

Identification of Four Genes as Prognosis Signatures in Lung Adenocarcinoma Microenvironment

This article was published in the following Dove Press journal:
Pharmacogenomics and Personalized Medicine

Yan Yao¹
Tingting Zhang²
Lingyu Qi²
Ruijuan Liu³
Gongxi Liu³
Jie Li²
Changgang Sun^{3,4}

¹Clinical Medical Colleges, Weifang Medical University, Weifang, Shandong Province, People's Republic of China;

²College of First Clinical Medicine, Shandong University of Traditional Chinese Medicine, Jinan, Shandong Province, People's Republic of China;

³Department of Oncology, Weifang Traditional Chinese Hospital, Weifang, Shandong Province, People's Republic of China; ⁴Innovative Institute of Chinese Medicine and Pharmacy, Shandong University of Traditional Chinese Medicine, Jinan, Shandong, People's Republic of China

Tel +86 13583690699
Email zhongliuyike@163.com

Background: Tumor microenvironment (TME) cells constitute a vital element of tumor tissues. Increasing evidence has shown that immune response in the microenvironment plays an active role in tumor invasion, metastasis, and recurrence, and is an important factor affecting tumor prognosis. Our study aimed to identify the gene signatures in lung adenocarcinoma (LUAD) microenvironment for prognosis and immunotherapy.

Methods: In this study, we evaluated, for the first time, the stromal and immune scores of 594 patients from The Cancer Genome Atlas (TCGA) database with LUAD using the ESTIMATE algorithm. Three hundred and sixty-seven dysregulated immune-related genes were identified. Then, we performed functional enrichment analysis of these genes, and found the best gene model and construct the signature through univariate, Lasso and multivariate COX regression analysis. To assess the independently prognostic ability of the signature, the Kaplan–Meier survival analysis and Cox's proportional hazards model were performed.

Results: Functional enrichment analysis and protein–protein interaction networks showed that the immune-related genes mainly played a role in immune response, activation/proliferation of immune-related cells, and chemokine activity. A prognostic model involving 6 genes was constructed and the signature was identified as an independent prognostic factor and significantly associated with the overall survival (OS) of LUAD. The area under curve (AUC) of the receiver operating characteristic curve (ROC curve) for the 6 genes signature in predicting the 3-year survival rate was 0.708. Finally, four genes (FOXN4, KLHL4, FAM83F and CCR2) can be used as candidate prognostic biomarkers for LUAD.

Conclusion: Our findings will help evaluate the prognosis of LUAD and provide new ideas for exploring the potential relationship between TME and LUAD treatment and prognosis.

Keywords: lung adenocarcinoma, tumor microenvironment, ESTIMATE algorithm, immune-related gene, prognosis signatures

Introduction

The tumor microenvironment (TME) is an integral part of the tumor, and is necessary in establishing a dynamic interaction with tumor cells. This interaction and dynamic communication acceleration can in turn affect the progress, metastasis, and recurrence of the tumor. TME is the cellular environment of tumor cells, and consists of extracellular matrix, soluble molecules, and tumor stroma cells.¹ In the TME, immune and stromal cells constitute the two major types of non-tumor components and have been proposed to be important in the diagnostic and prognostic assessment of the tumor.

Correspondence: Changgang Sun
Innovative Institute of Chinese Medicine and Pharmacy, Shandong University of Traditional Chinese Medicine, Jinan, Shandong, People's Republic of China
Tel +86 13583690699
Email zhongliuyike@163.com

Lung cancer is the leading cause of cancer-related death, whereas lung adenocarcinoma (LUAD), one of the main types of lung cancers, has an average 5-year survival rate of only 18%.^{2,3} Although molecular-targeted therapy has greatly enhanced the survival rate of patients, no targeted mutations have been identified in a large number of advanced LUAD patients. For these patients, the use of antibodies against immune checkpoints have demonstrated treatment activity and safety. A comprehensive immuno-genomic analysis of the TME based on the interaction between programmed death 1 (PD-1)/programmed cell death-ligand 1 (PD-L1) and tumor-infiltrating lymphocytes (TILs) is critical in deepening our understanding of the underlying mechanisms and in guiding us in tailoring optimal immunotherapeutic strategies.^{4,5} Therefore, it is particularly urgent to explore genetic markers of the LUAD microenvironment for prognosis and immunotherapy.

The treatment of tumors has evolved to become very accurate due to clinical research and evidence-based medicine. The advances in molecular biology research and the development of next-generation sequencing (NGS) technology have promoted our understanding of cell and gene changes in the process of tumorigenesis and development, and have yielded more targeted and individualized treatment choices. Researchers at home and abroad have used genome analysis as the main method to find new biological targets for LUAD. A growing body of literature suggests the crucial role of TME in cancer progression and therapeutic response,⁶ and the development of NGS provides an opportunity to systematically explore TME. Large collaborative projects, such as The Cancer Genome Atlas (TCGA) and International Cancer Genome Consortium (ICGC) have sequenced thousands of cancer genomes, thus improving our understanding of the molecular changes in the development of cancer.^{7,8} However, owing to small sample sizes and the lack of multi-omics data, few studies have been conducted on the genomic analysis of LUAD from the perspective of immunology. Since the RNA sequencing of tumor tissue usually also includes the sequencing of microenvironment cells, researchers have developed some expression profile-based estimations of the abundance of microenvironment cells in tumor tissue.^{9–11}

To date, only one method, referred to as “Estimation of Stromal and Immune cells in Malignant Tumors,” that uses Expression data (ESTIMATE) has been described and can be used to score the stromal and immune fraction in transcriptomic data of cancer tissues.^{12,13} ESTIMATE

algorithm is based on single sample Gene Set Enrichment Analysis (ssGSEA) and generates three scores: stromal score, immune score, and estimate score. This algorithm has been successfully applied to prostate cancer,¹⁴ breast cancer¹⁵ and colon cancer.¹⁶ However, the relationship between the proportion of immune and stromal fraction and the survival of LUAD patients has not been thoroughly investigated. Therefore, it is necessary to establish an immune marker that can represent the immune status of TME and predict the prognosis of LUAD based on a series of immune-related genes.

Our study is based on the large-scale mRNA-seq data of LUAD tumor samples obtained from the TCGA database, and for the first time, the fraction of stromal and immune cells in tumor tissues was scored using the ESTIMATE algorithm. The association between the ESTIMATE scores and patient survival data was assessed, and a list of microenvironment-associated genes that can predict poor outcomes in LUAD patients was extracted. Notably, we used the Cox model based on Lasso penalty and independent prognostic analysis to explore the core pathways and biomarkers involved in cancer progression.

Materials and Methods

Lung Adenocarcinoma Datasets and Pre-Processing

All the original data of lung adenocarcinoma were obtained from the TCGA database. RNA-sequencing data (FPKM values) were transformed into transcripts per kilobase million (TPM) values, since they more accurately resemble those resulting from microarrays and are more comparable between samples.¹⁷ The updated clinical data regarding gender, age, histological type, survival and outcome information of TCGA-LUAD samples were downloaded from the TCGA data using R package TCGAbiolinks.¹⁸ Genes were annotated using the Ensembl online database.¹⁹ This study is in accordance with the publication guidelines provided by TCGA, and did not require the approval of the ethics committee.

TME Analysis of Lung Adenocarcinoma

LUAD microenvironmental gene scores have specific genomic and transcriptomic spectrums within the LUAD samples. ESTIMATE algorithm exploits the unique properties of the transcriptional profiles of cancer samples in order to infer the abundance of immune and stroma cells,¹³ and is comprised of the following steps: 1) stromal and

immune scores were calculated using the ESTIMATE package in R (version 2.15.3). ESTIMATE outputs stromal, immune and ESTIMATE scores based on ssGSEA. 2) We integrated LUAD clinical data to explore the relationship between immune score, matrix score and clinical stage. 3) Using the survival package in R, we discussed the relationship between immune score, matrix score and survival. The ESTIMATE scores for each dataset were calculated and patients were divided into two equal groups of high or low ESTIMATE scores using median split. $P < 0.05$ was considered statistically significant.

Analysis of Differentially Expressed Genes (DEGs) Associated with the Immune and Stromal Scores in LUAD

Based on the immune and stromal scores, all samples were divided into either high- or low-content groups. DEGs among these two groups were determined using the R package limma,^{18,41} which implements an empirical

Bayesian approach to estimate gene expression changes using the moderated Wilcoxon test, and visualized as heat-maps using R software. Absolute log2-fold change > 1 and $\text{adj. } p < 0.05$ were set as the screening criterion. In addition, we used the Venn Diagram package of R to explore the TME-related genes that were dysregulated in both high- and low-score groups, and the results are presented as Venn Diagram.^{19,42}

Functional Enrichment Analysis

Gene-annotation enrichment analysis using the clusterProfiler R package was performed on TME signature genes,^{20,43} including the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) analysis. The results of GO analysis mainly consisted of three parts: biological process (BP), cellular component (CC) and molecular function (MF). GO and KEGG terms with $p < 0.05$ were considered statistically significant and we used the “Goplot” package in R to visualize the results.^{21,44}

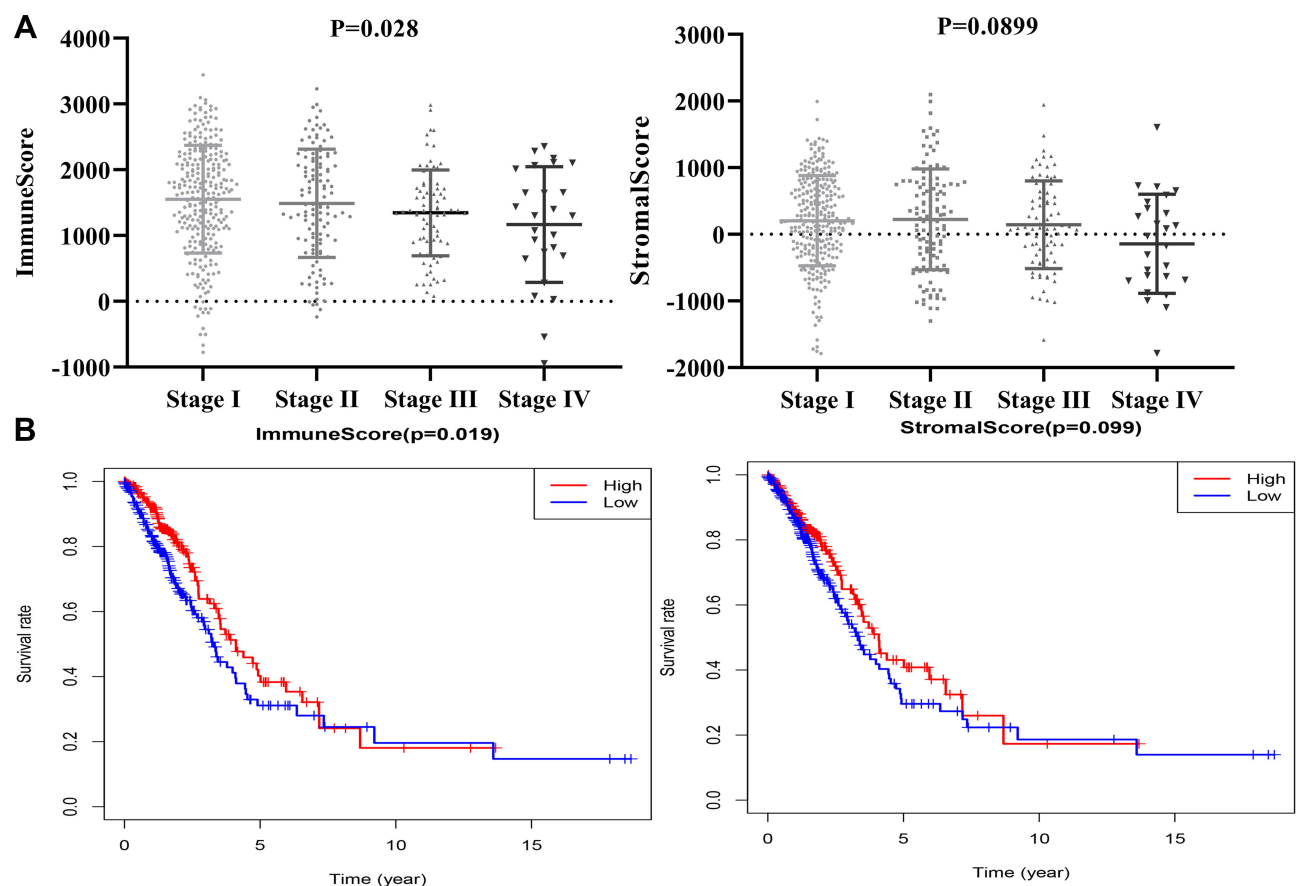


Figure 1 Immune scores and stromal scores are associated with LUAD stages and overall survival. (A) Distribution of immune scores and stromal scores in different stages. (B) Relationship between immune score, matrix score and survival of LUAD.

Abbreviation: LUAD, lung adenocarcinoma.

Protein–Protein Interaction (PPI) Network Construction and Topological Analysis

The PPI network was retrieved from the online STRING database^{22,45} and visualized using the Cytoscape software version 3.6.1.^{19,46} The cut-off criterion was a confidence score ≥ 0.90 . The degree of connectivity of each node of the network was calculated. The APP MCODE was used to identify the module with the highest connectivity and the hub genes from the PPI network were screened.

Dimension Reduction and Generation of TME Gene Signatures

The construction of DEGs associated with the TME meta-genes was performed as follows: 1) Integration of the survival time and status of LUAD clinical data with the expression level of DEGs. 2) Univariate Cox proportional hazard regression analysis was performed on DEGs using the survival package of R to obtain the prognostic value, and the genes with $p < 0.01$ were selected for further analysis. 3) To achieve variable selection and dimension reduction, we put the expression data of DEGs with $p < 0.01$ into the LASSO Cox

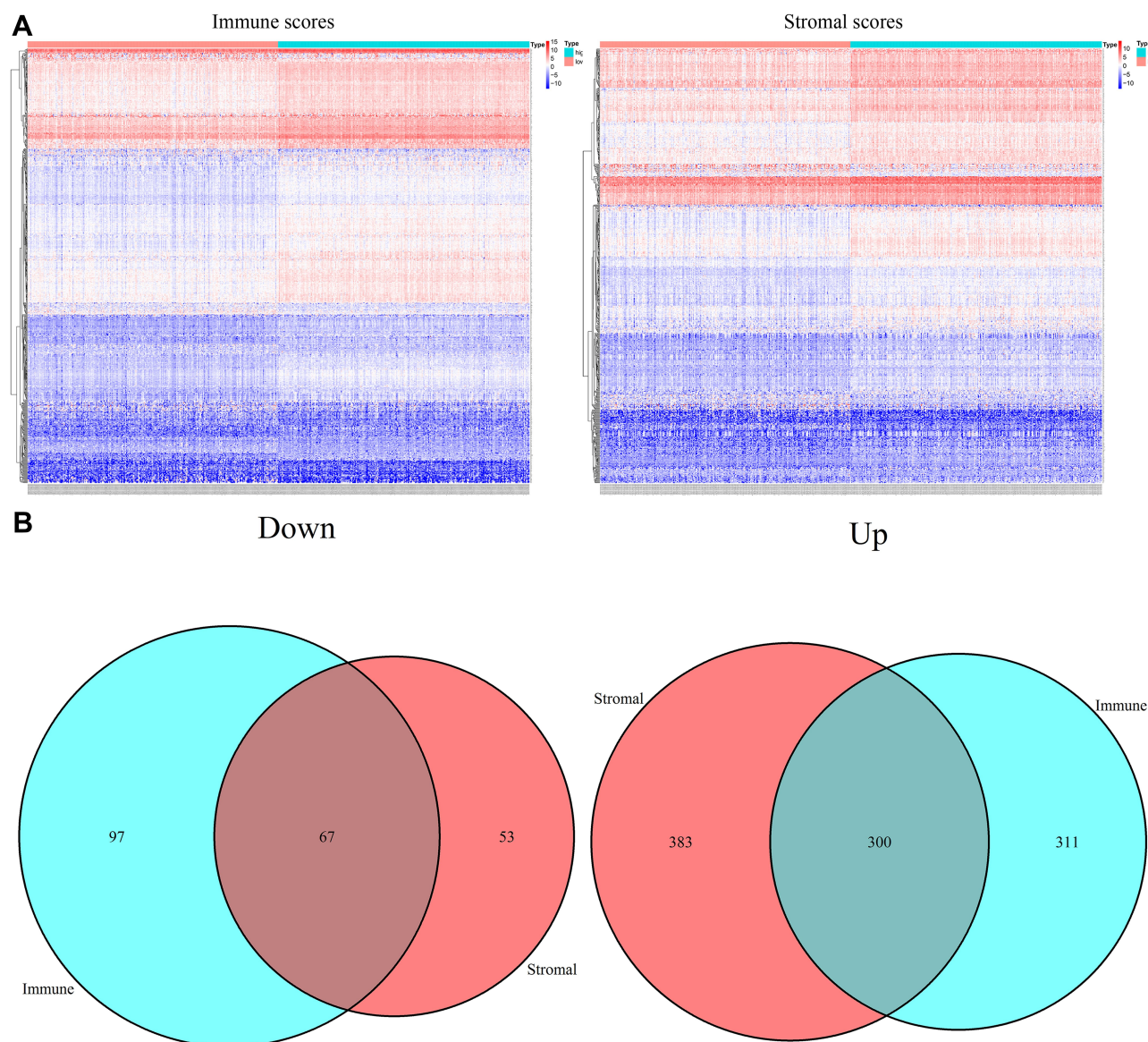


Figure 2 Comparison of gene expression profile with immune scores and stromal scores in LUAD. **(A)** Clustered heatmaps of the DEGs in LUAD. The rows represent genes and columns represent samples. $|\log_2FC| > 1$, $FDR < 0.05$. **(B)** Venn diagrams showing the number of commonly upregulated and downregulated DEGs in stromal and immune score groups.

regression. During this process, the penalty regularization parameter λ , was chosen using the cross-validation routine with a random 2000 and the R package “glmnet.” The value of lambda.min, the lambda value that yields the minimum mean cross-validated error, was calculated using R.²⁰ 4) The results of Lasso algorithm were used to perform a multivariate Cox proportional hazards regression analysis, and a prognostic model was established.

Prognostic Model Validation and Independent Prognostic Analysis

Each patient was given a risk score according to the risk model gene, and the corresponding median risk score was used as the boundary value, based on which the patients were divided into high-risk group and low-risk group. The survival differences between the two groups were calculated using the Kaplan–Meier and Log rank test. Moreover, using the survival ROC package in R, we constructed a 3-year time-dependent receiver operating characteristic (ROC) curve and estimated the area under the ROC curve (AUC) for the prognostic models. Finally, we combined clinical data with the scores in order to perform univariate and multivariate independent prognostic analysis.

Results

Landscape of LUAD TME-Associated Immune Scores and Stromal Scores

A total of 594 transcriptome specimens (535 cancer samples, 59 para-cancerous samples) were obtained from TCGA

database. Based on the ESTIMATE algorithm, we calculated the immune and stromal scores of all LUAD samples, and found that stromal scores ranged from -1789.62 to 2097.96 while immune score were between -942.51 and 3442.08 . Then, we analyzed the relationship between the scores and clinical stage. As shown in Figure 1A, with the progress to more advanced stages, the immune score and matrix score gradually decreased, and the correlation between immune scores and clinical stage was statistically insignificant ($p=0.028$). In order to explore the potential correlation between overall survival rate and immune score and/or matrix score, all patients were divided into two groups based on the median score. The results showed that the survival rate of the high score group was higher than that of low-score group (Figure 1B). The difference between immunization group score was statistically significant ($p=0.019$). It was thus revealed that immune score is a potential indicator of LUAD prognosis.

Verification of TME-Related DEGs Signature in LUAD

In order to reveal the correlation between global gene expression profiles with immune scores and/or stromal scores, and to search for genes related to LUAD microenvironment, we investigated the differences in high and low immune/matrix scores between different LUAD tumor samples. The results showed that the samples had strong heterogeneity in the immune scores group (Figure 2A). Venn diagrams showed that there were 300 upregulated genes and 67 downregulated

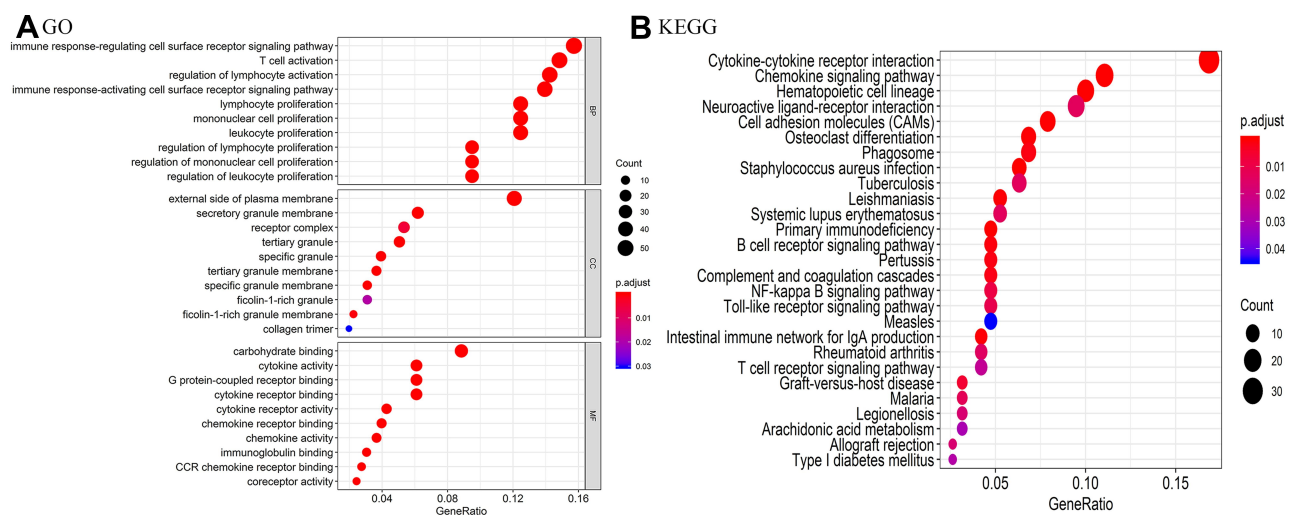


Figure 3 GO functional enrichment and KEGG pathway enrichment of DEGs in the stromal and immune score groups. **(A)** Top 10 biological process (BP), cellular components (CC) and molecular functions (MF) with the most significant P values. **(B)** All KEGG enrichment results of DEGs. $P<0.05$.

Abbreviations: GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

genes in the stromal scores group and immune scores group (Figure 2B). It is worth noting that these DEGs co-existed in immune and matrix scores group, and hence, we extracted the 367 DEGs for further analysis.

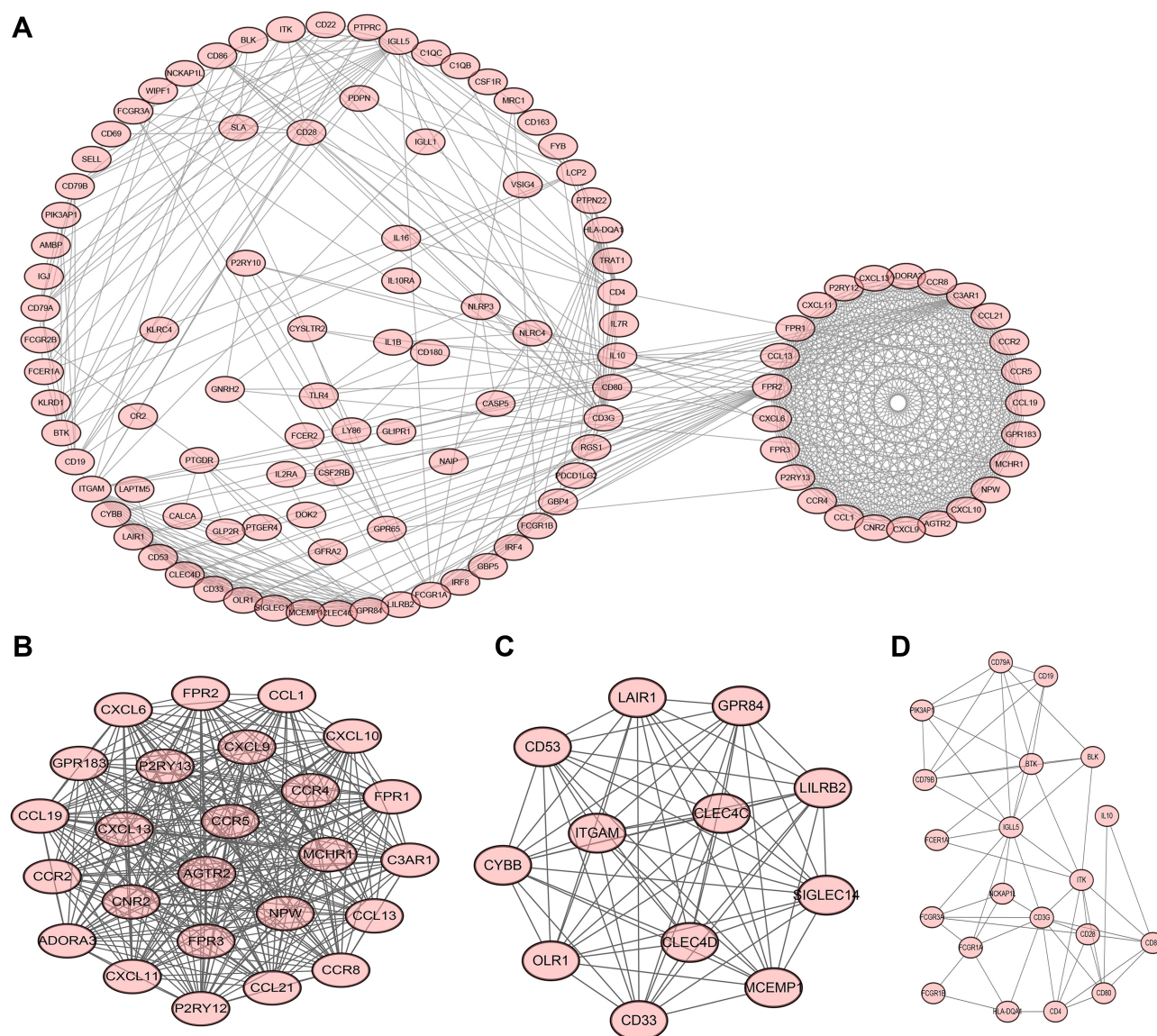
GO and KEGG Pathway Enrichment of DEGs

In order to further analyze the functional characteristics of DEGs in LUAD, we performed GO and KEGG pathway analysis of the 367 DEGs using the clusterProfiler package of R. Using the criteria of $p\text{-value} < 0.05$, 612 BP terms, 14 CC terms, 55 MF terms and 27 KEGG terms were enriched in our analysis (Supplementary Table S1). As shown in Figure 3, we drew the dotplot diagrams of the

top 10 BP, CC, MF and KEGG pathway terms. GO enrichment mainly included the immune response-regulating cell surface receptor signaling pathway, activation/proliferation of immune-related cells, external side of plasma membrane, immunoglobulin binding and chemokine activity. KEGG pathways mainly included cytokine-cytokine receptor interaction, chemokine signaling pathway and hematopoietic cell lineage. All results of functional analysis were closely related to the immune microenvironment.

PPI Network

Using the STRING online database and Cytoscape software, we filtered 367 DEGs and finally involved 113 nodes and 586 edges in the PPI network (Figure 4A).



Using MCODE to conduct module analysis on the network, we found that genes were mainly clustered in three modules, including 25, 12 and 20 points and 300, 66, and 58 edges (Figure 4B–D). The key part of the module were the related immune response genes, including mainly FPR3, CCL9, CXCL10, CD53 and CD3G.

The Prognostic Model

Next, we clarified the relationship between the expression level of 367 DEGs and the overall survival (OS), and constructed a prognostic model. The main results are as follows: 1) 22 genes were obtained from the univariate Cox proportional risk regression analysis with a threshold p value < 0.01 (Table 1); 2) The 22 genes were subjected to LASSO Cox regression analysis along with a 2000-fold cross validation (Figure 5A), and a set of 12 DEGs were identified (Figure 5B). 3) Multivariate Cox proportional risk regression analysis of the 12 DEGs yielded a prognostic model involving 6 genes (Figure 5C). The results showed that FOXN4, KLHL4 and FAM83F were risk factors and CCR2 was a protective factor.

Prognostic Model Evaluation

The LUAD patients were classified into low- and high-risk groups based on the multivariate Cox score result (Figure 6A); the expression heatmap of the 6 genes in the two groups are shown in Figure 6B. The survival analysis showed that patients with high risk had significantly shorter OS compared to those with low risk ($P = 6.302 \times 10^{-4}$) (Figure 6C). In addition, we constructed a 3-year time-dependent ROC curve, and the AUC value was 0.708 (Figure 6D). To further elucidate whether the score was an independent prognostic factor, we performed univariate and multivariate Cox regression analysis of clinically related factors. The risk score, metastasis, and clinical stage showed statistical differences in both univariate and multivariate analyses (Table 2). Therefore, we speculated that the prognostic model can be used as an independent prognostic factor for LUAD.

Discussion

The treatment of LUAD has made great progress in the past 30 years, but many obstacles remain in its molecular and biomedical research. In recent years, a large number of studies have shown that the immune response in the microenvironment plays an active role in tumor invasion, metastasis and recurrence, and is an important factor affecting

Table 1 Results of Univariate Cox Proportional Risk Regression Analysis of 367 DEGs ($P < 0.01$)

Genes	HR	HR.95L	HR.95H	p value
<i>INHA</i>	1.01	1	1.01	7.87E-05
<i>HAL</i>	1.06	1.03	1.09	8.22E-05
<i>FAM83F</i>	1.07	1.04	1.11	9.52E-05
<i>S100P</i>	1	1	1	5.49E-04
<i>CCR2</i>	0.82	0.74	0.92	7.28E-04
<i>APOC4-APOC2</i>	0.07	0.01	0.35	1.29E-03
<i>ADAMTS8</i>	0.8	0.69	0.92	2.03E-03
<i>BTK</i>	0.88	0.82	0.96	2.35E-03
<i>CD200R1</i>	0.57	0.4	0.83	3.05E-03
<i>FOXN4</i>	1.16	1.05	1.29	3.67E-03
<i>SPN</i>	0.9	0.85	0.97	3.77E-03
<i>TESC</i>	1	1	1	4.10E-03
<i>VEGFD</i>	0.91	0.85	0.97	4.58E-03
<i>RUBCNL</i>	0.67	0.5	0.89	5.27E-03
<i>P2RY13</i>	0.89	0.82	0.97	6.47E-03
<i>MS4A7</i>	0.96	0.93	0.99	6.54E-03
<i>NLRC4</i>	0.74	0.59	0.92	7.50E-03
<i>KLHL4</i>	1.23	1.06	1.44	8.00E-03
<i>ARHGEF6</i>	0.91	0.85	0.98	8.12E-03
<i>INSL4</i>	1.01	1	1.02	8.18E-03
<i>HLA-DQA1</i>	0.99	0.99	1	8.72E-03
<i>CD33</i>	0.73	0.58	0.93	9.33E-03

tumor prognosis.^{21,22} To capture the complex interactions between the LUAD genome and tumor immunity, explore the interactions between genes involved in the tumor immune microenvironment and their relationship to clinical outcome, we combined genomic, transcriptional, epigenetic, and clinical data. Our study indicated a microenvironment view of tissue-based tumor transcriptomic data and highlighted the contribution of stromal and immune cells to carcinogenesis. ESTIMATE algorithm was used to score a large number of LUAD transcriptional group samples in TCGA database, and a group of genes related to TME was obtained. Based on this, a prognostic model involving six genes was constructed. The six genes were significantly correlated with the prognosis of

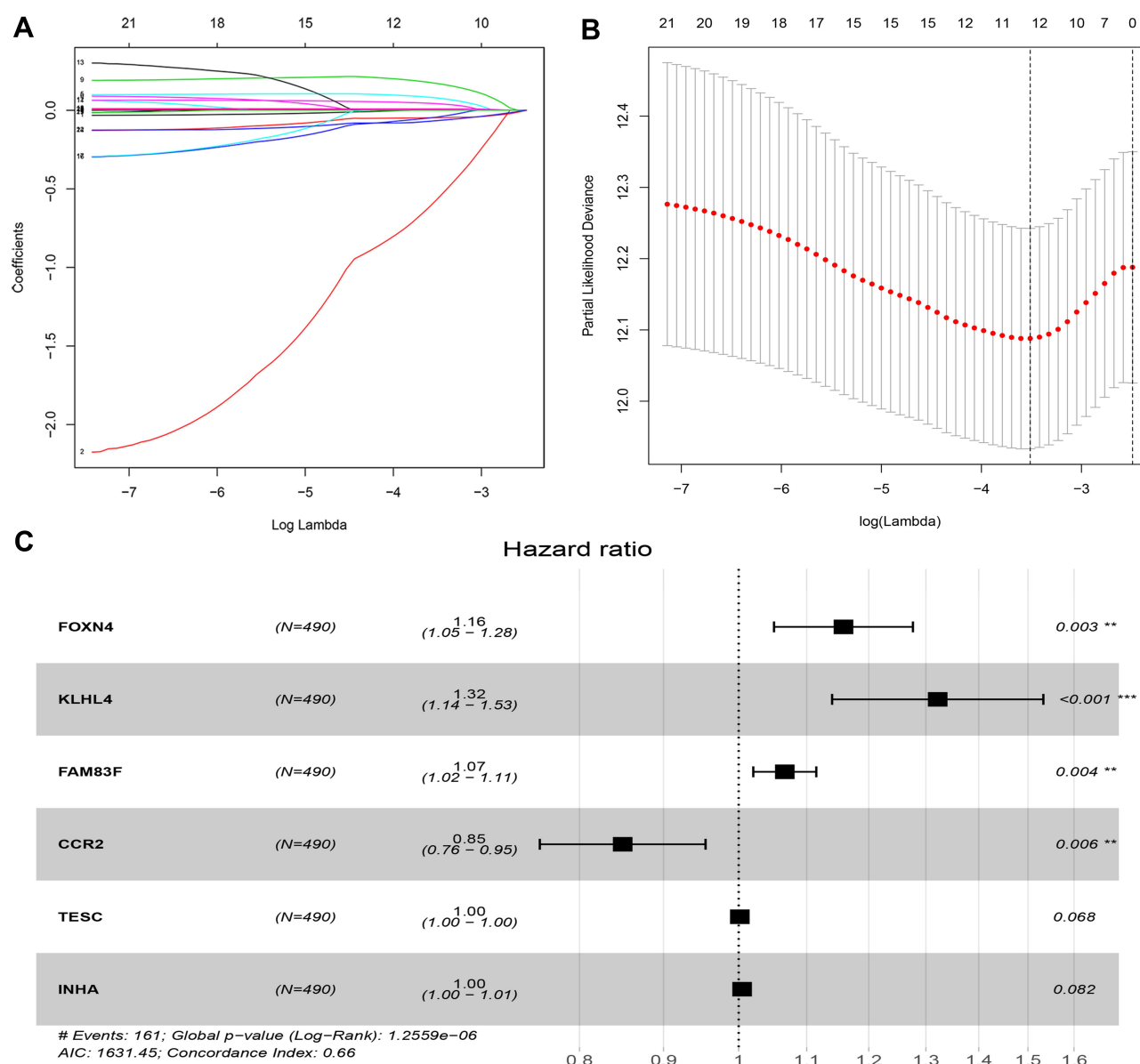
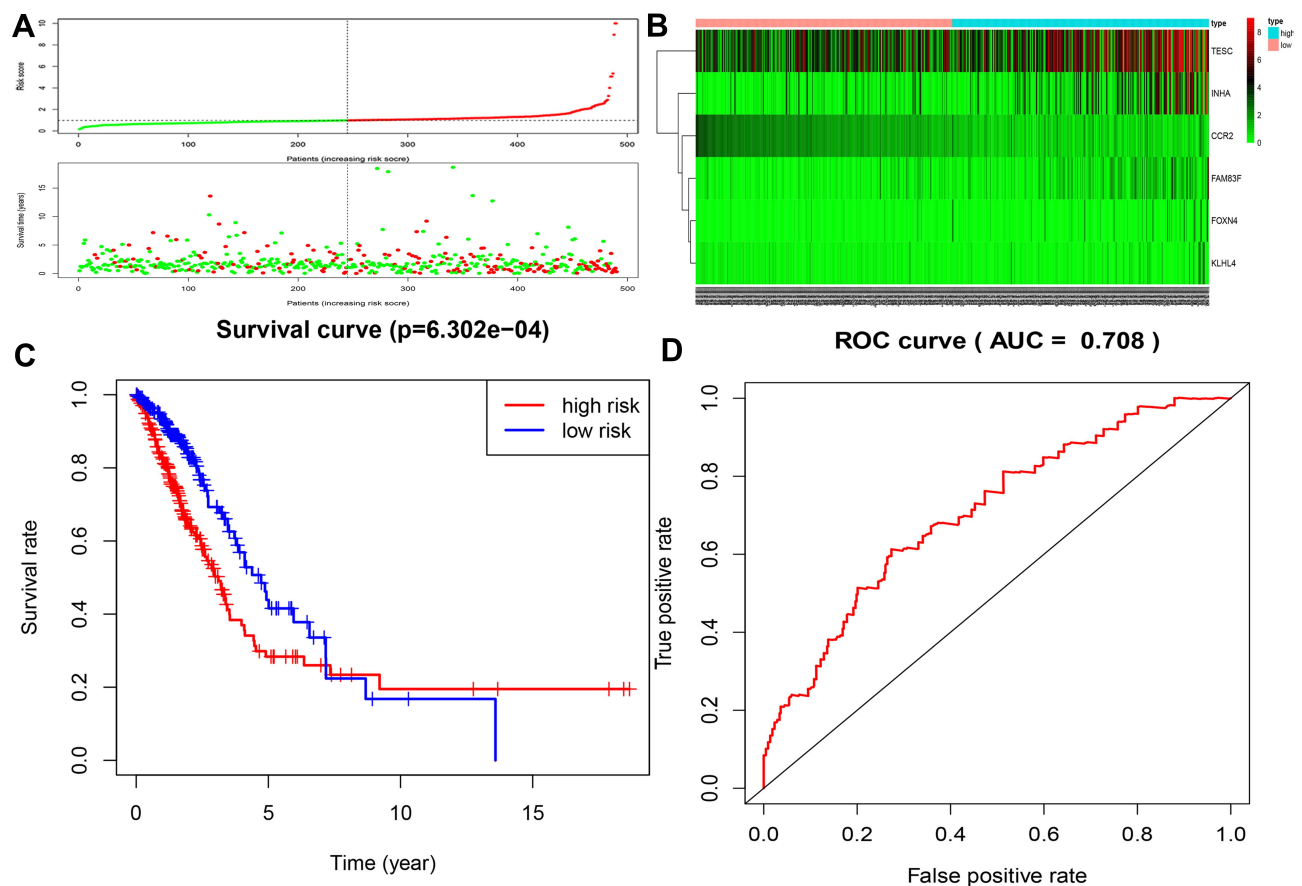


Figure 5 Identification of prognosis-related DEGs using LASSO regression model and multivariate Cox regression analysis. **(A)** LASSO coefficient profiles of the DEGs associated with disease-free survival of LUAD. **(B)** Plots of the cross-validation error rates. Each dot represents a lambda value along with error bars that represent the confidence interval for the cross-validated error rate. The top of the plot gives the size of each model. The vertical dotted line indicates the value with the minimum error and the largest lambda value where the deviance is within one SE of the minimum. **(C)** Forest plots of hazard ratios (HR) of survival-associated DEGs obtained using multivariate Cox regression analysis. A total of 6 DEGs were found to be prognostic factors. The genes with HR < 1 are protective factors, while the ones with HR > 1 are risk factors in CRC. The hazard ratio is the ratio of the HR corresponding to the conditions described by two levels of an explanatory variable. **P < 0.01, ***P < 0.001.

LUAD and are potential biomarkers and targets for immunotherapy.

We found that 367 genes were differentially expressed in immune and matrix scores, and investigated their complex biological functions using the GO and KEGG pathway analysis. The results showed that microenvironment-related DEGs were mainly enriched on the external side of the plasma membrane and receptor complex. They participated in immune response-regulating cell surface

receptor signaling pathway and T cell activation, and played a role in cytokine activity, G protein-coupled receptor binding, regulation of cytokine-cytokine receptor interaction and chemokine signaling pathway. Functional enrichment analysis showed that changes in the expression of these genes could lead to the activation, proliferation, invasion, and migration of immune-related cells in LUAD, which are crucial for tumor cell survival, invasion, and immune escape.^{23,24}



In our research, samples with a low immune score were closely associated with higher risk tendency, thus revealing poor prognosis for LUAD. LUAD is also closely related to smoking, age, pulmonary chronic diseases, and other

factors. TNM stage, clinical grade, and the method of tumor treatment can affect the prognosis assessment of patients. To investigate the adaptability of our model, we performed univariate and multivariate Cox regression

Table 2 Independent Prognostic Analysis of Six-Gene Signature by Univariate and Multivariate Cox Regression

Variables	Univariate Analysis			Multivariate Analysis		
	HR	95% CI of HR	P value	HR	95% CI of HR	P value
LUAD dataset (n=470)						
Six-gene risk score	1.201	1.139–1.266	9.62E-12	1.177	1.121–1.235	5.88E-11
Age	1.018	1.002–1.035	0.031	1.009	0.993–1.026	0.267
Gender(Female vs Male)	0.792	0.570–1.100	0.164	1.030	0.751–1.413	0.855
Stage	1.427	1.131–1.802	0.003	1.659	1.431–1.924	2.04E-11
T (Tumor)	1.116	0.896–1.391	0.326	1.496	1.223–1.828	8.57E-05
N (Lymph Node)	0.887	0.700–1.124	0.32	0.931	0.744–1.167	0.536
M (Metastasis)	1.343	1.075–1.679	0.009	1.701	1.121–1.235	2.08E-10

analyses and confirmed that the prognostic model can be used as an independent prognostic factor of LUAD. In this prognostic model, FOXP4, KLHL4, and FAM83F were the risk factors ($HR > 1$), CCR2 was the protective factor ($HR < 1$), and CCR2 was upregulated in immune and matrix scores, which was helpful in reducing the risk and confirming the treatment efficiency in LUAD patients. MCP-1/CCR2 signaling plays a crucial role in guiding immune cells in response to inflammatory stimulus.²⁵ Overexpression of CCR2 homologous receptor CCL2 has been proven to promote the production of Th2 cytokines and reduce the growth of metastatic lung tumors in wild C57BL/6 mice.²⁶ A large number of studies have found that CCL2-CCR2 signaling can maintain the proliferation and survival of cancer cells, stimulate the migration and invasion of cancer cells, and induce harmful inflammation and angiogenesis in prostate cancer,²⁷ breast cancer,^{28,29} colorectal cancer^{30,31} and other cancers. Foxn4, a member of the forkhead-box (FOX)/winged-helix transcription factor superfamily, plays an important role in the development of neurons, heart and alveoli.^{32,33} Its inactivation in mice can cause developmental defects. FOX transcription factors are implicated in carcinogenesis via gene amplification, retroviral integration, or chromosomal translocation, and the differential expression of FOXP4 has been demonstrated in nasopharyngeal carcinoma.³⁴ Studies have shown that KLHL4 is expressed in several fetal tissues, including the tongue, palate, and mandible. This result is consistent with the CPX phenotypes and upon regulation by the IGFBP5/Bcl-3 axis,^{35,36} its low expression is correlated with unfavorable outcomes in endocrinal treated ER α +PR+ breast cancer patients.³⁷ The FAM83 family of proteins is overexpressed in breast cancer, lung cancer, bladder cancer, cervical cancer, and other cancers, and has a carcinogenic effect.^{38,39} FAM83F is reported as a new carcinogenic protein overexpressed in papillary thyroid cancer, and is strongly immuno-positive in the cytoplasm. This is also consistent with our results.⁴⁰ Notably, our study shows that FOXP4, KLHL4, FAM83F and CCR2 have great potential as potential biomarkers of LUAD, but further experiments are needed to verify these results.

Conclusion

In conclusion, we performed a microenvironment-related gene analysis of LUAD using the TCGA database, based on ESTIMATE algorithm, and constructed a prognostic model. Further univariate and multivariate Cox regression analyses confirmed that the prognostic model can be used as an

independent prognostic factor for LUAD. This study provides a new and feasible method for evaluating the prognosis of LUAD and also a novel idea for exploring the potential relationship between TME and the prognosis of LUAD.

Abbreviations

LUAD, lung adenocarcinoma; TME, tumor microenvironment; TCGA, The Cancer Genome Atlas; ESTIMATE, Estimation of Stromal and Immune cells in Malignant Tumors; OS, overall survival; ssGSEA, single sample Gene Set Enrichment Analysis; PD-1, programmed death 1; PD-L1, programmed cell death-ligand 1; TILs, tumor-infiltrating lymphocytes; ICGC, International Cancer Genome Consortium; DEGs, differentially expressed gene; PPI, protein-protein interaction; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology; BP, biological process; CC, cellular component; MF, molecular function; ROC, receiver operating characteristic; AUC, area under curve.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

National Natural Science Foundation of China (81673799).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Hanahan D, Coussens LM. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell*. 2012;21(3):309–322. doi:10.1016/j.ccr.2012.02.022
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68(6):394–424. doi:10.3322/caac.21492
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;68(1):7–30. doi:10.3322/caac.21442
4. Hellmann MD, Rizvi NA, Goldman JW, et al. Nivolumab plus ipilimumab as first-line treatment for advanced non-small-cell lung cancer (CheckMate 012): results of an open-label, Phase 1, multi-cohort study. *Lancet Oncol*. 2017;18(1):31–41. doi:10.1016/S1470-2045(16)30624-6

5. Xu X, Huang Z, Zheng L, Fan Y. The efficacy and safety of anti-PD-1/PD-L1 antibodies combined with chemotherapy or CTLA4 antibody as a first-line treatment for advanced lung cancer. *Int J Cancer*. 2018;142(11):2344–2354. doi:10.1002/ijc.31252
6. Su S, Chen J, Yao H, et al. CD10(+)GPR77(+) cancer-associated fibroblasts promote cancer formation and chemoresistance by sustaining cancer stemness. *Cell*. 2018;172(4):841–856.e816. doi:10.1016/j.cell.2018.01.009
7. Hudson TJ, Anderson W, Artez A, et al. International network of cancer genome projects. *Nature*. 2010;464(7291):993–998.
8. Hoadley KA, Yau C, Wolf DM, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*. 2014;158(4):929–944. doi:10.1016/j.cell.2014.06.049
9. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–457. doi:10.1038/nmeth.3337
10. Angelova M, Charoentong P, Hackl H, et al. Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. *Genome Biol*. 2015;16:64. doi:10.1186/s13059-015-0620-6
11. Becht E, Giraldo NA, Lacroix L, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol*. 2016;17(1):218. doi:10.1186/s13059-016-1070-5
12. Yadav VK, De S. An assessment of computational methods for estimating purity and clonality using genomic data derived from heterogeneous tumor tissue samples. *Brief Bioinform*. 2015;16(2):232–241. doi:10.1093/bib/bbu002
13. Yoshihara K, Shahmoradgoli M, Martinez E, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612. doi:10.1038/ncomms3612
14. Sawyers CL, Shah N, Wang P, et al. Regulation of the glucocorticoid receptor via a BET-dependent enhancer drives antiandrogen resistance in prostate cancer. *Elife*. 2017;6:e27861. doi:10.7554/eLife.27861
15. Friedigkeit N, Watters RJ, Lucas PC, et al. Exome-capture RNA sequencing of decade-old breast cancers and matched decalcified bone metastases. *JCI Insight*. 2017;2(17). doi:10.1172/jci.insight.95703
16. Alonso MH, Ausso S, Lopez-Doriga A, et al. Comprehensive analysis of copy number aberrations in microsatellite stable colon cancer in view of stromal component. *Br J Cancer*. 2017;117(3):421–431. doi:10.1038/bjc.2017.208
17. Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci*. 2012;131(4):281–285. doi:10.1007/s12064-012-0162-3
18. Colaprico A, Silva TC, Olsen C, et al. TCGAAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*. 2016;44(8):e71. doi:10.1093/nar/gkv1507
19. Goldman M, Craft B, Swatloski T, et al. The UCSC cancer genomics browser: update 2015. *Nucleic Acids Res*. 2015;43:D812–817. doi:10.1093/nar/gku1073
20. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1–22. doi:10.18637/jss.v033.i01
21. Liu W, Ye H, Liu YF, et al. Transcriptome-derived stromal and immune scores infer clinical outcomes of patients with cancer. *Oncol Lett*. 2018;15(4):4351–4357. doi:10.3892/ol.2018.7855
22. Tan LY, Martini C, Fridlender ZG, Bonder CS, Brown MP, Ebert LM. Control of immune cell entry through the tumour vasculature: a missing link in optimising melanoma immunotherapy? *Clin Transl Immunol*. 2017;6(3):e134. doi:10.1038/cti.2017.7
23. Huang X, Sun W, Yan Z, et al. Integrative analyses of long non-coding RNA and mRNA involved in piglet ileum immune response to clostridium perfringens type C infection. *Front Cell Infect Microbiol*. 2019;9:130. doi:10.3389/fcimb.2019.00130
24. Yang K, Gao J, Luo M. Identification of key pathways and hub genes in basal-like breast cancer using bioinformatics analysis. *Oncotargets Ther*. 2019;12:1319–1331. doi:10.2147/OTT.S158619
25. Zhang K, Luo J. Role of MCP-1 and CCR2 in alcohol neurotoxicity. *Pharmacol Res*. 2019;139:360–366. doi:10.1016/j.phrs.2018.11.030
26. Hu K, Xiong J, Ji K, Sun H, Wang J, Liu H. Recombined CC chemokine ligand 2 into B16 cells induces production of Th2-dominant [correction of dominant] cytokines and inhibits melanoma metastasis. *Immunol Lett*. 2007;113(1):19–28. doi:10.1016/j.imlet.2007.07.004
27. Lin TH, Liu HH, Tsai TH, et al. CCL2 increases alphavbeta3 integrin expression and subsequently promotes prostate cancer migration. *Biochim Biophys Acta*. 2013;1830(10):4917–4927. doi:10.1016/j.bbagen.2013.06.033
28. Rego SL, Swamydas M, Kidiyoor A, et al. Soluble tumor necrosis factor receptors shed by breast tumor cells inhibit macrophage chemotaxis. *J Interferon Cytokine Res*. 2013;33(11):672–681. doi:10.1089/jir.2013.0009
29. Kitamura T, Qian BZ, Soong D, et al. CCL2-induced chemokine cascade promotes breast cancer metastasis by enhancing retention of metastasis-associated macrophages. *J Exp Med*. 2015;212(7):1043–1059. doi:10.1084/jem.20141836
30. Tsuyada A, Wang SE. Fibroblast-derived CCL2 induces cancer stem cells–response. *Cancer Res*. 2013;73(2):1032–1033. doi:10.1158/0008-5472.CAN-12-3194
31. Chun E, Lavoie S, Michaud M, et al. CCL2 promotes colorectal carcinogenesis by enhancing polymorphonuclear myeloid-derived suppressor cell population and function. *Cell Rep*. 2015;12(2):244–257. doi:10.1016/j.celrep.2015.06.024
32. Luo H, Jin K, Xie Z, et al. Forkhead box N4 (Foxn4) activates Dll4-Notch signaling to suppress photoreceptor cell fates of early retinal progenitors. *Proc Natl Acad Sci U S A*. 2012;109(9):E553–562. doi:10.1073/pnas.1115767109
33. Li S, Xiang M. Foxn4 influences alveologenesis during lung development. *Dev Dyn*. 2011;240(6):1512–1517. doi:10.1002/dvdy.22610
34. Lun SW, Cheung ST, Cheung PF, et al. CD44+ cancer stem-like cells in EBV-associated nasopharyngeal carcinoma. *PLoS One*. 2012;7(12):e52426. doi:10.1371/journal.pone.0052426
35. Braybrook C, Warry G, Howell G, et al. Identification and characterization of KLHL4, a novel human homologue of the Drosophila Kelch gene that maps within the X-linked cleft palate and Ankyloglossia (CPX) critical region. *Genomics*. 2001;72(2):128–136. doi:10.1006/geno.2000.6478
36. Frasor J, Weaver A, Pradhan M, et al. Positive cross-talk between estrogen receptor and NF-kappaB in breast cancer. *Cancer Res*. 2009;69(23):8918–8925. doi:10.1158/0008-5472.CAN-09-2608
37. Leyh B, Dittmer A, Lange T, Martens JW, Dittmer J. Stromal cells promote anti-estrogen resistance of breast cancer cells through an insulin-like growth factor binding protein 5 (IGFBP5)/B-cell leukemia/lymphoma 3 (Bcl-3) axis. *Oncotarget*. 2015;6(36):39307–39328. doi:10.18632/oncotarget.5624
38. Cipriano R, Graham J, Miskimen KL, et al. FAM83B mediates EGFR- and RAS-driven oncogenic transformation. *J Clin Invest*. 2012;122(9):3197–3210. doi:10.1172/JCI60517
39. Li Y, Dong X, Yin Y, et al. BJ-TSA-9, a novel human tumor-specific gene, has potential as a biomarker of lung cancer. *Neoplasia (New York, N Y)*. 2005;7(12):1073–1080. doi:10.1593/neo.05406
40. Fuziwar CS, Saito KC, Leoni SG, Waitzberg AFL, Kimura ET. The highly expressed FAM83F protein in papillary thyroid cancer exerts a pro-oncogenic role in thyroid follicular cells. *Front Endocrinol (Lausanne)*. 2019;10:134. doi:10.3389/fendo.2019.00134
41. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47. doi:10.1093/nar/gkv007

42. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinform.* 2011;12(1). doi:10.1186/1471-2105-12-35
43. Yu G, Wang L, Han Y, He Q. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284–287. doi:10.1089/omi.2011.0118
44. Walter W, Sanchez-Cabo F, Ricote M. GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics.* 2015;31(17):2912–2914. doi:10.1093/bioinformatics/btv300
45. Szklarczyk D, Morris J, Cook H, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 2017;45:D362–D368. doi:10.1093/nar/gkw937
46. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015;43(Database issue):D447–452. doi:10.1093/nar/gku1003

Pharmacogenomics and Personalized Medicine

Dovepress

Publish your work in this journal

Pharmacogenomics and Personalized Medicine is an international, peer-reviewed, open access journal characterizing the influence of genotype on pharmacology leading to the development of personalized treatment programs and individualized drug selection for improved safety, efficacy and sustainability. This journal is indexed

on the American Chemical Society's Chemical Abstracts Service (CAS). The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/pharmacogenomics-and-personalized-medicine-journal>