Identification of prognostic genes in lung adenocarcinoma immune microenvironment

Lian-Kui Han[1], Qi-Lin Huai[2], Wei Guo[3], Peng Song[3], De-Miao Kong[1], Shu-Geng Gao[3]

[1]Department of Thoracic Surgery, Guizhou Provincial People's Hospital, Guiyang, Guizhou 550001, China;

[2]Department of Graduate School, Zunyi Medical University, Zunyi, Guizhou 563000, China;

[3]Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China.

To the Editor: Lung adenocarcinoma (LUAD) with a high degree of malignancy is the most common histological subtype of lung cancer, and the leading cause of cancer-related death in the world.[1] Tumor immune microenvironment is mainly composed of infiltrating immune and stromal cells, which have an important influence on the prognosis of LUAD patients.[2] In recent years, rapid advances in high-throughput sequencing technology have profoundly changed our understanding of tumor research. It uses a large amount of public clinical data to help researchers more effectively to study the characteristics of tumors and improve our ability to diagnose, treat, and prevent cancer. The purpose of this study was to use the estimation of stromal and immune cells in malignant tumor tissues using expression data (ESTIMATE) algorithm to calculate the immune and stromal scores of LUAD patients from The Cancer Genome Atlas (TCGA) database, explore the correlation between the scores and the survival of LUAD patients, and identify differentially expressed genes (DEGs) based on the immune/stromal scores. In addition, we also judged the prognostic value of DEGs and further studied their potential molecular functions.

Recently, the focus of cancer-related research has been gradually turned to the tumor microenvironment based on cell biology and molecular mechanisms, which plays a major role in the progression of cancer including LUAD. Malignant tumor cells, mesenchymal cells, inflammatory mediators, endothelial cells, stromal cells, immune cells, and normal epithelial cells are involved to constitute the microenvironment of tumor.[3] For the development of prognostic and predictive biomarkers, an in-depth study on the expression of related genes in tumor immune microenvironment is of much significance. To assess the stromal and immune cells in tumor samples for their

infiltration, Yoshihara *et al*[4] designed an algorithm known as ESTIMATE to predict the number of stromal and immune cells in tumor samples based on the unique characteristics of the transcriptional profiles of cancer samples. To predict the infiltration of non-tumor cells, single-sample gene set enrichment analysis was conducted for calculating the stromal and immune scores. In the prediction of prostate cancer, glioblastoma, colon cancer, thyroid cancer, breast cancer, and acute myeloid leukemia, the ESTIMATE algorithm has been used to obtain a large number of genes with prognostic value. In this study, the immune/stromal scores of LUAD patients were analyzed in the TCGA database using the ESTIMATE algorithm in package edgeR. We identified DEGs by the "limma" package in R (Version 3.6.3). The prognostic value of DEGs was further verified by Gene Expression Omnibus (GEO) database, Kaplan-Meier (K-M) plotter database, and various bioinformatics tools, including Search Tool for the Retrieval of Interacting Genes/Proteins (STRING), and gene set enrichment analysis (GSEA) was used to explore the function and signaling pathway of these genes.

First, after the cases of missing clinical information were excluded, a total of 522 LUAD patients with clinical characteristics information were recruited. The clinicopathological factors extracted for analysis include age, gender, T stage, N stage, M stage, and tumor-node-metastasis (TNM) stage. Among them, the number of female patients is 280 (53.6%), while that of male patients is 242 (46.4%). On average, the age was 66 years with a range from 33 to 88 years. Ignoring the samples with undetermined TNM stage, the proportions of T stage were T1 (172, 33.0%), T2 (281, 53.8%), T3 (47, 9.0%), T4 (19, 3.6%); the proportions of N stage were N0 (335, 64.2%),

Lian-Kui Han, Qi-Lin Huai and Wei Guo contributed equally to this work.

Correspondence to: Shu-Geng Gao, Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Panjiayuannanli No. 17, Chaoyang District, Beijing 100021, China
E-Mail: gaoshugeng@vip.sina.com
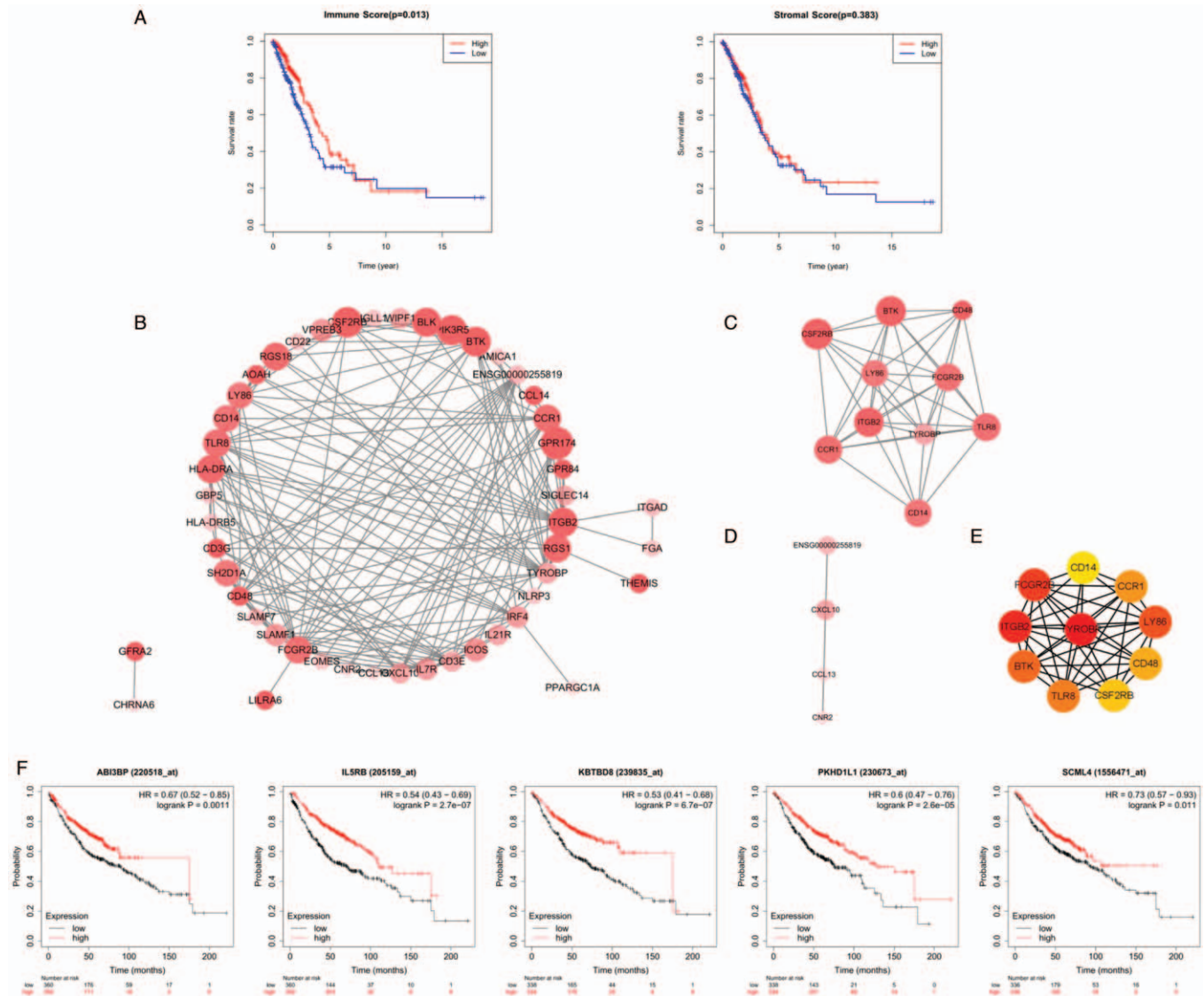
Figure 1: Immune/stromal score calculated based on the ESTIMATE algorithm and identification of DEGs with prognostic value in LUAD. (A) The correlation between immune/stromal score and OS of LUAD patients. (B) Based on the STRING database and Cytoscape software, the PPI network of common DEGs containing 49 nodes and 85 edges revealed the close relationship between these genes. (C) Module 1 contains ten nodes and 40 edges. (D) Module 2 contains four nodes and five edges. (E) PPI network of top ten DEGs; darker color represents higher scores. (F) Key genes with prognostic value in the GEO database and K-M plotter. DEGs: Differentially expressed genes; GEO: Gene Expression Omnibus; KEGG: Kyoto Encyclopedia of Genes and Genomes; LUAD: Lung adenocarcinoma; K-M: Kaplan-Meier; OS: Overall survival; PPI: Protein-protein interaction; STRING: Search Tool for the Retrieval of Interacting Genes/Proteins.

N1 (98, 18.8%), N2 (75, 14.4%), N3 (2, 0.4%); the proportions of M stage were M0 (353, 67.6%), and M1 (25, 4.8%). As shown in Supplementary Table 1, http://links.lww.com/CM9/A454, the patients in stages I and II with no metastasis account for over half. Besides, the ESTIMATE algorithm was applied to calculate the immune and stromal scores of all patients. Depending on their median, the stromal and immune scores were divided into two groups. The stromal scores of LUAD patients varied between −372.83 and 351.82, and the immune scores varied from −172.02 to 656.87. As shown in Figure 1A, the K-M survival curves were generated by high- and low-score groups. According to K-M survival analysis, the low immune score group indicates poor overall survival in LUAD cohort ($P = 0.013$). Moreover, the group of stromal score and overall survival time revealed no clear distinction ($P = 0.038$). The differential

expression analysis was conducted on the expression profile of LUAD from the TCGA database to investigate how gene expression profiles relate to immune and stromal scores. The genes with differentiated expression in the high and low immune score group can be displayed in the heatmap ($|$fold change$| > 1$, $P < 0.05$), and the Venn diagrams were plotted to show that there were 64 commonly up-regulated DEGs and eight commonly down-regulated DEGs. In addition, the gene ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of these 72 co-expressed DEGs showed that they are all related to the immune process [Supplementary Figure 1, http://links.lww.com/CM9/A454]. With the STRING database, a protein-protein interaction (PPI) network was obtained. As shown in Figure 1B, we discovered that the PPI network contained 49 nodes and 172 edges. Then, the MCODE was

employed to cluster the analysis of PPI network and obtain two significant modules [Figure 1C and 1D]. Additionally, the interacting genes were introduced into Cytoscape, based on which the top ten genes with the cytoHubba were identified [Figure 1E].

Next, we performed a survival analysis on 64 DEGs that were commonly up-regulated and eight DEGs that were commonly down-regulated. Of the 72 common DEGs in the immune score group and stromal score group, 22 genes were associated with the prognosis of LUAD patients as revealed by the log-rank test ($P < 0.05$). The typical genes are shown in Supplementary Figure 2, http://links.lww.com/CM9/A454. Functional enrichment analysis of 22 prognostic-related genes showed that these genes are related to immune response [Supplementary Figure 3, http://links.lww.com/CM9/A454]. Importantly, we determined that six of 22 genes also have prognostic value in the GEO database [Supplementary Figure 4, http://links.lww.com/CM9/A454]. And five of the above-mentioned six genes were also verified in the K-M plotter, which are ABI3BP, CSF2RB(IL5RB), KBTBD8, PKHD1L1, and SCML4 [Figure 1F]. These five genes are regarded as key genes. Finally, we used the biocarta gene sets to perform GSEA on five key genes, and the results showed that many gene sets in the samples with high expression of the five key genes were positively enriched. The five most significantly correlated pathways of five key genes are shown in order in Supplementary Figures 5 to 9, http://links.lww.com/CM9/A454.

Herein, we used the ESTIMATE algorithm to calculate the immune/stromal scores of LUAD patients from TCGA, found immune-related DEGs with prognostic value in the tumor immune microenvironment, and used multi-omics data mining technology to explore the potential molecular functions and pathways of these genes. As a result, five key genes (CSF2RB, PKHD1L1, ABI3BP, SCML4, and KBTBD8) are of great interest to better understand the prognosis of LUAD patients. ABI3BP, a potential biomarker of lung cancer, has been proven to be significantly suppressed in lung cancer cell lines.[5] The up-regulation of ABI3BP in gallbladder cancer has been reported to repress the development of gallbladder cancer via inhibiting H3K27 methylation induced by EZH2.[6] CSF2RB serves as an apoptosis inducer in cancer cells[7] and PKHD1L1 as an inhibitor of cell growth and invasion in thyroid cancer cells,[8] which are potential therapeutic targets for LUAD. The particular role of KBTBD8 and SCML4 in cancers has not been reported. Furthermore, the five most significant-related pathways of the five key genes by GSEA involve the TCR, IL2RB, FCER1, FMLP, CTCF, TOB1, CTLA4, GH, IL7, NO2IL12, IL12, and TH1TH2 signaling pathways. However, the particular mechanism of these signalings needs to be verified based on more *in vivo* and *in vitro* experiments.

In addition, this study focuses on the application of the ESTIMATE algorithm and various bioinformatics technologies. The value of key genes has not been further verified through experiments on human samples. These experiments will be resolved promptly in our future publications. In summary, the genes related to the immune microenvironment of LUAD were investigated using the ESTIMATE algorithm, based on which the function and prognostic values of these key genes were analyzed through integrated bioinformatics. The five key genes are likely to be potential therapeutic targets and prognostic markers, which make it necessary to conduct further research in this regard.

Conflicts of interest

None.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2018;68:7–30. doi: 10.3322/caac.21442.
2. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, et al. The immune landscape of cancer. Immunity 2018;48:812–830. e14. doi: 10.1016/j.immuni.2018.03.023.
3. Armani G, Madeddu D, Mazzaschi G, Bocchialini G, Sogni F, Frati C, et al. Blood and lymphatic vessels contribute to the impact of the immune microenvironment on clinical outcome in non-small-cell lung cancer. Eur J Cardiothorac Surg 2018;53:1205–1213. doi: 10.1093/ejcts/ezx492.
4. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun 2013;4:2612. doi: 10.1038/ncomms3612.
5. Terauchi K, Shimada J, Uekawa N, Yaoi T, Maruyama M, Fushiki S. Cancer-associated loss of TARSH gene expression in human primary lung cancer. J Cancer Res Clin Oncol 2006;132:28–34. doi: 10.1007/s00432-005-0032-1.
6. Lin N, Yao Z, Xu M, Chen J, Lu Y, Yuan L, et al. Long noncoding RNA MALAT1 potentiates growth and inhibits senescence by antagonizing ABI3BP in gallbladder cancer cells. J Exp Clin Cancer Res 2019;38:244. doi: 10.1186/s13046-019-1237-5.
7. Slattery ML, Mullany LE, Sakoda LC, Wolff RK, Samowitz WS, Herrick JS. Dysregulated genes and miRNAs in the apoptosis pathway in colorectal cancer patients. Apoptosis 2018;23:237–250. doi: 10.1007/s10495-018-1451-1.
8. Zheng C, Quan R, Xia EJ, Bhandari A, Zhang X. Original tumour suppressor gene polycystic kidney and hepatic disease 1-like 1 is associated with thyroid cancer cell progression. Oncol Lett 2019;18:3227–3235. doi: 10.3892/ol.2019.10632.