

Research Paper

Identification of the key genes and characterizations of Tumor Immune Microenvironment in Lung Adenocarcinoma (LUAD) and Lung Squamous Cell Carcinoma (LUSC)

Lemeng Zhang, Jianhua Chen[✉], Tianli Cheng, Hua Yang, Haitao Li, Changqie Pan

Thoracic Medicine Department 1, Hunan Cancer Hospital, Changsha, Hunan Province, P.R. China, 410013.

✉ Corresponding author: Chen Jianhua, Thoracic Medicine Department 1, Hunan Cancer Hospital, Changsha, Hunan Province, P.R. China, 410013, E-mail address: cjh_1000@163.com. Tel: +86-731-89762220; FAX: +86-731-89762220. Address: Tongzipo Rd 283#, Yuelu District, Changsha, Hunan Province, P.R. China, 410013.

© The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>). See <http://ivyspring.com/terms> for full terms and conditions.

Received: 2019.11.27; Accepted: 2020.05.29; Published: 2020.06.16

Abstract

This study aimed to investigate the key genes and immune microenvironment involved in different TNM stages of lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC). The gene expression and clinical characteristics data were downloaded from the genomic data commons (GDC) database. After initial data processing, the characteristics of the immune microenvironment were analyzed. The differentially expressed genes (DEGs) in tumor vs. normal, and in early vs. advanced stages were screened, followed by Spearman correlation test for tumor infiltrating immune cells (TIICs) to identify immune-related genes. Finally, functional enrichment, protein-protein interaction, and survival analyses were performed. In LUAD, early stage was with higher immune scores, greater number of memory B cells and M0 macrophages compared to advanced stage. M0 and M2 macrophages, and resting memory CD4+ T cells accounted for a large proportion of TIICs in LUAD. The abundance of M0 macrophage infiltration was significantly correlated with the TNM stage and survival. In LUSC, early stage was with higher cytolytic activity and neoantigen burden compared to advanced stage. M0 and M2 macrophages, and plasma cells accounted for a large proportion of TIICs in LUSC. The abundance of resting and activated mast cells was significantly correlated with TNM stage, while resting dendritic cells, eosinophils, activated memory CD4 T cells, and mast cells were significantly correlated with prognosis. Tumor mutation burden analysis revealed that the median of variants per sample decreased from stage I to IV in LUAD, while it increased in LUSC. Further, 83 and 9 immune-related DEGs were identified in LUAD and LUSC, respectively, of which 23 genes in LUAD and 2 genes in LUSC correlated with survival. In conclusion, we identified the key genes, and characterized the tumor immune microenvironment in LUAD and LUSC which may provide therapeutic targets for the treatment of NSCLC.

Key words: Immune microenvironment, Lung adenocarcinoma, Lung squamous cell carcinoma, Tumor mutation burden, differentially expressed genes

Introduction

Lung cancer remains the leading cause of cancer-related deaths worldwide [1]. Non-small cell lung cancer (NSCLC) accounts for approximately 85% of lung cancer and the 5-year survival range from 73% in stage IA to 15% in stage IV [2]. Besides clinical TNM stage, the histologic type and differentiation are also important factors associated with the prognosis

[3]. Lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) are the two major histologic subtype of NSCLC [4].

The therapeutic prospects for advanced NSCLC have changed remarkably with a deeper understanding of tumor and immune cells interactions and the development of immunological

checkpoint inhibitors [5, 6]. The immune surveillance can be co-opted by tumor cells to escape immune destruction [7, 8]. Taube et al. revealed that PD-L1 (a ligand of programmed death-1, PD-1) was expressed by cancer cells and immune infiltrating cells and its expression reflected an immune-active microenvironment [9]. The differences in recruitment and localization of immune cells in the tumor environment may represent different therapeutic and prognostic values [10]. High density of mature dendritic cells and CD8+ T lymphocytes were found to be related to an improved outcome in NSCLC [11].

The occurrence and development of lung cancer is a complex and dynamic process that relies on the synergy between gene mutations and tumor microenvironment [12]. The abundant communications among genes, proteins, as well as cells within tumor microenvironments, are the resource for biomarkers and targets in the diagnosis and treatment of lung cancer [10]. The modification and interaction of genes in the immune microenvironment among different histological types and clinical stages of lung cancer will provide a systematic therapeutic perspective for lung cancer. However, the immune microenvironment of NSCLC in different clinical stage and histological type has not been fully understood. Therefore, further comprehensive analysis of the tumor immune microenvironment is urgently needed.

However, it is difficult and laborious to identify potential therapeutic targets and biomarkers by comprehensively and systematically profiling various immune cells from heterogeneous tumor samples based only on experiments. Bioinformatics approaches can directly extract cell-type specific information using sophisticated computational approaches [13]. Studies of the tumor immune microenvironment increasingly involve complex datasets which mainly depend on sophisticated computational methods [14]. The bioinformatics approaches were used in our study to investigate the immune microenvironment and the genes involved in different stages of LUAD and LUSC. Based on the gene expression and clinical data in genomic data commons (GDC, <https://gdc.cancer.gov>) database, the immune microenvironment characteristics, including immune score, cytolytic activity score, tumor mutation burden (TMB), neoantigens, and immune cells infiltration were investigated and the immune-related differential expressed genes (DEGs) were identified and analyzed. This study will provide a comprehensive perspective for the treatment of LUAD and LUSC.

Materials and Methods

Data acquisition and preprocessing

The RNA-seq fragments per kilobase million (FPKM), including corresponding clinical phenotype data of the two histologic subtype of NSCLC, including LUAD and LUSC were downloaded from the GDC database. The platform for the RNAseq data was Illumina HiSeq 2000 RNA Sequencing platform, and all the data were downloaded in August 2019. The data were re-annotated based on the annotation information (hg39, V22) in Gencode (<https://www.gencodegenes.org/>) database [15]. The Ensembl_ID was converted into Symbol_ID, and the Ensembl_ID with the highest expression value was selected when multiple Ensembl_IDs matched the same Symbol_ID.

Immune microenvironment analysis

Based on the formula,

$$TPM_i = \left(\frac{FPKM_i}{\sum_j FPKM_j} \right) \cdot 10^6$$

the gene expression data in FPKM format were converted into transcripts per million (TPM), which were then used in the analysis of immune microenvironment. The immune scores were calculated using the ESTIMATE (version 1.0.13) [16] in R package. The immune cytolytic activity score was calculated by the log-average of GZMA and PRF1 expression values in TPM [17]. Then, the TMB was analyzed using the Maftools (version 2.0.16) [18] in R package based on the somatic mutation files called by using muTect software, followed by plotting of summary and oncoplots. Information about the neoantigens and neoantigen origin protein were downloaded from The Cancer Immunome Atlas (TCIA, <https://tcia.at/home>) database. The amount of neoantigens and neoantigen origin protein for each sample were counted, followed by standardization by log10. The *t* test was performed on the difference of immune scores, immune cytolytic activity score, neoantigens and neoantigen origin protein between tumor stages, and boxplot was created using ggpubr (version 0.2.2)[19] in R package. The tumor infiltrating immune cells (TIICs) in LUAD and LUSC were analyzed based on the CIBERSORT algorithm [20], to estimate the abundance value of 22 immune cell types between different tumor stages.

Identification of immune-related DEGs

After filtering the genes with low expression values (genes with a count of 0 in more than 50 % samples), the differentially expressed analysis between early stages and advanced stages were performed for the tumor samples with clinical stages

information. The differentially expressed analysis of tumor vs. normal were also performed using limma package [21] (version: 3.40.6) after data preprocessing using edgeR (version: 3.26.6) [22]. The DEGs were screened with the threshold of P value < 0.05 and $|\log_2 \text{fold change (FC)}| > 0.585$, and the volcano plot and heatmaps were plotted using ggplot2 (version: 3.2.1) and pheatmap (version: 1.0.12), respectively. The Spearman correlation test between DEGs and abundance of TIICs were conducted, and the immune-related DEGs were selected with the threshold of P value < 0.05 and $|r| > 0.15$.

Functional enrichment analysis

To investigate the functional involvement of the DEGs, clusterProfiler [23] (version: 3.12.0) in R package was used to conduct the gene ontology (GO) terms and KEGG pathways enrichment analysis for all the DEGs based on Fisher exact test. GO terms included three categories: biological process (BP), cellular component (CC), and molecular function (MF). The significantly enriched GO terms and KEGG pathways were selected with the cut-off of P value < 0.05.

Protein-protein interaction (PPI) network

The interactions between DEGs were retrieved from STRING (version: 10.0, <http://string-db.org/>) database with a minimum required interaction score setting of 0.900 (highest confidence). Then, the PPI network was visualized using Cytoscape software, in which the functional modules were identified using MCODE plugin [24] of Cytoscape. The functional modules with score > 10 were selected, and the immune related DEGs in the functional modules were marked.

Survival analysis of DEGs

The prognosis-related clinical data including overall survival (OS) and smoking history were retrieved. The samples were divided into high-expression and low-expression groups based on the median value of each gene expression, together with log-rank statistical test. The significant threshold was set as P value < 0.05 to select the genes associated with prognosis, and then Kaplan-Meier (K-M) survival curves were plotted.

Results

Basic characteristics of samples

Detailed clinical and pathological characteristics of the 519 LUAD patients and 550 LUSC patients are shown in Table 1A and Table 1B, respectively. The 519 LUAD patients smoked 41.32 ± 26.93 years with OS of 31.86 ± 30.58 months, of which 184 patients (35.45%)

died. The 550 LUSC patients smoked 52.66 ± 31.08 years with OS of 36.90 ± 38.09 months, of which 286 patients (52.00%) died.

A total of 496 LUAD and 490 LUSC samples had data pertaining to both gene expression and clinical phenotype. A total of 19,712 genes were obtained in both LUAD and LUSC samples after re-annotation, and finally 17,522 and 17,751 genes were obtained in the LUAD and LUSC samples, following filtering out of genes with low expression values.

Table 1. The clinical and pathological characteristics of 519 LUAD patients and 550 LUSC patients

Clinical	Describe
(A) Basic characteristics of LUAD samples	
pathologic_M (M1/M0)	26/346
pathologic_N (N3/N2/N1/N0)	32/72/96/335
pathologic_T (T4/T3/T2/T1)	17/45/278/174
Stage (IV / III / II / I)	27/81/124/287
Group (advance/ early)	108/411
OS (months)	31.86±30.58
Survival_status (Dead/Alive)	184/335
Smoked years	41.32±26.93
(B) Basic characteristics of LUSC samples	
pathologic_M (M1/M0)	7/407
pathologic_N (N3/N2/N1/N0)	23/46/131/293
pathologic_T (T4/T3/T2/T1)	23/69/292/115
Stage (IV / III / II / I)	7/86/161/245
Group (advance / early)	93/406
OS (months)	36.90±38.09
Survival_status (Dead/Alive)	213/286
Smoked years	52.66±31.08

Immune score and immune cytolytic activity score in LUAD and LUSC

The immune cells from the tumor immune microenvironment play an important role in tumor progression. The density and location of such immune cells can be quantified by an immune score which is considered to be a tangible indicator of prognosis of the tumor. The cytolytic activity score has been reported to be related to improved prognosis and anti-regulatory activities that limit the immune response [17]. In this study, for the LUAD subtype, there was significant difference in the immune scores between the early and advanced stages, and stage I showed the highest immune score out of the four stages (Figure 1A). Further, the immune cytolytic activity score among the stages was also calculated, however, no significant difference was found among the stages (Figure 1B). In the LUSC subtype, there was no significant difference in the immune score between the early and advanced stages (Figure 1C). Conversely, the distribution of immune cytolytic activity score was different from the immune score. The early stages showed higher immune cytolytic activity score than the advanced stage (Figure 1D).

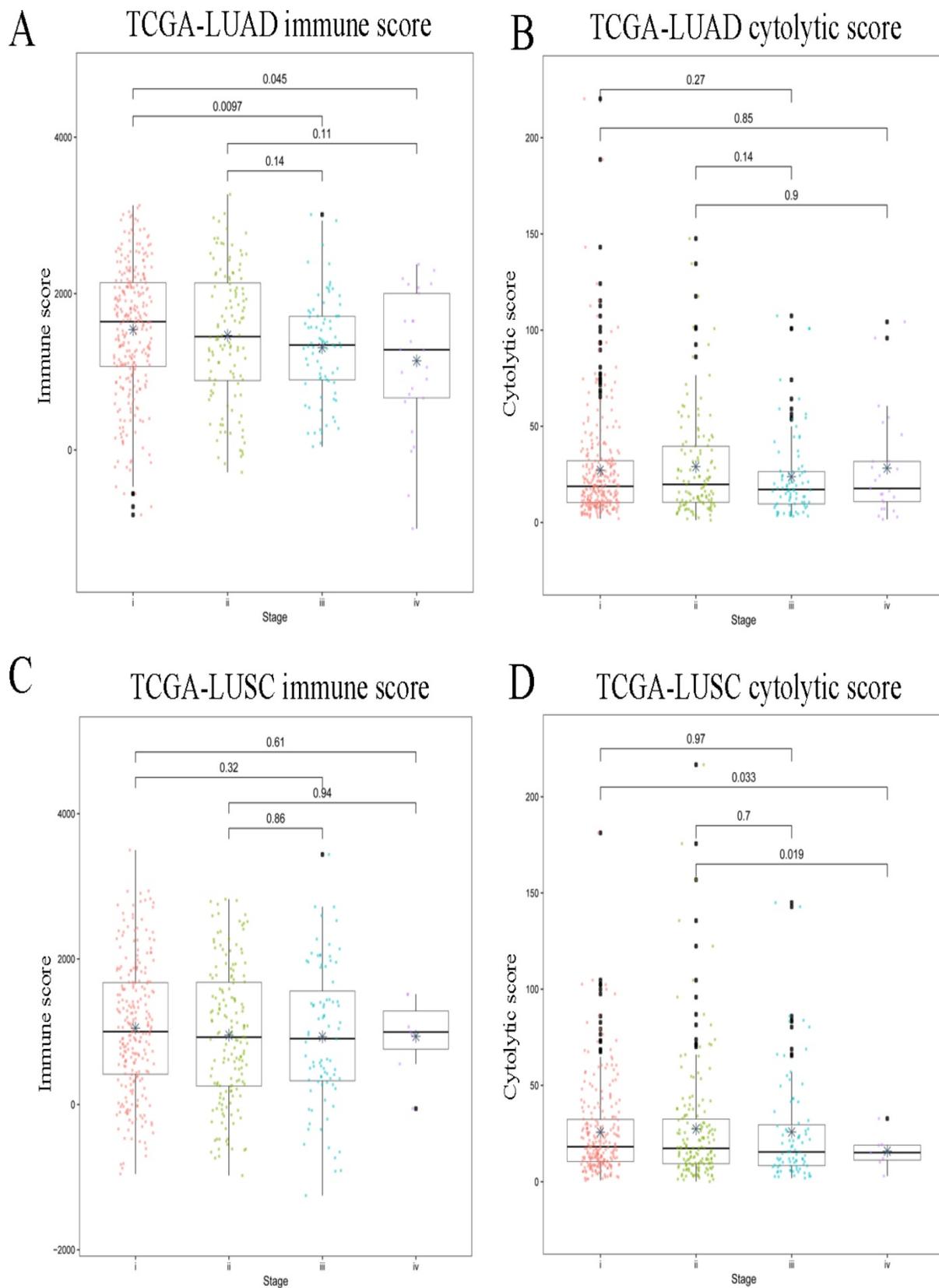


Figure 1. Immune scores and immune cytolytic activity scores between early and advanced stages in LUAD and LUSC. (A) Distribution of immune scores in LUAD. **(B)** Distribution of immune cytolytic activity scores in LUAD. **(C)** Distribution of immune scores in LUSC. **(D)** Distribution of immune cytolytic activity scores in LUSC. The horizontal line represents the median value; asterisks represent the mean value. The number represents P value, and P < 0.05 shows significant statistical difference between two stages.

Table 2. The top 10 mutated genes in samples with different tumor stages

(A) The top 10 mutated genes in LUAD samples with different tumor stages							
Stage I	Proportion	Stage II	Proportion	Stage III	Proportion	Stage IV	Proportion
TP53	46%	TTN	50%	TTN	50%	TP53	56%
TTN	43%	TP53	50%	TP53	46%	KEAP1	44%
MUC16	41%	CSMD3	41%	MUC16	41%	CSMD3	40%
CSMD3	38%	RYR2	38%	RYR2	36%	SPTA1	40%
RYR2	35%	MUC16	37%	LRP1B	35%	TTN	32%
LRP1B	33%	USH2A	34%	USH2A	30%	RYR2	28%
ZFHX4	32%	KRAS	31%	KRAS	28%	RYR3	28%
USH2A	31%	ZFHX4	31%	CSMD3	25%	KRAS	24%
KRAS	25%	LRP1B	30%	FLG	24%	RPIL1	24%
FLG	25%	SPTA1	30%	COL11A1	22%	ZFHX4	24%
(B) The top 10 mutated genes in LUSC samples with different tumor stages							
Stage I	Proportion	Stage II	Proportion	Stage III	Proportion	Stage IV	Proportion
TP53	81%	TP53	78%	TP53	73%	MUC16	86%
TTN	71%	TTN	71%	TTN	71%	TP53	86%
CSMD3	39%	CSMD3	44%	CSMD3	49%	TTN	86%
MUC16	38%	MUC16	38%	RYR2	43%	ADGRL3	71%
RYR2	36%	RYR2	34%	MUC16	40%	CUBN	71%
LRP1B	34%	SYNE1	32%	USH2A	38%	ZFHX4	71%
USH2A	31%	USH2A	31%	LRP1B	33%	CSMD3	57%
SYNE1	30%	LRP1B	31%	SYNE1	29%	LRP1B	57%
ZFHX4	28%	ZFHX4	30%	NAV3	28%	PCDH15	57%
KMT2D	23%	CDH10	23%	XIRP2	27%	SYNE1	57%

TMB and neoantigens in LUAD and LUSC

TMB is considered as a promising indicator to predict the effect of immune checkpoint inhibitors, and it is associated with the neoantigen burden [25]. Chen et al. demonstrated that high TMB and neoantigen burden was significantly correlated with improved immunotherapeutic effect in NSCLC [26]. Therefore, TMB and neoantigens were analyzed for all the samples and the different stages in LUAD and LUSC.

For all the LUAD samples, the most frequent variant was missense mutation, followed by nonsense mutation. Single nucleotide polymorphism (SNP) was responsible for such variants, and single nucleotide variants (SNVs) mostly occurred as C > A and C > T (Figure 2A). Similar results were found in samples of different tumor stages (data not shown). The top 10 mutated genes in all the LUAD samples were TP53, KRAS, XIRP2, ZFHX4, USH2A, LRP1B, CSMD3, RYR2, MUC16, and TTN (Figure 2B), and these are listed in Table 2A. Similarly, for all the LUSC samples and samples with different tumor stages, the most frequent variant was missense mutation, followed by nonsense mutation. SNP was responsible for such variants, and the SNV mostly occurred as C > A and C > T (Figure 3A). The top 10 mutated genes in all the LUSC samples were TP53, FAM135B, ZFHX4, USH2A, SYNE1, LRP1B, RYR2, CSMD3, MUC16, and TTN (Figure 3B), and these are listed in Table 2B. The top 10 overlapping mutated genes were TP53, ZFHX4, USH2A, LRP1B, CSMD3, RYR2, MUC16 and TTN.

The mutated genes KRAS and XIRP2 were identified only in LUAD, while FAM135B and SYNE1 were identified only in LUSC. Investigation of the neoantigens and neoantigen origin proteins revealed that there were no significant differences in the neoantigen burden in early vs. advanced stages in LUAD (Figure 2C-D). However, in LUSC, stage III had significantly higher neoantigen burden compared to stage I (Figure 3C-D).

Immune cells infiltration in LUAD and LUSC

The 22 different immune cell types among different tumor stages in LUAD and LUSC subtypes were analyzed using the CIBERSORT algorithm. The bar charts of immune cell subset proportions (Supplemental Figure S1) and immune cell subset heatmaps (Figure 4A) revealed that the M0 macrophages, M2 macrophages, and resting memory CD4 T cells accounted for a large proportion of immune cell infiltration in LUAD. Further, the differences in the proportion of TIICs in early (stages I and II) vs. advanced stages (stages III and IV) were also investigated. Early stage displayed greater number of memory B cells and M0 macrophages compared to the advanced stage in LUAD (Figure 4B). In addition, the associations between OS and TIIC abundance, and between tumor stage and TIIC abundance, were also investigated (Table 3). We found that the abundance of M0 macrophage infiltration was significantly correlated with both the tumor stage ($P = 0.016$) and OS ($P = 0.049$) (Supplemental Figure S2).

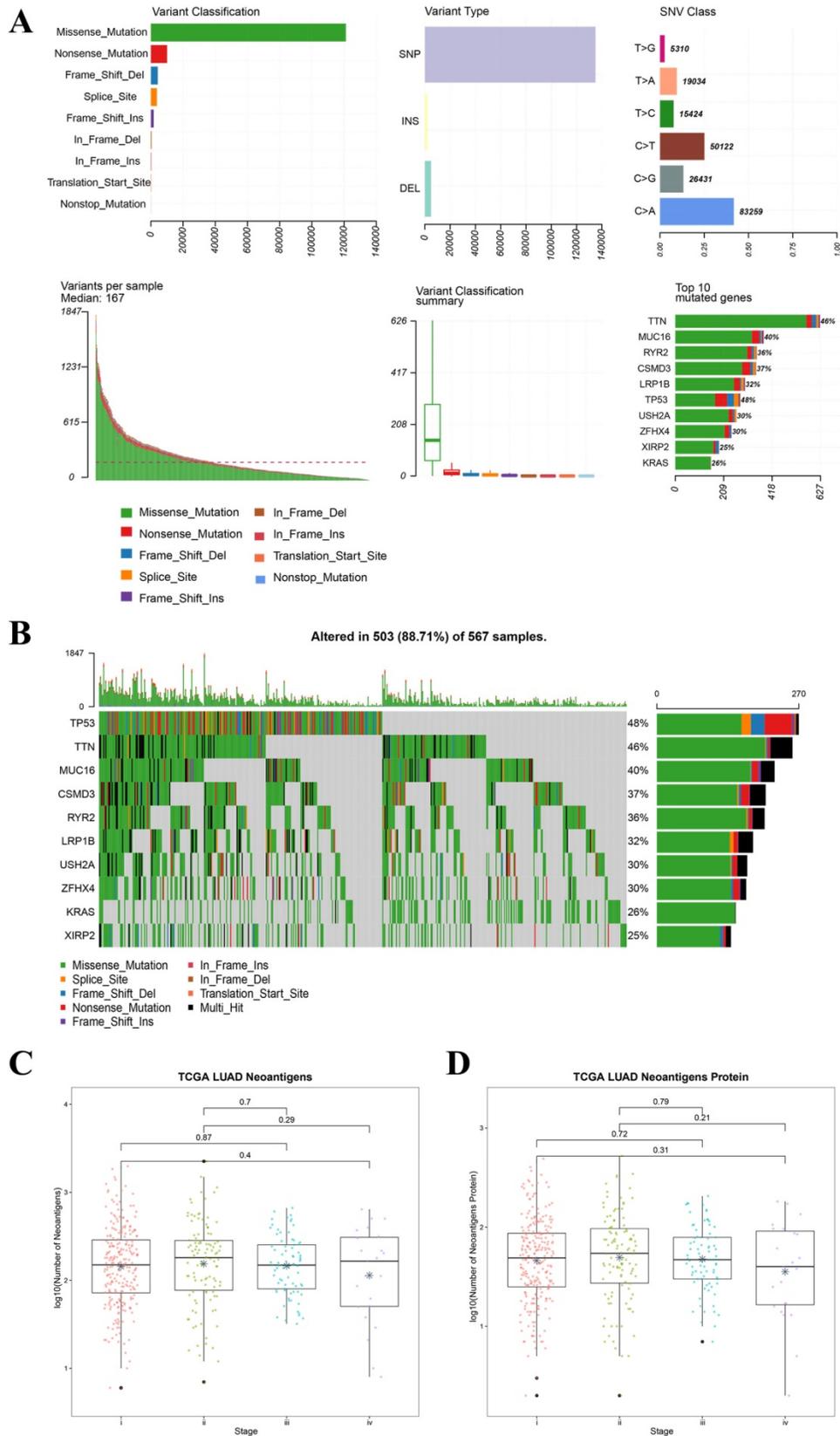


Figure 2. Landscape of TMB and neoantigens in LUAD. (A) The summary plot of TMB in LUAD. **(B)** The oncoplot of TMB in LUAD. **(C)** Distribution of neoantigen burden in LUAD. **(D)** Distribution of neoantigen origin protein in LUAD. Missense mutation account for most of the variant classification, followed by nonsense mutation; single nucleotide polymorphisin (SNP) is the major variant type; C>A is the major site of single nucleotide variation (SNV), followed by C>T; median of variants per sample is 167; List of top 10 mutated genes in summary plot shows the top 10 genes ordered by total number of variants in each gene, and the percentage following each gene represent the ratio of samples with this gene variation to total samples. Oncoplot shows the list of top 10 gene ordered by the the number of samples with the gene variants, and the percentage represent the ratio of samples with this gene variation to total samples. The horizontal line and asterisks in panel C and D represent median value and mean value, respectively. The number represents P value, and P < 0.05 shows significant statistical difference between two stages

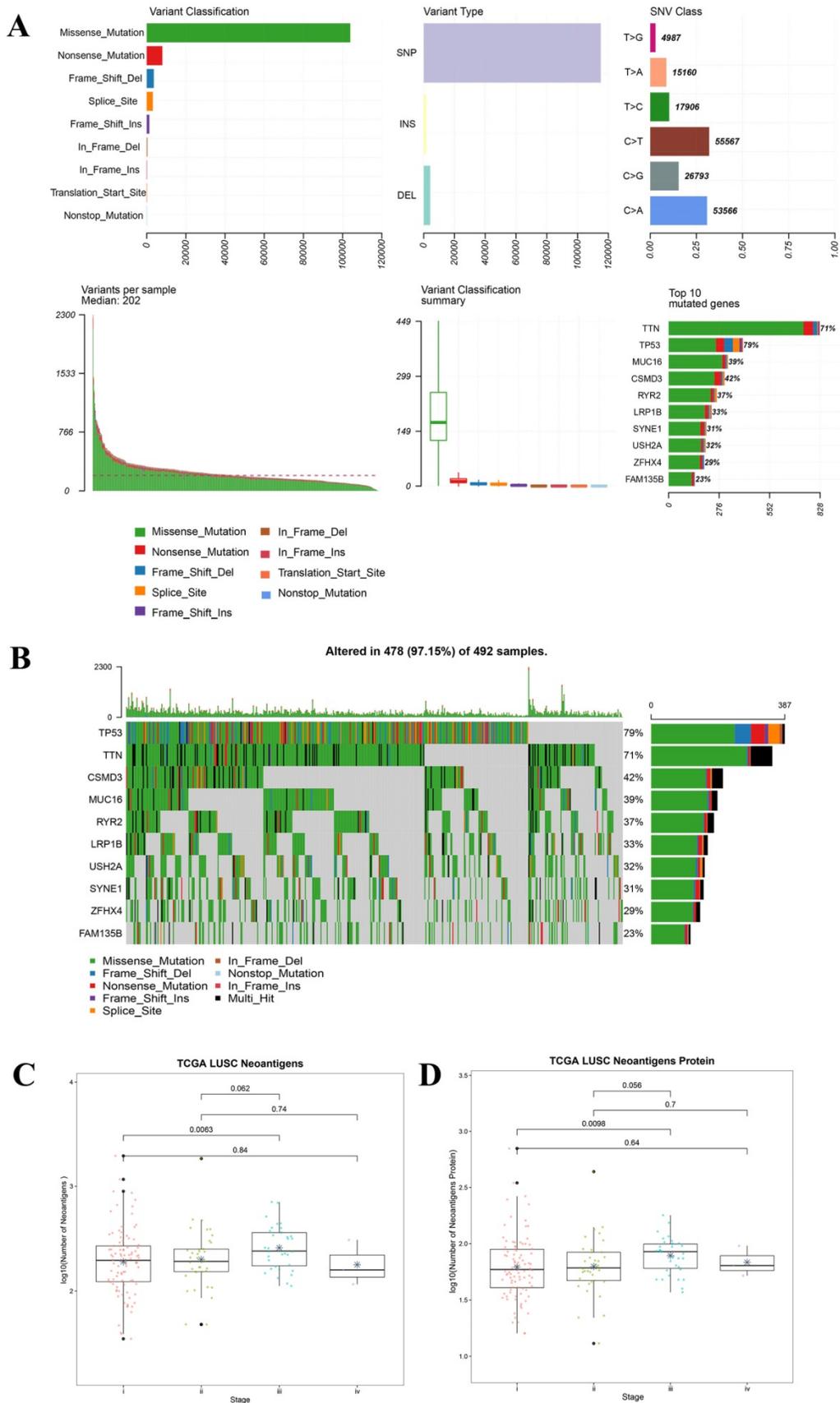


Figure 3. Landscape of TMB and neoantigens in LUSC. (A) The summary plot of TMB in LUSC. **(B)** The oncoplot of TMB in LUSC. **(C)** Distribution of neoantigen burden in LUSC. **(D)** Distribution of neoantigen origin protein in LUSC. Missense mutation account for most of the variant classification, followed by nonsense mutation; single nucleotide polymorphisin (SNP) is the major variant type; C>T is the major site of single nucleotide variation (SNV), followed by C>A; median of variants per sample is 202; List of top 10 mutated genes in summary plot shows the top 10 genes ordered by total number of variants in each gene, and the percentage following each gene represent the ratio

of samples with this gene variation to total samples. Oncoplot shows the list of top 10 gene ordered by the number of samples with the gene variants, and the percentage represent the ratio of samples with this gene variation to total samples. The horizontal line and asterisks in panel C and D represent median value and mean value, respectively. The horizontal line and asterisks in panel C and D represent median value and mean value, respectively. The number represents P value, and $P < 0.05$ shows significant statistical difference between two stages.

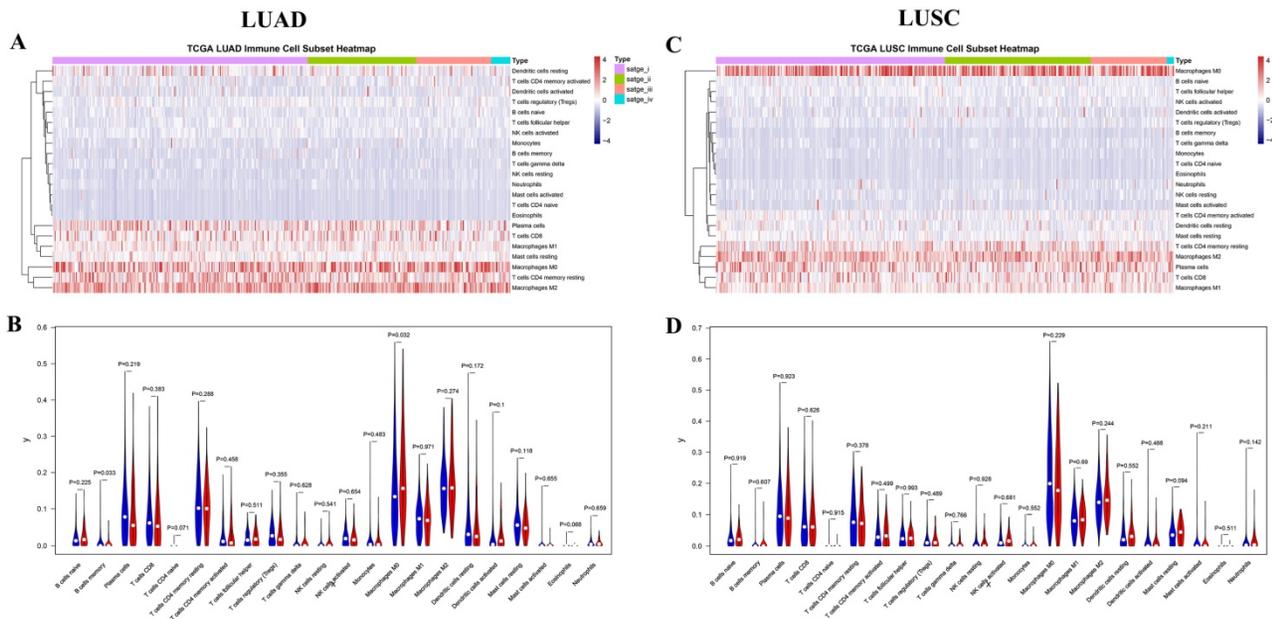


Figure 4. The landscape of immune infiltration in LUAD and LUSC. (A) Immune cell subset heatmap in LUAD. **(B)** Violin plot of early and advanced stage in LUAD. **(C)** Immune cell subset heatmap in LUSC. **(D)** Violin plot of early and advanced stage in LUSC. The dots in Violin plot represent the median value, and the blue color and red color represent early stage and advanced stage, respectively. $P < 0.05$ shows significant statistical difference between two stages.

Table 3. Analysis of the abundance of tumor infiltrating immune cells

	LUAD		LUSC	
	Immune cell types	P value	Immune cell types	P value
Stages	Macrophages M0	0.016	Mast cells resting	0.042
	/	/	Mast cells activated	0.032
Overall survival	Macrophages M0	0.049	Dendritic cells resting	0.024
	/	/	Eosinophils	0.017
	/	/	T cells CD4 memory activated	0.047
Immune cell clustering	Macrophages M0	/	Macrophages M0	/
	T cells CD4 memory resting	/	Macrophages M2	/
Stage I vs III	Macrophages M2	/	Plasma cells	/
	B cells memory	0.015	/	/
	Macrophages M0	0.04	/	/
	Mast cells resting	0.011	/	/
	Eosinophils	0.043	/	/
Stage I vs IV	Mast cells activated	0.015	B cells memory	0.001
Stage II vs III	/	/	Mast cells resting	0.01
	/	/	Mast cells activated	0.033
Stage II vs IV	/	/	B cells memory	0.006

Similarly, M0 macrophages, M2 macrophages, and plasma cells accounted for a large proportion of the immune cell infiltration in LUSC (Figure 4C, Supplemental Figure S3). However, no significant difference was observed in the proportion of TIICs in the early vs. advanced stages in LUSC (Figure 4D). Although, the abundance of resting ($P = 0.042$) and activated mast cells ($P = 0.032$) were found to be significantly correlated with the tumor stages. Four

immune cell types including resting dendritic cells, eosinophils, activated memory CD4 T cells, and mast cells were significantly correlated with OS (Table 3, Supplemental Figure S4).

Screening and functional enrichment of DEGs

A total of 495 DEGs were screened between stages I and IV in LUAD, including 232 upregulated and 263 downregulated genes. Volcano plots of the DEGs are shown in Figure 5A. These DEGs were significantly enriched in the various GO terms and KEGG pathways, such as regulation of lymphocyte activation, leukocyte proliferation, and hematopoietic cell lineage (Figure 5B-C). Then, the DEGs that correlated with the abundance of M0 macrophage infiltration were identified using Spearman correlation test. A total of 83 immune-related DEGs were identified in LUAD which were significantly enriched in mitosis-related biological processes, such as mitotic nuclear division, mitotic sister chromatid segregation, and nuclear division, and enriched in several KEGG pathways, such as hematopoietic cell lineage, cell cycle, and tight junction (Figure 6A-B).

A total of 621 DEGs (492 upregulated and 129 downregulated) were screened between stages I and IV in LUSC (Figure 7A). These DEGs were significantly enriched 8 molecular function terms (Figure 7B), including, G protein-coupled receptor binding. None of the KEGG pathways were enriched.

A total of 9 immune-related DEGs were identified in LUSC, and they were enriched in antibacterial-related biological processes and 3 KEGG pathways, including

protein digestion and absorption, glutamatergic synapse, and natural killer cell mediated cytotoxicity (Figure 7C-D).

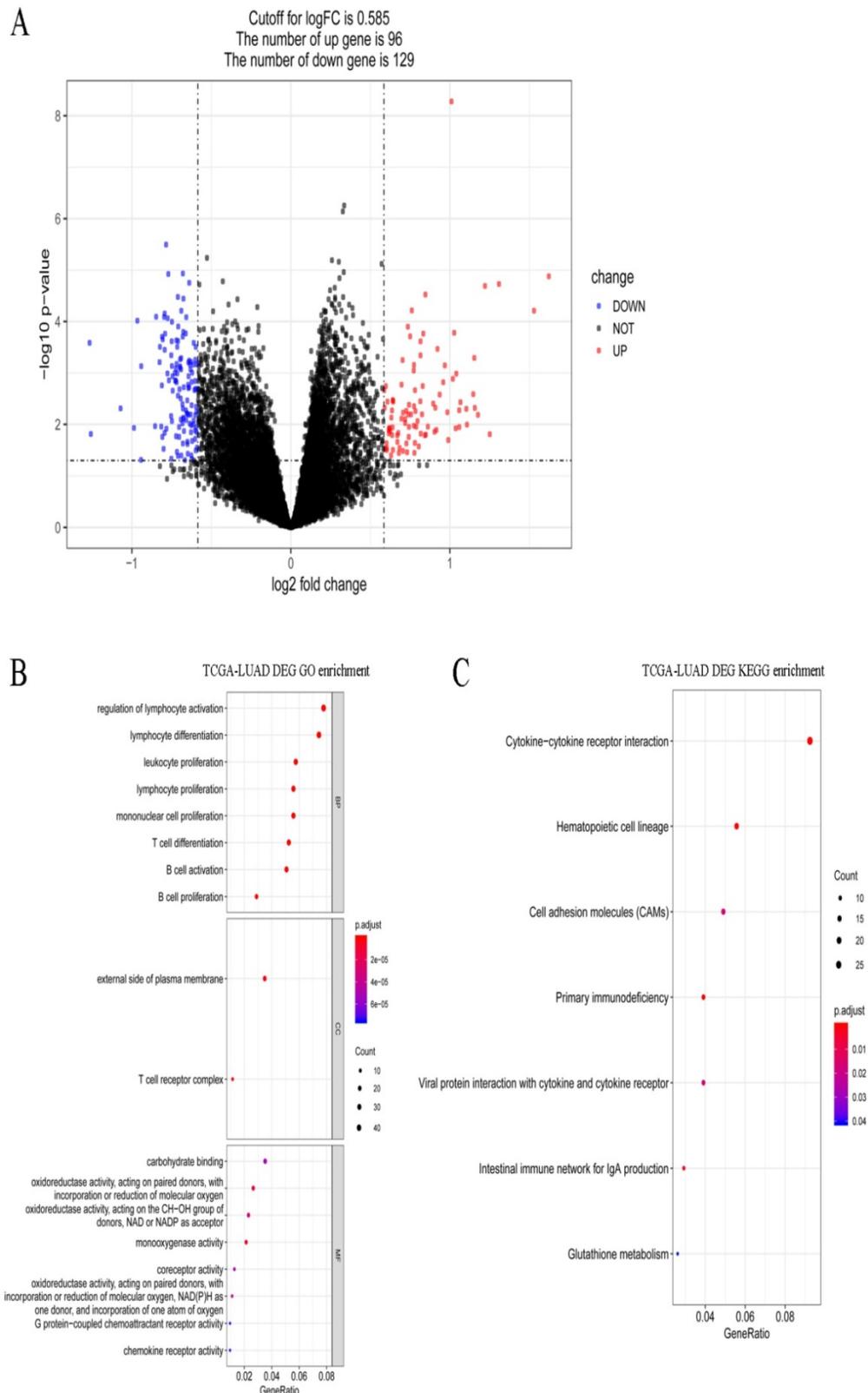


Figure 5. Gene expression spectrum and functional enrichment in LUAD. (A) The volcano plots of DEGs between stages I and IV in LUAD. Red dots and blue dots represent the significant up-regulated and down-regulated DEGs, respectively. **(B)** The significantly enriched GO terms for DEGs between stages I and IV in LUAD. **(C)** The significantly enriched KEGG pathways for DEGs between stages I and IV in LUAD.

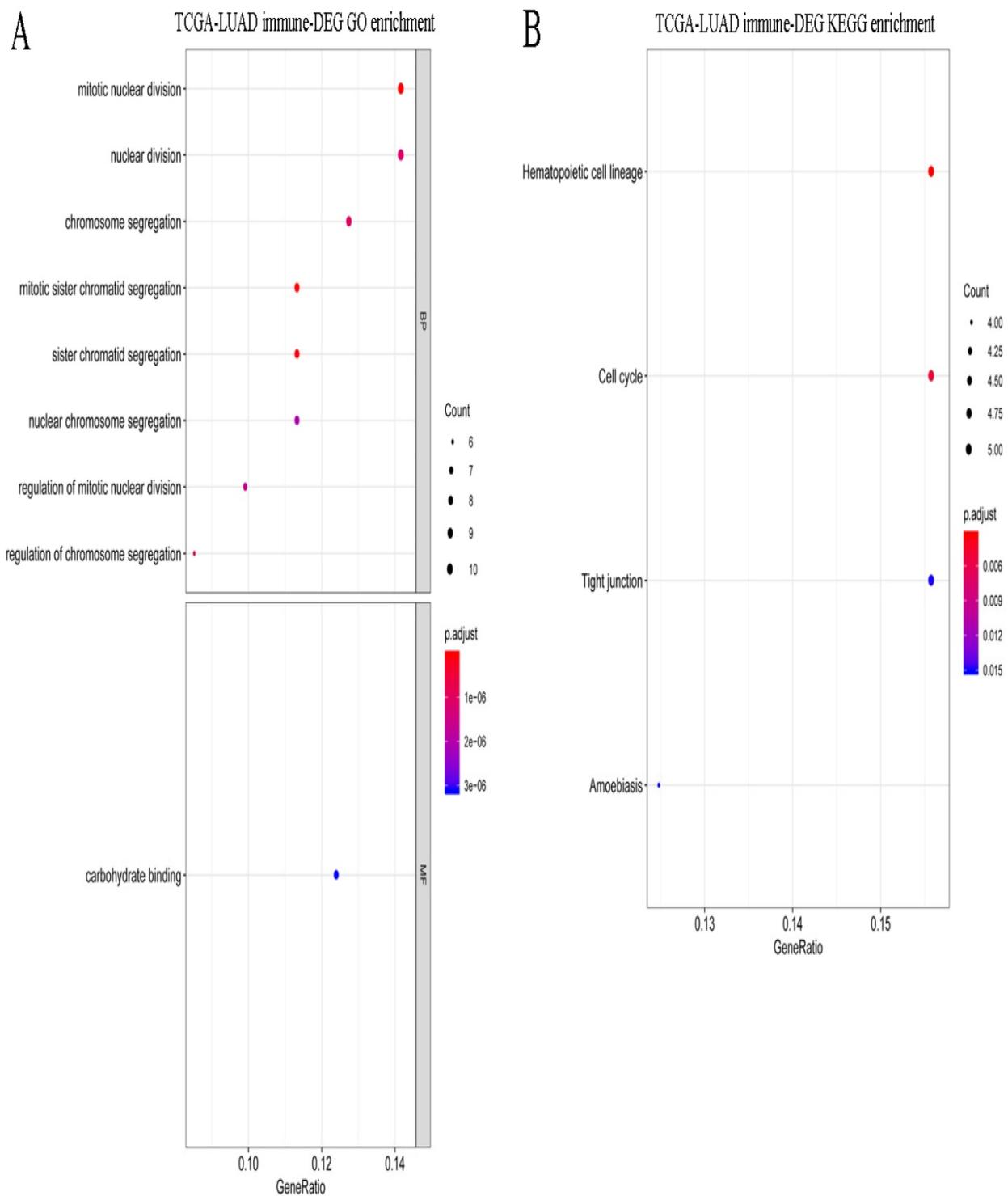


Figure 6. Functional enrichment for immune-related DEGs in LUAD. (A) The significantly enriched GO terms for immune-related DEGs in LUAD. **(B)** The significantly enriched KEGG pathways for immune-related DEGs in LUAD.

PPI network and functional modules

Proteins and their functional interactions form the backbone of the cellular machinery, and their connectivity networks will further the understanding of biological phenomena [27]. Therefore, PPI networks were constructed based on the interactions

in STRING database. In LUAD, the PPI network contained 233 genes, of which 35 were immune-related genes, and one module with score > 10 was identified from the PPI network (Figure 8A-B). The module contained one immune-related gene (GAL), one upregulated gene (GNG4), and 17 downregulated genes.

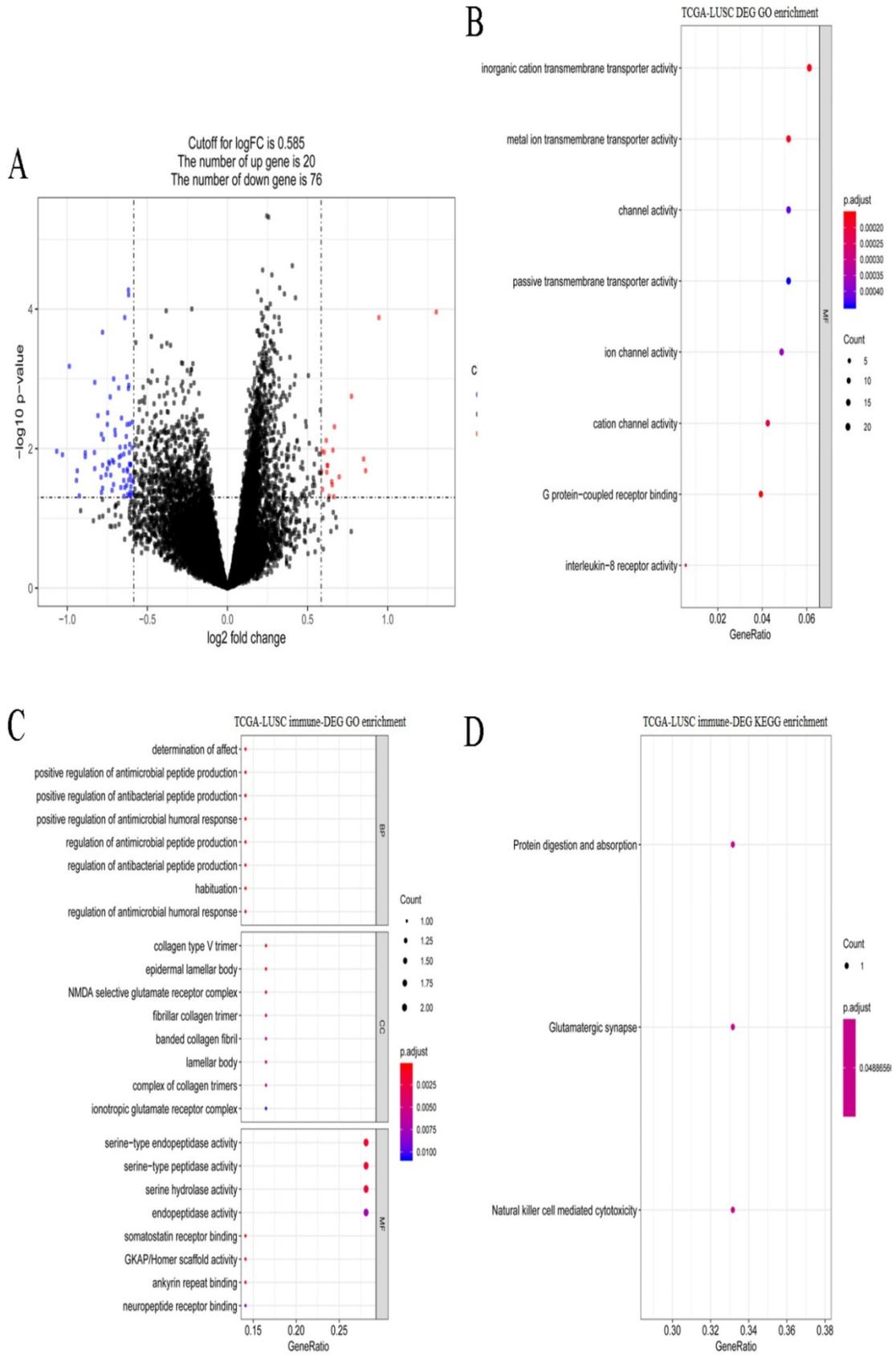


Figure 7. Gene expression spectrum and functional enrichment in LUSC. (A) The volcano plots of DEGs between stages I and IV in LUSC. Red dots and blue dots represent the significant up-regulated and down-regulated DEGs, respectively. **(B)** The significantly enriched GO terms for DEGs between stages I and IV in LUSC. **(C)** The significantly enriched GO terms for immune-related DEGs in LUSC. **(D)** The significantly enriched KEGG pathways for immune-related DEGs in LUSC.

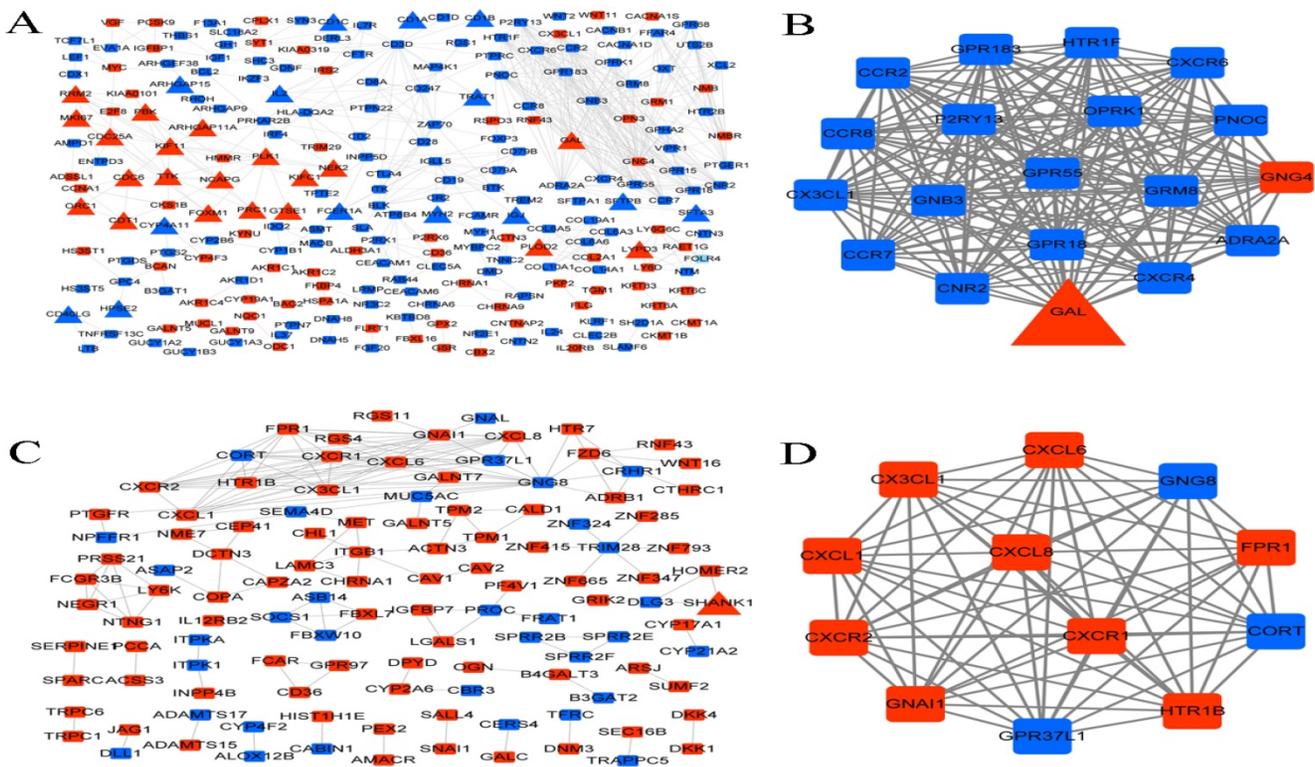


Figure 8. PPI network and functional modules. (A) PPI network of DEGs in LUAD. **(B)** The identified functional module with score > 10 in LUAD. **(C)** PPI network of DEGs in LUSC. **(D)** The identified functional module with score > 10 in LUSC. Square and triangle represent the DEGs and immune-related DEGs, respectively. Red nodes and blue nodes represent the significant up-regulated and down-regulated genes, respectively.

Table 4. Immune-related DEGs correlated with survival

LUAD		LUSC	
Gene name	P value	Gene name	P value
LYPD3	0.001	SHANK1	0.007
PLK1	0.002	COL5A3	0.009
RRM2	0.002		
PLOD2	0.003		
KIF11	0.004		
EXO1	0.006		
SPOCK1	0.006		
PIPRN	0.007		
GTSE1	0.008		
PRC1	0.009		
CDC6	0.010		
EGLN3	0.013		
FAM72D	0.014		
FOXM1	0.014		
CYP4A11	0.018		
FAM64A	0.018		
NEK2	0.021		
CDC25A	0.024		
TTK	0.028		
ARHGAP11A	0.029		
CA10	0.031		
ORC1	0.041		
NCAPG	0.046		

In LUSC, the PPI network contained 116 genes, of which only one was an immune-related gene (SHANK1), and one module with score > 10 was identified from the PPI network (Figure 8C-D). The module contained 12 genes, including 3

downregulated and 9 upregulated genes.

Survival analysis

Survival analysis was performed for the identified immune-related DEGs. Among the 83 immune-related DEGs in LUAD, 23 genes were found to have significant correlation with survival. Only two genes, including SHANK1 and COL5A3, were significantly correlated with survival among the 9 immune-related DEGs in LUSC (Table 4).

Discussion

In the current study, the gene expression and clinical phenotype data of LUAD and LUSC samples were used to analyze the immune microenvironment among different histological types and clinical stages. We found that there were some similarities and differences in the immune microenvironment of LUAD and LUSC. In the LUAD subtype, the early stage had a higher immune score than the advanced stage, but there was no significant difference in the immune cytolytic activity score or neoantigen burden. While in LUSC, the early stage was with a higher immune cytolytic activity score and neoantigen burden than the advanced stage, but there was no significant difference in the immune score. Roufas et

al. demonstrated that the levels of immune cytolytic activity varied greatly in different cancer types, and the cytolytic index along with complex associations among different TIICs was able to promote evasion from immune surveillance [28].

Our results indicated that TIICs and cell types were significantly correlated with progression and survival, of which the abundance of macrophage infiltration was associated with clinical stages and prognosis in LUAD, while the abundance of mast cells infiltration was associated with clinical stages in LUSC. Mast cells are well known for their roles in allergic disorders [29]. Recently, mast cells were found to play an important role in controlling inflammatory responses. They could not only produce a variety of inflammatory mediators, but also ameliorate inflammation by producing immune regulatory factors (IL-10) [30]. Mast cell infiltration has been implicated in metastasis and angiogenesis in several human malignancies [31]. Tryptase mast cell infiltration was found to greatly reduce the OS and disease-free survival in various solid tumors, and was significantly related to worse OS in NSCLC [32]. Mast cells infiltration and its number in the tumor microenvironment in lung cancer was also associated with poor prognosis [33]. Zhang et al. demonstrated that tumor infiltration by macrophages was associated with tumor lymph angiogenesis and unfavorable prognosis in LUAD [34]. Kawai et al. showed that macrophages infiltration could indicate prognosis of patients with stage IV NSCLC [35]. The results indicated that mast cell and macrophages infiltration in the tumor microenvironment was associated with prognosis. The densities of mast cells and macrophages should be evaluated by immunohistochemical staining in clinical samples, and their relationships with clinicopathological factors and prognosis should be further investigated.

TMB has been considered as a useful biomarker for response to immune therapy and prognosis in lung cancer [36]. In TMB analysis, the median of variants per sample showed a decrease from stage I to IV in LUAD, while it increased from stage I to IV in LUSC. Of the top 10 mutated genes KRAS and XIRP2 were only present in LUAD, while FAM135B and SYNE1 were only present in LUSC. However, in both LUAD and LUSC, the most frequent variant was missense mutation, followed by nonsense mutation. SNP was responsible for such variants, and the SNV mostly occurred as C > A and C > T. The main mutated genes were TP53, ZFH4, USH2A, LRP1B, CSMD3, RYR2, MUC16, and TTN. Owada-Ozaki et al. reported that high TMB in post-operation NSCLC patients might indicate unfavorable prognosis, and stage I NSCLC patients with high TMB showed

higher recurrence rate [37]. While Devarakonda et al. showed that high TMB could be a better prognosis for patients with resected NSCLC [38]. Such differences in the two studies might be associated with histologic subtypes. The distribution of TMB and the subsets of patients with high TMB had not been well described in most cancer types. Based on the analysis of human cancer genomes, Chalmers et al. found the median mutations for LUAD was 6.3 Mb and for LUSC was 9.0 Mb, and suggested that many cancer types have a substantial portion of patients with high TMB who might benefit from immunotherapy [39]. For the clinical application of TMB, validated assays and proper thresholds are indispensable. To make TMB function as a more reliable predictive biomarker, there were requirements for extra biologic guidance or computational assistance to predict associated-neoantigens [40]. It was suggested that highly mutated tumors were more likely to harbor neoantigens which make them targets of activated immune cells. Neoantigen is bound to the major histocompatibility complex (MHC) molecules and exists on the surface of tumor cells in the form of a protein complex. It can be specifically recognized by the cytotoxic T cell receptor (TCR) to activate the immune response of T cells. Studies had reported that tumor neoantigen plays a pivotal role in immune escape, anti-tumor immune response and immunotherapy [41-43].

Further, a total of 495 and 621 DEGs were identified between stages I and IV in LUAD and LUSC, respectively. The DEGs in LUAD were significantly enriched in T lymphocyte cell differentiation, etc. While the DEGs in LUSC were mainly enriched in metal ion transmembrane transporter activity, G protein-coupled receptor binding, etc. This suggested that the DEGs in LUAD and LUSC were different, and could be used to distinguish two subtypes. Of these, 83 immune-related DEGs were identified in LUAD of which 23 genes (including, PLK1 and RRM2) were found to be correlated with survival, while 9 immune-related DEGs were identified in LUSC and 2 (including, SHANK1) were found to be correlated with survival. PLK1, Polo Like Kinase 1, encodes a Ser/Thr protein kinase of the CDC5/Polo subfamily that is essential in mitotic progression [44]. Jolien et al. demonstrated that the level of PLK1 was increased in LUAD, and the combined evaluation of PLK1, carbonic anhydrase IX, and TP53 could predict prognosis of LUAD patients [45]. Noboru et al. showed that the expression of karyopherin beta 1 could be decreased by inhibiting PLK1, and such a decrease could inhibit cell proliferation via apoptosis in LUAD cells [46]. Ribonucleotide reductase (RR) plays a crucial role in

DNA replication and repair. It consists of RRM1 and RRM2 subunits, and its enzymatic activity is regulated mainly by RRM2 subunit [47]. RRM2 has been reported to have an active role in the progression of multiple cancers, including NSCLC [48]. Previous studies have reported that low expression of RRM1 and RRM2 could increase the response of NSCLC patients undergoing platinum-based chemotherapy, and could indicate a better outcome in NSCLC [49, 50]. RRM2, but not RRM1, could be a valuable biomarker to predict survival outcome of women, non-smokers, and former smokers who had stopped smoking for at least 10 years [51]. SHANK1, SH3 and multiple ankyrin repeat domains 1, a member of the SHANK family, are scaffold proteins that function in the establishment of dendritic spines [52]. SHANK1 is found to be expressed at high levels in lung cancer tissues compared to para cancerous tissues, however, its high expression had no significant correlation with gender, age, pathological grade or classification except with T stage [53]. In our study, PLK1 and RRM2 were identified as immune-related DEGs, and were found to be significantly associated with survival in LUAD. SHANK1 was the only immune-related DEG in the PPI network in LUSC, and it was significantly associated with survival in LUSC. Hence, we concluded that PLK1 and RRM2 may be useful biomarkers to predict survival outcome in LUAD, and SHANK1 was a crucial biomarker in LUSC. Further clinical study, both in vitro and in vivo experiments are necessary to validate the expressions of PLK1, RRM2, and SHANK1 and their roles in proliferation, metastasis and invasion. In addition, further clinical studies are required to determine whether those genes were independent prognosis biomarkers as well as its associations with immunotherapy efficacy.

Despite of the several novel findings, there still remained some limitations. (1) Multiple hypothesis test in statistical analyses should be taken into account. Moreover, TCGA clinical data for staging include different AJCC staging editions. (2) Tumor immune microenvironment characteristics, including immune score, immune cytolytic activity score, neoantigen burden, were analyzed. The clinical application in immunotherapy of these characteristics should be further investigated with large sample size. (3) Small cell lung cancer (SCLC) accounts for approximately 15% of lung cancer, however the study of immune microenvironment characteristics and involved key genes were still limited and largely unknown. Further studies in SCLC are still needed.

In conclusion, the characteristics of immune microenvironment and genes involved in the different stages of LUAD and LUSC were investigated. These

findings may provide theoretical basis in further studies and in accurate personalized immunotherapy for NSCLC patients.

Supplementary Material

Supplementary figures and tables.

<http://www.jcancer.org/v11p4965s1.pdf>

Acknowledgement

This study was supported by grants from the Hunan Provincial Key Research and Development Program for Social Development (2016SK2006), National Natural Science Foundation of Hunan Province (2019JJ50353), Natural Science Foundation of Hunan Province National Health Commission (B2019091), Natural Science Foundation of Changsha Science and technology Bureau (kp1901084), Wu Jieping Medical Foundation (320.6750.18368) and Cancer Foundation of China (NCC2018B58).

Author's Contributions

Lemeng Zhang and Jianhua Chen contribute to clinical data collection. Tianli Cheng, Hua Yang, Haitao Li and Changqie Pan contribute to data statistics analyze. Lemeng Zhang contributed to manuscript preparation. Lemeng Zhang and Jianhua Chen contribute to the study design, manuscript writing and data analyze. All authors read and approved the final manuscript.

Competing Interests

The authors have declared that no competing interest exists.

References

- Torre LA, Siegel RL, Jemal A. Lung cancer statistics. Lung cancer and personalized medicine: Springer. 2016. p. 1-19.
- Woodard GA, Jones KD, Jablons DM. Lung Cancer Staging and Prognosis. Cancer Treatment & Research. 2016; 170: 47.
- Beadsmoore CJ, Sreaton NJ. Classification, staging and prognosis of lung cancer. European Journal of Radiology. 2003; 45: 8-17.
- Ciborowski M, Kisluk J, Pietrowska K, et al. Development of LC-QTOF-MS method for human lung tissue fingerprinting. A preliminary application to non-small cell lung cancer. Electrophoresis. 2017; 38.
- Brahmer JR, Govindan R, Anders RA, et al. The Society for Immunotherapy of Cancer consensus statement on immunotherapy for the treatment of non-small cell lung cancer (NSCLC). Journal for Immunotherapy of Cancer. 2018; 6: 75.
- Remon J, Besse B, Soria JC. Successes and failures: what did we learn from recent first-line treatment immunotherapy trials in non-small cell lung cancer? Bmc Medicine. 2017; 15: 55.
- Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. Nature Reviews Cancer. 2012; 12: 252-64.
- Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. cell. 2011; 144: 646-74.
- Song M, Chen X, Wang L, et al. Future of anti-PD-1/PD-L1 applications: Combinations with other therapeutic regimens. Chinese Journal of Cancer Research. 2018; 30: 157-72.
- Wang L, Zhu B, Zhang M, et al. Roles of immune microenvironment heterogeneity in therapy-associated biomarkers in lung cancer. Seminars in Cell & Developmental Biology. 2016; 64: 90-7.
- Goc J, Germain C, Vo-Bourgais TKD, et al. Dendritic Cells in Tumor-Associated Tertiary Lymphoid Structures Signal a Th1 Cytotoxic Immune Contexture and License the Positive Prognostic Value of Infiltrating CD8⁺ T Cells. Cancer Research. 2014; 74: 705.

12. Wood SL, Pernemalm M, Crosbie PA, et al. The role of the tumor-microenvironment in lung cancer-metastasis and its relationship to potential therapeutic targets. *Cancer Treatment Reviews*. 2014; 40: 558-66.
13. Clancy T, Dannenfels R, Troyanskaya O, et al. Bioinformatics approaches to profile the tumor microenvironment for immunotherapeutic discovery. *Curr Pharm Des*. 2017; 23.
14. Liu CC, Steen CB, Newman AM. Computational approaches for characterizing the tumor immune microenvironment. *Immunology*. 2019; 158: 70-84.
15. Harrow J, Frankish A, Gonzalez JM, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome research*. 2012; 22: 1760-74.
16. Yoshihara K, Shahmoradgoli M, Martínez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature communications*. 2013; 4: 2612.
17. Rooney MS, Shukla SA, Wu CJ, et al. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell*. 2015; 160: 48-61.
18. Mayakonda A, Lin D-C, Assenov Y, et al. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome research*. 2018; 28: 1747-56.
19. Kassambara A. Ggpubr: 'Ggplot2' Based Publication Ready Plots, R Package Version 0.2. 2018. 2019.
20. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nature methods*. 2015; 12: 453.
21. Smyth GK, Ritchie M, Thorne N, et al. LIMMA: linear models for microarray data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Statistics for Biology and Health. 2005.
22. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26: 139-40.
23. Yu G, Wang L-G, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology*. 2012; 16: 284-7.
24. Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC bioinformatics*. 2003; 4: 2.
25. Owada-Ozaki Y, Muto S, Takagi H, et al. Prognostic impact of tumor mutation burden in patients with completely resected non-small cell lung cancer: brief report. *Journal of Thoracic Oncology*. 2018; 13: 1217-21.
26. Chen H, Chong W, Teng C, et al. The immune response-related mutational signatures and driver genes in non-small-cell lung cancer. *Cancer Science*. 2019; 110: 2348.
27. Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2018; 47: D607-D13.
28. Roufas C, Chasiotis D, Makris A, et al. The Expression and Prognostic Impact of Immune Cytolytic Activity-Related Markers in Human Malignancies: A Comprehensive Meta-analysis. *Frontiers in oncology*. 2018; 8: 27-.
29. Peng X, Wang J, Li X, et al. Targeting Mast Cells and Basophils with Anti-FcεR1α Fab-Conjugated Celastrol-Loaded Micelles Suppresses Allergic Inflammation. *Journal of Biomedical Nanotechnology*. 2015; 11: 2286.
30. Milling S. Adipokines and the control of mast cell functions: from obesity to inflammation? *Immunology*. 2019; 158: 1-2.
31. Ribatti D. Mast cells as therapeutic target in cancer. *European Journal of Pharmacology*. 2016; 778: 152-7.
32. Hu G, Wang S, Cheng P. Tumor-infiltrating tryptase+ mast cells predict unfavorable clinical outcome in solid tumors. *International journal of cancer*. 2018; 142: 813-21.
33. Shemesh R, Gorzalczany Y, Peled N, et al. 10P The micro-environmental cross talk between mast cells and lung cancer cells through cell-to-cell contact. *Journal of Thoracic Oncology*. 2018; 13: S5-S6.
34. Cheng ZB, Juan G, Jun W, et al. Tumor-associated macrophages infiltration is associated with peritumoral lymphangiogenesis and poor prognosis in lung adenocarcinoma. *Medical Oncology*. 2011; 28: 1447-52.
35. Osamu K, Genichiro I, Kaoru K, et al. Predominant infiltration of macrophages and CD8(+) T Cells in cancer nests is a significant predictor of survival in stage IV nonsmall cell lung cancer. *Cancer*. 2010; 113: 1387-95.
36. Wang X, Kong C, Xu W, et al. Decoding tumor mutation burden and driver mutations in early stage lung adenocarcinoma using CT-based radiomics signature. *Thoracic cancer*. 2019.
37. Owada-Ozaki Y, Muto S, Takagi H, et al. Prognostic Impact of Tumor Mutation Burden in Patients with Completely Resected Non-Small Cell Lung Cancer: Brief Report. *Journal of Thoracic Oncology*. 2018; 13: 1217-21.
38. Devarakonda S, Rotolo F, Tsao MS, et al. Tumor Mutation Burden as a Biomarker in Resected Non-Small-Cell Lung Cancer. *Journal of Clinical Oncology*. 2018; 36: JCO2018781963.
39. Chalmers ZR, Connelly CF, Fabrizio D, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome medicine*. 2017; 9: 34.
40. Goto Y. Tumor mutation burden: is it ready for the clinic? : American Society of Clinical Oncology. 2018.
41. Jiang T, Shi T, Zhang H, et al. Tumor neoantigens: from basic research to clinical applications. *Journal of hematology & oncology*. 2019; 12: 93.
42. Zhou C, Zhu C, Liu Q. Toward in silico Identification of Tumor Neoantigens in Immunotherapy. *Trends in molecular medicine*. 2019; 25: 980-92.
43. Laumont CM, Vincent K, Hesnard L, et al. Noncoding regions are the main source of targetable tumor-specific antigens. *Science translational medicine*. 2018; 10.
44. Yao D, Gu P, Wang Y, et al. Inhibiting Polo-like Kinase 1 (PLK1) enhances radiosensitization via modulating DNA repair proteins in non-small cell lung cancer. *Biochemistry & Cell Biology-biochimie Et Biologie Cellulaire*. 2017; 96.
45. Van den Bossche J, Deben C, Op de Beeck K, et al. Towards Prognostic Profiling of Non-Small Cell Lung Cancer: New Perspectives on the Relevance of Polo-Like Kinase 1 Expression, the TP53 Mutation Status and Hypoxia. *Journal of Cancer*. 2017; 8: 1441-52.
46. Sekimoto N, Suzuki Y, Sugano S. Decreased KPNB1 Expression is Induced by PLK1 Inhibition and Leads to Apoptosis in Lung Adenocarcinoma. *Journal of Cancer*. 2017; 8: 4125-40.
47. Vincenzo DA, Valerio D, Forrester FM, et al. Cyclin F-mediated degradation of ribonucleotide reductase M2 controls genome integrity and DNA repair. *Cell*. 2012; 149: 1023-34.
48. Wang L, Meng L, Wang X-w, et al. Expression of RRM1 and RRM2 as a novel prognostic marker in advanced non-small cell lung cancer receiving chemotherapy. *Tumor Biology*. 2014; 35: 1899-906.
49. Zhao H, Zhang H, Du Y, et al. Prognostic significance of BRCA1, ERCC1, RRM1, and RRM2 in patients with advanced non-small cell lung cancer receiving chemotherapy. *Tumor Biology*. 2014; 35: 12679-88.
50. Wang L, Meng L, Wang XW, et al. Expression of RRM1 and RRM2 as a novel prognostic marker in advanced non-small cell lung cancer receiving chemotherapy. *Tumor Biology*. 2014; 35: 1899-906.
51. Mah V, Alavi M, Márquez-Garbán DC, et al. Ribonucleotide reductase subunit m2 predicts survival in subgroups of patients with non-small cell lung carcinoma: effects of gender and smoking status. *Plos One*. 2015; 10: e0127600.
52. Falley K, Schütt J, Iglauer P, et al. Shank1 mRNA: dendritic transport by kinesin and translational control by the 5' untranslated region. *Traffic*. 2010; 10: 844-57.
53. ChenBo, WeiYunyan, Wujianqing. Expression and clinical significance of SHANK1 in lung cancer. *Int J Respir*. 2017; 37: 1364-8.