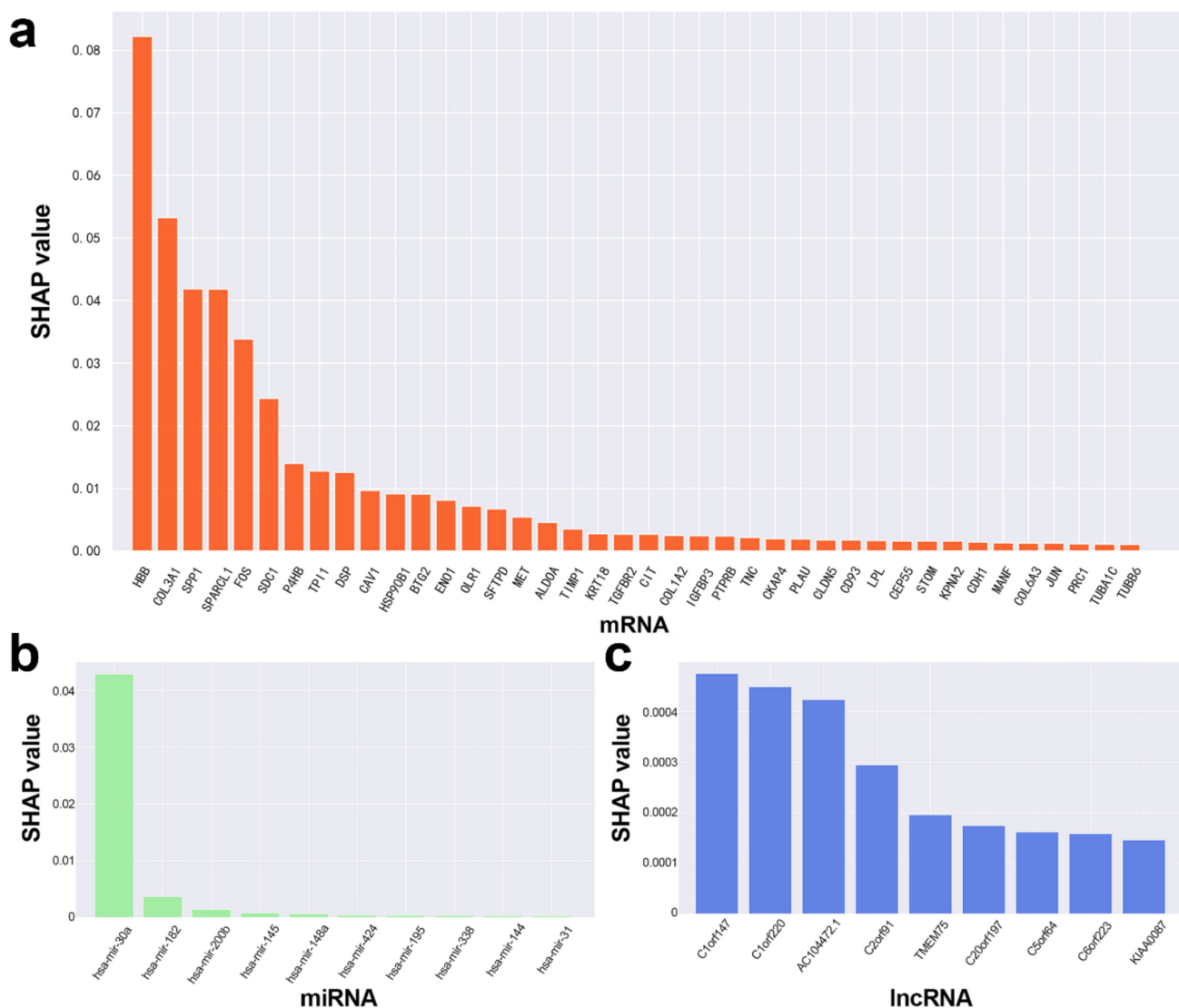


Figure 7. The SHAP values of top RNAs with SHAP value > 0.0001 in the ceRNA DNN model: (a) Top 40 mRNAs, (b) top 10 miRNA, and (c) top 9 lncRNA.

With FPKM of these potential ceRNAs as features for the input layer, another ceRNA DNN model was established and trained. The accuracy increased and the loss decreased as the training progressed (Figure 9), indicating no overfitting during the process. The learning process conformed to the law of deep learning, and the learning efficiency of deep learning ensured the reliability of these potential ceRNAs.

In the DisGeNet database (<https://disgenet.org/>, accessed on 9 April 2021), 37 potential mRNAs/genes were cancer/tumor related genes, among which 25 were directly related to lung tumors and complications. GO functional enrichment analysis showed that potential mRNAs mainly enriched in GO-BP functions such as regulated exocytosis, transforming growth factor beta receptor signaling, muscle structure development, and response to extracellular stimulus, referring to the phenomenon of metastases and lung tissue deterioration during cancer development. The functions of GO-MF were mainly cell adhesion molecule binding, SMAD binding and other membrane binding protein activity. These potential mRNAs were mainly enriched in functions directly related to the

proliferation, metastasis and diffusion of cancer cells, and were highly correlated with the occurrence and development of lung adenocarcinoma.

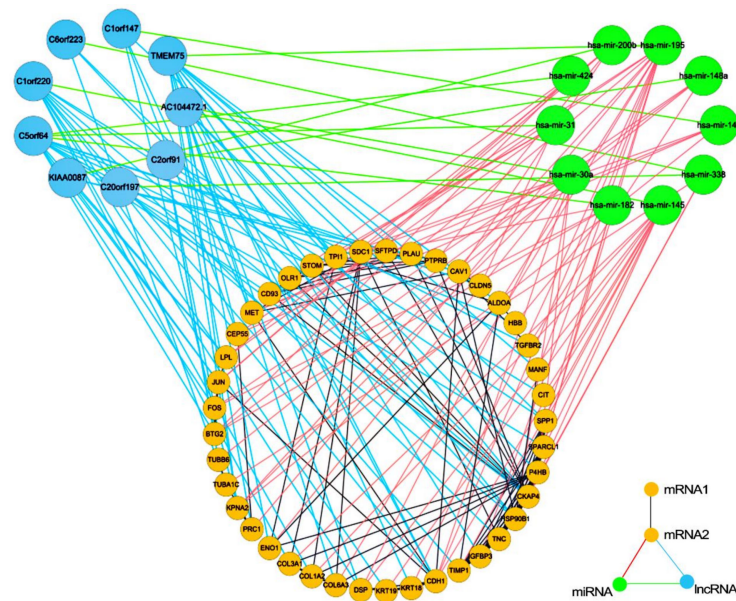


Figure 8. The potential ceRNA subnetwork with ceRNA regulatory and mRNA interaction relationships.

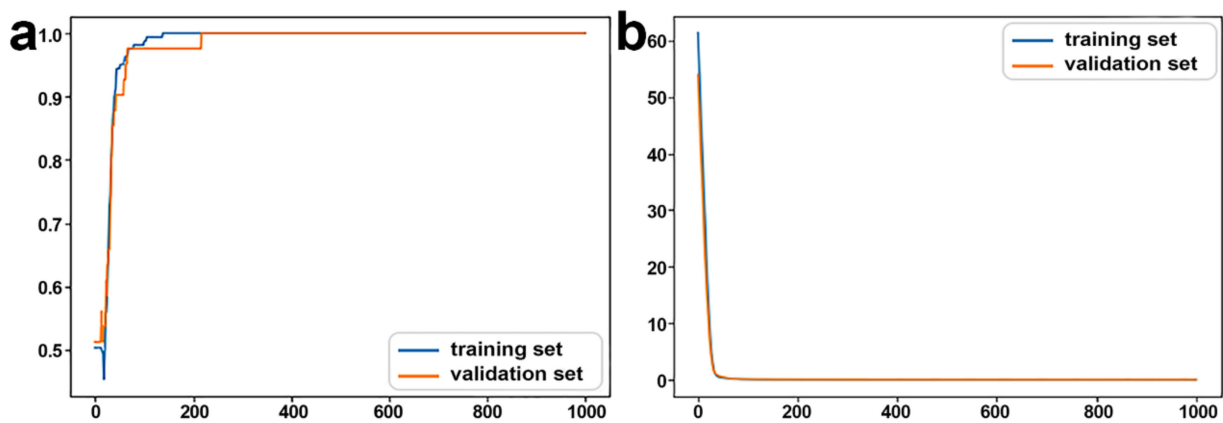
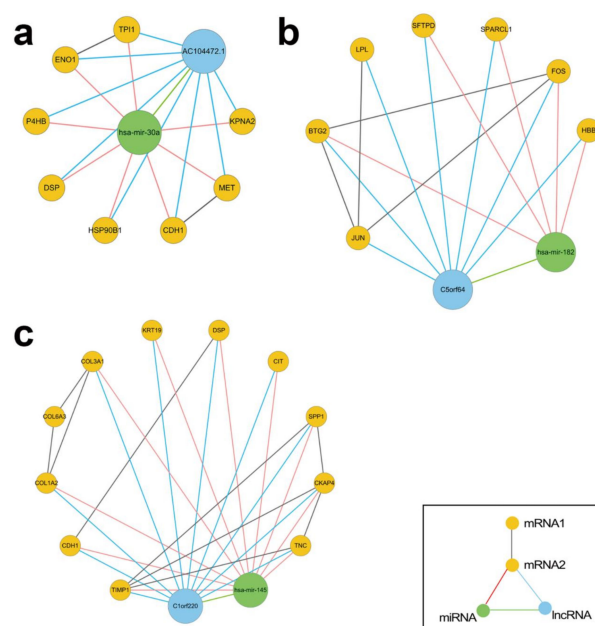


Figure 9. The accuracy and loss of the ceRNA DNN model for potential ceRNAs: (a) Accuracy curve and (b) loss curve.

### 3.3. Regulatory Modules for Lung Adenocarcinoma

Modules comprised of one miRNA and its regulated lncRNAs and mRNAs with the same up/down regulation direction were identified as regulatory modules for lung adenocarcinoma (Figure 10). MiRNA hsa-mir-30a and lncRNA AC104472.1 constituted regulatory module for lung adenocarcinoma a with TPI1, KPNA2, MET, HSP90B1, P4HB, DSP, CDH1, and ENO1 (Figure 10a). Among them, hsa-mir-30a had the highest contribution in potential miRNAs, and its expression level was significantly down-regulated. Hsa-mir-182, which ranked second in potential miRNA contribution, and lncRNA C5orf64 formed regulatory module for lung adenocarcinoma b (Figure 10b). Among the RNAs regulated by hsa-mir-182, the hemoglobin  $\beta$  coding gene HBB has the highest contribution in potential mRNAs. Hsa-mir-145 formed the third regulatory module for lung adenocarcinoma c with lncRNA C1orf220, which ranked second in potential lncRNAs, and mRNAs (COL1A2, COL3A1, SPP1, TIMP1 and CDH1) (Figure 10c).



**Figure 10.** Three regulatory modules for lung adenocarcinoma. (a) regulatory module for lung adenocarcinoma a: hsa-mir-30a with its regulated lncRNA AC104472.1 and mRNAs, (b) regulatory module for lung adenocarcinoma b: hsa-mir-182 with its regulated lncRNA C5orf64 and mRNAs, and (c) regulatory module for lung adenocarcinoma c: hsa-mir-145 with its regulated lncRNA C1orf220 and mRNAs.

These regulatory modules for lung adenocarcinoma were further validated by literature review, functional enrichment analysis and an independent dataset.

### 3.3.1. Literature Review

Related studies have shown that hsa-mir-30a, as an important regulator of tumor suppressors, could inhibit cell proliferation, migration, and invasion in vitro [22]. When its expression is down-regulated, cancer is more likely to become worse. The lncRNA AC104472.1 is one of the cancer lncRNAs of the immune-related function, and is a potential prognostic marker for the treatment of breast cancer [23]. Additionally, hsa-mir-30a has a potential role in regulating autophagy in cancer cells. Autophagy-related genes (e.g., P4HB) are overexpressed if hsa-mir-30a regulation is inhibited. Accelerated autophagy behavior of normal cells provides a microenvironment enriched with nutrients for cancer cells [24]. Current studies have shown that P4HB is overexpressed in all kinds of tumor cells and is an important indicator to detect the tumor progression level [25]. In addition to P4HB, the mRNA targets regulated by hsa-mir-30a also included many cancer therapeutic targets. For example, DSP is related to the growth and metastasis of cancer cells [26]. CDH1 is a widely known proto oncogene in the occurrence and development of cancer [27], while a relatively new oncogenic determinant MET regulates the occurrence, progression, and malignancy of epithelial carcinomas including lung adenocarcinoma [28].

Hsa-mir-182 was upregulated in lung and other cancers to promote cancer cell migration and invasion, and was found to have good potential for cancer diagnosis [29]. lncRNA C5orf64, which was regulated by hsa-mir-182, has recently been confirmed by a large number of bioinformatics methods to be significantly positively correlated with the abundance of immune neutrophils, and has the potential to regulate tumor microenvironment and help to reshape mutant patterns [30]. In lung adenocarcinoma, the dysfunction of HBB can directly lead to different degrees of anemia in patients, which has become a major problem in the treatment of lung adenocarcinoma [31]. In addition, BTG2 is a gene enriched in the Hallmark of angiogenesis and platelets, whose downregulation is directly related to the invasion of cancer cells [32]. SPARCL1, which was also down-regulated

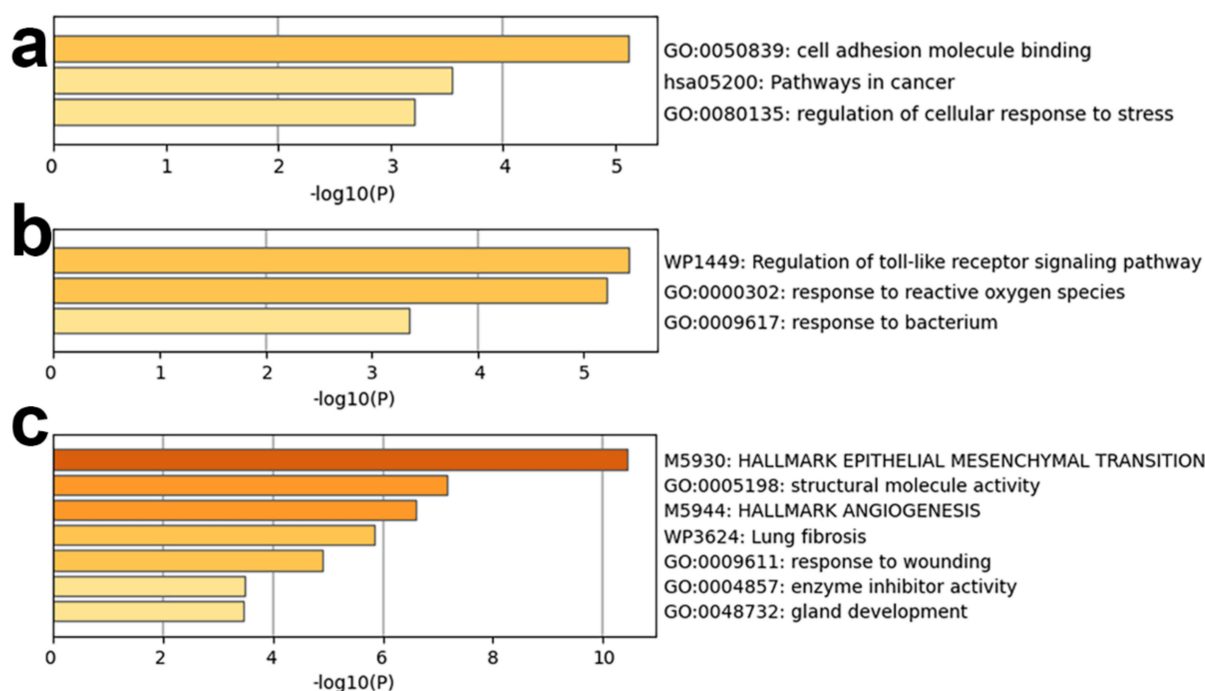
by hsa-mir-182, was proved to be able to optimize clinical efficacy by preventing tumor invasion and angiogenesis [33].

LncRNA C1orf220 was targeted by hsa-mir-145. Bioinformatics studies have shown that C1orf220 plays an important role in central gene regulation of lung squamous cell carcinoma [34]. COL3A1 and COL1A2 higher expressed in tumor samples were hub genes in a miRNA–gene interaction network and related to the survival time of lung adenocarcinoma [35]. Silencing SPP1 was found to reduce EGFR resistance to tyrosine kinase inhibitors and reduce its invasiveness in lung adenocarcinoma [36]. Tumor-derived protein tissue inhibitor of metalloproteinases-1 (TIMP1) correlates with poor prognosis in many cancers [37]. Wang et al. confirmed that TIMP1 regulated metabolism in metastases by activating the PI3K/Akt pathway and found TIMP1 as a potential biomarker for understanding lung adenocarcinoma pathogenesis [38].

These regulatory modules for lung adenocarcinoma provided a better insight into the regulatory role they play in the initiation of lung adenocarcinoma.

### 3.3.2. Functional Enrichment Analysis

GO, KEGG, wikiPathways and Hallmark functional enrichment analysis was performed for mRNAs in regulatory modules for lung adenocarcinoma (Figure 11). GO functional enrichment analysis showed that mRNAs in each module mainly enriched in GO functions such as cell adhesion molecule binding, response to reactive oxygen species, enzyme inhibitor activity, etc. These functions are all critical factors involved in cancers. The regulation roles of them in lung adenocarcinoma has been studied [39–41].



**Figure 11.** Functional enrichment analysis results of mRNAs in three regulatory modules for lung adenocarcinoma: (a–c) are results for three regulatory modules for lung adenocarcinoma a, b, and c, respectively.

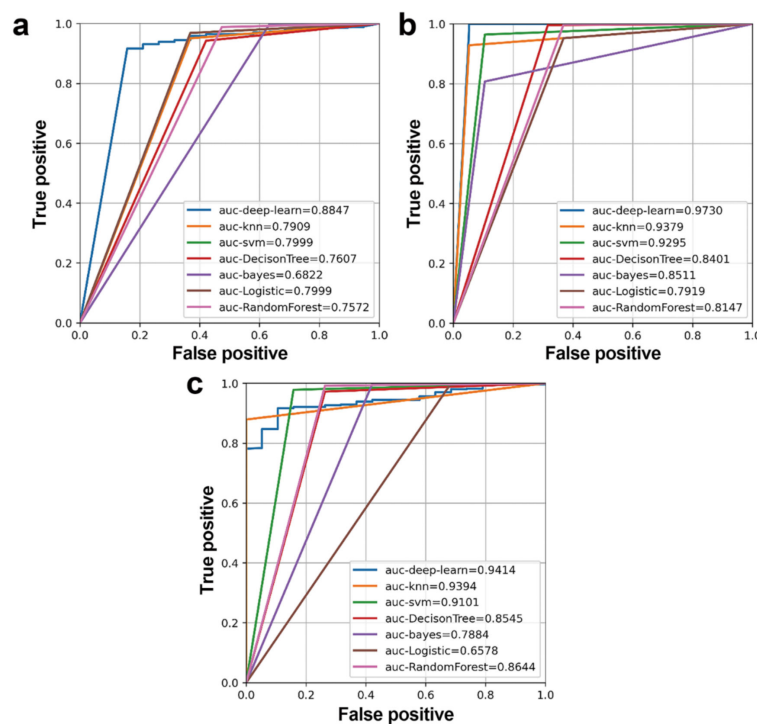
For KEGG pathway enrichment analysis, mRNAs in the first regulatory module for lung adenocarcinoma were mainly enriched in pathways in cancer. In addition, the results of enrichment by wikiPathways showed that mRNAs in the other two regulatory modules for lung adenocarcinoma were mainly enriched in Regulation of toll-like receptor signaling pathway and Lung fibrosis. Toll-like receptors (TLRs), such as Toll-Like Receptor 2 and 4, are pillars of the immune system that have been linked to several forms of malignancy

including lung adenocarcinoma [42,43]. Lung fibrosis has been reported to be a risk factor for developing lung carcinogenesis [44].

MRNAs in the third regulatory module for lung adenocarcinoma were enriched mainly in the Hallmark gene set of EPITHELIAL MESENCHYMAL TRANSITION and ANGIOGENESIS. Lung adenocarcinoma could cause severe epithelial-mesenchymal transition disorder, leading to pulmonary epithelial cell dysfunction [45]. Angiogenesis is the process of capillary sprouting from pre-existing vessels and plays a critical role in the carcinogenic process of lung adenocarcinoma [46].

### 3.3.3. Independent Dataset Validation

To further demonstrate the effectiveness of identified regulatory modules for lung adenocarcinoma, the DNN model was applied to classify samples of an independent dataset TCGA-LUAD using three regulatory modules for lung adenocarcinoma, respectively. In order to further prove that the DNN model was superior to traditional machine learning methods, KNN, SVM, decision tree, Multi-feature Bayesian, Logistic regression, and random forest were applied to the independent dataset, respectively. The ROC curve was drawn according to a series of different cut-off values with true positive as the ordinate and false positive as the abscissa (Figure 12).



**Figure 12.** ROC curves and AUC values of regulatory modules for lung adenocarcinoma for the independent dataset using different machine learning methods. (a–c) are results for three regulatory modules for lung adenocarcinoma a, b, and c, respectively.

Most machine learning methods had good classification performance ( $AUC > 0.75$ ), while the DNN model had the best performance, demonstrating the effectiveness and diagnostic values of all three regulatory modules for unpaired lung adenocarcinoma samples.

## 4. Discussion

In this paper, three regulatory modules for lung adenocarcinoma were identified from multiple DNN models using expression data in interaction and ceRNA networks. They participated in the carcinogenesis of lung adenocarcinoma by regulation of miRNAs to mRNAs and lncRNAs. These modules were further validated in literature and an independent dataset and were expected to be used for lung adenocarcinoma diagnosis. The main

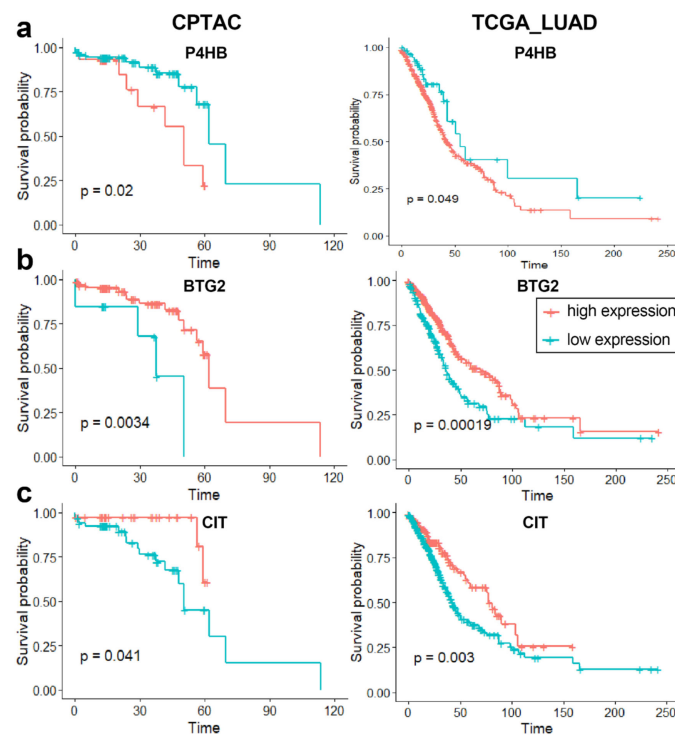
advantage of DNN is that it can modify the multidimensional weight of each feature during the learning process. Regulatory modules for lung adenocarcinoma identified using DNN models from paired samples of CPTAC could distinguish disease samples from normal ones for unpaired samples of the TCGA-LUAD dataset, indicating the effectiveness for both paired and unpaired samples. The higher accuracy might come from the power of DNN in describing sophisticated relationships between genes, while the simple classification rules that traditional machine learning methods used may be not capable enough. The novelty of our study was summarized as follows: (1) The expression data was acquired from CPTAC and TCGA, which were all large multi-omic datasets for various cancers. This made our results covered most scenario.; (2) The ceRNA network constructed based on experimentally validated miRNA targets and correlations could better explain the respective roles of different RNAs in biological processes; (3) Potential RNAs were screened by multiple mRNAs and ceRNAs DNN models, which could intelligently recognize the sophisticated relationships between RNAs appropriately.

Regulatory modules for lung adenocarcinoma regulated the initiation process of lung adenocarcinoma through ceRNAs relationships. All mRNAs of regulatory modules for lung adenocarcinoma have been validated to be associated with cancer in DisGeNet, of which 15 were directly related to lung cancer and its complications. lncRNAs and miRNAs were further searched in ncRNA-disease association databases to further exhibit their disease association. Two lncRNAs of regulatory modules for lung adenocarcinoma were stored in lncRNADisease v2.0 as non-small cell lung cancer associated, which integrated comprehensive experimentally supported and predicted ncRNA-disease associations curated from manual literatures and other resources [47]. All miRNAs of regulatory modules for lung adenocarcinoma were associated with lung adenocarcinoma in HMDD (the Human microRNA Disease Database) v3.2 curating experimentally supported miRNA and disease association data [48]. Experimental validation is needed for other RNAs in regulatory modules for lung adenocarcinoma in the future.

The regulatory modules for lung adenocarcinoma we identified were not population specific, since the CPTAC data contained 66 white, 2 black or African American, 2 Asian, and 130 not reported samples according to the race data from the clinical information. To study the regulatory relationships of RNAs in the three regulatory modules for specific population, we recalculated the correlations between RNAs for different races respectively. Most RNA pairs for white samples showed the same regulatory relationships (the black or African American and Asian samples were omitted because the number of them were too small to be correlated, Table S1). Then similar progress was performed to the TCGA-LUAD dataset according to the race data, i.e., 456 white, 58 black or African American, 8 Asian and 29 not reported samples. Similar results were obtained for most cases of white and black or African American samples (Table S1). Regulatory relationships for Asian samples (the correlation analysis was not performed for hsa-mir-30a since its expression value in Asian samples is zero) were not the same as those in the regulatory modules, probably due to the small sample size. As a result, we searched for other datasets with miRNA, mRNA and lncRNA expression data of Asian samples in the GEO database, but none was obtained. Only one expression profile with mRNA and miRNA expression data, GSE128311, was found. This dataset had 77 Asian samples. It was found that the regulatory relationships of GSE128311 were consistent with the regulatory relationships in the three regulatory modules (only the correlations between miRNAs and mRNAs were calculated, as no lncRNA was in this dataset, Table S1). The above results did not illustrate the population specificity of the regulatory modules for lung adenocarcinoma.

The prognosis value of regulatory modules for lung adenocarcinoma was then evaluated using the univariate Kaplan–Meier survival analysis and multivariate Cox regression. Kaplan–Meier curve showed similar results for some mRNAs of regulatory modules for lung adenocarcinoma in CPTAC and TCGA-LUAD datasets. For example, high expression of P4HB as well as low expression of BTG2 and CIT was associated with poor overall survival of lung adenocarcinoma patients ( $p$ -value of log-rank test < 0.05, Figure 13).

Clinical application of these genes was worth further study due to their diagnostic and prognostic value.



**Figure 13.** Kaplan–Meier survival analysis results for (a) P4HB, (b) BTG2 and (c) CIT in CPTAC and TCGA-LUAD datasets.

Other RNAs of regulatory modules for lung adenocarcinoma showed better results in the TCGA-LUAD dataset. Kaplan–Meier survival analysis in the TCGA-LUAD dataset showed that the overall survival of patients with the low expression hsa-mir-30a was poorer than that with the high expression ( $p < 0.05$ ). Moreover, multivariate Cox analysis confirmed that each module was a risk factor for overall survival among patients in the TCGA-LUAD cohort ( $p < 0.05$ ). A probable reason for better TCGA-LUAD prognostic results was that it was a more balanced dataset for different stages, while most patients of CPTAC were in their early stage. In spite of this, regulatory modules for lung adenocarcinoma could still be used for survival analysis in more balanced datasets for different stages. Further experimental validation for the regulatory relationships in lung adenocarcinoma would also be necessary.

In this paper, ceRNA networks were constructed according to experimentally validated miRNA targets from multiple databases, making the size of ceRNA networks small. The interactions predicted by confident computational tools should be taken into the analysis to improve the effectiveness and reliability of regulatory modules for lung adenocarcinoma.

## 5. Conclusions

To sum up, three regulatory modules for lung adenocarcinoma were identified using expression data by multiple DNN models in biological networks. MiRNAs hsa-mir-30a, hsa-mir-182, hsa-mir-145 and their regulated lncRNAs and mRNAs participated in the carcinogenesis of lung adenocarcinoma through ceRNAs relationships. These regulatory modules showed the relationship with lung adenocarcinoma in terms of expression levels, functions, pathways, and literature. They could distinguish disease samples from normal ones, and thus had potential for lung adenocarcinoma diagnosis. The regulatory relationships between RNAs were validated in various datasets, including CPTAC, TCGA and an expression profile from the GEO database. These regulatory modules had potential values for clinic prognosis. Our study will contribute to improving the understanding of



the ceRNA network regulatory mechanisms in the carcinogenesis of lung adenocarcinoma and provide schemes for identifying novel regulatory modules of other cancers.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biology11091291/s1>, Table S1: The regulatory relationships of the regulatory modules for different races in CTPAC, TCGA and GSE128311.

**Author Contributions:** Conceptualization, W.L., L.C. and W.H.; Data Curation, J.L. and X.W. (Xinyan Wang); Formal Analysis, K.L.; Funding Acquisition, W.L. and L.C.; Investigation, L.F. and W.L.; Methodology, L.F. and K.L.; Project Administration, L.C.; Software, K.L.; Supervision, L.C.; Validation, S.Q., Z.Z., S.S., X.W. (Xu Wang), B.Y. and Y.H.; Visualization, L.F. and W.L.; Writing—Original Draft Preparation, W.L.; Writing—Review and Editing, W.L. and L.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was funded by the National Natural Science Foundation of China [grant numbers 61702141 and 81627901]; the Natural Science Foundation of Heilongjiang Province [grant number LH2021F043]; the Heilongjiang Postdoctoral Funds for Scientific Research Initiation [grant number LBH-Q17132].

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. This data can be found here: <https://proteomics.cancer.gov/programs/cptac> (accessed on 20 March 2021).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Hutchinson, B.D.; Shroff, G.S.; Truong, M.T.; Ko, J.P. Spectrum of Lung Adenocarcinoma. *Semin. Ultrasound CT MR* **2019**, *40*, 255–264. [[CrossRef](#)] [[PubMed](#)]
- Denisenko, T.V.; Budkevich, I.N.; Zhivotovsky, B. Cell death-based treatment of lung adenocarcinoma. *Cell Death Dis.* **2018**, *9*, 117. [[CrossRef](#)] [[PubMed](#)]
- Kuhn, E.; Morbini, P.; Cancellieri, A.; Damiani, S.; Cavazza, A.; Comin, C.E. Adenocarcinoma classification: Patterns and prognosis. *Pathologica* **2018**, *110*, 5–11. [[PubMed](#)]
- Zhang, Y.; Huang, Y.X.; Wang, D.L.; Yang, B.; Yan, H.Y.; Lin, L.H.; Li, Y.; Chen, J.; Xie, L.M.; Huang, Y.S.; et al. LncRNA DSCAM-AS1 interacts with YBX1 to promote cancer progression by forming a positive feedback loop that activates FOXA1 transcription network. *Theranostics* **2020**, *10*, 10823–10837. [[CrossRef](#)] [[PubMed](#)]
- Athanasios, A.; Charalampos, V.; Vasileios, T.; Ashraf, G.M. Protein-Protein Interaction (PPI) Network: Recent Advances in Drug Discovery. *Curr. Drug Metab.* **2017**, *18*, 5–10. [[CrossRef](#)]
- Al-Harazi, O.; Kaya, I.H.; El Allali, A.; Colak, D. A Network-Based Methodology to Identify Subnetwork Markers for Diagnosis and Prognosis of Colorectal Cancer. *Front. Genet.* **2021**, *12*, 721949. [[CrossRef](#)]
- Khedkar, H.N.; Wang, Y.C.; Yadav, V.K.; Srivastava, P.; Lawal, B.; Mokgautsi, N.; Sumitra, M.R.; Wu, A.T.H.; Huang, H.S. In-Silico Evaluation of Genetic Alterations in Ovarian Carcinoma and Therapeutic Efficacy of NSC777201, as a Novel Multi-Target Agent for TTK, NEK2, and CDK1. *Int. J. Mol. Sci.* **2021**, *22*, 5895. [[CrossRef](#)]
- Roudi, R.; Beikzadeh, B.; Roviello, G.; D'angelo, A.; Hadizadeh, M. Identification of hub genes, modules and biological pathways associated with lung adenocarcinoma: A system biology approach. *Gene Rep.* **2022**, *27*, 101638. [[CrossRef](#)]
- Qi, X.; Zhang, D.H.; Wu, N.; Xiao, J.H.; Wang, X.; Ma, W. ceRNA in cancer: Possible functions and clinical implications. *J. Med. Genet.* **2015**, *52*, 710–718. [[CrossRef](#)]
- Jafarinejad-Farsangi, S.; Jazi, M.M.; Rostamzadeh, F.; Hadizadeh, M. High affinity of host human microRNAs to SARS-CoV-2 genome: An in silico analysis. *Noncoding RNA Res.* **2020**, *5*, 222–231. [[CrossRef](#)] [[PubMed](#)]
- Zhang, H.D.; Jiang, L.H.; Sun, D.W.; Li, J.; Ji, Z.L. The role of miR-130a in cancer. *Breast Cancer* **2017**, *24*, 521–527. [[CrossRef](#)] [[PubMed](#)]
- Li, F.; Huang, C.; Li, Q.; Wu, X. Construction and Comprehensive Analysis for Dysregulated Long Non-Coding RNA (lncRNA)-Associated Competing Endogenous RNA (ceRNA) Network in Gastric Cancer. *Med. Sci. Monit.* **2018**, *24*, 37–49. [[CrossRef](#)] [[PubMed](#)]
- Hu, Z.; Chen, J.; Tian, T.; Zhou, X.; Gu, H.; Xu, L.; Zeng, Y.; Miao, R.; Jin, G.; Ma, H.; et al. Genetic variants of miRNA sequences and non-small cell lung cancer survival. *J. Clin. Investig.* **2008**, *118*, 2600–2608. [[CrossRef](#)] [[PubMed](#)]
- Kather, J.N.; Pearson, A.T.; Halama, N.; Jager, D.; Krause, J.; Loosen, S.H.; Marx, A.; Boor, P.; Tacke, F.; Neumann, U.P.; et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat. Med.* **2019**, *25*, 1054–1056. [[CrossRef](#)]

15. Hirasawa, T.; Aoyama, K.; Tanimoto, T.; Ishihara, S.; Shichijo, S.; Ozawa, T.; Ohnishi, T.; Fujishiro, M.; Matsuo, K.; Fujisaki, J.; et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. *Gastric Cancer* **2018**, *21*, 653–660. [[CrossRef](#)]
16. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [[CrossRef](#)]
17. Huys, Q.J.; Maia, T.V.; Frank, M.J. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat. Neurosci.* **2016**, *19*, 404–413. [[CrossRef](#)]
18. Varet, H.; Brilllet-Gueguen, L.; Coppee, J.Y.; Dillies, M.A. SARTools: A DESeq2- and EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data. *PLoS ONE* **2016**, *11*, e0157022. [[CrossRef](#)]
19. Nguyen, L.C.; Nguyen-Xuan, H. Deep learning for computational structural optimization. *ISA Trans.* **2020**, *103*, 177–191. [[CrossRef](#)]
20. Park, D.J.; Park, M.W.; Lee, H.; Kim, Y.J.; Kim, Y.; Park, Y.H. Development of machine learning model for diagnostic disease prediction based on laboratory tests. *Sci. Rep.* **2021**, *11*, 7567. [[CrossRef](#)]
21. Luan, H.; Zhang, C.; Zhang, T.; He, Y.; Su, Y.; Zhou, L. Identification of Key Prognostic Biomarker and Its Correlation with Immune Infiltrates in Pancreatic Ductal Adenocarcinoma. *Dis. Markers* **2020**, *2020*, 8825997. [[CrossRef](#)] [[PubMed](#)]
22. Saleh, A.D.; Cheng, H.; Martin, S.E.; Si, H.; Ormanoglu, P.; Carlson, S.; Clavijo, P.E.; Yang, X.; Das, R.; Cornelius, S.; et al. Integrated Genomic and Functional microRNA Analysis Identifies miR-30-5p as a Tumor Suppressor and Potential Therapeutic Nanomedicine in Head and Neck Cancer. *Clin. Cancer Res.* **2019**, *25*, 2860–2873. [[CrossRef](#)] [[PubMed](#)]
23. Liu, Z.; Mi, M.; Li, X.; Zheng, X.; Wu, G.; Zhang, L. lncRNA OSTN-AS1 May Represent a Novel Immune-Related Prognostic Marker for Triple-Negative Breast Cancer Based on Integrated Analysis of a ceRNA Network. *Front. Genet.* **2019**, *10*, 850. [[CrossRef](#)] [[PubMed](#)]
24. Cheng, Y.; Chen, G.; Hu, M.; Huang, J.; Li, B.; Zhou, L.; Hong, L. Has-miR-30a regulates autophagic activity in cervical cancer upon hydroxycamptothecin exposure. *Biomed. Pharm.* **2015**, *75*, 67–74. [[CrossRef](#)]
25. Xie, L.; Li, H.; Zhang, L.; Ma, X.; Dang, Y.; Guo, J.; Liu, J.; Ge, L.; Nan, F.; Dong, H.; et al. Autophagy-related gene P4HB: A novel diagnosis and prognosis marker for kidney renal clear cell carcinoma. *Aging (Albany NY)* **2020**, *12*, 1828–1842. [[CrossRef](#)]
26. Wang, H.; Wu, M.; Lu, Y.; He, K.; Cai, X.; Yu, X.; Lu, J.; Teng, L. lncRNA MIR4435-2HG targets desmoplakin and promotes growth and metastasis of gastric cancer by activating Wnt/beta-catenin signaling. *Aging (Albany NY)* **2019**, *11*, 6657–6673. [[CrossRef](#)]
27. Ye, T.; Li, J.; Sun, Z.; Liu, D.; Zeng, B.; Zhao, Q.; Wang, J.; Xing, H.R. Cdh1 functions as an oncogene by inducing self-renewal of lung cancer stem-like cells via oncogenic pathways. *Int. J. Biol. Sci.* **2020**, *16*, 447–459. [[CrossRef](#)]
28. Yao, H.P.; Hudson, R.; Wang, M.H. Progress and challenge in development of biotherapeutics targeting MET receptor for treatment of advanced cancer. *Biochim Biophys Acta. Rev. Cancer* **2020**, *1874*, 188425. [[CrossRef](#)]
29. Lin, G.; Li, J.; Cai, J.; Zhang, H.; Xin, Q.; Wang, N.; Xie, W.; Zhang, Y.; Xu, N. RNA-binding Protein MBNL2 regulates Cancer Cell Metastasis through MiR-182-MBNL2-AKT Pathway. *J. Cancer* **2021**, *12*, 6715–6726. [[CrossRef](#)]
30. Pang, Z.; Chen, X.; Wang, Y.; Wang, Y.; Yan, T.; Wan, J.; Wang, K.; Du, J. Long non-coding RNA C5orf64 is a potential indicator for tumor microenvironment and mutation pattern remodeling in lung adenocarcinoma. *Genom.* **2021**, *113*, 291–304. [[CrossRef](#)]
31. Pirker, R.; Wiesenberger, K.; Pohl, G.; Minar, W. Anemia in lung cancer: Clinical impact and management. *Clin. Lung Cancer* **2003**, *5*, 90–97. [[CrossRef](#)] [[PubMed](#)]
32. Chen, Z.; Chen, X.; Lu, B.; Gu, Y.; Chen, Q.; Lei, T.; Nie, F.; Gu, J.; Huang, J.; Wei, C.; et al. Up-regulated LINC01234 promotes non-small-cell lung cancer cell metastasis by activating VAV3 and repressing BTG2 expression. *J. Hematol Oncol.* **2020**, *13*, 7. [[CrossRef](#)] [[PubMed](#)]
33. Gagliardi, F.; Narayanan, A.; Gallotti, A.L.; Pieri, V.; Mazzoleni, S.; Cominelli, M.; Rezzola, S.; Corsini, M.; Brugnara, G.; Altabella, L.; et al. Enhanced SPARCL1 expression in cancer stem cells improves preclinical modeling of glioblastoma by promoting both tumor infiltration and angiogenesis. *Neurobiol. Dis.* **2020**, *134*, 104705. [[CrossRef](#)] [[PubMed](#)]
34. Shi, Y.; Li, Y.; Yan, C.; Su, H.; Ying, K. Identification of key genes and evaluation of clinical outcomes in lung squamous cell carcinoma using integrated bioinformatics analysis. *Oncol. Lett.* **2019**, *18*, 5859–5870. [[CrossRef](#)] [[PubMed](#)]
35. Yu, D.H.; Ruan, X.L.; Huang, J.Y.; Liu, X.P.; Ma, H.L.; Chen, C.; Hu, W.D.; Li, S. Analysis of the Interaction Network of Hub miRNAs-Hub Genes, Being Involved in Idiopathic Pulmonary Fibers and Its Emerging Role in Non-small Cell Lung Cancer. *Front. Genet.* **2020**, *11*, 302. [[CrossRef](#)] [[PubMed](#)]
36. Wang, X.; Zhang, F.; Yang, X.; Xue, M.; Li, X.; Gao, Y.; Liu, L. Secreted Phosphoprotein 1 (SPP1) Contributes to Second-Generation EGFR Tyrosine Kinase Inhibitor Resistance in Non-Small Cell Lung Cancer. *Oncol. Res.* **2019**, *27*, 871–877. [[CrossRef](#)]
37. Schoeps, B.; Eckfeld, C.; Prokopchuk, O.; Bottcher, J.; Haussler, D.; Steiger, K.; Demir, I.E.; Knolle, P.; Soehnlein, O.; Jenne, D.E.; et al. TIMP1 Triggers Neutrophil Extracellular Trap Formation in Pancreatic Cancer. *Cancer Res.* **2021**, *81*, 3568–3579. [[CrossRef](#)]
38. Wang, Z.; Li, Z.; Zhou, K.; Wang, C.; Jiang, L.; Zhang, L.; Yang, Y.; Luo, W.; Qiao, W.; Wang, G.; et al. Deciphering cell lineage specification of human lung adenocarcinoma with single-cell RNA sequencing. *Nat. Commun.* **2021**, *12*, 6500. [[CrossRef](#)]
39. Yoshimoto, T.; Matsubara, D.; Soda, M.; Ueno, T.; Amano, Y.; Kihara, A.; Sakatani, T.; Nakano, T.; Shibano, T.; Endo, S.; et al. Mucin 21 is a key molecule involved in the incohesive growth pattern in lung adenocarcinoma. *Cancer Sci.* **2019**, *110*, 3006–3011. [[CrossRef](#)]

40. Kasiri, S.; Chen, B.; Wilson, A.N.; Reczek, A.; Mazambani, S.; Gadhvi, J.; Noel, E.; Marriam, U.; Mino, B.; Lu, W.; et al. Stromal Hedgehog pathway activation by IHH suppresses lung adenocarcinoma growth and metastasis by limiting reactive oxygen species. *Oncogene* **2020**, *39*, 3258–3275. [[CrossRef](#)] [[PubMed](#)]
41. Hsu, H.L.; Lee, C.H.; Chen, C.H.; Zhan, J.F.; Wu, S.Y. Angiotensin-converting enzyme inhibitors and angiotensin II receptor blockers might be associated with lung adenocarcinoma risk: A nationwide population-based nested case-control study. *Am. J. Transl. Res.* **2020**, *12*, 6615–6625. [[CrossRef](#)] [[PubMed](#)]
42. Gergen, A.K.; Kohtz, P.D.; Halpern, A.L.; Li, A.; Meng, X.; Reece, T.B.; Fullerton, D.A.; Weyant, M.J. Activation of Toll-Like Receptor 2 Promotes Proliferation of Human Lung Adenocarcinoma Cells. *Anticancer Res.* **2020**, *40*, 5361–5369. [[CrossRef](#)] [[PubMed](#)]
43. Li, Y.; Yang, W.; Wu, B.; Liu, Y.; Li, D.; Guo, Y.; Fu, H.; Li, Y. KDM3A promotes inhibitory cytokines secretion by participating in TLR4 regulation of Foxp3 transcription in lung adenocarcinoma cells. *Oncol. Lett.* **2017**, *13*, 3529–3537. [[CrossRef](#)] [[PubMed](#)]
44. Karampitsakos, T.; Tzilas, V.; Tringidou, R.; Steiropoulos, P.; Aidinis, V.; Papiris, S.A.; Bouros, D.; Tzouveleki, A. Lung cancer in patients with idiopathic pulmonary fibrosis. *Pulm. Pharmacol. Ther.* **2017**, *45*, 1–10. [[CrossRef](#)] [[PubMed](#)]
45. Sarode, P.; Mansouri, S.; Karger, A.; Schaefer, M.B.; Grimminger, F.; Seeger, W.; Savai, R. Epithelial cell plasticity defines heterogeneity in lung cancer. *Cell Signal.* **2020**, *65*, 109463. [[CrossRef](#)]
46. Cai, S.; Guo, X.; Huang, C.; Deng, Y.; Du, L.; Liu, W.; Yang, C.; Zhao, H.; Ma, K.; Wang, L.; et al. Integrative analysis and experiments to explore angiogenesis regulators correlated with poor prognosis, immune infiltration and cancer progression in lung adenocarcinoma. *J. Transl. Med.* **2021**, *19*, 361. [[CrossRef](#)]
47. Bao, Z.; Yang, Z.; Huang, Z.; Zhou, Y.; Cui, Q.; Dong, D. LncRNADisease 2.0: An updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res.* **2019**, *47*, D1034–D1037. [[CrossRef](#)]
48. Li, Y.; Qiu, C.; Tu, J.; Geng, B.; Yang, J.; Jiang, T.; Cui, Q. HMDD v2.0: A database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res.* **2014**, *42*, D1070–D1074. [[CrossRef](#)]