Mathematical Biosciences and Engineering

MBE, 20(6): 9818–9838.
DOI: 10.3934/mbe.2023430
Received: 05 January 2023
Revised: 15 February 2023
Accepted: 01 March 2023
Published: 24 March 2023

http://www.aimspress.com/journal/MBE

Research article

Systematic analysis of lncRNA gene characteristics based on PD-1 immune related pathway for the prediction of non-small cell lung cancer prognosis

Hejian Chen[1], Shuiyu Xu[2], Yuhong Zhang[2] and Peifeng Chen[1,*]

[1] Department of Respiratory and Critical Care Medicine, Zhuji People's Hospital of Zhejiang Province, Zhuji 311800, China
[2] Department of Oncology, HaploX Biotechnology, Shenzhen 518035, China

* **Correspondence:** Email: chenpeifengsxzj@sina.com.

Abstract: *Background:* Non-small cell lung cancer (NSCLC) is heterogeneous. Molecular subtyping based on the gene expression profiles is an effective technique for diagnosing and determining the prognosis of NSCLC patients. *Methods:* Here, we downloaded the NSCLC expression profiles from The Cancer Genome Atlas and the Gene Expression Omnibus databases. ConsensusClusterPlus was used to derive the molecular subtypes based on long-chain noncoding RNA (lncRNA) associated with the PD-1-related pathway. The LIMMA package and least absolute shrinkage and selection operator (LASSO)-Cox analysis were used to construct the prognostic risk model. The nomogram was constructed to predict the clinical outcomes, followed by decision curve analysis (DCA) to validate the reliability of this nomogram. *Results:* We discovered that PD-1 was strongly and positively linked to the T-cell receptor signaling pathway. Furthermore, we identified two NSCLC molecular subtypes yielding a significantly distinctive prognosis. Subsequently, we developed and validated the 13-lncRNA-based prognostic risk model in the four datasets with high AUC values. Patients with low-risk showed a better survival rate and were more sensitive to PD-1 treatment. Nomogram construction combined with DCA revealed that the risk score model could accurately predict the prognosis of NSCLC patients. *Conclusions:* This study demonstrated that lncRNAs engaged in the T-cell receptor signaling pathway played a significant role in the onset and development of NSCLC, and that they could influence the sensitivity to PD-1 treatment. In addition, the 13 lncRNA model was effective in assisting clinical treatment decision-making and prognosis evaluation.

Keywords: non-small cell lung cancer; PD-1; long non-coding RNAs; molecular subtypes; risk model

1. Introduction

Lung cancer is responsible for the world's highest mortality rate [1], accounting for > 25% of cancer death; 82% of those deaths are caused by smoking [2]. Approximately 85% of the lung cancer tissue types are non-small cell lung cancer (NSCLC) [3]. As lung cancer does not display obvious symptoms in its early stages, a majority of the patients have progressed to the advanced stages. Metastasis occurs early in these patients with a poor overall survival rate [4]. Despite the recent advances in early detection, radiotherapy, chemotherapy, and surgical biopsy and treatment, lung cancer patients still show a poor prognosis.

Some of the NSCLC treatment strategies include the development of particular antibodies against the cytotoxic T lymphocyte-related protein 4 receptor, programmed death (PD-1) receptor and programmed death ligand-1 (PD-L1), and the implementation of these strategies in the first and second lines of treatment has significantly improved the survival rate of NSCLC patients [5]. For patients with advanced NSCLC and without molecular derivatives, conventional dual drug-based chemotherapy has been replaced by combined or non-combined chemotherapy with immunotherapy as the first treatment strategy [6].

The earlier reports have indicated that the implementation of the PD-1/PD-L1 strategy, in particular, has considerably increased the survival rate of the NSCLC patients. It was noted that the treatment with anti-PD-1/PD-L1 interrupted the association between PD-1 and its ligand, disrupted inhibitory signal transduction, restored T cell viability, thereby reactivating the anti-tumor immunological response [7]. However, the efficacy of PD-1/PD-L1 inhibitors differs due to high variation in the molecular immune subtypes and immunological microenvironment [8]. Several factors like the neoantigens, tumor infiltrating lymphocytes, PD-L1 expression levels, tumor mutation burden, and driver gene mutations, influence the success of the anti-PD-1/PD-L1 therapy. Hence, additional investigation into the biomarkers could help to select the most appropriate form of therapy for the patients and to correctly predict the efficiency of anti-PD-1/PD-L1 treatment [8].

Long noncoding RNA (lncRNA) is described as the type of RNA that has a length of > 200 nucleotide units [9]; however, it cannot encode complete proteins. Owing to the success of the Human Genome Project and wide application of second-generation sequencing, many researchers have begun to recognize the significant role of lncRNAs in many human diseases, particularly in cancers [10]. A few studies have identified their involvement in the pathophysiological processes of different cancers or their roles as diagnostic or clinical prognostic markers [11]. LncRNAs can regulate gene expression at RNA transcription, post transcription and epigenetic levels. Abnormal expression of lncRNAs in NSCLC patients is closely related to the onset and development of lung cancer. It can alter the drug resistance and radiation sensitivity of NSCLC patients by acting as an oncogene or tumor suppressor gene, regulating the proliferation, invasion and migration of lung cancer cells via various mechanisms. Currently, the predictive role of lncRNA gene characteristics remains poorly elucidated in patients with NSCLC. Hence, further investigation is needed to discover novel lncRNA gene characteristics for prognosis and treatments for NSCLC patients [12].

Here, we have derived the gene expression data from different public databases like the Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA). The PD-1-related immune pathways were identified by using single sample gene set enrichment analysis (ssGSEA), and we determined the molecular subtypes of NSCLC based on the lncRNA of the PD-1-associated immune pathways. Also, we assessed the association between the different molecular subtypes and the clinical features and prognosis of this disease. To assess the NSCLC prognosis, a prognostic risk model was developed by using the differential expressed genes (DEGs) among NSCLC subtypes. Furthermore, we constructed a nomogram to assist clinical decision-making and prognosis evaluation. Variations between the molecular subtypes and immunotherapy/ chemotherapy sensitivity were analyzed.

2. Materials and methods

2.1. Collection of the gene expression profile data and their preprocessing

In this study, we downloaded the RNA-Seq data, clinical survival data and all of the information regarding the characteristics of 993 TCGA-LUAD and LUSC samples, as based on the TCGA GDC API. Thereafter, we downloaded the GSE31210 dataset (226 samples), GSE19188 (82 samples) and GSE50081 (181 samples) chip datasets from the GEO database.

For the TCGA-LUAD and LUSC data, the fragments per kilobase of transcript per million fragments mapped values were converted to transcripts per million values. We eliminated samples that did not have any clinical follow-up data, survival times, or status. Human GFF3 annotation files from the Gencode website (https://www.gencodegenes.org/) were downloaded to extract the corresponding relationship between ensemble IDs and gene symbols: we then converted ensemble IDs to the gene symbols. Expression of multiple gene symbols was considered as their mean value. We used the removeBatchEffect function of the LIMMA package (Version 3.44.3) [13] to eliminate the batch effect between various datasets (Figure S1).

The following steps were implemented to process the NSCLC-GEO dataset: a) Retain the NSCLC tissue samples; b) Eliminate samples that did not present any clinical follow-up data; c) Eliminate samples that had no information regarding the survival time; d) Eliminate samples with no status; e) Convert the probe to symbols based on the annotation data file. The sample clinical statistics after data pre-processing is described in Table S1.

2.2. ssGSEA

The ssGSEA algorithm analyses for 29 immune gene sets include all genes associated with various cells involved in the immune system, as well as their pathways, functions and checkpoints. In this study, we applied the ssGSEA algorithm via the R package (GSVA (Version 1.30.0), GSEABase (Version 1.38.0) and LIMMA) for systematical assessment of the immunological features of each sample used in this study [14,15].

2.3. *Identification of molecular subtypes*

We used the TCGA expression profile to derive the expression levels of the 479 immune-associated genes. The univariate Cox analysis ($P < 0.01$) was utilized to acquire the genes associated with the NSCLC prognosis by using the Cox function in the Survival R package (Version

2.38-2) (https://mran.microsoft.com/web/packages/survival/index.html). We consistently clustered the 993 TCGA-NSCLC samples by using the ConsensusClusterPlus (Version 1.48.0) R package (parameters used: pFeature = 1, pItem = 0.8, reps = 100, distance = Minkowski) [16]. The accurate cluster number was determined based on the cumulative distribution function (CDF) and CDF delta area. The Pam algorithm was applied as a clustering algorithm and the Minkowski distance was the distance measure in the study.

2.4. Identification and functional analysis of the DEGs

We extracted the data related to the DEGs from the different molecular subtypes, i.e., C1 and C2 groups, of the TCGA dataset by using the LIMMA package [13]. The following filter criteria were set: false discovery rate (FDR) < 0.05 and |log2fold change (FC)| > 1.5. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis and gene ontology (GO) functional enrichment analysis were performed by applying the DEG data for subtype grouping using the WebGestaltR package (Version 0.4.2).

2.5. Establishment and assessment of the immune subtype-associated lncRNA prognostic risk scoring system

We randomly sampled and divided the samples from the TCGA dataset into training and testing datasets at the ratio of 3:2. Univariate Cox regression and LASSO-Cox regression analyses were used to develop a novel prognostic risk model based on the DEGs in the various subtypes. The penalty parameter (λ, lambda) for the model was determined via tenfold cross-validation following the minimum criteria (i.e., the value of λ corresponding to the lowest partial likelihood deviance). When lambda = 0.02160913, genes were selected for further analysis. For decreasing gene numbers contained in the model, the stepwise Akaike information criterion (stepAIC) process was applied. This step-by-step regression considers the statistical fitting degree of this model and determines the parameters used for the data fit. In the stepAIC process, which was implemented by using the MASS package (Version 7.3-54), a complicated model is initially generated, and one variable is deleted to decrease the AIC. This model was developed by using the “glmnet” R package (Version 3.0-2) based on the formula for the risk score: Σβi × Expi, where β, i.e., the Cox regression coefficient is the respective gene, I denotes the prognosis-related lncRNA and Exp is the expression level of the prognostic lncRNA. The efficiency of the model was improved when the value was smaller. This indicated that the model acquired an appropriate fitting degree with fewer parameters [17]. Thereafter, we calculated the risk scores for every sample included in the TCGA training, TCGA testing, TCGA dataset and external GEO datasets, respectively. Then, the optimal cut-off value was determined by using the “surv_cutpoint” function in the “survminer” R package. Finally, Kaplan-Meier (KM) curves were generated for survival analysis. The “survivalROC” R package was used to generate time-dependent receiver operating characteristic (ROC) curves to assess the predictive power.

2.6. Development and validation of the nomogram

The nomogram could visually and effectively present all results of the above risk model and prediction. This nomogram used the line length to define the effects of various factors and their values on the final results. We constructed the nomograph model based on the univariate and multivariate analytical results [18].

2.7. Predicting the benefit of each subclass from immunotherapy

We used the published the data of the lung cancer patients who received immunotherapy to predict the efficiency of the immunotherapy in their subclass. Based on the SubMap algorithms (https://cloud.genepattern.org/gp) in Gene Pattern 2.0, we compared the gene expression profiles of the same significant genes in our subtypes with another published dataset containing lung cancer patients who received immunotherapy [19].

2.8. Statistical analysis

We conducted the data visualization and the statistical analysis using R software (Version 3.6.3, https://www.r-project.org/ ver. 3.6.3). Bonferroni correction was performed to compare the differences in the SubMap analysis of anti-PD-L1. The distributions of immune scores or IC50 values between the two subtypes were compared by using Wilcoxon testing. Log-rank testing was employed to determine the significant differences for KM curves. Univariate and the multivariate Cox regression analyses were used to screen features that were significantly associated with prognosis in the risk model. $P < 0.05$ was considered to be significant.

3. Results

3.1. Identifying the immune lncRNAs related to PD-1

We used the ssGSEA process to score the KEGG pathways using the TCGA data, and we estimated the relationship between the PD-1 (i.e., PDCD1) and the KEGG pathways. The results indicated that the PD-1 and KEGG-T cell receptor signaling pathways were significantly related (Figure 1A). Then, the TCGA dataset was applied to screen 479 lncRNAs that were significantly related to the score of the KEGG-T cell receptor signaling pathway. We noted that 301 lncRNAs were positively related to each other, whereas 178 lncRNAs were not related. Then, a heat map was developed for the gene expression changes of the 479 lncRNAs based on the score of the KEGG-T cell receptor signaling pathway, as arranged from high to low values (Figure 1B).

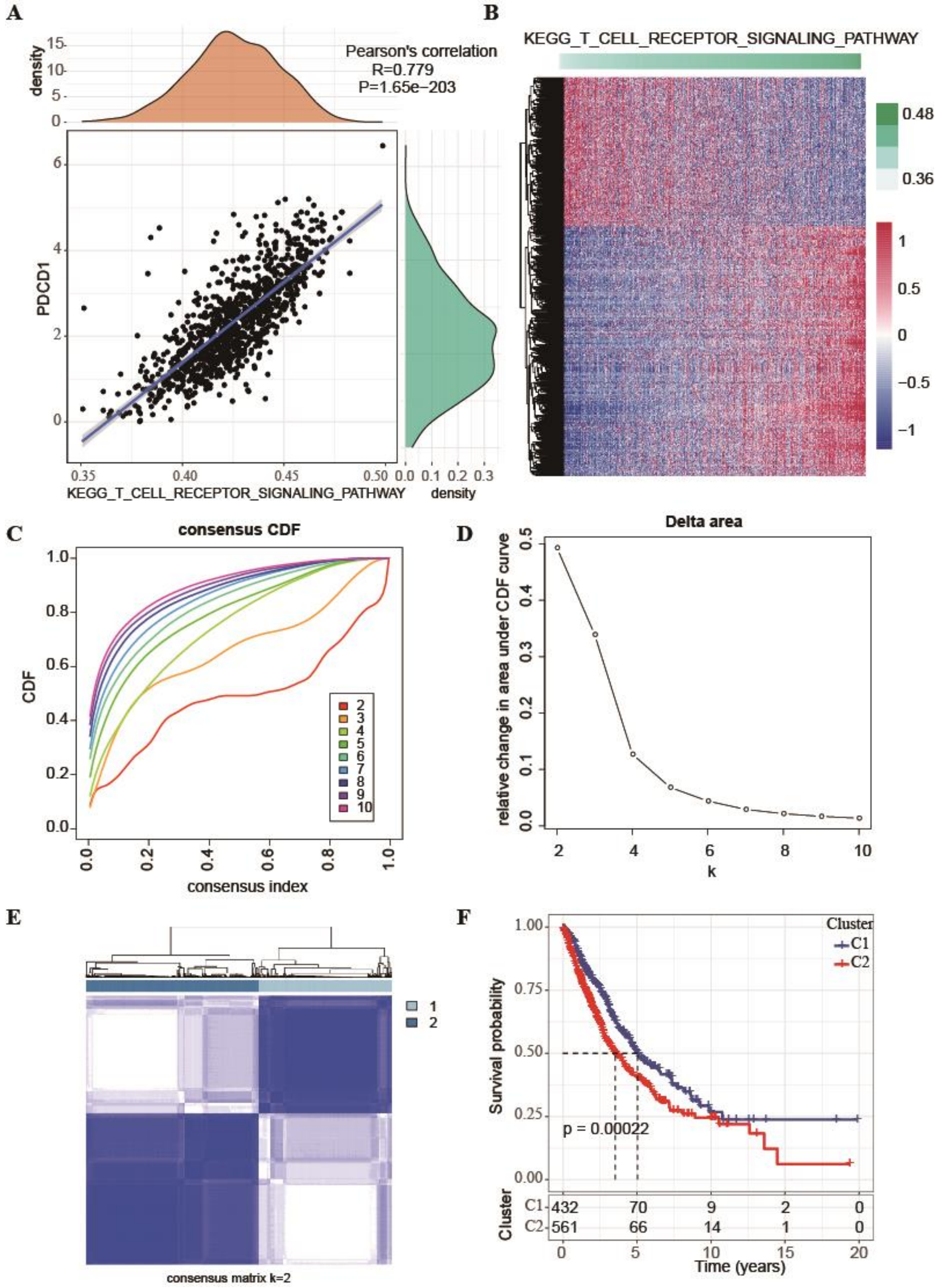

Figure 1. Identification and clustering of the PD-1 related immune lncRNAs. (A): Correlation between the PD-1 and KEGG T-cell receptor signaling pathways in the TCGA dataset; (B): Changes in lncRNA gene expression that are significantly related to the KEGG-T cell receptor signaling pathway; (C-D): CDF curve and the CDF delta area curve for the TCGA queue samples, respectively, where the delta area curve for the consensus clustering indicates the difference in the area under the CDF curve for every category number, k, compared to k-1; (E): A sample clustering heat map if k = 2; (F): KM curve for the prognosis relationship between two subtypes in the TCGA dataset.

3.2. Molecular subtyping of immune-related lncRNAs

The TCGA dataset was first used to calculate the univariate analysis of 479 lncRNA genes. There were 59 lncRNA genes related to the prognosis in TCGA, according to a univariate survival analysis. ConsensusClusterPlus was used to cluster 993 NSCLC samples based on 59 lncRNAs from the TCGA cohort. By using the CDF, we determined the accurate cluster number. The CDF delta area curve showed that the cluster had relatively stable clustering results when the value of 2 was selected (Figure 1C,D). Thus, k = 2 was used and we obtained two related subtypes (Figure 1E). Analysis of the prognostic features of the 2 molecular subtypes showed significant prognostic variations. Thus, we concluded that C2 showed a worse prognosis than C1 (P = 0.00022, Figure 1F). Additionally, using the TCGA dataset, the distribution of various clinical characteristics in these two molecular subtypes was compared. The findings revealed significant gender disparities among the TCGA subtypes (Figure S2), with a majority of the male patients displaying the C1 subtype. Other clinical characteristics were not significantly different among subgroups.

Here, we identified six types of immune infiltrations, which included C1 (wound healing), C2 (predominance of INF-r), C3 (inflammation), C4 (lymphocyte depletion), C5 (immunological silence) and C6 (TGF-beta predominance), have been identified as the tumor suppressors from the corresponding tumor promoters in human malignancies. C1, C2 and C6 were associated with a poor prognosis [20]. A majority of NSCLC patients from the TCGA dataset had immune subtypes C1 and C2, while no patient had the C5 immune subtype (Figure S3A). Further comparison of the distribution of the samples between both the molecular subtypes and existing subtypes (Figure S3B) showed that the C1, C2 and C4 immune subtype had poor prognosis (Figure S3C).

3.3. Differential gene analysis and functional identification of molecular subtypes

A total of 642 DEGs between C1 and C2 groups were screened by using the TCGA dataset. Additionally, the KEGG pathway and GO functional enrichment of 642 DEGs in both subtype groups were analyzed. For the GO functional annotation of DEGs, we annotated 768 items with significantly different biological processes (BP) (FDR < 0.05), 84 cellular component (CC) items (FDR < 0.05) and 71 molecular function (MF) items (FDR < 0.05). Figure 2A-C shows the top 10 items of BP, CC and MF. The differential genes were enriched by the KEGG pathway, and 52 distinct pathways were found to be significantly enriched (FDR 0.05), for instance, the immune pathways like Th17 cell differentiation and cytokine-cytokine receptor interaction, as well as the tumor-related pathways like the cell adhesion molecules (CAMs) and the hematopoietic cell lineages (Figure 2D), indicating that DEGs were involved in NSCLC through these immune-related pathways.

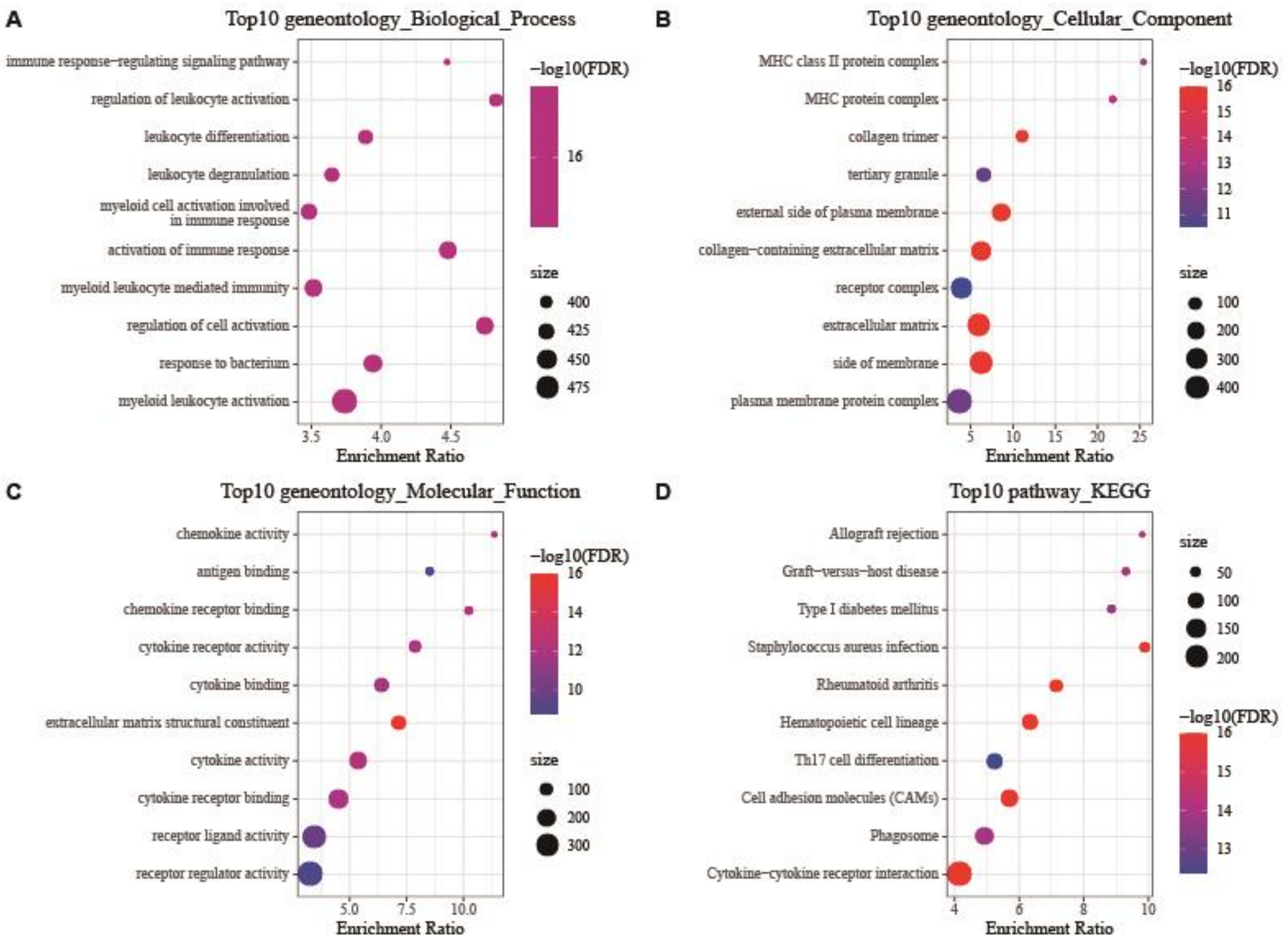

Figure 2. Functional differences of the molecular subtypes. (A): Biological process (BP) annotation map of the upregulated genes involved in different molecular subtypes; (B): cellular component (CC) annotation map of the upregulated genes involved in different molecular subtypes; (C): molecular function (MF) annotation map of the upregulated genes involved in different molecular subtypes; (D): KEGG annotation map of the upregulated genes involved in different molecular subtypes.

3.4. *Comparative analysis of immune microenvironment and immunotherapy / chemotherapy sensitivity of molecular subtypes*

As described in the heat map presented in Figure 3A, we used various techniques like MCP-counter, EPIC, quantiseq, ESTIMATE, and ssGSEA to analyze the distribution of the immune scores in various groups. Most of the five different types of immune scores had significant differences, and the expression was higher in the C2 group (Figure 3A). A literature review revealed 47 immune checkpoint genes [21]. We assessed the differential expression of all genes in the two molecular subtype groups; we noted that a majority of the immune checkpoint (IC) genes showed significant variations, and that the C2 subtype showed a higher variation than C1 (Figure 3B).

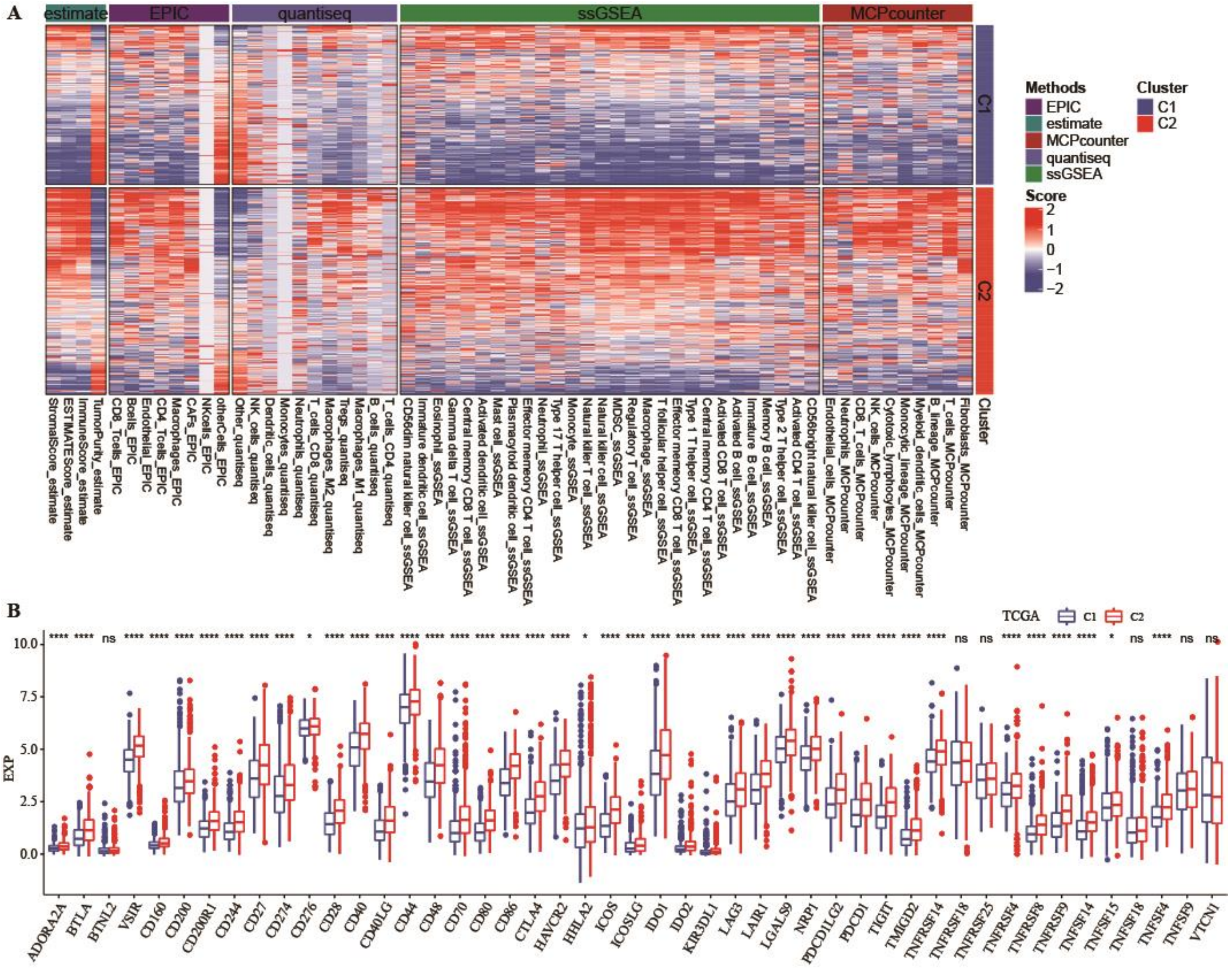

Figure 3. Comparative analysis of the immune status of the two subtypes. (A): Distribution of five scoring algorithms in the TCGA dataset; (B): Expression differences in IC genes derived from the TCGA dataset. ANOVA was used to determine significance level (* P < 0.05; ** P < 0.01; *** P < 0.001; **** P < 0.0001).

Then, we determined the differences between the sensitivity of the C1 and C2 molecular subtypes to chemotherapy and immunotherapy. Here, we used the subclass mapping technique to compare the similarities between both the defined subtypes and the immunotherapy patients in the GSE78220 and IMvigor210 datasets. A lower p-value indicated a higher similarity (GSE78220 was treated with anti-PD-1, while IMvigor210 was treated with PD-L1). The results showed that, in the TCGA dataset, the C2 subtype was similar to anti-PD-1-NR, and it is similar to sd/pd treated by PD-L1 (Figure 4A,B). At the same time, the response degree of different subtypes to traditional chemotherapy drugs was also analyzed. It was found that subtype C1 was more sensitive to rapamycin, pyrimethamine, bortezomib, vinorelbine (Figure 4C), whereas sunitinib, bexarotene, midostaurin and bleomycin were more influential on the C2 subtype (Figure 4D).

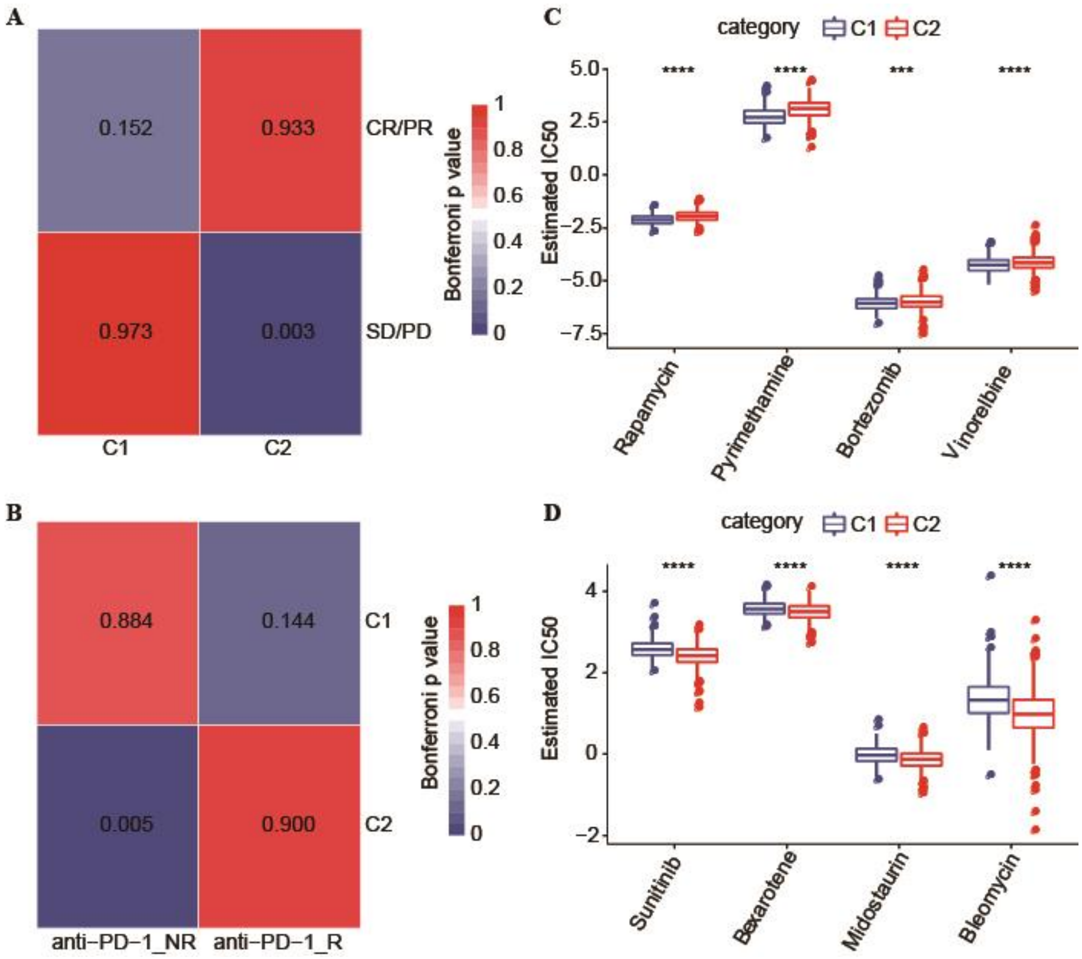

Figure 4. Responsiveness to immunotherapy / chemotherapy of the two molecular subtypes. (A): TCGA SubMap analysis using anti-PD-L1 (GSE78220) (Bonferroni-corrected P < 0.05); (B): TCGA SubMap analysis of anti-PD-1 (IMvigor210) (Bonferroni-corrected P < 0.05); (C): Drugs that are more sensitive to the C1 subtype; (D): Drugs that are more sensitive to C2 subtypes, with a complete response (CR), partial response (PR), response (R), no response (NR), progressive disease (PD) and stable disease (SD). ANOVA was used to determine the significance level (* P < 0.05; ** P < 0.01; *** P < 0.001; **** P < 0.0001).

3.5. Development of the prognostic risk model based on the immune subtype-related lncRNA genes

First, according to the TCGA dataset, we screened 792 DEGs between C1 and C2 groups, and then we conducted the univariate analysis to obtain 52 genes based on their prognosis (P < 0.05). Then, LASSO was used to select the best gene, and 28 genes were obtained according to the minimum lambda = 0.02160913 (Figure 5A). These 28 genes were analyzed based on multiple factors. To reduce the number of genes, the stepAIC technique was applied to obtain 13 lncRNA genes (LINC00944, TRG-AS1, AC133552.5, ZNF674-AS1, AP006621.3, LINC01607, APTR, ITGB1-DT, AC022144.1, AC025569.1, LINC00857, LINC01806, and XIST) and calculate the risk coefficient for the associated genes (Figure 5B). Then, we estimated the risk scores for every sample in the TCGA training and validation datasets, respectively, categorized them into the high-risk and low-risk groups using the optimal cutoff score and then developed their KM and ROC curves, respectively. We noted that the high-risk patients showed a worse prognosis, while the ROC curve of the risk score showed a higher AUC (Figure 5C–G).

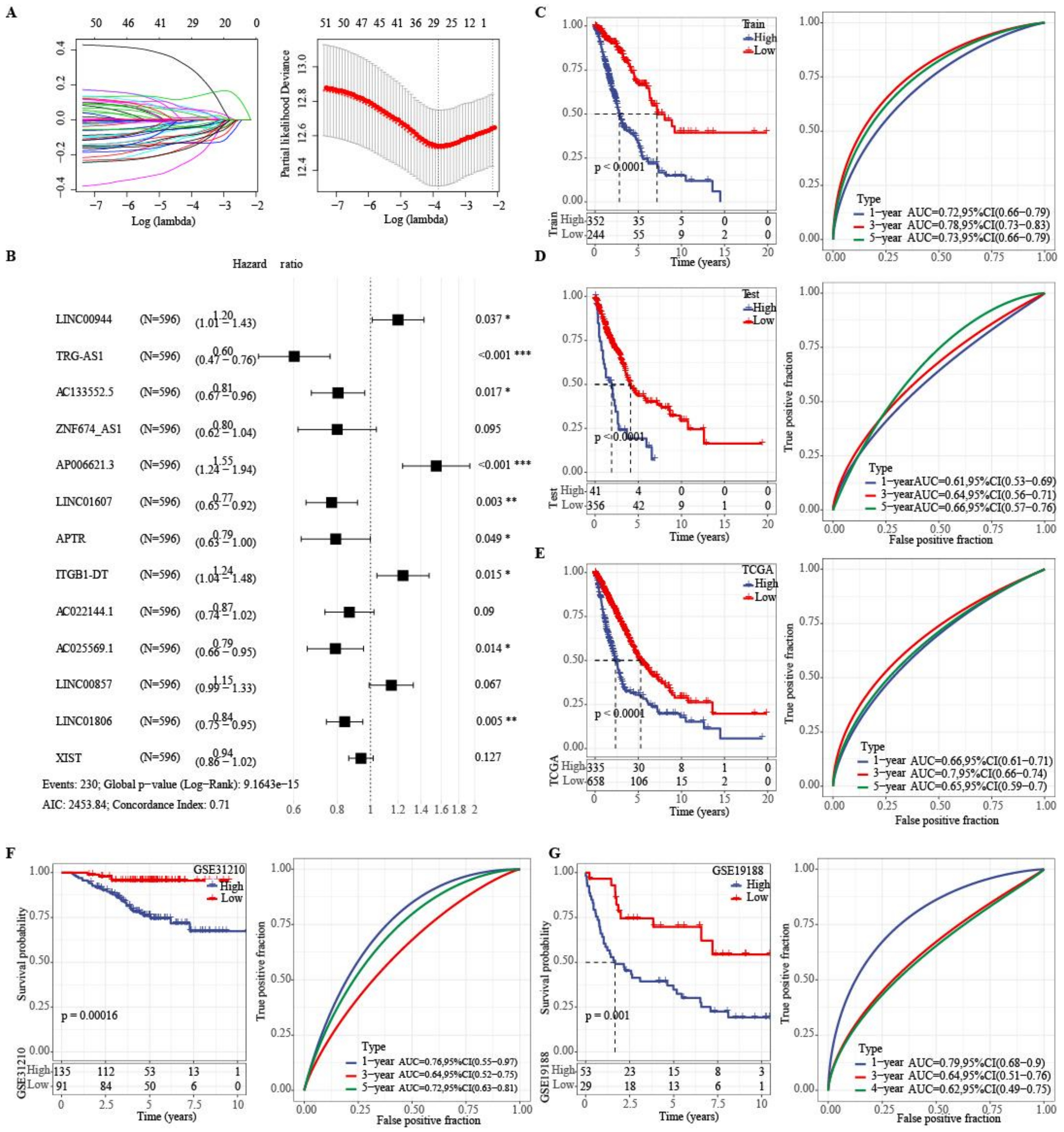

Figure 5. Establishment and evaluation of a prognostic risk model. (A): LASSO coefficient profile for 40 prognostic RNAs derived from the TCGA training dataset. We developed the coefficient profile plot by using a log (Lambda) sequence; (B): Results of the multifactor analysis of the genes in the final model; (C): KM and ROC curves for the model derived from a TCGA training dataset (596 samples); (D): KM and ROC curves for the model derived from the TCGA testing dataset (397 samples); (E): KM and ROC curves for the model derived from all TCGA datasets (993 samples); (F): KM and ROC curves for the model derived from the GSE31210 dataset (226 samples); (G): KM and ROC analysis of the model derived from all GSE19188 datasets (82 samples). ANOVA was used to determine the significance level (* P < 0.05; ** P < 0.01; *** P < 0.001; **** P < 0.0001).

We also compared the risk score distribution for the TCGA among the various clinical feature groups, and we noted that the subtypes, gender groups and the N and T stages showed significant difference ($P < 0.05$). The risk score is higher in the later stage of the tumor, and the C2 risk score with poor prognosis in the subtype is higher. The C1 subtype with better prognosis has a lower risk score. Furthermore, the male patients showed a higher risk score than the female ones. No significant difference was noted between other clinical feature groups (Figure S4). TCGA datasets were grouped according to clinical characteristics, and the survival effects of our risk groups on different clinical characteristics were plotted. The results showed that our risk group had fine results in different clinical groups. Additionally, the high-risk patients showed a worse prognosis than the low-risk patients, which proved the reliability of our risk group (Figure S5).

3.6. *Risk Score and clinical features to construct a nomogram and forest map*

To determine the independent application of the risk scoring model in clinical applications, univariate and multivariate survival analysis was conducted to examine the factors that were significantly related to patient's survival by using the complete clinical data in the TCGA dataset. Also, the clinical data of the patients in the TCGA dataset, like sex, age, N stage, M stage, T stage, and the risk score were examined (Figure 6). The univariate COX regression analysis showed that the N stage, T stage, M stage, stage, age, and risk score from the TCGA dataset, were significantly correlated to the patient's survival. Based on multivariate analysis, risk score (1.81, 95% CI = 1.52-2.15, $p < 1e\text{-}5$) was influential on patient's survival (Figure 6A,B).

Nomography is described as a technique used to effectively display the risk model results. It can also be applied to predict the outcome of an application. The nomogram presents the effects of diverse variables, in addition to their values, on the final outcome, as based on the length of a straight line. We developed the nomogram by using the data of the clinical features T stage, age, N stage and risk score as derived from the TCGA datasets (Figure 6C). This model showed that the risk score feature significantly affected the prediction of the survival rate of the patients, which implies that the risk score model could offer a better prediction of the final prognosis. Furthermore, the nomograms (data for 1, 3 and 5 years) used to determine the model's performance (Figure 6D) were analyzed. Decision curve analysis (DCA) refers to a basic technique for assessing clinical prediction models, molecular markers and diagnostic tests. The DCA diagrams for T stage, age, N stage, risk score and the nomogram were generated. The results showed that the nomogram had a greater effect (Figure 6E).

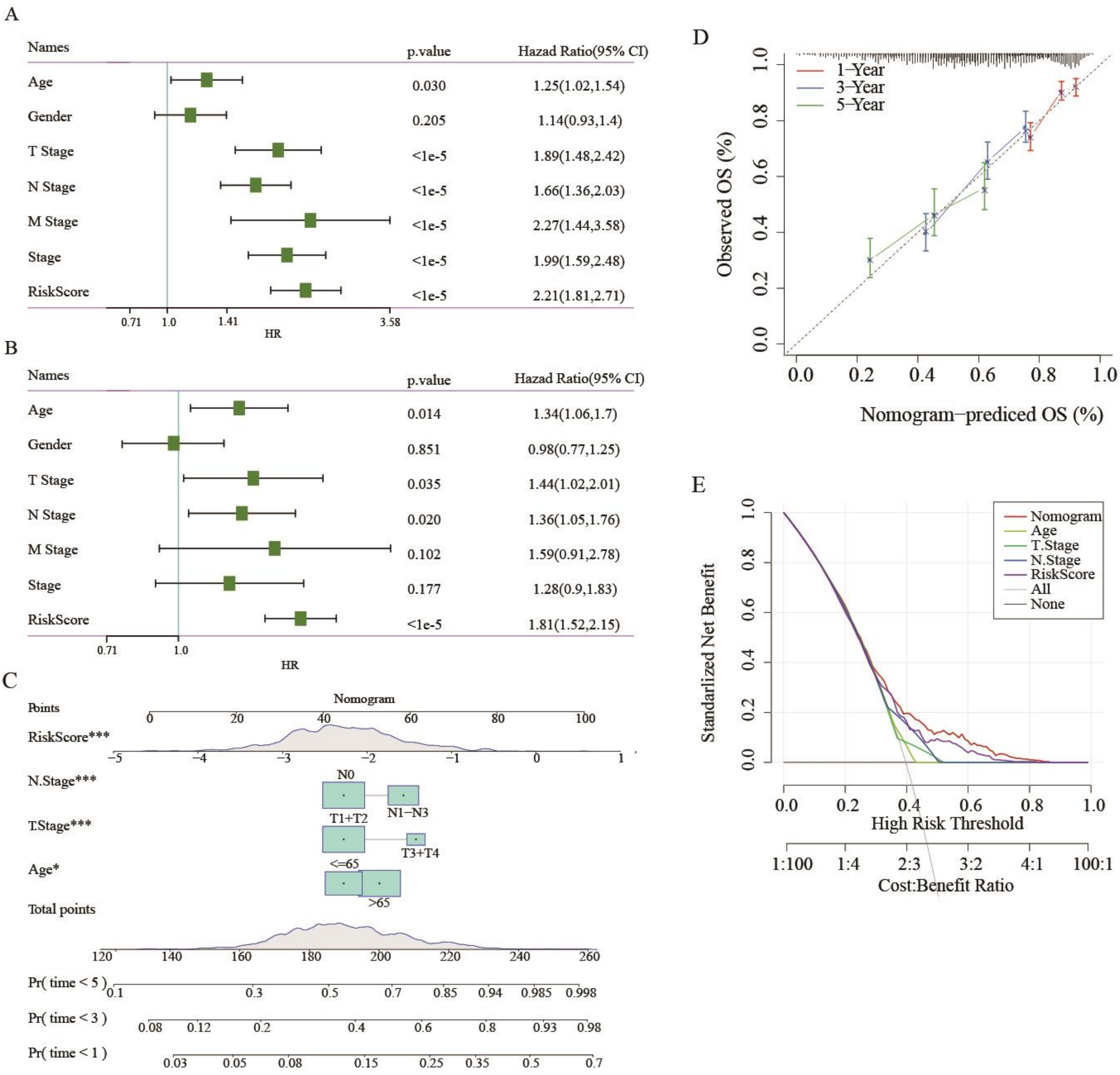

Figure 6. Development of a personalized prognostic nomogram for NSCLC. (A): Univariate Cox regression analysis of the risk score and clinicopathological features; (B): Multivariate Cox regression analysis of the risk score and clinicopathological features. (C): Nomograms for forecasting the 1-, 3- and 5-year patient survival rate based on the TCGA dataset; (D): Calibration chart used to forecast the 1-, 3- and 5-year NSCLC patient survival rate; (E): DCA diagram for the factors of age, stage, risk score and nomogram. ANOVA was used to determine the significance level (* P < 0.05; ** P < 0.01; *** P < 0.001; **** P < 0.0001).

3.7. Difference analysis of the risk score on immunotherapy

The differences in immunotherapy among different risk groups in the TCGA dataset were analyzed. Here, the subclass mapping technique was applied to compare the similarity scores between the various defined risk groups and the immunotherapy patients included in the GSE78220 and IMvigor210 data sets. A lower p-value was proportional to a higher similarity (GSE78220 was treated with anti-PD-1, while IMvigor210 was treated with PD-L1). The results showed that, in the TCGA datasets, the high group and anti-PD-1-NR were similar to the group of SD/PD treated with PD-L1 (Figure 7).

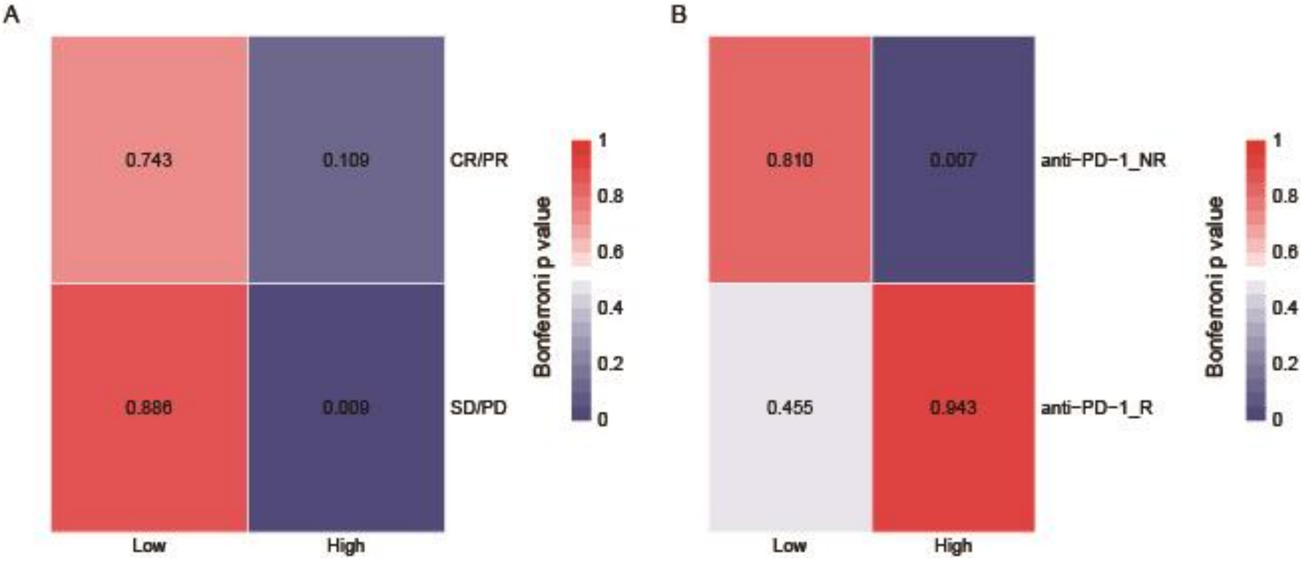

Figure 7. Different immunotherapies used for different risk groups. (A): TCGA SubMap analysis of anti-PD-L1 (Bonferroni-corrected P < 0.05); (B): TCGA SubMap analysis of anti-PD-1 (Bonferroni-corrected P < 0.05).

4. Discussion

In this study, we first identified the T-cell receptor signaling pathway associated with PD-1, and we further screened 479 lncRNAs significantly related to the score of the T cell receptor signaling pathway. Additionally, 993 NSCLC samples derived from the TCGA dataset were genotyped as per the 59 lncRNAs. These samples were categorized into two subtypes with significantly different prognoses. The C2 subtype showed a significantly poor prognosis. Further analysis indicated the presence of significant differences in the gender-based clinical features among the different subtypes, where a large proportion of the C1 subtypes included male patients. Then, we conducted GO functional enrichment and KEGG pathway analyses on the DEGs and found that the immune-related pathways including the Th17 cell differentiation and cytokine-cytokine receptor interaction, as well as the tumor-related pathways such as the CAMs and hematopoietic cell lineage were significantly enriched. The results of this study suggest that the T cell receptor pathway-related lncRNAs played an important role in the invasion, development, immune response and the treatment of NSCLC patients, which may provide new insight into the progress of NSCLC and immunotherapy.

T cell receptors (TCRs) present on the T-cell surfaces can identify the antigen peptide presented by the major histocompatibility complex (MHC) on the antigen-presenting cell surface, thereby activating c-Jun N-terminal kinase (JNK), ERK, NF-kappa B and other signaling pathways in the T cells [22]. Many transcription factors involved in cell division and differentiation are activated by a variety of signal pathways to control T cell proliferation, differentiation, cytokine production, apoptosis and other cellular processes. The typical intracellular signals of TCR activation also include MAPK, PKC and calcium signaling pathways. A previous experimental study has demonstrated that caspase cascade regulation, cell cycle arrest, the activation of MAPK and PI3K/Akt/mTOR signal modification contribute to the apoptosis of human leukemia Jurkat T cells [23]. Another study has revealed that the activation of mitochondrial reactive oxygen species-mediated Src and PKC pathways using low-level laser provokes MHC class II-restricted T cell immunity to tumors [24]. The activation of TCR signaling will not only cause the proliferation of T cells and the production of cytokines, but it will also encourage T cells to differentiate into effector T cells and perform functions [22]. In some cases, it is noted that tumor-specific immune activation can increase the survival rate of cancer patients. An assorted TCR library has been widely acknowledged as a necessary requirement for accurately predicting the adaptive immune response [25]. These findings verified that the activation of TCR pathways contribute to immunity in tumors.

Immunotherapy is an important tumor therapy, widely used to treat NSCLC patients. Immune checkpoint (IC) inhibitors targeting the CTLA-4, PD-1, and PD-L1 molecules have been developed and used to treat patients with NSCLS and other types of tumors, showing promising clinical activity, effective toxicity features and a long-lasting response [26]. PD-1/PD-L1 is a set of important ICs that are involved in the tumors' immune escape processes. Development of anti PD-1/PD-L1 immune molecules could help to deter the tumor cells from inhibiting the operation of the immune system through the PD-1/PD-L1 axis, as well as to restore its negative effects on the malignant cells. A recent study has confirmed that PD-1 signaling attenuated the death of leukemic stem cells induced by TCRs in T cell acute lymphoblastic leukemia, which provides a potential therapeutic strategy for treating acute lymphoblastic leukemia via PD-1 blockade [27]. The anti-PD-1/PD-L1 immunotherapy mechanism was closely associated with the JNK, nuclear factor kappa B subunit 1, phosphatidylinositol 3-kinase (PI3K)/AKT (AKT serine / threonine kinase 1) pathways and other complicated signal pathways [28]. Moreover, as previously reported, tonsil-derived mesenchymal stem cells inhibit Th17-mediated autoimmune response through regulation of the PD-1/PD-L1 pathway [29]. Therefore, regulation of immune-related pathways represents a promising approach for immunotherapy.

Then, we identified 13 lncRNA genes (LINC00944, TRG-AS1, AC133552.5, ZNF674-AS1, AP006621.3, LINC01607, APTR, ITGB1-DT, AC022144.1, AC025569.1, LINC00857, LINC01806 and XIST). It was noted that lncRNA-LINC00944 promoted tumorigenesis and inhibited Akt phosphorylation in renal cell carcinoma [30]. Immune-related lncRNA LINC00944 was seen to regulate the ADAR1 expression in breast cancer and to be related to its prognosis [31]. Some studies have shown that the TRG-AS1 gene was significantly up-regulated in lung cancer and promoted the invasion and propagation of lung cancer cells through the miR-224-5p/SMAD4 axis [32]. TRG-AS1 has been shown to be an effective mediator of carcinogenicity in the tongue squamous cell carcinoma that was regulated by the microRNA-543/Yes-associated protein 1 axis [33]. Earlier studies showed the therapeutic efficacy of the ZNF674-AS1 gene in NSCLC. ZNF674-AS1 suppressed the growth of NSCLC, by inducing P2 and downregulating the miR-423-3p expression [34]. Additionally, it was noted that ZNF674-AS1 inhibited the invasion and the migration of NSCLC, as it regulated the miR-23a/E-cadherin axis [35]. It has been reported that Wnt signaling is associated with tumor heterogeneity, growth, immunity and drug resistance [36].

APTR is engaged in the onset and development of gynecological cancer and it has been shown to promote the proliferation of the uterine leiomyoma cells as it targeted the Erα for activation of the Wnt/β catenin pathway [37]. ITGB1-DT was seen to be a novel biomarker related to lung adenocarcinoma immunity [38,39]. It promotes the progression of lung adenocarcinoma through the establishment of a positive feedback loop with the ITGB1/Wnt/β-catenin/MYC pathway [40]. In the past, the LINC00857 gene was considered to be involved in the onset and development of various types of cancers [41,42] because it was significantly up-regulated in lung adenocarcinoma. The LINC00857 gene primarily regulates the proliferation, glycolysis and the apoptosis of LUAD cells by targeting the miR-1179/SPAG5 axis, offering a potential target for clinically treating LUAD [43]. The oncogene XIST was significantly up-regulated in the NSCLC tissues. Several studies presented that XIST was involved in the invasion, proliferation, apoptosis, migration and chemosensitivity of NSCLC cells [44–46]. XIST played a vital role as the ceRNA in NSCLC development, even at the transcriptional level [47]. Collectively, these 13 lncRNA genes might reveal their contribution to the development and progression of NSCLC.

Risk factors have been selected for predictive model construction in many previous studies, and these risk features have been verified by clinical trials or manually selected by medical experts [48]. However, manually selecting and validating each risk factor has been considered as tedious work. A machine learning approach has become a novel strategy to select robust features and improve the long-term prognosis in cancers or diseases. For examples, Khosla A and colleagues have automatically selected robust features by using a conservative mean, which stacked with support vector machines (SVMs) generates a greater AUC compared to classical methods [49]. Fang et al employed recursive feature elimination with cross-validation, which incorporated linear support vector classifier, a random forest classifier, an extra tree classifier, an AdaBoost classifier and a multinomial naïve-Bayes classifier to select robust features for stroke prognosis prediction, which can accurately infer the long-term prognosis of acute stroke [50]. Recently, a random forest-based algorithm incorporating a nested 10-fold cross-validation has been utilized to assess the presence of congenital heart disease (CHD). And, six robust features that have potential values for screening CHD have been selected [51]. Moreover, the incorporation of a segment-based convolutional neural network with an SVM has been adopted to recognize atrial fibrillation from electrocardiogram records [50]. Random forest-based benchmark models incorporating K-nearest neighbor, SVMs, logistic regression, stochastic gradient descent and AdaBoost have been used for risk classification in breast cancer patients [52]. Therefore, a machine learning approach can be used to select robust features and improve prognosis prediction for NSCLC patients.

One of the advantages of this model is that compared with whole genome sequencing, targeted sequencing based on specific genes can significantly reduce medical costs. Secondly, we selected lncRNA significantly related to the PD-1 related T cell receptor signaling pathway as the target gene, which was seen to be important for forecasting the immunotherapeutic sensitivity of NSCLC. Furthermore, in clinical diagnosis and the development of a treatment strategy, we noted that the treatment plan and prognosis of the patients was primarily dependent on the pathological stage of the disease. The nomogram clinical model constructed in this study provides a basis for designing an individualized treatment plan for NSCLC patients. Our proposed modelling approach may be applicable for other diseases such as ophthalmic diseases including age-related macular degeneration and Stargardt diseases. These molecular subtypes and their changes in such diseases could be identified by using optical coherence tomography, and they could be used for modeling [53–56]. Compared with single-gene models, multi-gene models can comprehensively reflect the patient's status.

However, there are some limitations in the current study. First, we downloaded retrospective transcriptome data from TCGA and GEO datasets and validated the robustness of the prognostic model on TCGA datasets and external datasets. The predictive value of this prognostic model should be validated by performing prospective studies with larger samples. Second, this prognostic model should be iteratively verified following long-term clinical use. Besides, although we have identified 13 lncRNAs associated with NSCLC, their potential mechanism in the development and progression should be revealed by further in vivo and in vitro studies. Meanwhile, classical methods were used for our model development. Novel machine learning algorithms, such as random forest-based algorithm can be incorporated with SVMs, k-nearest neighbor, or AdaBoost to select robust features, which can infer the long-term prognosis for NSCLC patients.

5. Conclusions

Based on the enrichment of the PD-1 related T cell receptor signaling pathway, we identified two lncRNA-related molecular subtypes of NSCLC. These molecular classifications facilitate the understanding of lncRNA in NSCLC and could be used to supplement the existing tumor TNM staging system. In addition, we developed and validated a novel 13-lncRNA-related prognostic model that could be applied to predict the prognosis and the sensitivity to PD-1 immunotherapy for NSCLC patients.

Acknowledgments

The present study was supported by the Shenzhen Special Fund for Innovation and Entrepreneurship of Overseas High-Level Talents (KQJSCX20180330124428928) and Shenzhen Technology R&D Fund (JSGG20180508152646606). The datasets generated and/or analyzed during the current study are available in the [GSE31210] repository, [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31210], [GSE19188] repository, [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE19188] and [GSE50081] repository, [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE50081].

Conflict of interests

The authors report that there are no competing interests to declare.

References

1. R. L. Siegel, K. D. Miller, H. E. Fuchs, A. Jemal, Cancer Statistics, 2021, *CA Cancer J. Clin.*, **71** (2021), 7–33. https://doi.org/10.3322/caac.21654
2. F. Islami, A. G. Sauer, K. D. Miller, R. L. Siegel, S. A. Fedewa, E. J. Jacobs, et al., Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States, *CA Cancer J. Clin.*, **68** (2018), 31–54. https://doi.org/10.3322/caac.21440
3. M. Zheng, Classification and pathology of lung cancer, *Surg. Oncol. Clin.*, **25** (2016), 447–468. https://doi.org/10.1016/j.soc.2016.02.003

4. M. Wang, R. S. Herbst, C. Boshoff, Toward personalized treatment approaches for non-small-cell lung cancer, *Nat. Med.*, **27** (2021), 1345–1356. https://doi.org/10.1038/s41591-021-01450-2
5. M. MacManus, F. Hegi-Johnson, Overcoming immunotherapy resistance in NSCLC, *Lancet Oncol.*, **23** (2022), 191–193. https://doi.org/10.1016/S1470-2045(21)00711-7
6. A. Insa, P. Martín-Martorell, R. D. Liello, M. Fasano, G. Martini, S. Napolitano, et al., Which treatment after first line therapy in NSCLC patients without genetic alterations in the era of immunotherapy? *Crit. Rev. Oncol. Hematol.*, **169** (2022), 103538. https://doi.org/10.1016/j.critrevonc.2021.103538
7. F. Xie, M. Xu, J. Lu, L. Mao, S. Wang, The role of exosomal PD-L1 in tumor progression and immunotherapy, *Mol. Cancer*, **18** (2019), 146. https://doi.org/10.1186/s12943-019-1074-3
8. M. Niu, M. Yi, N. Li, S. Luo, K. Wu, Predictive biomarkers of anti-PD-1/PD-L1 therapy in NSCLC, *Exp. Hematol. Oncol.*, **10** (2021), 18. https://doi.org/10.1186/s40164-021-00211-8
9. P. Yu, X. He, F. Lu, L. Li, H. Song, X. Bian, Research progress regarding long-chain non-coding RNA in lung cancer: A narrative review, *J. Thorac. Dis.*, **14** (2022), 3016. https://doi.org/10.21037/jtd-22-897
10. W. Sun, Y. Shi, Z. Wang, J. Zhang, H. Cai, J. Zhang, et al., Interaction of long-chain non-coding RNAs and important signaling pathways on human cancers, *Int. J. Oncol.*, **53** (2018), 2343–2355. https://doi.org/10.3892/ijo.2018.4575
11. C. C. Sun, W. Zhu, S. J. Li, W. Hu, J. Zhang, Y. Zhuo, et al., FOXC1-mediated LINC00301 facilitates tumor progression and triggers an immune-suppressing microenvironment in non-small cell lung cancer by regulating the HIF1α pathway, *Genome Med.*, **12** (2020), 77. https://doi.org/10.1186/s13073-020-00773-y
12. M. M. Balas, A. M. Johnson, Exploring the mechanisms behind long noncoding RNAs and cancer, *Noncoding RNA Res.*, **3** (2018), 108–117. https://doi.org/10.1016/j.ncrna.2018.03.001
13. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, et al., limma powers differential expression analyses for RNA-sequencing and microarray studies, *Nucleic Acids Res.*, **43** (2015), e47. https://doi.org/10.1093/nar/gkv007
14. S. Hänzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data, *BMC Bioinf.*, **14** (2013), 7. https://doi.org/10.1186/1471-2105-14-7
15. D. Merico, R. Isserlin, O. Stueker, A. Emili, G. D. Bader, Enrichment map: a network-based method for gene-set enrichment visualization and interpretation, *PloS One*, **5** (2010), e13984. https://doi.org/10.1371/journal.pone.0013984
16. M. D. Wilkerson, D. N. Hayes, ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking, *Bioinformatics*, **26** (2010), 1572–1573. https://doi.org/10.1093/bioinformatics/btq170
17. Z. Zhang, Variable selection with stepwise and best subset approaches, *Ann. Transl. Med.*, **4** (2016), 136. https://doi.org/10.21037/atm.2016.03.35
18. V. P. Balachandran, M. Gonen, J. J. Smith, R. P. DeMatteo, Nomograms in oncology: more than meets the eye, *Lancet Oncol.*, **16** (2015), e173–180. https://doi.org/10.1016/S1470-2045(14)71116-7
19. F. Ay, M. Kellis, T. Kahveci, SubMAP: aligning metabolic pathways with subnetwork mappings, *J. Comput. Biol.*, **18** (2011), 219–235. https://doi.org/10.1089/cmb.2010.0280

20. V. Thorsson, D. L. Gibbs, S. D. Brown, D. Wolf, D. S. Bortone, T. H. O. Yang, et al., The Immune Landscape of Cancer, *Immunity*, **48** (2018), 812–830.e14. https://doi.org/10.1016/j.immuni.2018.03.023
21. L. Danilova, W. J. Ho, Q. Zhu, T. Vithayathil, A. D. Jesus-Acosta, N. S. Azad, et al., Programmed cell death Ligand-1 (PD-L1) and CD8 expression profiling identify an immunologic subtype of pancreatic ductal adenocarcinomas with favorable survival, *Cancer Immunol. Res.*, **7** (2019), 886–895. https://doi.org/10.1158/2326-6066.CIR-18-0822
22. T. N. Schumacher, T-cell-receptor gene therapy, *Nat. Rev. Immunol.*, **2** (2002), 512–519. https://doi.org/10.1038/nri841
23. W. Xu, X. Wang, Y. Tu, H. Masaki, S. Tanaka, K. Onda, et al., Tetrandrine and cepharanthine induce apoptosis through caspase cascade regulation, cell cycle arrest, MAPK activation and PI3K/Akt/mTOR signal modification in glucocorticoid resistant human leukemia Jurkat T cells, *Chem. Biol. Interact.*, **310** (2019), 108726. https://doi.org/10.1016/j.cbi.2019.108726
24. H. Chang, Z. Zou, J. Li, Q. Shen, L. Liu, X. An, et al., Photoactivation of mitochondrial reactive oxygen species-mediated Src and protein kinase C pathway enhances MHC class II-restricted T cell immunity to tumours, *Cancer Lett.*, **523** (2021), 57–71. https://doi.org/10.1016/j.canlet.2021.09.032
25. X. Wang, B. Zhang, Y. Yang, J. Zhu, S. Cheng, Y. Mao, et al., Characterization of distinct T cell receptor repertoires in tumor and distant non-tumor tissues from lung cancer patients, *Genom. Proteom. Bioinf.*, **17** (2019), 287–296. https://doi.org/10.1016/j.gpb.2018.10.005
26. N. Seetharamu, D. R. Budman, K. M. Sullivan, Immune checkpoint inhibitors in lung cancer: past, present and future, *Future Oncol.*, **12** (2016), 1151–1163. https://doi.org/10.2217/fon.16.20
27. X. Xu, W. Zhang, L. Xuan, Y. Yu, W. Zheng, F. Tao, et al., PD-1 signalling defines and protects leukaemic stem cells from T cell receptor-induced cell death in T cell acute lymphoblastic leukaemia, *Nat. Cell Biol.*, **25** (2023), 170–182. https://doi.org/10.1038/s41556-022-01050-3
28. F. Bie, H. Tian, N. Sun, R. Zang, M. Zhang, P. Song, et al., Research progress of Anti-PD-1/PD-L1 immunotherapy related mechanisms and predictive biomarkers in NSCLC, *Front. Oncol.*, **12** (2022), 769124. https://doi.org/10.3389/fonc.2022.769124
29. J. Y. Kim, M. Park, Y. H. Kim, K. H. Ryu, K. H. Lee, K. A. Cho, et al. Tonsil-derived mesenchymal stem cells (T - MSCs) prevent Th17 - mediated autoimmune response via regulation of the programmed death - 1/programmed death ligand - 1 (PD - 1/PD - L1) pathway, *J. Tissue Eng. Regen. Med.*, **12** (2018), e1022–e1033. https://doi.org/10.1002/term.2423
30. C. Chen, H. Zheng, LncRNA LINC00944 promotes tumorigenesis but suppresses akt phosphorylation in renal cell carcinoma, *Front. Mol. Biosci.*, **8** (2021), 697962. https://doi.org/10.3389/fmolb.2021.697962
31. P. R. de Santiago, A. Blanco, F. Morales, K. Marcelain, O. Harismendy, M. S. Herrera, et al., Immune-related IncRNA LINC00944 responds to variations in ADAR1 levels and it is associated with breast cancer prognosis, *Life Sci.*, **268** (2021), 118956. https://doi.org/10.1016/j.lfs.2020.118956
32. M. Zhang, W. Zhu, M. Haeryfar, S. Jiang, X. Jiang, W. Chen, et al., Long non-coding RNA TRG-AS1 promoted proliferation and invasion of lung cancer cells through the miR-224-5p/SMAD4 Axis, *Oncol. Targets Ther.*, **14** (2021), 4415–4426. https://doi.org/10.2147/OTT.S297336

33. S. He, X. Wang, J. Zhang, F. Zhou, L. Li, X. Han, TRG-AS1 is a potent driver of oncogenicity of tongue squamous cell carcinoma through microRNA-543/Yes-associated protein 1 axis regulation, *Cell Cycle*, **19** (2020), 1969–1982. https://doi.org/10.1080/15384101.2020.1786622
34. Y. Liu, R. Huang, D. Xie, X. Lin, L. Zheng, ZNF674-AS1 antagonizes miR-423-3p to induce G0/G1 cell cycle arrest in non-small cell lung cancer cells, *Cell Mol. Biol. Lett.*, **26** (2021), 6. https://doi.org/10.1186/s11658-021-00247-y
35. J. Wang, S. Liu, T. Pan, M. Wang, L. Li, X. Weng, et al., Long non-coding RNA ZNF674-AS1 regulates miR-23a/E-cadherin axis to suppress the migration and invasion of non-small cell lung cancer cells, *Transl. Cancer Res.*, **10** (2021), 4116–4124. https://doi.org/10.21037/tcr-21-1499
36. H. Zhao, T. Ming, S. Tang, S. Ren, H. Yang, M. Liu, et al., Wnt signaling in colorectal cancer: Pathogenic role and therapeutic target, *Mol. Cancer*, **21** (2022), 144. https://doi.org/10.1186/s12943-022-01616-7
37. W. Zhou, G. Wang, B. Li, J. Qu, Y. Zhang, LncRNA APTR promotes uterine leiomyoma cell proliferation by targeting ERα to activate the Wnt/β-catenin pathway, *Front. Oncol.*, **11** (2021), 536346. https://doi.org/10.3389/fonc.2021.536346
38. B. Q. Qiu, X. H. Lin, S. Q. Lai, F. Lu, K. Lin, X. Long, et al., ITGB1-DT/ARNTL2 axis may be a novel biomarker in lung adenocarcinoma: A bioinformatics analysis and experimental validation, *Cancer Cell Int.*, **21** (2021), 665. https://doi.org/10.1186/s12935-021-02380-2
39. C. He, H. Yin, J. Zheng, J. Tang, Y. Fu, X. Zhao, Identification of immune-associated lncRNAs as a prognostic marker for lung adenocarcinoma, *Transl. Cancer Res.*, **10** (2021), 998–1012. https://doi.org/10.21037/tcr-20-2827
40. R. Chang, X. Xiao, Y. Fu, C. Zhang, X. Zhu, Y. Gao, ITGB1-DT facilitates lung adenocarcinoma progression via forming a positive feedback loop with ITGB1/Wnt/β-Catenin/MYC, *Front. Cell Dev. Biol.*, **9** (2021), 631259. https://doi.org/10.3389/fcell.2021.631259
41. Y. Huang, Y. Lin, X. Song, D. Wu, LINC00857 contributes to proliferation and lymphomagenesis by regulating miR-370-3p/CBX3 axis in diffuse large B-cell lymphoma, *Carcinogenesis*, **42** (2021), 733–741. https://doi.org/10.1093/carcin/bgab013
42. D, Zhou, S. He, D. Zhang, Z. Lv, J. Yu, Q. Li, et al., LINC00857 promotes colorectal cancer progression by sponging miR-150-5p and upregulating HMGB3 (high mobility group box 3) expression, *Bioengineered*, **12** (2021), 12107–12122. https://doi.org/10.1080/21655979.2021.2003941
43. L. Wang, L. Cao, C. Wen, J. Li, G. Yu, C. Liu, LncRNA LINC00857 regulates lung adenocarcinoma progression, apoptosis and glycolysis by targeting miR-1179/SPAG5 axis, *Hum. Cell*, **33** (2020), 195–204. https://doi.org/10.1007/s13577-019-00296-8
44. J. Liu, L. Yao, M. Zhang, J. Jiang, M. Yang, Y. Wang, Downregulation of LncRNA-XIST inhibited development of non-small cell lung cancer by activating miR-335/SOD2/ROS signal pathway mediated pyroptotic cell death, *Aging*, **11** (2019), 7830–7846. https://doi.org/10.18632/aging.102291
45. P. Katopodis, Q. Dong, H. Halai, C. I. Fratila, A. Polychronis, V. Anikin, et al., In silico and in vitro analysis of lncRNA XIST reveals a panel of possible lung cancer regulators and a five-gene diagnostic signature, *Cancers*, 1**2** (2020), 3499. https://doi.org/10.3390/cancers12123499

46. X. Xu, X. Zhou, Z. Chen, C. Gao, L. Zhao, Y. Cui, Silencing of lncRNA XIST inhibits non-small cell lung cancer growth and promotes chemosensitivity to cisplatin, *Aging*, **12** (2020), 4711–4726. https://doi.org/10.18632/aging.102673
47. Y. Shen, Y. Lin, K. Liu, J. Chen, J. Zhong, Y. Gao, et al., XIST: A meaningful long noncoding RNA in NSCLC process, *Curr. Pharm. Des.*, **27** (2021), 1407–1417. https://doi.org/10.2174/1381612826999201202102413
48. J. Song, S. Zhang, Y. Sun, J. Gu, Z. Ye, X. Sun, et al., A radioresponse-related lncRNA biomarker signature for risk classification and prognosis prediction in non-small-cell lung cancer, *J. Oncol.*, (2021), 4338838. https://doi.org/10.1155/2021/4338838
49. A. Khosla, Y. Cao, C. C. Y. Lin, H. K. Chiu, J. Hu, H. Lee, An integrated machine learning approach to stroke prediction, Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining, 2010. Available from: https://dl.acm.org/doi/abs/10.1145/1835804.1835830
50. G. Fang, W. Liu, L. Wang, A machine learning approach to select features important to stroke prognosis, *Comput. Biol. Chem.*, **88** (2020), 107316. https://doi.org/10.1016/j.compbiolchem.2020.107316
51. V. T. Truong, B. P. Nguyen, T. H. Nguyen-Vo, W. Mazur, E. S. Chung, C. Palmer, et al., Application of machine learning in screening for congenital heart diseases using fetal echocardiography, *Int. J. Cardiovasc Imaging*, **38** (2022), 1007–1015. https://doi.org/10.1007/s10554-022-02566-3
52. Q. H. Nguyen, T. T. Do, Y. Wang, S. S. Heng, K. Chen, W. H. M. Ang, et al., Breast cancer prediction using feature selection and ensemble voting, in *2019 International Conference on System Science and Engineering (ICSSE)*, 2019, 250–254. Available from: https://ieeexplore.ieee.org/abstract/document/8823106
53. R. K. Meleppat, K. E. Ronning, S. J. Karlen, K. K. Kothandath, M. E. Burns, E. N. Pugh, et al., In situ morphologic and spectral characterization of retinal pigment epithelium organelles in mice using multicolor confocal fluorescence imaging, *Invest. Ophthalmol. Vis. Sci.*, **61** (2020), 1. https://doi.org/10.1167/iovs.61.13.1
54. R. K. Meleppat, P. Zhang, M. J. Ju, S. K. Manna, Y. Jian, E. N. Pugh, et al., Directional optical coherence tomography reveals melanin concentration-dependent scattering properties of retinal pigment epithelium, *J. Biomed. Opt.*, **24** (2019), 1–10. https://doi.org/10.1117/1.JBO.24.6.066011
55. S. H. Chung, T. N. Sin, B. Dang, T. Ngo, T. Lo, D. Lent-Schochet, et al., CRISPR-based VEGF suppression using paired guide RNAs for treatment of choroidal neovascularization, *Mol. Ther. Nucleic Acids*, **28** (2022), 613–622. https://doi.org/10.1016/j.omtn.2022.04.015
56. S. H. Chung, I. N. Mollhoff, U. Nguyen, A. Nguyen, N. Stucka, E. Tieu, et al., Factors impacting efficacy of AAV-mediated CRISPR-based genome editing for treatment of choroidal neovascularization, *Mol. Ther. Methods Clin. Dev.*, **17** (2020), 409–417. https://doi.org/10.1016/j.omtm.2020.01.006