Identification of Immune-Related lncRNA Biomarkers Associated With Prognosis In LUAD

Zihu Guo
Northwest University
Lijuan Shao
Jinan University
Liang Han
Beijing University of Chinese Medicine
Yaohua Ma
Northwest University
Fei He
Jinan University
Shengnan Liang
Northwest A&F University
Liujing Wu
Jinan University
Linzhi Li
Fudan University
Li Jiang
Shihezi University
Yueping Li
Shihezi University
Haiqing Wang
Northwest University
Furong Li
Jinan University
Yonghua Wang
Northwest University
Jingbo Wang (wangjingbo@sztu.edu.cn)
Shenzhen Technology University

Research Article

Keywords: Biomarker, LncRNA, immue regulation, prognosis, lung adenocarcinoma

Posted Date: June 15th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-590465/v1

Abstract

Background: Long non-coding RNAs (lncRNAs) are involved in regulation of immune response and could serve as biomarkers for tumors.

Methods: WGCNA were used to construct an immune-related gene interaction network and elucidated possible functions of the lncRNAs by gene enrichment analyses. 6 lncRNAs were selected and used to construct an immune-related lncRNAs risk score (ILRS) classifier.

Results: Patients with high and low ILRS showed significant differences in microenvironment and immune cell infiltration type. Furthermore, enrichment analyses showed that these lncRNAs regulated immune-related pathways. Finally, four clusters derived from LUAD patients showed a significant correlation with overall survival of patients.

Conclusions: This work provide new insights for subsequent development of clinical strategies against the disease.

Background

Lung cancer is the most common cause of cancer-related deaths in both men and women [1], with lung adenocarcinoma (LUAD) reported to be main histological type [2]. Advanced metastasis occurs in more than 50% of LUAD cases with a 5-year survival rate of 4% reported [3]. Despite research efforts into early diagnosis including development of high-throughput technologies, limited success has been reported in the overall survival rate. This is because most LUAD cases are identified during advanced stage [4, 5]. Tumorigenesis in LUAD cases is complicated and active mutations of oncogenic kinase such as EGFR, MET and ERBB2 genes frequently occurred [6]. Consequently, suitable prognostic markers are urgently needed to guide early diagnosis and develop strategies to fight this fatal disease.

Multiple lines of evidence suggest that malignant cancer phenotypes are not only defined by intrinsic activities of cancer cells but also through tumor microenvironment components, especially tumor-infiltrating immune cells [7, 8]. Lung cancer cells can evade the immune system through various processes, such as inhibition of dendritic cell antigen presentation, promotion of regulatory T cell subpopulation and low expression of MHC (Major Histocompatibility Complex) molecules [9–11]. Based on all these interactions between immune and cancer cells, the state of the immune systems can reflect that of cancer cells. For instance, Chen et al. [12] found that stratification of the type of tumor microenvironment immune types could serve as an approach for tailoring immunotherapeutic strategies in certain types of tumors including lung squamous cell carcinoma. Researches also proved that during tumor treatment, components of the local immune infiltrate shift from a pre-existing to a therapy-induced immune response. As a result, the immune contexture of the tumor could yield information to aid prognosis [13].

Reports indicate that only 2% of the human DNA encodes proteins while 90% of them could not be transcripted [14]. Transcripts that do not encode proteins are referred as non-coding RNAs [15] with those ranging from 200 nucleotides to 100 kilobases in length considered long non-coding RNAs (lncRNAs) [16, 17]. Studies have demonstrated that are involved in many biological processes including normal developmental processes and cancers, as well as immune responses [18]. For instance, a host-derived lnc-Lsm3b was found to compete with viral RNAs at the late stage of innate immunity [19]. Similarly, LincR-Ccr2-5'AS was shown to act as a Th2-specific lncRNA, regulating expression of immune-related genes in Th2 cells and movement of mouse Th2 cells into the lungs [20, 21]. With regards to tumor information and development, an altered expression of lncRNAs has been shown to affect development, invasion and metastasis of many cancers due to regulation of gene expression is regulated by lncRNAs at epigenetic, transcriptional, and post-transcriptional levels [22]. Moreover, different lncRNAs have been shown to enhance the chemoresistance of cancer cells by improving DNA repair, modulating cellular apoptosis and changing drug metabolism [23]. According to previous studies, several lncRNAs related to LUAD process have been identified. These include, DGCR5 that promotes LUAD progression by inhibiting hsa-mir-22-3p and HOXA11-AS that drives cisplatin resistance of human LUAD cells by modulating miR-454-3p/Stat3 [24, 25]. In addition, immune-related lncRNAs have been identified as potential therapeutic targets in anaplastic gliomas, breast and colorectal cancer [26–28]. However, specific roles played by lncRNAs in LUAD-infiltrating immune cells remain unknown.

In this study, we investigated the relationship between immune-related lncRNAs and clinical outcomes in LUAD patients using genome-wide comparative analysis of RNA expression profiles. We constructed an immune-related gene interaction network based on lncRNA using WGCNA analysis, and further modularized the network. We then chose a prognosis-related module for LASSO COX analysis and selected the lncRNA set which was significantly related to the overall survival. We constructed the immune lncRNA risk score (ILRS) classifier based on these lncRNA sets. The relationship between ILRS and LUAD was evaluated and proved that ILRS was influenced by the immune invasion score. Furthermore, enrichment analysis revealed that immune-related lncRNAs participate in many immune regulation processes. Finally, we stratified TCGA-LUAD population into 4 clusters that had significant relationships with clinical outcomes based on the immune-related lncRNA profiles.

Methods

Expression profiles and acquisition of clinical data

The RNA expression profiles for LUAD patients was obtained from Gene Expression Omnibus (GEO) database [29]. This dataset was sequenced on the GPL16791 (Illumina HiSeq 2500 (Homo sapiens)) platform and can be accessed using accession number GSE81089. The database contained RNA expression profiles of 199 LUAD patients as well as other relevant clinical information such as age, gender, tumor stage, life style and ps who (WHO performance status). This data-set was mainly used for subsequent acquisition of immune-related co-expression modules. TCGAbiolinks [30] package was then used to download RNA-seq data and related clinical information of LUAD patients from the TCGA

database. The dataset contained expression profiles of 524 patients and was mainly used for construction of prognostic models based on immune-related lncRNAs. RNA sequence data for these samples were generated using the IlluminaHiSeq platform. Clinical metadata of these patients included their age, TNM stage, stage, overall survival data and outcome.

Acquisition of immune-related genes

Immune-related genes were downloaded from innateDB[31] and ImmPort[32] (Immunology Database and Analysis Portal). Innate immune-related genes from 35,747 publications were included in the two databases (innateDB and ImmPort) and included all genes involved in different immune processes from a variety of experimental studies. We combined the immune-related RNAs from these two databases to obtain a total of 1,238 genes (Table S1).

Weighted gene co-expression network analysis

We employed the Weighted gene co-expression network analysis (WGCNA) [33, 34] to obtain the candidate immune-related lncRNAs and GSE81089 to construct a co-expression network. In order to establish a reliable RNA co-expression network, we first screened genes and lncRNAs according to the following criteria: 1) for each gene and lncRNA, RNA with zero expression on more than 60% of LUAD samples was deleted. 2) RNA expression data with a variance equal to 0 was deleted. Then, the WGCNA package [34] was used to construct a topological overlap matrix (TOM) of immune-related RNA and lncRNAs which represents the relative interconnectedness between each pair of genes. To enable better characterization of the biological network in the constructed TOM, we selected the optimal soft threshold here to satisfy the scale-free topology of the network as previously described [34]. Then, we proceeded to do modular analysis of immune-related gene and lncRNAs, and calculated Pearson correlation coefficients between different modules as well as clinically related indices. If the proportion of immune-related genes contained in a module was greater than 1/4, and the module was significantly correlated to prognostic indicators (p-value < 0.05), then the module was selected for immune-related prognostic module. LncRNAs in these modules are defined as candidate immune-related lncRNAs.

Identification of immune-related lncRNAs

To further screen the immune-related lncRNAs controlling LUAD prognosis, the Cox proportional-hazards (Coxph) model, Least Absolute Shrinkage and Selection Operator (LASSO) regression models were applied during WGCNA analysis. Briefly, the TCGA-LUAD samples were split into a training and testing set in the ratio 2:1 for developing and validating the prognostic model. The univariate Coxph model, adjusting for age, gender, stage, T stage, N stage and M stage of the patients, was used to select the lncRNAs with prognostic significance ($p < 0.05$). Then these prognostic biomarkers were screened using multivariate LASSO regression via the glment [35] and survival[36] packages implemented in the R version 3.5.2 [37].

Establishment of a risk assessment model

Multivariate Coxph analysis was first performed on each immune-related lncRNA based on the TCGA-LUAD training dataset. For further construction of the prognostic risk score model, a coefficient for each immune-related lncRNA was extracted and the model constructed based on linear combination of their expression levels as follows:

$$risk\ \ score = \sum_{i=1}^{N} \gamma_i * E_i$$

where γ_i is the coefficient of lncRNA i obtained from multivariate Coxph regression, E_i is the expression of RNA *i*, and *N* is the number of lncRNAs in the Coxph model. Then, the risk score for each patient was calculated based on expression of the selected lncRNAs. The patients in the TCGA-LUAD training, testing and entire dataset were divided into high and low-risk subgroups for further analysis according to the median cut-off value obtained from the risk scores. Association between the risk score and each clinical feature was assessed using Fisher's exact test, with $p < 0.05$ considered a significant cut-off. TimeROC package [38] was further used to separately draw receiver operator characteristics (ROC) curves for the datasets with the area under the ROC curve (AUC) calculated to examine and compare performance of the classifier.

Determination of immune cell infiltration and leukocyte subtypes in the LUAD samples

To investigate the relationship between ILRS and the degree of immune cell infiltration, we used the ESTIMATE package [39] to generate stromal and immune scores as well as tumor purity of TCGA-LUAD patients. The stromal and immune scores represented presence of stromal and immune cells, respectively in tumor tissues while tumor purity was calculated based on the combination of these scores. They were obtained on the basis of RNA expression profiles of the LUAD samples. Furthermore, we used CIBERSORTx [40] to assess the relative fraction of 22 leukocyte subtypes based on all mRNA transcripts in the TCGA-LUAD cohort, with samples showing statistical significance ($p < 0.05$) kept for further analysis. A student's *t* test was conducted using the ggpubr package [41] to compare differences in immune cell infiltration and leukocyte fractions across the high and low ILRS groups.

Enrichment analysis

We used the Limma package [42] to obtain differentially expressed genes (DEGs) at FDR< 0.01, and |fold change| > 1.5 between patients with a high and low risk scores. We then used the DAVID Functional Annotation Bioinformatics Microarray Analysis [43] to perform gene ontology (GO) and KEGG enrichment analyses on the DEGs. GO enrichment analysis included biological process (BP), molecular function (MF) and subcellular localization (CC). Enrichment analysis and visualization were conducted using the TCGAanalyze_EAcomplete and TCGAvisualize_EAbarplot2 functions implemented in TCGAbiolinck package [30] respectively.

Survival analysis

To evaluate the prognostic value of different variables for LUAD, including lncNRA expression level, risk score and immune infiltration, we performed the Kaplan-Meier (KM) analysis and considered p <0.05, obtained by log-rank test to be statistically significant. KM analysis and log-rank test were mainly implemented by the "sruvival" package [36] with the "survminer" package [44] used for drawing the KM curve.

Patient stratification

We used expression profiles from immune-related lncRNAs to classify the LUAD patients. First, a z-score transformation was used to normalize immune-related lncRNA expression in the TCGA-LUAD cohort as previously described [45]. Then, a Pearson's correlation coefficient (PCC) between samples was computed to assess their similarity followed by the use of partition around medoids algorithm to subsequently divide the samples into *k* subgroups [63]. The maximizing cluster reliability approach was adopted to determine the optimal number of subgroups as previously described. Finally, a consensus clustering analysis was carried out using the "ConsensusClusterPlus" package[46].

To further validate and visualize the immune-related lncRNA derived clustering, a principal component analysis (PCA) was performed among the TCGA-LUAD tumor samples. We conducted the PCA analysis based on normalized expression data using the "prcomp" function from the "stats" module implemented in R 3.5.2 software. We then selected the first two principal components that had most variance and projected each sample into two-dimensional space. Finally, "ggplot2" package [47] was used to visualize the expression pattern of all samples as well as the immune-related lncRNA -derived clusters.

Results

Identification of candidate immune-related RNA modules

WGCNA was successfully used to obtain immune-related lncRNAs. First, the similarity of RNA expression profiles in 108 LUAD patients (GSE81089) was calculated and unsupervised clustering further used to detect outliers. Three obviously abnormal samples were deleted (Figure S1). Then, we constructed a regulatory network containing lncRNAs and immune-related genes. To ensure that the distribution of the regulatory network degree obeys the power-law distribution, we chose a soft threshold of 2 (Fig. 1A, Figure S2). LncRNAs and mRNAs were clustered into 11 modules using WGCNA (Fig. 1B). The numbers of RNAs (lncRNA and mRNA) contained in each module are outlined in table S1. Pearson correlation coefficients of each module associated with the prognosis of patients' clinical phenotypes was calculated with the results showing that red, black and turquoise modules were respectively related with the patients' status, ps who, overall survival days (Fig. 1C, p = 0.05, 0.02 and 0.05 respectively). Furthermore, we found that the proportion of immune-related genes in these modules was greater than 25% (Fig. 1D), which suggested that lncRNAs in these modules may be highly related to immune responses. Consequently, resultant lncRNAs from the three modules were selected as candidates for further analysis.

Construction of prognostic classifiers based on candidate immune-related lncRNAs

We first integrated 156 lncRNAs from three modules; 42, 27 and 87 in the red, black and turquoise module respectively. The data from LUAD patients downloaded from TCGA were then divided into training and testing datasets in the ratio of 2:1, and respectively used to construct the prognostic classifier model for further validation. In the training dataset, the univariate Coxph regression analysis adjusting for age, gender, stage and TNM stage was used to find 7 lncRNAs ($p < 0.05$) associated with overall survival. Then, LASSO Coxph regression was employed to obtain 6 immune-related prognostic biomarkers. A multivariate Coxph regression model was constructed and relevant parameters calculated based on the training dataset. A forest map showed that 3 of these immune-related lncRNAs (RP6-91H8.3, RP11-504A18.1 and AC006129.2) were potential risk while the other 3 (RP11-1275H24.1, LINC01003 and AC006129.2) were potential protection factors (Fig. 2A). This result was further confirmed by KM curve, in which these lncRNAs were found to be significantly correlated with overall survival (logrank test $p < 0.05$) (Fig. 2B – 2G). WGCNA analysis was used to evaluate the interaction between these lncRNAs and immune-related genes and further construct an interaction network (Fig. 2H). Results showed that these lncRNAs came from different modules, including RP691H8.3 from the black, AC006129.2 from the red, and the remaining lncRNA from the turquoise module. This indicated that these lncRNAs may control the prognosis of LUAD patients from different aspects. More importantly, each lncRNA interacts with multiple immune-related genes, suggesting their potential roles in immune regulation.

Furthermore, a prognostic classifier model was constructed using the multivariate Coxph regression coefficient of lncRNA and the immune lncRNA-based risk score (ILRS) for each patient in the training dataset calculated. Patients were divided into high and low risk score groups (Fig. 3A). We found that patients with a high ILRS had a significantly (Fisher's test $p < 0.05$) increased risk of dying (Fig. 3B). In addition, we found that immune-related lncRNAs showed different expression patterns between the groups and ILRS was also related to various clinical information, including Stage (Fisher's test $p < 0.05$) (Fig. 3C). This phenomenon was consistent in the testing (Fig. 3D – 3F) as well as the entire datasets (Fig. 3G – 3I). In conclusion, the classifier based on immune-related lncRNAs has potential to be a prognostic indicator for patients.

We evaluated ILRS performance using ROC curve in order to verify its efficacy in predicting prognosis of LUAD patients. In the test dataset, a 1, 3, and 5-year survival of LUAD patients of 0.71, 0.61, and 0.54, respectively was predicted (Fig. 4A). The KM curve showed a significant correlation (logrank-test $p < 0.05$) between ILRS and the overall survival (Fig. 4B). We also observed the predictive effect of ILRS on patient survival in the entire dataset (Fig. 4C, 4D). Notably, ILRS outperformed traditional clinical indicators, including stage and T Stage, in predicting 3 and 5-year survival (Fig. 4E, 4F). Furthermore, results from the univariate ($p = 4.60E-10$) and multivariate ($p = 3.42E-06$) Coxph analyses indicated that ILRS was the most significant and independent risk factor in the entire TCGA cohort (Table 1).

Table 1
Univariate and multivariate Cox proportional hazards analyses of prognostic factors for overall survival in TCGA-LUAD cohort

Variables		Univariate		Multivariate	
		HR (95% CI)	*P-value*	HR (95% CI)	*P-value*
Age (days)		1.00 (1.00–1.00)	0.4151	1.00 (1.00–1.00)	0.1991
Gender	Female	Ref		Ref	
	Male	1.09 (0.82–1.45)	0.5596	0.86 (0.63–1.17)	0.3465
Stage	Stage I	Ref		Ref	
	Stage II	2.4 (1.68–3.42)	1.41E-06	0.99 (0.54–1.82)	0.9838
	Stage III	3.5 (2.41–5.08)	4.22E-11	1.55 (0.63–3.82)	0.3392
	Stage IV	3.75 (2.17–6.46)	2.09E-06	7.82 (0.42–144.7)	0.1672
T Stage	T1	Ref		Ref	
	T2	1.51 (1.06–2.15)	0.0224	1.25 (0.86–1.82)	0.2356
	T3	2.90 (1.72–4.88)	5.95e-05	2.28 (1.23–4.23)	0.0091
	T4	3.29 (1.73–6.25)	0.0003	1.23 (0.57–2.64)	0.5980
	TX	4.82 (1.16–20.04)	0.0306	1.74 (0.14–21.51)	0.6643
N Stage	N0	Ref		Ref	
	N1	2.36 (1.68–3.31)	7.51E-07	2.04 (1.15–3.61)	0.0143
	N2	3.08 (2.11–4.47)	4.22E-09	1.57 (0.69–3.61)	0.2846
	N3	0 (0-Inf)	0.9943	0 (0-Inf)	0.9947
	NX	1.21 (0.44–3.29)	0.7155	0.88 (0.21–3.64)	0.8627
M Stage	M0	Ref		Ref	
	M1	2.12 (1.24–3.63)	0.0059	0.28 (0.02–4.67)	0.3730
	MX	0.87 (0.61–1.24)	0.4362	0.91 (0.62–1.33)	0.6289
ILRS		36.61 (11.8-113.6)	4.60E-10	17.59 (5.24–9.01)	3.42E-06

Differences in immune infiltration scores between ILRS groups

To examine the relationship between ILRS and immune infiltration, we used Estimate package[39] to first determine the stromal and immune scores as well as tumor Purity of each LUAD patient based on data from TCGA. The results showed no significant relationship (logrank test $p > 0.05$) between Stromal score and survival. However, the extension trend of high stromal scores in the overall survival was still observed (Fig. 5A). Immune Score and tumor purity were both significantly (logrank test $p < 0.05$)correlated with the overall survival of patients (Fig. 5B,5C). We then investigated the effects of high and low ILRS on stromal, and immune scores as well as tumor purity and found significant differences in the training, testing and entire datasets (Fig. 5D -5F). Among them, the stromal and immune scores of patients in the high ILRS group were significantly lower than those of patients in the low ILRS group across the three datasets (training, testing, entire). On the other hand, tumor purity in the high ILRS group was significantly higher than that in the low group. These results indicated that ILRS is linked to immune infiltration. In addition, this process may to some extent be regulated by the immune-related lncRNAs from LUAD.

Differences in leukocyte cell subsets between ILRS groups

The abundance of different types of immune response cells is directly related to prognosis in tumor patients [48]. To gain insights into the relationship between ILRS and different types of immune cell infiltration, we first estimated the proportion of 22 types of immune cells in each LUAD patients identified from the TCGA database using CIBERSORTx [40], and selected samples with a p-value < 0.05 for further analysis. From the analysis, we observed that the types (total > 70%) of infiltrating immune cells in the testing dataset were mainly T, and B cells as well as macrophages (Fig. 6A). All the three types of immune cells were found to be closely related to prognosis of tumor patients. Consistent results were observed in the entire dataset. Next, we compared abundance of the 22 types of immune cells in LUAD patients with high and low ILRS (Fig. 6C – 6D). It is worth noting that the tumor killing immune cells, such as CD8 T, plasma and NK cells, were significantly lower in high than in low ILRS LUAD patients. However, immunosuppression-related cells were more abundant in high ILRS. This indicates that immune-related lncRNAs may be mainly involved in the regulation of these immune cells (CD8 T cells, plasma cells, NK cells).

GO and pathway enrichment analyses identify ILRS associated genes

GO and pathway enrichment analyses were carried out to explore the potential mechanism of action of different prognosis and immune infiltration in high and low ILRS patients. Differential gene expression analysis revealed a total of 1,214 genes in patients with high and low ILRS. In Go analysis, the Biological Process (GOBP), Cellular Component (GOCC), Molecular Function (GOMF) and pathway enrichment analyses were performed for these genes. From these, we identified 15 most significant items for each enrichment analyses and these are outlined in Fig. 7. Most of these items were directly related to immune response. Based on the GOBP enrichment, we observed that these DEGs were related to immune system

regulation, T cell activation and antigen presentation among other functions (Fig. 7A). In the enrichment of GOCC and GOMF, we found that these genes are involved in production of MHCII class molecules, as well as cytokines (Fig. 7B, 7C). Consistent results were observed in pathway enrichment analysis, which indicated that these genes were primarily enriched in T helper, and cytotoxic T cells as well as dendritic cell-related pathways (Fig. 7D).

Stratification of LUAD patients based on candidate immune-related lncRNA

We used molecular cancer profiles to stratify patients in cancer informatics studies due to the complexity and heterogeneous of this disease [49]. Based on our earlier results that demonstrated the remarkable prognostic values of immune-related lncRNAs, we investigated whether these immune-related lncRNAs are capable of classifying LUAD patients into clinically-relevant subtypes. Here, we used expression profiles of the immune-related lncRNAs of LUAD patients from TCGA database as genomic signatures with unsupervised consensus clustering approach for discovery of distinct subgroups. According to the results, LUAD patients clustered into four distinct clusters with different patient numbers of 113, 186,121 and 104 (Fig. 8A). We further validated the separation between different patient groups by using a principal component analysis (PCA) [29]. The PCA map showed lower intra-cluster patient-to-patients compared to the inter-cluster similarities with the first two principal components contributing up to 74% of the total variation. In addition, patients in different subgroups exhibited unique expression profiles (Fig. 8B). To assess whether the stratification determined by the immune-related lncRNAs were associated with clinical outcomes, a KM survival analysis was adopted for evaluation of prognostic performance in the clusters with respect to overall survival. The immune-related lncRNA-based subtypes were significantly (log-rank test, $p < 0.01$) associated with overall survival in this LUAD population (Fig. 8C). Taken together, these findings demonstrate that the immune-related lncRNAs could be used as probes to identify novel patient subgroups with significant clinical outcomes expected.

Discussion

The immune system plays a vital role in development of LUAD. Technological advancement such as high-throughput technology has enabled partial characterization of biological features of LUAD infiltrating immune cells at gene and mRNA levels [50, 51]. Exploring more regulatory mechanisms of the immune system is of primary importance towards designing and innovating immunotherapeutic strategies. In recent years, the discovery of lncRNAs has provided a new perspective on gene regulation in diverse biological processes [52] although the role played by these factors during LUAD remains unclear. In order to provide insights into mechanisms regulating immune responses during LUAD, we obtained and analyzed RNA expression profiles of 101 LUAD patients from GEO database. A total of 11 modules were clustered using WGCNA and 156 lncRNAs in 3 immune-related modules were selected (Fig. 1). Immune-related clusters showed a significant correlation to the overall survival of patients.

To further verify the candidate immune-related lncRNAs in LUAD, 524 more patients from the TCGA database were divided into training and testing datasets and used to construct a prognostic classifier model. From the LASSO Coxph regression results, we identified RP6-91H8.3, RP11-504A18.1 and AC092171.4 as potential protection factors while RP11-1275H24.1, LINC01003 and AC006129.2 were found to be potential risk factors (Fig. 2). Four of these lncRNAs, except RP6-91H8.3 and RP11-1275H24.1 have been studied. Rong-quan He *et al.* reporting RP11-504A18.1 and AC006129.2 as independent prognostic indicators for soft-tissue sarcoma in patients [53]. Similarly, LINC01003 was reported to be a prognostic predictor for acute myeloid leukemia [54] while high expression of AC092171.4 in hepatocellular carcinoma patients correlated with poor survival [55]. However, the role of these immune-related lncRNAs during LUAD has not been studied. The results this, therefore, provides insights into this little-understood phenomenon.

Patients from TCGA database were divided into high and low ILRS score groups based on multivariate Coxph regression coefficient analysis of each patient. Recently, Huang et al. [56] constructed a lncRNA-based risk score model to predict prognosis of lung squamous cell carcinoma and confirmed that 9 lncRNAs were strongly related to this disease. In the current study, our results showed that a risk score was highly correlated to various clinical information in LUAD patients (Fig. 3). ILRS based ROC curve further exhibited a 1, 3, and 5-year survival of LUAD patients with AUC values of 0.71, 0.61, and 0.54, respectively, which were superior to traditional clinical indicators in predicting 3 and 5-year survival (Fig. 4). These results indicate that ILRS can serve as a new and independent risk factor for prediction of prognosis in LUAD patients.

Immune infiltration of the tumor can be directly related to prognosis of patients. In addition, tumor-infiltrating immune cells acquire unique characteristics in LUAD [57, 58]. To determine whether ILRS is related to immune infiltrating cells, we estimated the stromal, and immune scores as well as tumor Purity of each LUAD patient. We found that the immune score and tumor purity were significantly correlated with overall survival (Fig. 5), and up to 70% of the infiltrating immune cell types in the test dataset were T, B cells as well as macrophages (Fig. 6). Reports have demonstrated that activation of cytotoxic T cell responses is essential for tumor cell elimination and the adoptive immunotherapy requires isolation and expansion of specific CD8 + T cells [59, 60]. Consequently, prominent and distinct associations have been discovered between tumor-infiltrating B cell (TIB) - produced antibodies and survival in LUAD [61]. Moreover, the exhausted TIBs have been found to be associated with a regulatory T cell phenotype in non-small cell lung cancer patients. This is indicative of the protective nature of B cells and could form a basis for future targeted immunotherapies [62]. Accumulating evidence suggests that M2 polarization of tumor-associated macrophages (TAMs) play a vital role in cancer progression [63]. For instance, Wang DH et al. [64] found that targeting alveolar macrophages in EGFR-mutant LUAD patients showed potential to mitigate disease progression. In addition, several other studies have shown that lncRNAs could modulate activation of macrophages, B and T cells [18, 65–68]. To gain more insights into the functional roles of the immune-related lncRNAs, GO and pathway enrichment analyses performed in this study revealed that most of the significant items for each enrichment analysis were directly related to immune response (Fig. 7). The participation of immune-related lncRNAs in dendritic cell maturation, T cell

activation and cytotoxic T lymphocyte associated antigen-4 (CTLA-4) signaling indicates their importance in regulation of immune responses to LUAD. CTLA-4 antibodies were the first immune checkpoint inhibitors approved by the US Food and Drug Administration [69]. A combination of checkpoint inhibition therapy is being extensively evaluated for potential clinical benefits in tumor-related histology [70]. Therefore, it is possible that lncRNAs related to CTLA-4 signaling may guide the application of checkpoint inhibitors.

Conclusions

From the findings of this study, we confirm outstanding prognostic values of the immune-related lncRNAs following analysis of the role of immune-related lncRNAs in classification of LUAD patients. The unsupervised consensus clustering approach reveled four distinct clusters with unique expression profiles. In addition, the study revealed that immune-related lncRNA-based subtypes were significantly associated with overall survival in the LUAD population (Fig. 8). Taken together, our results present evidence that different immune-related lncRNA patterns are related to LUAD progression, implying that these specific lncRNAs may provide additional information for LUAD classification, prognosis as well as immunotherapy.

Abbreviations

PCA : principal component analysis; LUAD: Lung adenocarcinoma; lncRNAs: Long non-coding RNAs; ILRS: immune-related lncRNAs risk score; MHC: Major Histocompatibility Complex; WGCNA: weighted gene co-expression network analysis; TIB: tumor-infiltrating B cell; KM: Kaplan-Meier; PCC: Pearson's correlation coefficient; BP: biological process; MF: molecular function; CC: subcellular localization; DEGs: differentially expressed genes; ROC: receiver operator characteristics; TOM: topological overlap matrix; TAMs: tumor-associated macrophages; CTLA-4: T lymphocyte associated antigen-4; GEO: Gene Expression Omnibus.

Declarations

Ethics approval and Consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The gene expression data and clinical data in this study can be found online at the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/, GSE81089) and TCGA (https://www.genome.gov/Funded-

Programs-Projects/Cancer-Genome-Atlas) databases under accession numbers.

Competing Interest

The authors declare that there are no conflicts of interest with this work.

Funding

This work was supported by the Science and Technology Project of Shenzhen (NO. JCYJ20180228164407689), the National Natural Science Foundation of China (NO. U1603285) and China Postdoctoral Science Foundation (NO. 2018M631046).

Author Contributions

JBW, YHW and FRL supervised the project, ZHG interpreted the data, JBW and ZHG wrote the article. ZHG, LH, YHM, SNL and LJS designed the research, ZHG, LJS, LJW, YPL, LZL and FH analyzed all the data. YHM and HQW pervaded the technical support.

Acknowledgements

Not applicable.

Supplementary materials

Supplementary material associated with this article can be found in the online version.

Code availability

Not applicable.

References

1. Peng, Z., et al., *The long noncoding RNA LINC00312 induces lung adenocarcinoma migration and vasculogenic mimicry through directly binding YBX1*. Mol Cancer, 2018. **17**(1): p. 167.
2. Ding, F., et al., *Overexpression of S100A14 contributes to malignant progression and predicts poor prognosis of lung adenocarcinoma*. Thorac Cancer, 2018. **9**(7): p. 827–835.
3. Richardson, A.M., et al., *Vimentin Is Required for Lung Adenocarcinoma Metastasis via Heterotypic Tumor Cell-Cancer-Associated Fibroblast Interactions during Collective Invasion*. Clin Cancer Res, 2018. **24**(2): p. 420–432.
4. Ding, C., et al., *Circulating tumor cell levels and carcinoembryonic antigen: An improved diagnostic method for lung adenocarcinoma*. Thorac Cancer, 2018. **9**(11): p. 1413–1420.
5. Wei, X., et al., *GMDS knockdown impairs cell proliferation and survival in human lung adenocarcinoma*. BMC Cancer, 2018. **18**(1): p. 600.

6. Alam, H., et al., *HP1gamma Promotes Lung Adenocarcinoma by Downregulating the Transcription-Repressive Regulators NCOR2 and ZBTB7A*. Cancer Res, 2018. **78**(14): p. 3834–3848.
7. Liu, X., et al., *The prognostic landscape of tumor-infiltrating immune cell and immunomodulators in lung cancer*. Biomed Pharmacother, 2017. **95**: p. 55–61.
8. Mantovani, A., et al., *Cancer-related inflammation.* Nature, 2008. **454**(7203): p. 436–44.
9. Wang, J.B., X. Huang, and F.R. Li, *Impaired dendritic cell functions in lung cancer: a review of recent advances and future perspectives*. Cancer Commun (Lond), 2019. **39**(1): p. 43.
10. Bremnes, R.M., et al., *The Role of Tumor-Infiltrating Lymphocytes in Development, Progression, and Prognosis of Non-Small Cell Lung Cancer*. J Thorac Oncol, 2016. **11**(6): p. 789–800.
11. Sharma, P. and J.P. Allison, *The future of immune checkpoint therapy*. Science, 2015. **348**(6230): p. 56–61.
12. Chen, Y.P., et al., *Genomic Analysis of Tumor Microenvironment Immune Types across 14 Solid Cancer Types: Immunotherapeutic Implications*. Theranostics, 2017. **7**(14): p. 3585–3594.
13. Fridman, W.H., et al., *The immune contexture in cancer prognosis and treatment*. Nat Rev Clin Oncol, 2017. **14**(12): p. 717–734.
14. Djebali, S., et al., *Landscape of transcription in human cells*. Nature, 2012. **489**(7414): p. 101–8.
15. Zhou, W., et al., *Galectin-3 activates TLR4/NF-kappaB signaling to promote lung adenocarcinoma cell proliferation through activating lncRNA-NEAT1 expression*. BMC Cancer, 2018. **18**(1): p. 580.
16. Peng, F., et al., *Differential expression analysis at the individual level reveals a lncRNA prognostic signature for lung adenocarcinoma*. Mol Cancer, 2017. **16**(1): p. 98.
17. Iyer, M.K., et al., *The landscape of long noncoding RNAs in the human transcriptome*. Nat Genet, 2015. **47**(3): p. 199–208.
18. Huang, D., et al., *NKILA lncRNA promotes tumor immune evasion by sensitizing T cells to activation-induced cell death*. Nat Immunol, 2018. **19**(10): p. 1112–1125.
19. Jiang, M., et al., *Self-Recognition of an Inducible Host lncRNA by RIG-I Feedback Restricts Innate Immune Response*. Cell, 2018. **173**(4): p. 906–919 e13.
20. Hu, G., et al., *Expression and regulation of intergenic long noncoding RNAs during T cell development and differentiation*. Nat Immunol, 2013. **14**(11): p. 1190–8.
21. Chen, J., L. Ao, and J. Yang, *Long non-coding RNAs in diseases related to inflammation and immunity*. Ann Transl Med, 2019. **7**(18): p. 494.
22. Chen, J., et al., *LncRNA LINC01512 Promotes the Progression and Enhances Oncogenic Ability of Lung Adenocarcinoma*. J Cell Biochem, 2017. **118**(10): p. 3102–3110.
23. Sun, J., et al., *LncRNA XIST promotes human lung adenocarcinoma cells to cisplatin resistance via let-7i/BAG-1 axis*. Cell Cycle, 2017. **16**(21): p. 2100–2107.
24. Dong, H.X., et al., *LncRNA DGCR5 promotes lung adenocarcinoma (LUAD) progression via inhibiting hsa-mir-22-3p.* J Cell Physiol, 2018. **233**(5): p. 4126–4136.

25. Zhao, X., et al., *LncRNA HOXA11-AS drives cisplatin resistance of human LUAD cells via modulating miR-454-3p/Stat3.* Cancer Sci, 2018. **109**(10): p. 3068–3079.
26. Wang, W., et al., *An immune-related lncRNA signature for patients with anaplastic gliomas*. J Neurooncol, 2018. **136**(2): p. 263–271.
27. Liu, Z., et al., *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**: p. 850.
28. Xu, J., et al., *Long Noncoding RNA MIR17HG Promotes Colorectal Cancer Progression via miR-17-5p.* Cancer Res, 2019. **79**(19): p. 4882–4895.
29. Edgar, R., M. Domrachev, and A.E. Lash, *Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.* Nucleic Acids Research, 2002. **30**(1): p. 207–210.
30. Colaprico, A., et al., *TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data*. Nucleic Acids Research, 2015. **44**(8): p. e71-e71.
31. Lynn, D.J., et al., *InnateDB: facilitating systems-level analyses of the mammalian innate immune response*. Molecular systems biology, 2008. **4**(1).
32. Bhattacharya, S., et al., *ImmPort: disseminating data to the public for the future of immunology*. Immunologic research, 2014. **58**(2–3): p. 234–239.
33. Zhang, B. and S. Horvath, *A general framework for weighted gene co-expression network analysis*. Statistical applications in genetics and molecular biology, 2005. **4**(1).
34. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis*. BMC bioinformatics, 2008. **9**(1): p. 559.
35. Friedman, J., T. Hastie, and R. Tibshirani, *glmnet: Lasso and elastic-net regularized generalized linear models*. R package version, 2009. **1**(4).
36. Therneau, T.M. and T. Lumley, *Package 'survival'.* Survival analysis Published on CRAN, 2014.
37. Ihaka, R. and R. Gentleman, *R: a language for data analysis and graphics*. Journal of computational and graphical statistics, 1996. **5**(3): p. 299–314.
38. Blanche, P., *Package'timeROC': time-dependent ROC curve and AUC for censored survival data*. Vienna: R Foundation for Statistical Computing, 2013.
39. Yoshihara, K., et al., *Inferring tumour purity and stromal and immune cell admixture from expression data*. Nature communications, 2013. **4**: p. 2612.
40. Newman, A.M., et al., *Determining cell type abundance and expression from bulk tissues with digital cytometry*. Nature biotechnology, 2019: p. 1.
41. Kassambara, A., *ggpubr:"ggplot2" based publication ready plots.* R package version 0.1, 2017. **6**.
42. Ritchie, M.E., et al., *limma powers differential expression analyses for RNA-sequencing and microarray studies*. Nucleic Acids Research, 2015. **43**(7): p. e47-e47.
43. Huang, D.W., et al., *DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists*. Nucleic acids research, 2007. **35**(suppl_2):

p. W169-W175.

44. Kassambara, A., M. Kosinski, and P. Biecek, *survminer: Drawing Survival Curves using'ggplot2'*. R package version 0.3, 2017. **1**.
45. Alvarez, M.J., et al., *A precision oncology approach to the pharmacological targeting of mechanistic dependencies in neuroendocrine tumors*. Nature genetics, 2018. **50**(7): p. 979.
46. Wilkerson, M.D. and D.N. Hayes, *ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking*. Bioinformatics, 2010. **26**(12): p. 1572–1573.
47. Ginestet, C., *ggplot2: elegant graphics for data analysis*. Journal of the Royal Statistical Society: Series A (Statistics in Society), 2011. **174**(1): p. 245–246.
48. Solomon, B., et al., *Identification of an excellent prognosis subset of human papillomavirus-associated oropharyngeal cancer patients by quantification of intratumoral CD103 + immune cell abundance*. Ann Oncol, 2019. **30**(10): p. 1638–1646.
49. Choi, S.K., et al., *Heterogeneous oncologic outcomes according to surgical pathology in high-risk prostate cancer: implications for better risk stratification and preoperative prediction of oncologic outcomes*. J Cancer Res Clin Oncol, 2017. **143**(9): p. 1871–1878.
50. Calvert, R.J., et al., *K-ras 4A and 4B mRNA levels correlate with superoxide in lung adenocarcinoma cells, while at the protein level, only mutant K-ras 4A protein correlates with superoxide*. Lung Cancer, 2013. **80**(3): p. 263–9.
51. Yu, N., et al., *Identification of tumor suppressor miRNAs by integrative miRNA and mRNA sequencing of matched tumor-normal samples in lung adenocarcinoma*. Mol Oncol, 2019. **13**(6): p. 1356–1368.
52. Atianand, M.K., D.R. Caffrey, and K.A. Fitzgerald, *Immunobiology of Long Noncoding RNAs.* Annu Rev Immunol, 2017. **35**: p. 177–198.
53. He, R.Q., et al., *Prediction of clinical outcome and survival in soft-tissue sarcoma using a ten-lncRNA signature*. Oncotarget, 2017. **8**(46): p. 80336–80347.
54. Zeng, H., et al., *Characterization of a 4 lncRNAs-based prognostic risk scoring system in adults with acute myeloid leukemia*. Leuk Res, 2019. **88**: p. 106261.
55. Liao, H.T., et al., *Identification of The Aberrantly Expressed LncRNAs in Hepatocellular Carcinoma: A Bioinformatics Analysis Based on RNA-sequencing*. Sci Rep, 2018. **8**(1): p. 5395.
56. Huang, G., et al., *A nine-long non-coding RNA signature for prognosis prediction of patients with lung squamous cell carcinoma*. Cancer Biomark, 2019. **26**(3): p. 239–247.
57. Meng, J., et al., *The establishment of immune infiltration based novel recurrence predicting nomogram in prostate cancer*. Cancer Med, 2019. **8**(11): p. 5202–5213.
58. Lavin, Y., et al., *Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses*. Cell, 2017. **169**(4): p. 750–765 e17.
59. Atzin-Mendez, J.A., et al., *Expansion of quiescent lung adenocarcinoma CD8 + T cells by MUC1-8-mer peptide-T2 cell-beta2 microglobulin complexes*. Oncol Rep, 2016. **35**(1): p. 33–42.

60. Flores-Villanueva, P.O., et al., *An Isolated TCR alphabeta Restricted by HLA-A*02:01/CT37 Peptide Redirecting CD8(+) T Cells To Kill and Secrete IFN-gamma in Response to Lung Adenocarcinoma Cell Lines*. J Immunol, 2018. **200**(8): p. 2965–2977.
61. Isaeva, O.I., et al., *Intratumoral immunoglobulin isotypes predict survival in lung adenocarcinoma subtypes*. J Immunother Cancer, 2019. **7**(1): p. 279.
62. Bruno, T.C., et al., *Antigen-Presenting Intratumoral B Cells Affect CD4(+) TIL Phenotypes in Non-Small Cell Lung Cancer Patients*. Cancer Immunol Res, 2017. **5**(10): p. 898–907.
63. Xu, F., et al., *Astragaloside IV inhibits lung cancer progression and metastasis by modulating macrophage polarization through AMPK signaling*. J Exp Clin Cancer Res, 2018. **37**(1): p. 207.
64. Wang, D.H., et al., *Progression of EGFR-Mutant Lung Adenocarcinoma is Driven By Alveolar Macrophages*. Clin Cancer Res, 2017. **23**(3): p. 778–788.
65. Cao, J., et al., *LncRNA-MM2P Identified as a Modulator of Macrophage M2 Polarization*. Cancer Immunol Res, 2019. **7**(2): p. 292–305.
66. Li, X., et al., *Regulation of Macrophage Activation and Polarization by HCC-Derived Exosomal lncRNA TUC339*. Int J Mol Sci, 2018. **19**(10).
67. Brazao, T.F., et al., *Long noncoding RNAs in B-cell development and activation*. Blood, 2016. **128**(7): p. e10-9.
68. Chen, Y.G., A.T. Satpathy, and H.Y. Chang, *Gene regulation in the immune system by long noncoding RNAs*. Nat Immunol, 2017. **18**(9): p. 962–972.
69. Pardoll, D.M., *The blockade of immune checkpoints in cancer immunotherapy*. Nat Rev Cancer, 2012. **12**(4): p. 252–64.
70. Chae, Y.K., et al., *Current landscape and future of dual anti-CTLA4 and PD-1/PD-L1 blockade immunotherapy in cancer; lessons learned from clinical trials with melanoma and non-small cell lung cancer (NSCLC)*. J Immunother Cancer, 2018. **6**(1): p. 39.

Figures

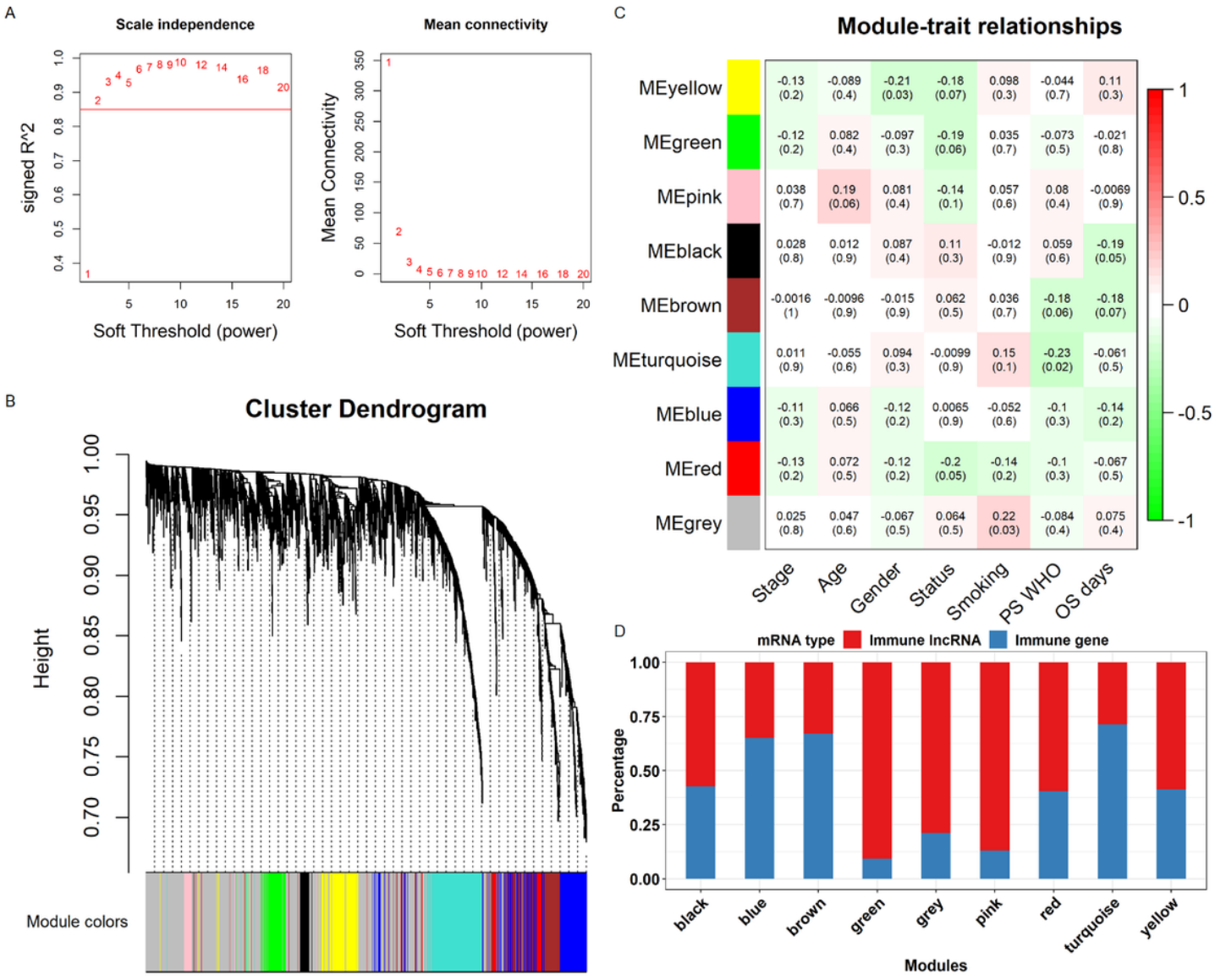

Figure 1

Identification of prognosis related modules with WGCNA. (A) Analysis of the scale-free fit index and mean connectivity for various soft-thresholding powers. (B) Dendrogram of all differentially expressed genes clustered based on a dissimilarity measure (1-TOM). (C) Association between modules and clinical features. Pearson correlation coefficient and p-value are provided for each cell. (D) Percentage of immune-related genes and lncRNAs in each module.

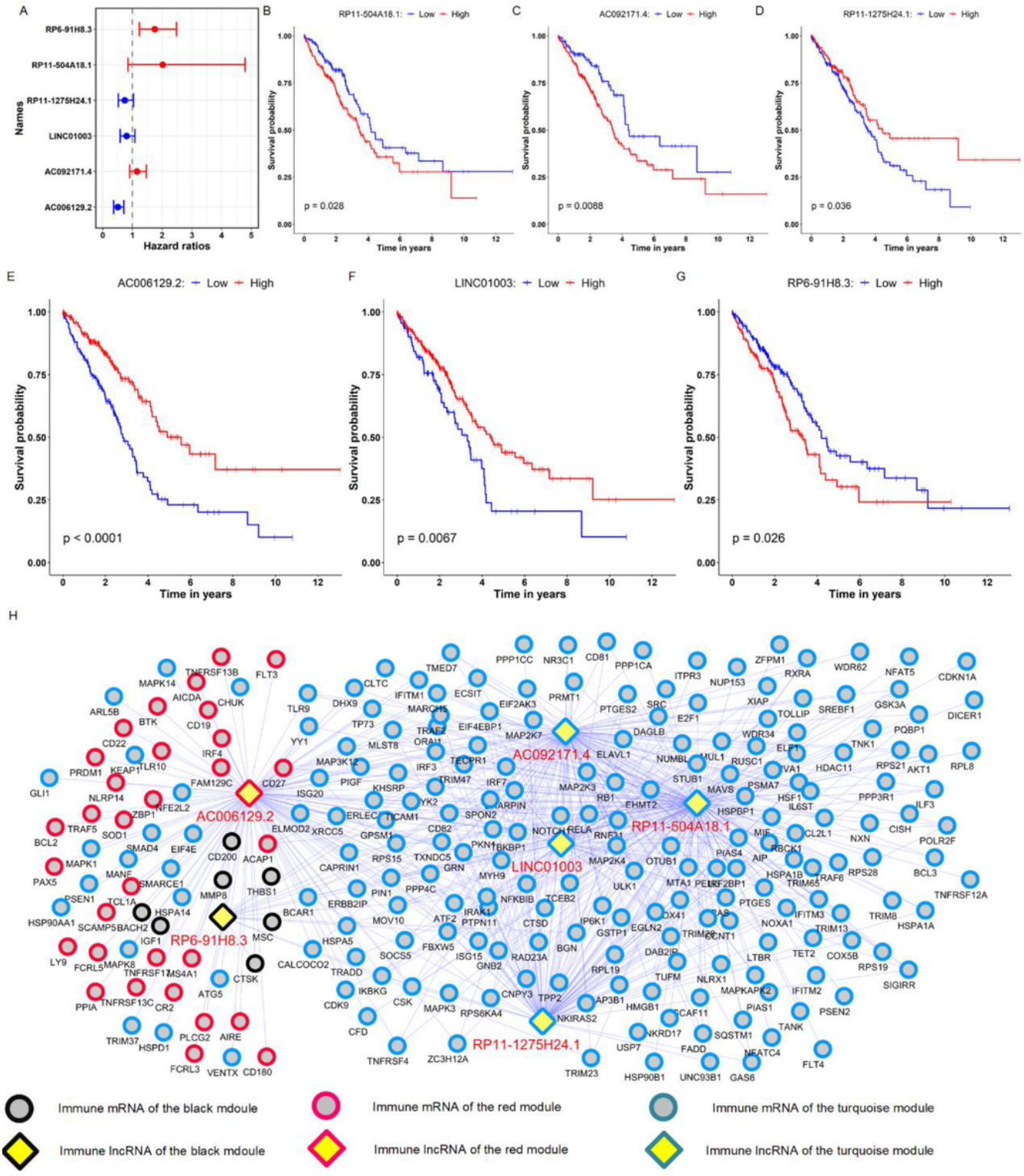

Figure 2

Immune-related lncRNA as a prognostic biomarker. (A) Hazard ratios and 95% confidence intervals of 6 immune-related lncRNAs in the training set. (B-G) Relationship between immune-related lncRNA expression and overall survival. (H) Interaction network between immune-related lncRNAs and immune related genes.

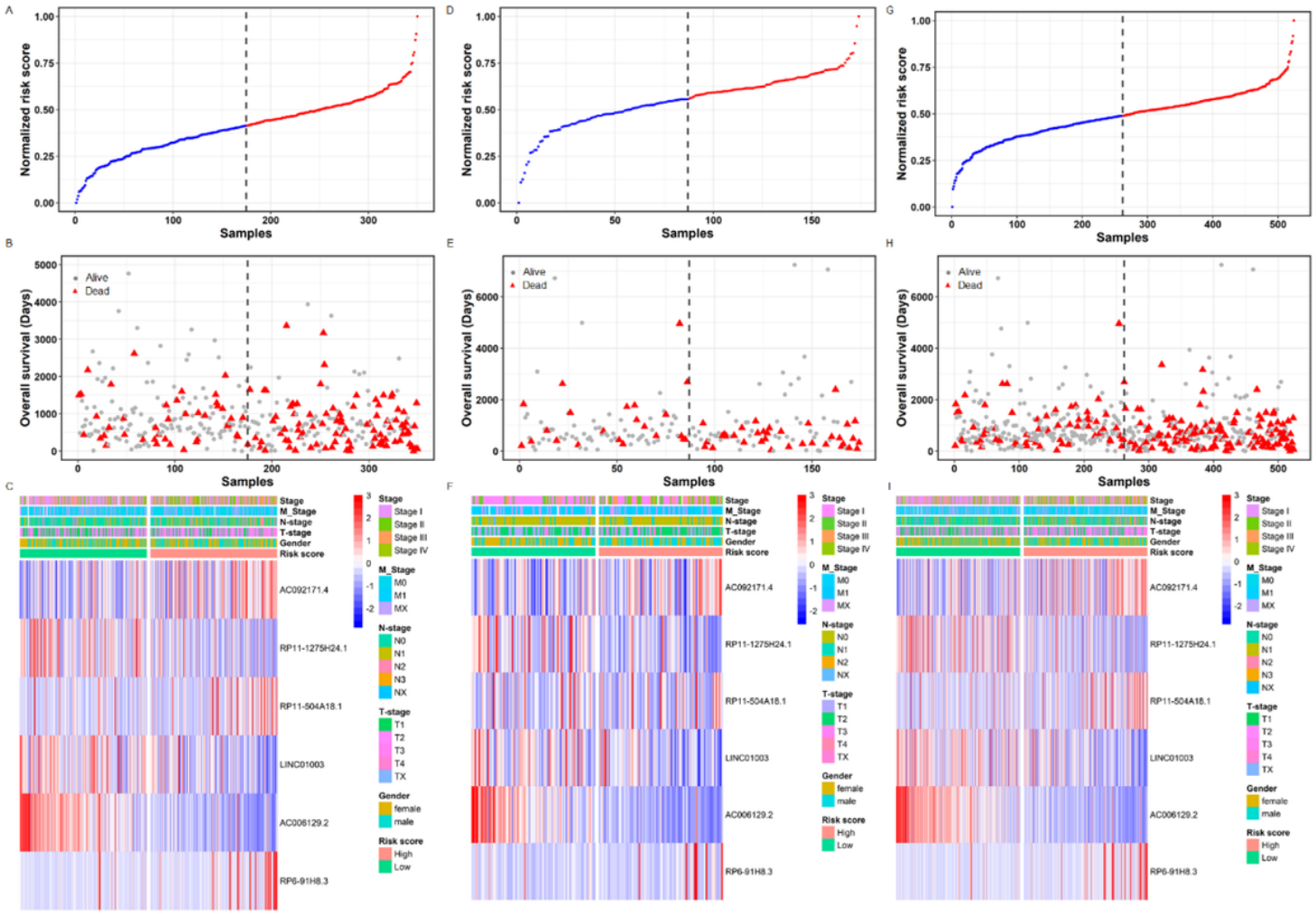

Figure 3

Relationship between ILRS and clinical outcomes of LUAD patients. ILRS rank (A), distribution of overall survival rate (B) and expression pattern of immune-related lncRNAs (C) of LUAD patients in training dataset. ILRS rank (D), distribution of overall survival rate (E) and expression pattern of immune-related lncRNAs (F) of LUAD patients in testing dataset. ILRS rank (G), distribution of overall survival rate (H) and expression pattern of immune-related lncRNAs (I) of LUAD patients in entire dataset.

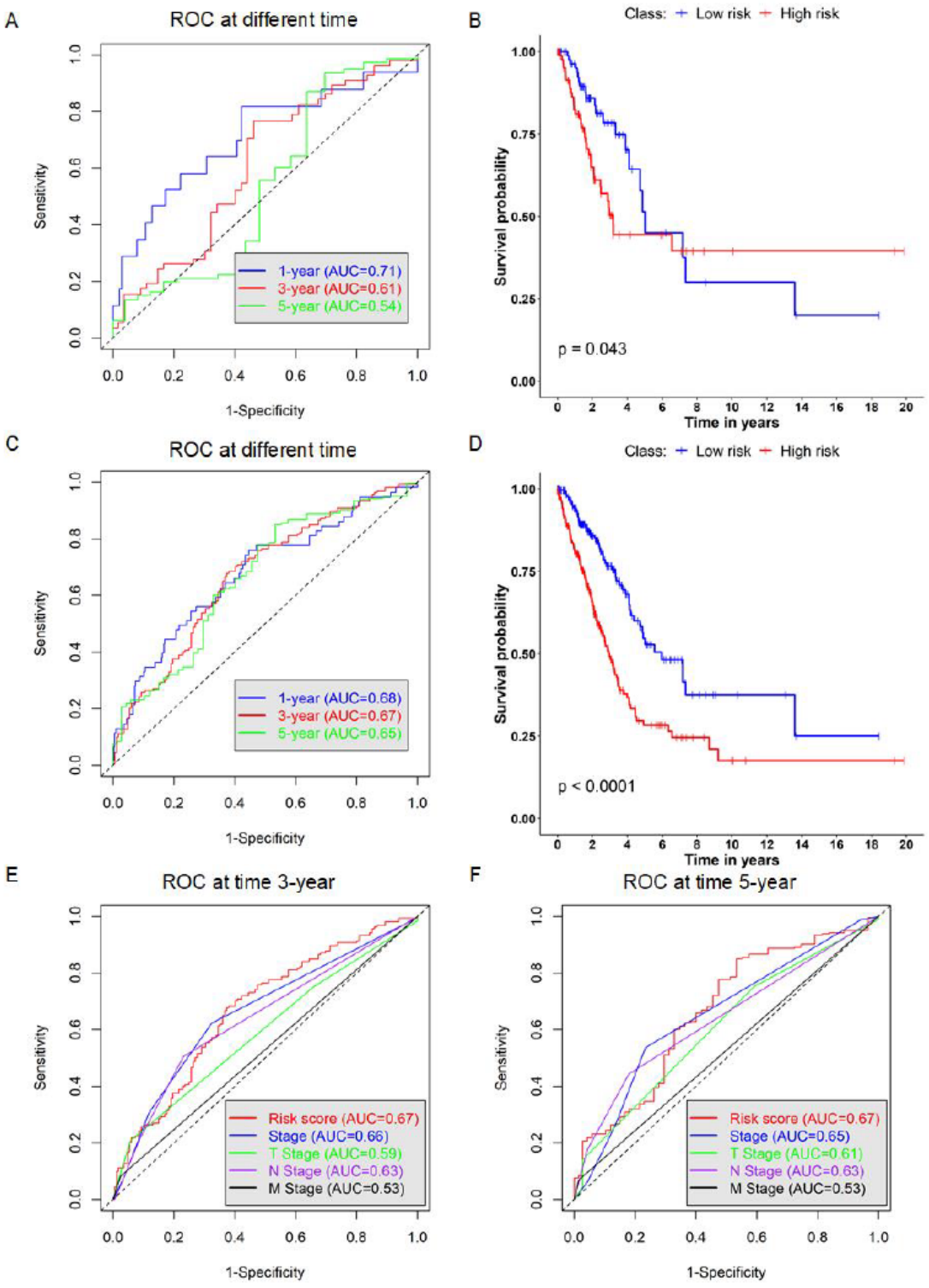

Figure 4

Evaluation of prognostic value of ILRS. (A) The ability of ILRS to predict 1, 3 and 5-year survival of LUAD patients in testing dataset. (B) Relationship between ILRS and overall survival of LUAD patients in testing dataset. (C) Prediction of 1, 3 and 5-year survival rate in entire dataset based on ILRS. (D) Relationship between ILRS and overall survival in the entire dataset. The ability of ILRS, stage, T stage, N stage and M stage to predict the 3-year (E) and 5-year (F) survival in the entire dataset.

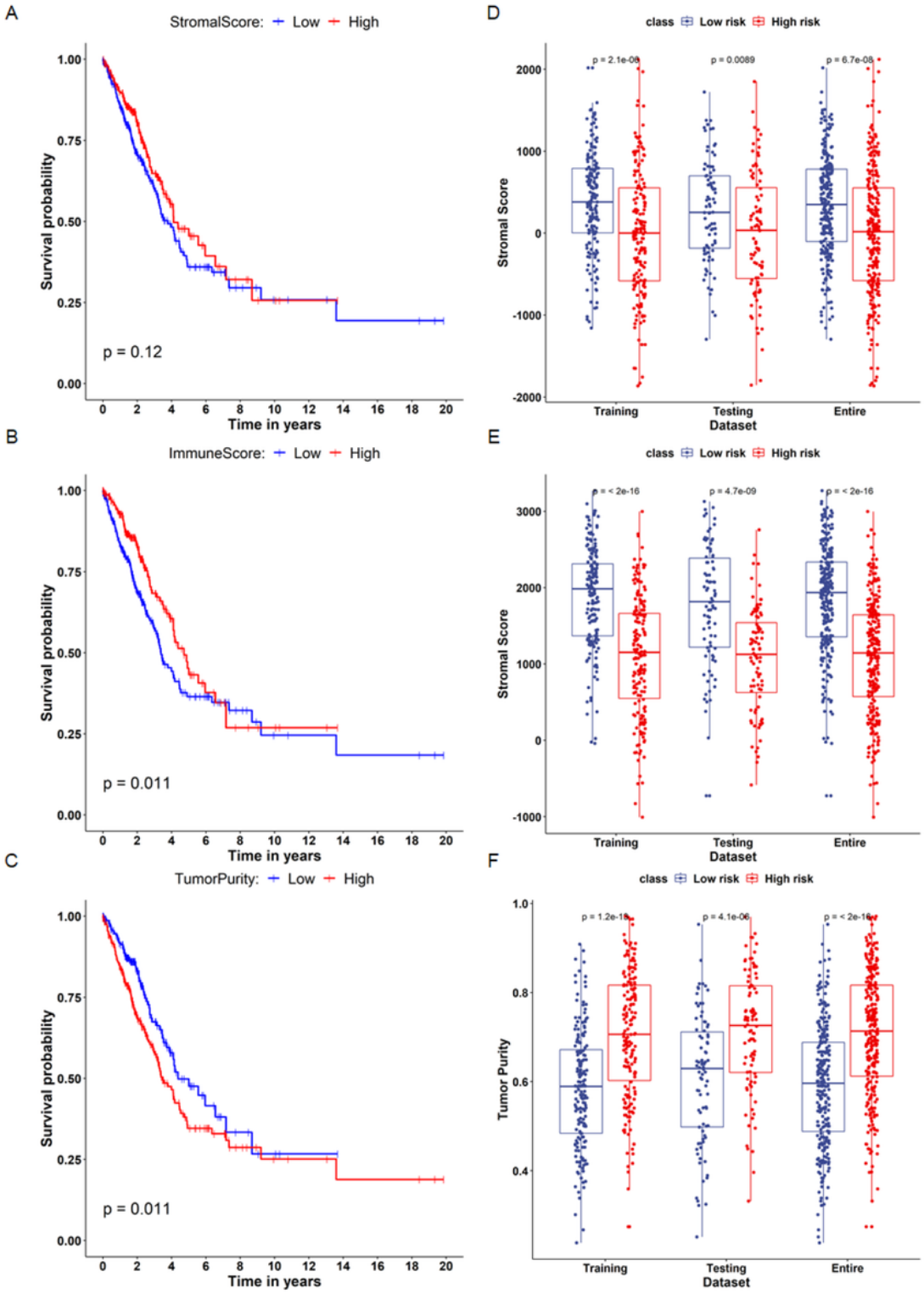

Figure 5

Stromal score and immune infiltration score in the high ILRS and low ILRS groups. Relationship between stromal score (A), immune score (B), tumor purity (C) and overall survival. Stromal score (D), immune score (E) and tumor purity (F) in high ILRS and low ILRS classes of TCGA-LUAD training, testing and entire dataset.

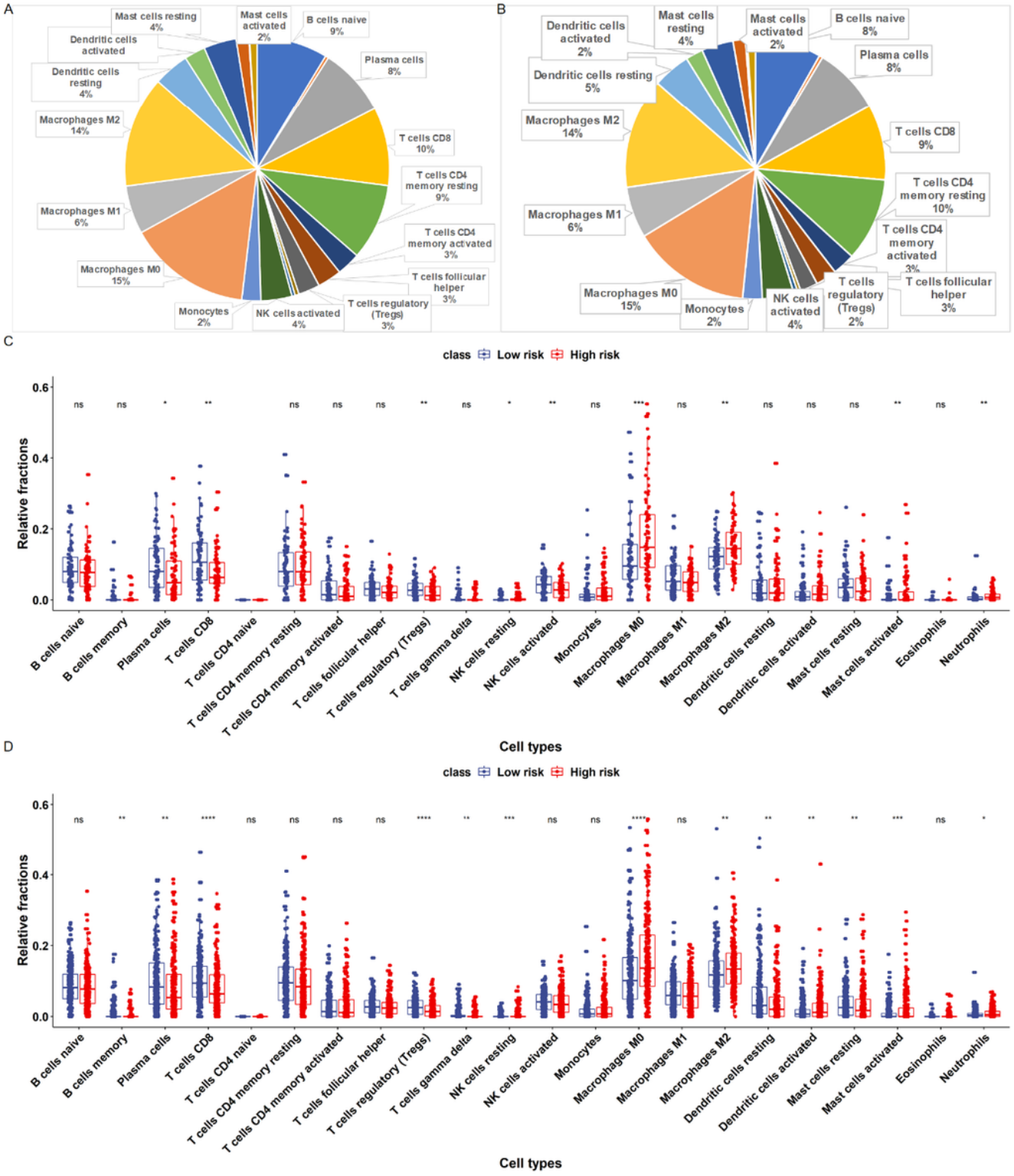

Figure 6

Dissecting the leukocyte cell subsets between high ILRS and low ILRS groups. (A) Mean relative abundance of leukocyte cell in TCGA-LUAD testing dataset. (B) Mean relative abundance of leukocyte cell in TCGA-LUAD entire dataset. (C) Relative abundance of leukocytes in high ILRS and low ILRS classes of TCGA-LUAD testing samples. (D) Relative abundance of leukocytes in high ILRS and low ILRS classes of TCGA-LUAD entire samples. ns: $p > 0.05$; *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$; ****: $p \leq 0.0001$.

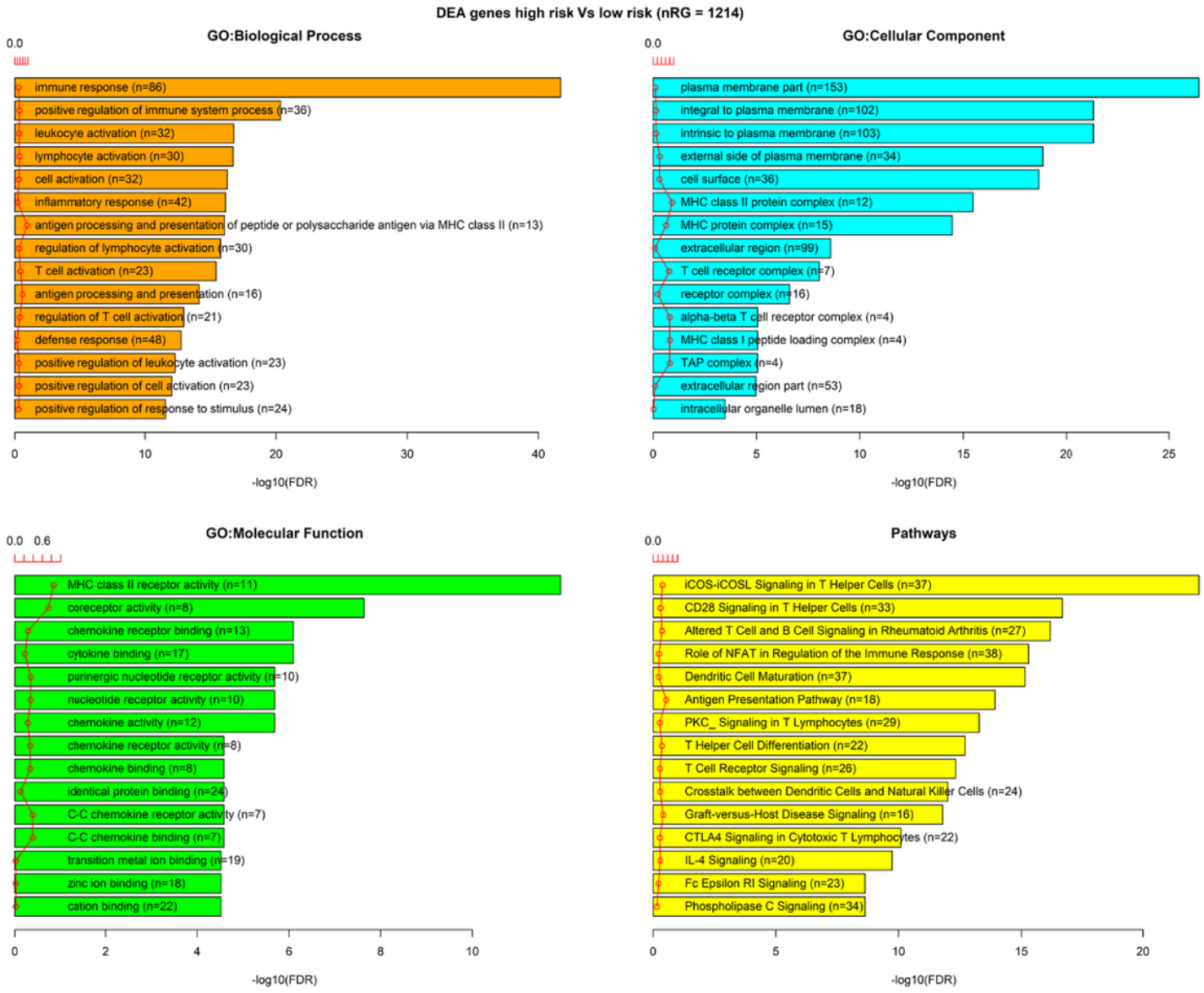

Figure 7

GO and KEGG enrichment analysis for the differentially expressed genes (DEG) between the high ILRS and low ILRS group. (A) GO: Biological Process enrichment analysis on the DEGs. (B) GO: Cellular Component enrichment analysis on the DEGs. (C) GO: Molecular Function enrichment analysis on the DEGs. (D) Pathway enrichment analysis for the DEGs. n is the number of DEGs involved in this term.

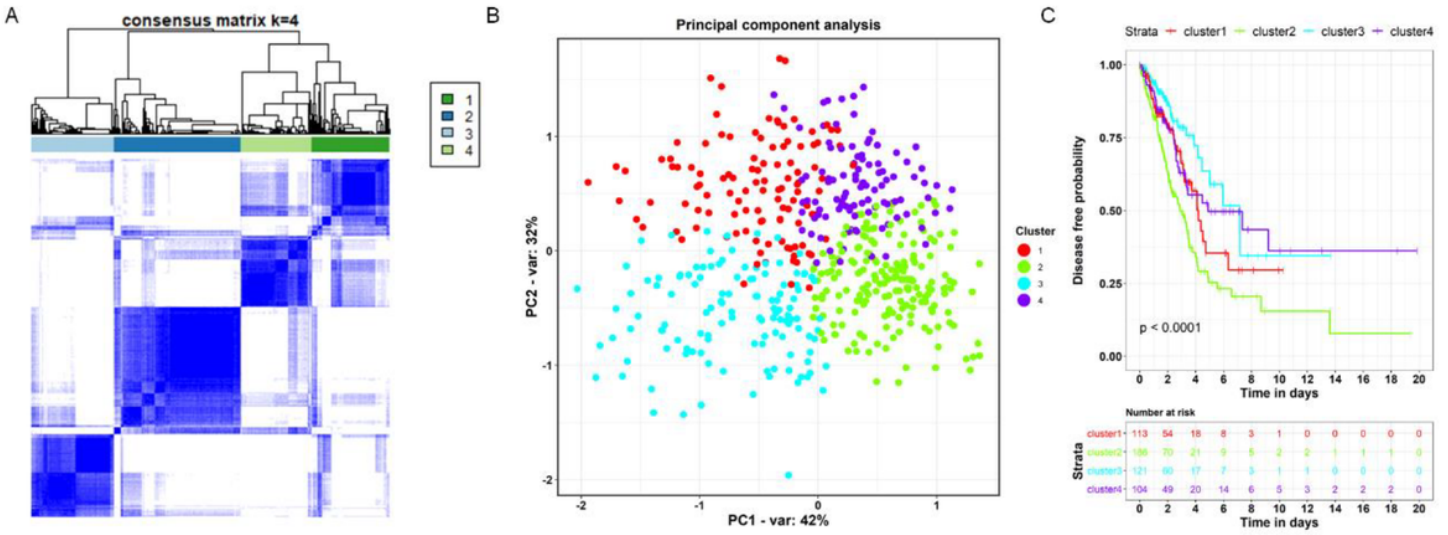

Figure 8

Stratification of TCGA-LUAD patients based on the expression of immune-related lncNRAs. (A) Consensus clustering analysis of 524 TCGA-LUAD samples based on the expression profile of the RNAs in the ceRNA network. The heatmap shows the Pearson's correlation coefficient of sample-by-sample. (B) Identification of TCGA-LUAD population by unsupervised clustering (A). Each point depicts a single patient, colored according to cluster designation. The principal component analysis (PCA) was used to reduce dimensionality, and the first and second principal components (x-axis and y-axis) were employed to visualize the expression patterns of different clusters. (C) Kaplan-Meier curves and risk table showing the association between the immune-related lncNRA biomarker subtypes and the overall survival time.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.csv
- supplementaryfigure.docx