



OPEN ACCESS

EDITED BY

Ming Jun Zheng,
Ludwig Maximilian University of Munich,
Germany

REVIEWED BY

Songwei Feng,
Southeast University, China
Yaling Li,
China Medical University, China
Yi Li,
Nanjing Medical University, China

*CORRESPONDENCE

Leilei Liang,
liangleilei10006@163.com
Nan Song,
songnan@tongji.edu.cn

[†]These authors have contributed equally
to this work

SPECIALTY SECTION

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

RECEIVED 22 June 2022

ACCEPTED 22 July 2022

PUBLISHED 15 August 2022

CITATION

Xu Y, Wang Y, Liang L and Song N (2022),
Single-cell RNA sequencing analysis to
explore immune cell heterogeneity and
novel biomarkers for the prognosis of
lung adenocarcinoma.
Front. Genet. 13:975542.
doi: 10.3389/fgene.2022.975542

COPYRIGHT

© 2022 Xu, Wang, Liang and Song. This
is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Single-cell RNA sequencing analysis to explore immune cell heterogeneity and novel biomarkers for the prognosis of lung adenocarcinoma

Yong Xu[†], Yao Wang[†], Leilei Liang* and Nan Song*

Department of Thoracic Surgery, Shanghai Pulmonary Hospital, School of Medicine, Tongji University, Shanghai, China

Background: Single-cell RNA sequencing is necessary to understand tumor heterogeneity, and the cell type heterogeneity of lung adenocarcinoma (LUAD) has not been fully studied.

Method: We first reduced the dimensionality of the GSE149655 single-cell data. Then, we statistically analysed the subpopulations obtained by cell annotation to find the subpopulations highly enriched in tumor tissues. Monocle was used to predict the development trajectory of five subpopulations; beam was used to find the regulatory genes of five branches; qual was used to screen the key genes; and cellchart was used to analyse cell communication. Next, we used the differentially expressed genes of TCGA-LUAD to screen for overlapping genes and established a prognostic risk model through univariate and multivariate analyses. To identify the independence of the model in clinical application, univariate and multivariate Cox regression were used to analyse the relevant HR, 95% CI of HR and *p* value. Finally, the novel biomarker genes were verified by qPCR and immunohistochemistry.

Results: The single-cell dataset GSE149655 was subjected to quality control, filtration and dimensionality reduction. Finally, 23 subpopulations were screened, and 11-cell subgroups were annotated in 23 subpopulations. Through the statistical analysis of 11 subgroups, five important subgroups were selected, including lung epithelial cells, macrophages, neuroendocrine cells, secret cells and T cells. From the analysis of cell trajectory and cell communication, it is found that the interaction of five subpopulations is very complex and that the communication between them is dense. We believe that these five subpopulations play a very important role in the occurrence and development of LUAD. Downloading the TCGA data, we screened the marker genes of these five subpopulations, which are also the differentially expressed genes in tumorigenesis, with a total of 462 genes, and constructed 10 gene prognostic risk models based on related genes. The 10-gene signature has strong robustness and can achieve stable prediction efficiency in datasets from different platforms. Two new molecular markers related to LUAD, HLA-DRB5 and CCDC50, were verified by qPCR and immunohistochemistry. The results showed that *HLA-DRB5* expression was negatively correlated with the

risk of LUAD, and CCDC50 expression was positively correlated with the risk of LUAD.

Conclusion: Therefore, we identified a prognostic risk model including CCL20, CP, HLA-DRB5, RHOV, CYP4B1, BASP1, ACSL4, GNG7, CCDC50 and SPATS2 as risk biomarkers and verified their predictive value for the prognosis of LUAD, which could serve as a new therapeutic target.

KEYWORDS

lung adenocarcinoma, single-cell sequencing, tumor heterogeneity, tumor immunity, prognosis

Introduction

Lung cancer is still one of the main types of cancer, and its mortality is still the highest of all cancers (Siegel et al., 2021). Lung adenocarcinoma (LUAD) is the most common histological subtype of lung cancer, accounting for almost half of all lung cancer deaths (Kenfield et al., 2008; Houston et al., 2014). Because of the decline in smoking rates, the incidence and mortality of many other types of lung cancer, such as squamous cell lung cancer and small cell lung cancer, have been decreasing. The incidence rate and incidence rate of LUAD are increasing (Remen et al., 2018; Barta et al., 2019; Choi et al., 2019; Tseng et al., 2019). At present, the treatment of patients with advanced LUAD is still limited to targeted therapy and radiotherapy and chemotherapy, and the prognosis is still very poor. Therefore, finding accurate prognostic biomarkers and effective therapeutic targets is still of great significance to improve the poor prognosis of LUAD patients.

In recent decades, high-throughput sequencing technology has been widely used in various fields of biology and medicine, which has greatly promoted related research and applications. However, traditional transcriptome sequencing technology (bulkRNA-seq) is based on tissue samples or cell populations, which reflect the average expression level of genes in the cell population, but there is extensive heterogeneity between cells, which is of great significance for targeted therapy of tumors (Dagogo-Jack and Shaw, 2018). In recent years, single-cell RNA SEQ (scRNA-seq) technology has developed vigorously. ScRNA-seq can reveal the expression of all genes in the whole genome at the single-cell level and study cell heterogeneity more intuitively (Lavin et al., 2017). At present, scRNA-seq has been widely used in different types of tissues and cell lines of various species (especially human and mouse), including normal and diseased cells. Single-cell sequencing has been used in the study of pancreatic cancer, colon cancer, and so on, but (Moncada et al., 2020; Zhang L. et al., 2020; Liang et al., 2021) it has not been widely studied in lung cancer. We found and defined the cell subsets of LUAD by single-cell analysis and explored their predictive ability in the prognosis of LUAD.

This study screened cell types with significant differences in subpopulation abundance through single-cell analysis and

screened cell types with different subpopulation abundance. At the same time, combined with LUAD bulkRNA-seq in TCGA data, the marker genes related to prognosis were screened, and the risk model was constructed accordingly. Finally, we identified a prognostic risk model and verified its predictive value for the prognosis of LUAD, which could serve as a new therapeutic target.

Materials and methods

Data acquisition and preprocessing

The single-cell sequencing data GSE149655 were downloaded from the GEO database. A total of four samples were detected, including two LUAD samples and two normal samples. The bulkRNA-seq data of LUAD were downloaded from the TCGA database and further processed and transformed into TPM data.

The clinical phenotype data of TCGA-LUAD were downloaded, and the samples lacking survival time and survival status were eliminated. GSE31210 of LUAD was downloaded from the GEO database. By transforming the probe into a gene symbol, multiple gene symbols corresponding to one probe were removed, and the average value of one symbol corresponding to the probe was taken.

Clustering dimensionality reduction of single-cell data

First, the single-cell data were filtered by setting each gene to be expressed in at least three cells and each cell to express at least 250 genes, calculating the proportion of mitochondria and rRNA through the percentagefeatureset function, and ensuring that the gene expressed by each cell was greater than 500 and the mitochondrial content was less than 35%. Then, we counted the number of cells in each sample before and after filtration. Then, the merged data are standardized through log normalization. Find highly variable genes through the findvariablefeatures function (identify variable

characteristics based on variance stabilization transformation (“VST”) and then scale all genes by using the `scaledata` function and PCA dimensionality reduction to find anchor points. We selected $\text{dim} = 40$ and clustered the cells through the `findneighbors` and `findclusters` functions (set $\text{resolution} = 0.5$).

Subgroup definition

We downloaded the marker genes of human cells from the official website of CellMarker (<http://biocc.hrbmu.edu.cn/CellMarker/>). At the same time, the corresponding subgroups of cluster profilers are selected through the corresponding functions of cluster profilers.

Subgroup statistical analysis

A total of 11 subpopulations were obtained through cell annotation. We counted the number of cells in tumor samples and normal samples, constructed a 2×2 contingency table, calculated the p value using Fisher’s test (bilateral test), and calculated the corresponding difference multiple (TVSN). To define the development trajectories of the five cell subsets, we used Monocle to predict the development trajectories of the five subsets. Then, we used the BEAM (branched expression analysis modelling) method to find the regulatory genes of five branches, screened the key genes with `qval` (corrected P), screened the 100 genes with the smallest `qval`, drew the heatmap and enriched the pathway.

Screening of key genes and construction, evaluation and validation of the prognostic risk model

The expression profile data of FPKM of TCGA were downloaded and further transformed into TPM. The standard deviation of each gene expressed in all samples was greater than 0.5 for filtering. The expression profile matrix of LUAD was analysed by the `limma` package, and the differentially expressed genes were screened by $|\log_{2}FC| > 1$ and $\text{FDR} < 0.05$.

First, single-factor risk analysis is carried out. Using the expression profile data of TCGA, for the related genes and survival data, the univariate Cox proportional hazards regression model was carried out by using the R-package survival Cox function, and $p < 0.01$ was selected as the threshold for filtering.

Next, multivariate analysis is carried out. Lasso regression was used to further compress the genes to reduce the number of genes in the risk model. Next, 10-fold cross validation was used to

build the model and analyse the confidence interval under each λ .

The risk score of each sample was calculated according to the expression level of the sample, and the risk score distribution of the sample was drawn. ROC analysis of the prognostic classification of the risk score was carried out by using the R software package `timeROC`, and the prognostic prediction and classification efficiency at 1, 3 and 5 years were analysed. Finally, we calculated the z score for the risk score, divided the samples with risk scores greater than zero into a high-risk group and a low-risk group, and drew a KM curve. Lasso regression and the risk score were performed as previously described (Yu et al., 2021).

Finally, we used the GEO dataset (GSE31210) to verify the model.

Univariate and multivariate analysis of the 10-gene signature and its relationship with pathways.

To identify the independence of the 10-gene signature model in clinical application, the relevant HR, 95% CI of HR and p value were analysed by univariate and multivariate Cox regression in the clinical information carried by all TCGA data. The clinical information recorded by TCGA patients was systematically analysed, including sex, stage and risk type.

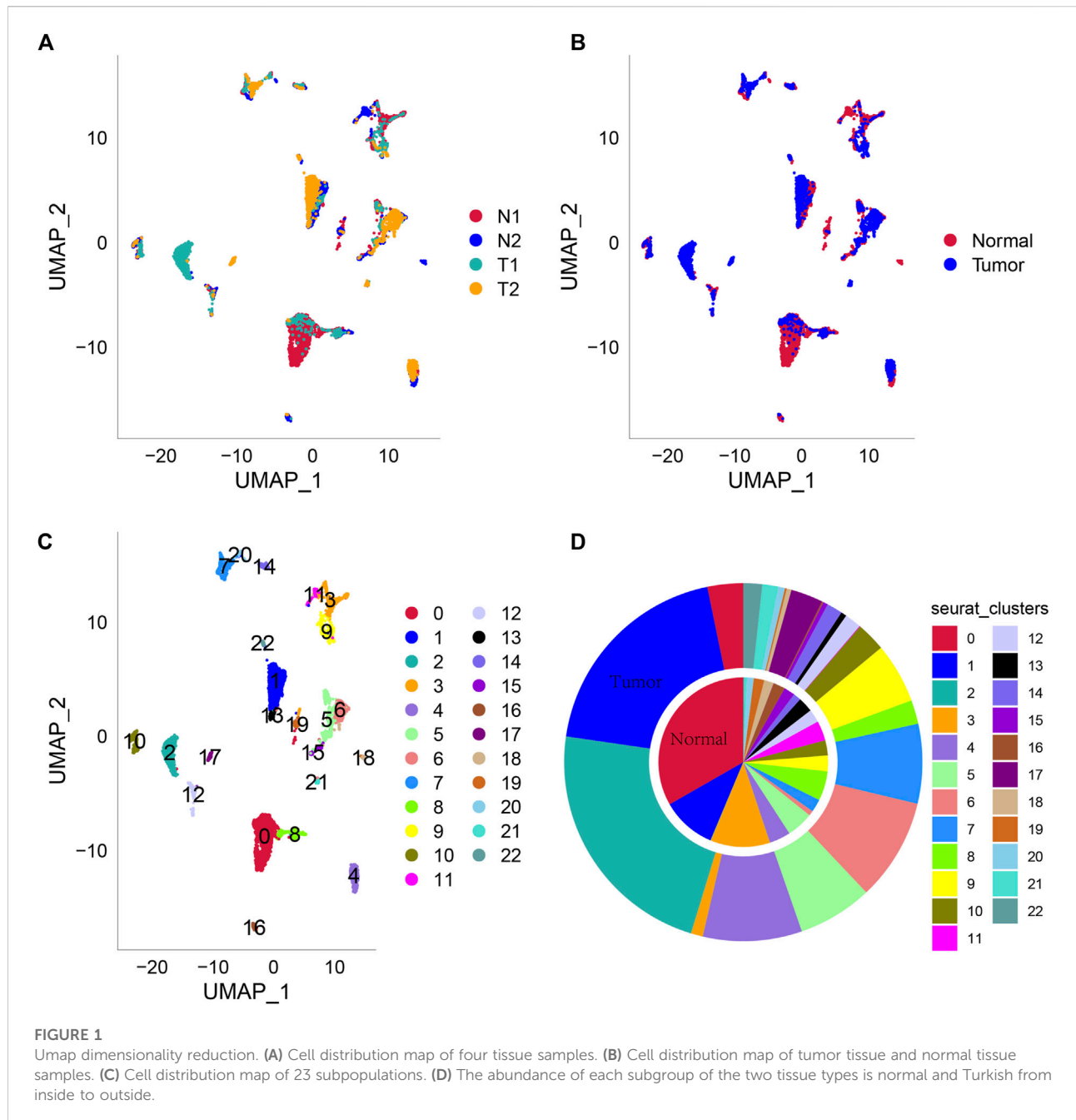
To further observe the relationship between the risk scores of different samples and biological functions, the expression profiles corresponding to TCGA samples were analysed by single-sample GSEA (ssGSEA) with the R software package `GSVA`, and the scores of each sample on different functions were calculated; that is, the ssGSEA scores of each sample corresponding to each function were obtained, and the correlation between these functions and risk scores was further calculated. A function with a correlation of no less than 0.3 was selected.

Tissue samples

Samples of LUAD and normal tissues were collected from 15 patients (all >16 years of age), immediately placed in liquid nitrogen and preserved at -80°C . None of the LUAD patients received preoperative antitumor therapies. Patients and their families in this study were fully informed, and informed consent was obtained from all participants. This study was approved by the Ethics Committee of Shanghai Pulmonary Hospital (K20-148Y).

RNA isolation and quantitative real-time PCR

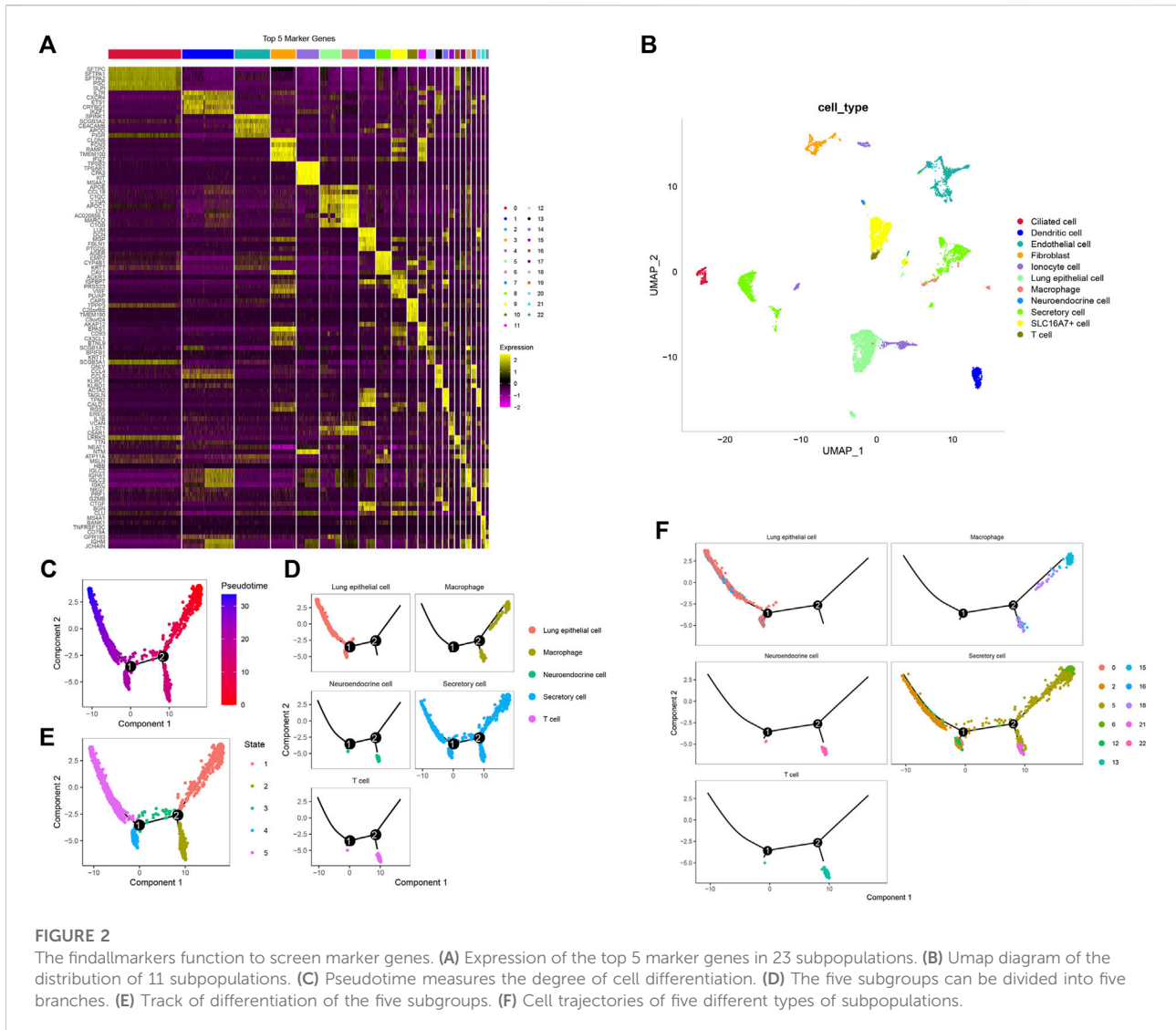
Briefly, total RNA was isolated from tissues and cells by using TRIzol[®] reagent (Thermo Fisher Scientific, Inc.) and then reverse transcribed using a QuantiTect Reverse Transcription Kit



(QIAGEN, Valencia, CA) according to the manufacturer’s specifications. qPCR amplification was performed by using SYBR-Green PCR mix (Takara), and the expression levels of target genes were normalized to the level of GAPDH. The primer sequences were as follows: HLA-DRB5 Forward Sequence-GAACAGCCAGAAGGACTTCCTG and Reverse Sequence-GCAGGATACACAGTCACCTTAGG. CCDC50 Forward Sequence-AGTGATGAACCTCACCATTCTAAG and Reverse Sequence-GAAATGCCGTGTGGAAGCTCTGC.

Immunohistochemistry

Each group of sarcoma samples was fixed in 10% formalin, embedded in paraffin, and processed into 5 μm continuous sections. Samples were incubated in rabbit anti-HLA-DRB5 (Origene, OTI4G7; 1:1,200) anti-CCDC50 (Abcam, ab127169; 1:1,200) overnight at 4°C, followed by incubation with horseradish peroxidase-coupled goat anti-rabbit secondary antibody at 37°C for 30 min. The experimental procedure was



performed according to strict adherence to the manufacturers’ instructions. The IHC quantitation analysis was calculated by ImageJ software.

Results

Screening and definition of single-cell subsets of LUAD

First, the genes were screened. [Supplementary Figure S1](#) is the quality control chart before filtration, and [Supplementary Figure S2](#) is the quality control chart after filtration. Then, we counted the number of cells in each sample before and after filtration, as shown in [Supplementary Figure S3A](#).

[Supplementary Figure S3B](#) (left) shows the distribution of hypervariable genes and non-hypervariable genes, and the top 20 hypervariable genes are shown in [Supplementary Figure S3B](#) (right).

Then, all genes were scaled by using the scaledata function, and anchor points were found by PCA dimensionality reduction ([Supplementary Figure S3C](#)). Next, the cells were clustered, and 23 subpopulations were obtained. Then, we select the first 40 PCs and use umap to further reduce the dimension. The distribution of the four samples is shown in [Figure 1A](#). Two of the four samples were tumor tissues, and two were normal tissues. The distribution of cells in tumor tissues and normal tissues is shown in [Figure 1B](#), and [Figure 1C](#) shows the distribution of 23 cell subsets. At the same time, we counted the abundance of these 23 subpopulations in each sample ([Figure 1D](#)). Next, we used the findallmarkers function to screen marker genes of 23 subgroups by logfc = 0.5 (differential multiple), minpct = 0.3 (minimum expression ratio of differential genes) and screened them

TABLE1 Subgroups define information.

cell_type	seraut_cluster
Lung epithelial cell	0
SLC16A7+ cell	1
Secretory cell	2
Endothelial cell	3
Dendritic cell	4
Secretory cell	5
Secretory cell	6
Fibroblast	7
Ionocyte cell	8
Endothelial cell	9
Ciliated cell	10
Endothelial cell	11
Secretory cell	12
T cell	13
Ionocyte cell	14
Macrophage	15
Lung epithelial cell	16
Ionocyte cell	17
Macrophage	18
SLC16A7+ cell	19
Fibroblast	20
Secretory cell	21
Neuroendocrine cell	22

with corrected $p < 0.05$. Here, we only show the expression of the top 5 marker genes with the most prominent contribution in each subgroup (Figure 2A). The results of marker genes are shown in Supplementary Table S1.

We downloaded the marker genes of human cells from the official website of CellMarker (<http://biocc.hrbmu.edu.cn/CellMarker/>), selected the corresponding organization of the lung, and Supplementary Table S2 was the list of marker genes of cells. At the same time, through the enricher function of the clusterprofiler package, the definition of 23 subsets is finally completed. As shown in Table 1, 23 clusters are annotated to 11 subgroups. By merging these subgroups, 11 subgroups (ciliated cells, dendritic cells, endothelial cells, fibroblasts, monocyte cells, lung epithelial cells, macrophages, neuroendocrine cells, secretory cells, SLC16A7+ cells and T cells) were obtained. Figure 2B is the umap diagram of the distribution of these 11 subgroups.

Cell trajectory analysis of single-cell subsets in LUAD

A total of 11 subpopulations were obtained through cell annotation, which were counted according to the method. The

statistical results are shown in Table 2. On the premise of $p < 0.05$, five subpopulations of lung epithelial cells, macrophages, neuroendocrine cells, secret cells and T cells were highly enriched in LUAD tissues. We used Monocle to predict the developmental trajectories of five subpopulations.

In previous studies, LUAD tumor samples contained 18.2% tumor cells and 53.4% T cells, while normal samples contained 10.4% epithelial cells and 44.1% T cells, indicating that T cells are the dominant cell type in tumor and normal samples. Tumor-associated macrophages have strong plasticity and, if reprogrammed, can clear tumor cells and regulate the adaptive immune system for cancer immunotherapy. These studies suggested that the five cell subpopulations obtained from our analysis are potentially valuable in identifying tumors from normal tissue and in tumor treatment.

First, pseudotime was used to measure the degree of cell differentiation (Figure 2C). Next, we show the trajectory of differentiation of the five subpopulations (Figure 2E). The five subpopulations could be divided into five branches and states (Figure 2D).

The seraut_cluster of the four subgroup trajectory diagrams. Figure 2F shows that lung epithelial cells were mainly in the 0 and 16 subgroups. In the trajectory diagram, these two subgroups are on the branches of state 4 and state 5. In these subgroups, macrostat1 and macrostat15 were mainly enriched. Neuroendocrine cells were mainly enriched in subgroup 22, which was mainly on the branch of State 2 in the trajectory diagram. Secret cells were mainly enriched in subgroups 2, 5, 6, 12 and 21. In the trajectory diagram, 2 and 12 are mainly on the branches of state 4, states 5, 5 and 6 are mainly on the branches of state 1 and state 2, and 21 is mainly on the branch of state 2. T cells were mainly enriched in subgroup 13. In the trajectory diagram, 13 is mainly on the State2 branch (Supplementary Figure S4A).

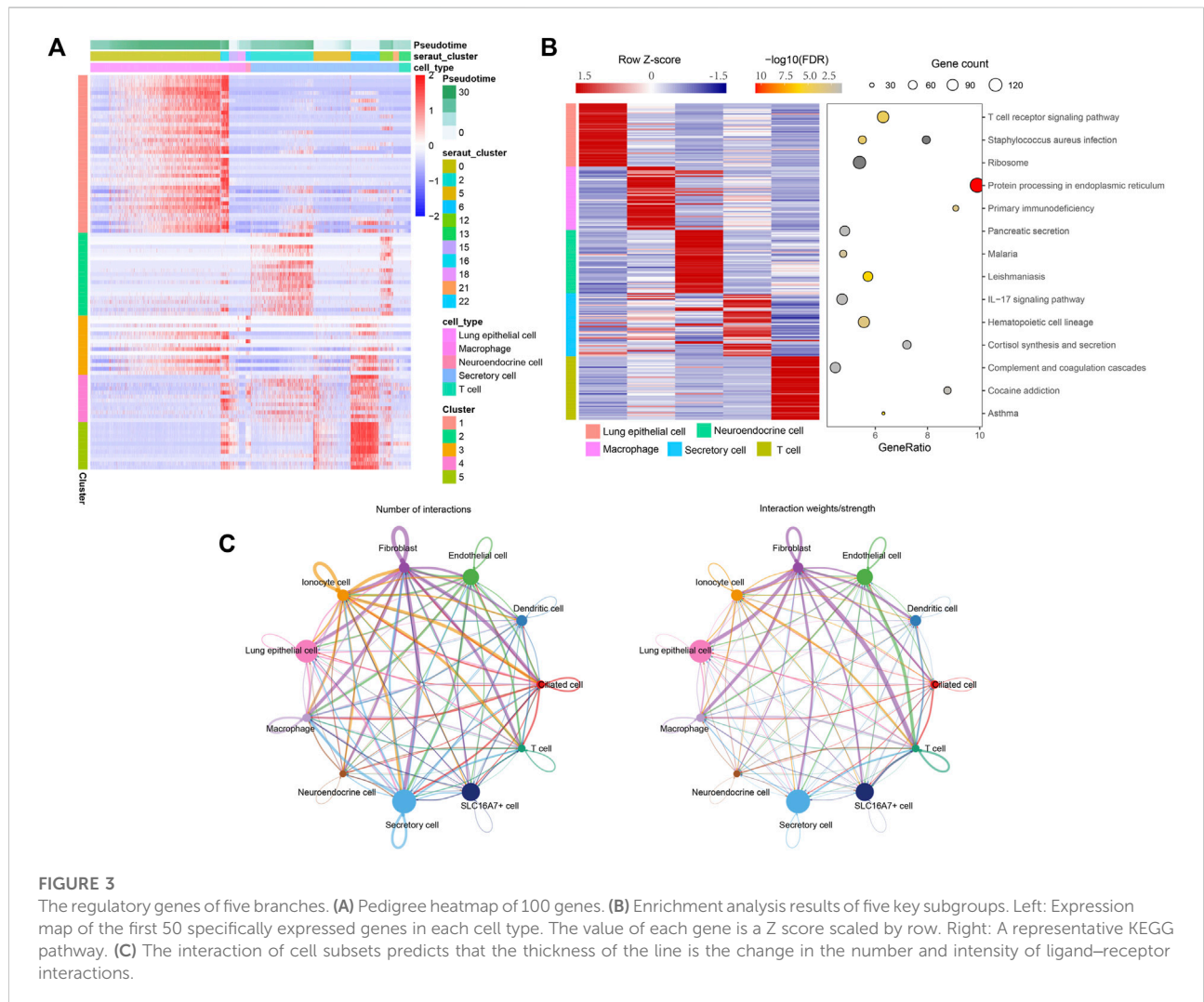
Then, we used branched expression analysis modelling (BEAM) to find the regulatory genes of five branches, screened the key genes with qval (corrected P), and screened the 100 genes with the smallest qval. Therefore, we used these 100 genes to draw the pedigree heatmap (of which 100 genes are listed as cells) (Figure 3A). Information on these 100 genes is shown in Supplementary Table S3.

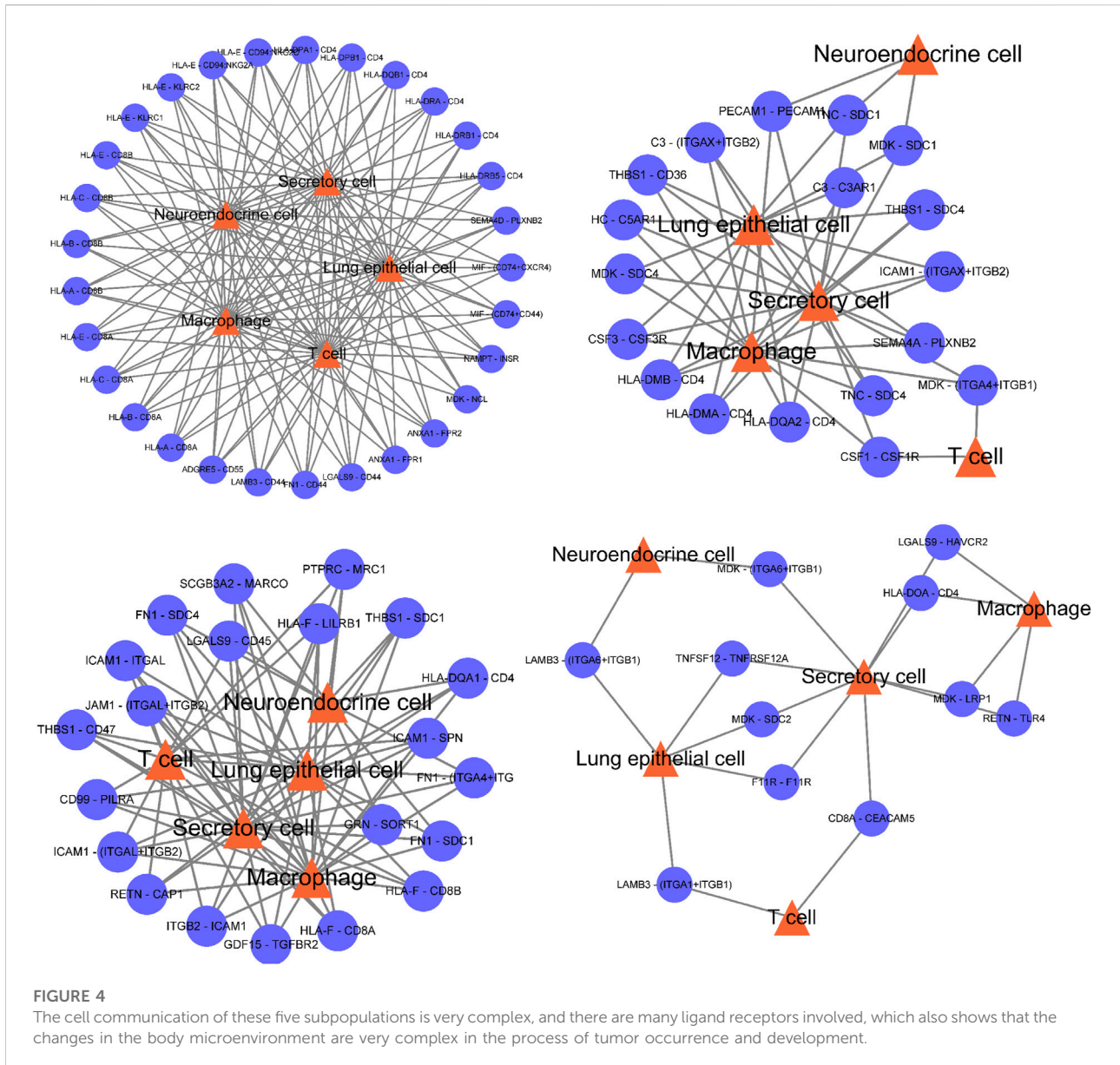
Pathway enrichment of subpopulations

To further study the functions of these five subpopulations, we extracted the marker genes of these five subpopulations, conducted KEGG enrichment analysis through the webgestalt package, and screened the key pathways with FDR <0.05. The enrichment results are shown in Supplementary Table S4. The first three pathways were screened by enrichment ratio, the first 50 marker genes were screened for each subgroup, and the expression heatmap was drawn (Figure 3B).

TABLE2 Cell subpopulation statistics.

cell_name	T_celltype	T_no_celltype	N_celltype	N_no_celltype	p.val	fc
Ciliated cell	94	3,439	134	4,503	0.542639269	0.918528902
Dendritic cell	316	3,217	188	4,449	1.87E-19	2.324559025
Endothelial cell	232	3,301	842	3,795	9.42E-57	0.316768618
Fibroblast	275	3,258	175	4,462	5.70E-15	2.152152942
Ionocyte cell	229	3,304	331	4,306	0.250804955	0.901657242
Lung epithelial cell	119	3,414	1,647	2,990	7.36318877704002e-318	0.063279208
Macrophage	30	3,503	183	4,454	3.41E-20	0.208439604
Neuroendocrine cell	60	3,473	6	4,631	4.15E-16	13.33429312
Secretory cell	1,466	2067	424	4,213	4.41E-264	7.047240555
SLC16A7+ cell	694	2,839	575	4,062	5.17E-19	1.726895876
T cell	18	3,515	132	4,505	9.57E-17	0.174770464





Communication analysis of cell subsets

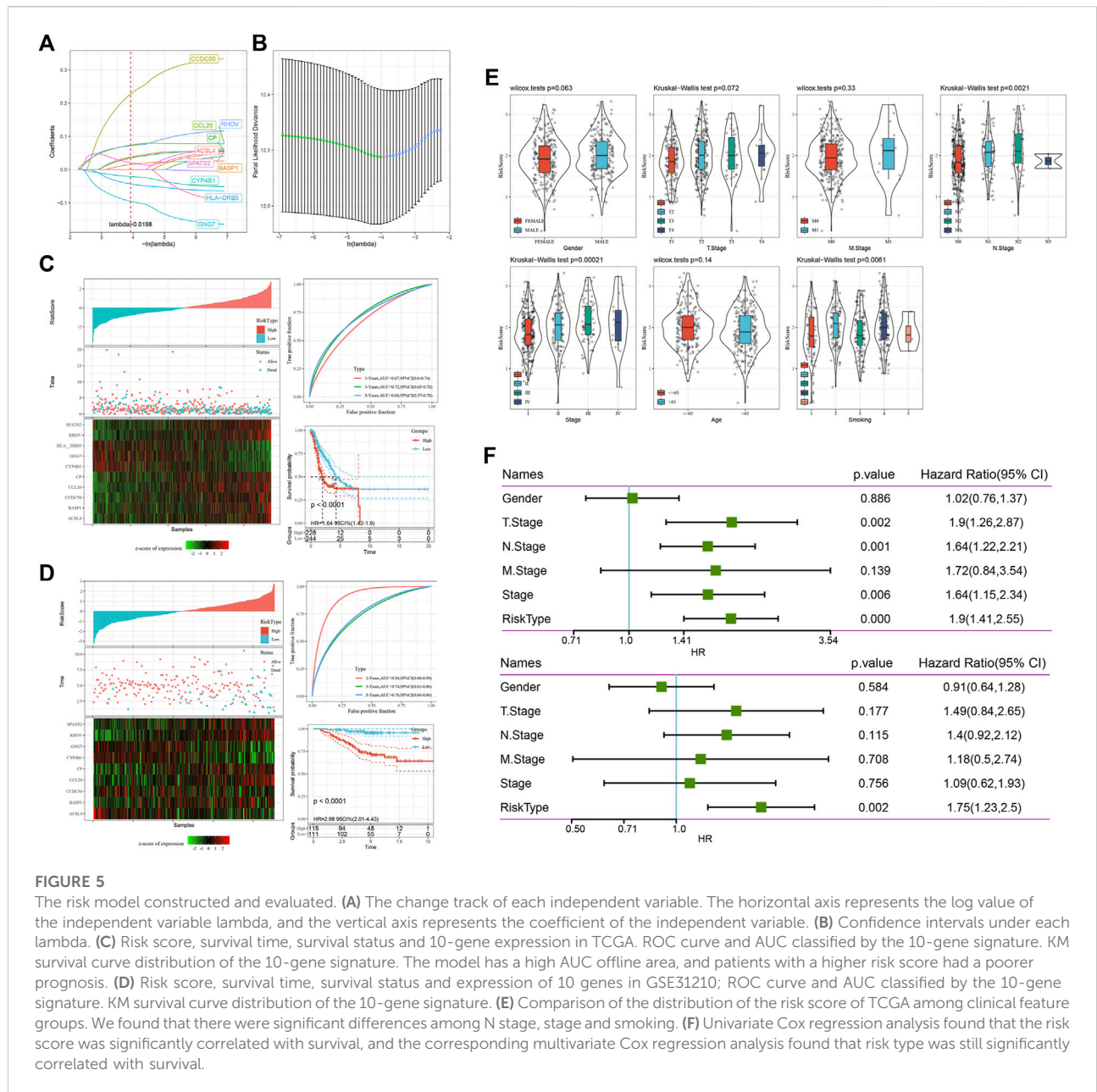
Cell communication of 11-cell subsets was analysed by CellChart. [Supplementary Table S5](#) shows the results of the cell communication analysis. From [Figure 3C](#), it can be seen that the interaction of these 11 subsets changes in the number and intensity of ligand–receptor interactions.

The predicted ligand receptor interactions of the five important subpopulations screened above were used to draw the interaction network between cell subpopulations and ligand receptors. [Figure 4](#) shows that the cell communication of these five subpopulations is very complex, and many ligand receptors are involved, which also shows that the changes in the body

microenvironment are very complex in the process of tumor occurrence and development. These five subpopulations may all play a very important role in tumorigenesis and development.

Screening of key genes and construction and validation of prognostic risk model

The expression profile data of FPKM of TCGA were downloaded and further converted into TPM. First, we filtered through the standard deviation of the expression of each gene in all samples greater than 0.5, used the limma package to analyse the difference in the expression profile



matrix of LUAD, and screened the differentially expressed genes with $|\log_2 FC| > 1$ and $FDR < 0.05$. A total of 2,812 differentially expressed genes were screened, of which 1,110 genes were upregulated and 1,702 genes were downregulated. [Supplementary Table S6](#) shows the results of all gene difference analyses. Through overlap analysis, we found that 462 differentially expressed genes were marker genes of these five subgroups ([Supplementary Figure S4B](#)).

Using the expression profile data of TCGA, for the relevant genes and survival data, the R package survival coxph was used to carry out the univariate Cox proportional hazards regression model, and $p < 0.01$ was selected as the threshold for filtering.

Finally, there were 16 genes with differences. The univariate Cox analysis results are shown in [Supplementary Table S7](#).

At present, 16 genes related to prognosis in TCGA have been identified, but the large number of these genes is not conducive to clinical detection, so we need to further narrow the range of immune genes under the condition of maintaining high accuracy. We further compressed these 16 genes using lasso regression to reduce the number of genes in the risk model. First, we analysed the change trajectory of each independent variable. It can be seen that with the gradual increase in lambda, the number of independent variable coefficients tending to 0 also gradually increases. We used 10-fold cross validation to build the

model. Analyse the confidence interval under each lambda. Figure 5A,B shows that when lambda = 0.0198, the model reached the optimum. Therefore, we selected 10 genes when lambda = 0.0198 as the target genes in the next step.

Finally, we obtained 10 genes: *CCL20*, *CP*, *HLA-DRB5*, *RHOV*, *CYP4B1*, *BASP1*, *ACSL4*, *GNG7*, *CCDC50*, and *SPATS2*. The final 10-gene signature formula is as follows:

$$\begin{aligned} \text{RiskScore} = & 0.089 * \text{CCL20} + 0.054 * \text{CP} - 0.058 * \text{HLA} - \text{DRB5} \\ & + 0.084 * \text{RHOV} - 0.029 * \text{CYP4B1} + 0.027 * \text{BASP1} \\ & - 0.055 * \text{ACSL4} - 0.118 * \text{GNG7} + 0.306 * \text{CCDC50} \\ & + 0.002 * \text{SPATS2} \end{aligned}$$

We calculated the risk score of each sample according to the expression level of the sample and drew the risk score distribution of the sample (Figure 5C). We used the R software package *timeroc* to analyse the ROC curve of the prognostic classification of the risk score. We analysed the classification efficiency of prognosis prediction at 1, 3, and 5 years, from which we can see that the model has a high AUC offline area. Finally, we calculated the z score for the risk score, divided the samples with risk scores greater than zero into a high-risk group and a low-risk group, and drew Kaplan–Meier survival curves, from which we can see that patients with a higher risk score had a poorer prognosis ($p < 0.0001$).

To better evaluate the risk model constructed in this study, we used GSE31210 for verification. We calculated the risk score of each sample according to the expression level of the sample and drew the risk score distribution of the sample (Figure 5D). Similarly, we used the R software package *timeROC* to analyse the ROC of prognosis classification of risk score. We analysed the classification efficiency of prognosis prediction at 1, 3, and 5 years, from which we can see that the model has a high AUC offline area. Finally, we calculated the z score for the risk score, divided the samples with risk scores greater than zero into a high-risk group and a low-risk group, and drew a KM curve, from which we can see that there was a very significant difference between them ($p < 0.0001$).

The risk score suggests that LUAD is related to smoking

By comparing the distribution of the risk score of TCGA among clinical feature groups, we found that there were significant differences among N stage, stage and smoking ($p < 0.05$) (Figure 5E).

Univariate and multivariate analysis of the 10-gene signature

To identify the independence of the 10-gene signature risk model in clinical application, we used univariate and multivariate

Cox regression to analyse the relevant HR, 95% CI of HR and p value in the clinical information carried by all TCGA data. We systematically analysed the clinical information recorded by TCGA patients, including sex, stage and risk type. In the TCGA datasets, univariate Cox regression analysis found that the risk score was significantly correlated with survival, and the corresponding multivariate Cox regression analysis found that risk type (HR = 1.75, 95% CI = 1.23–2.5, $p < 0.05$) was still significantly correlated with survival (Figure 5F). The above situation shows that our model 10-gene signature risk model has good prediction performance in clinical application value.

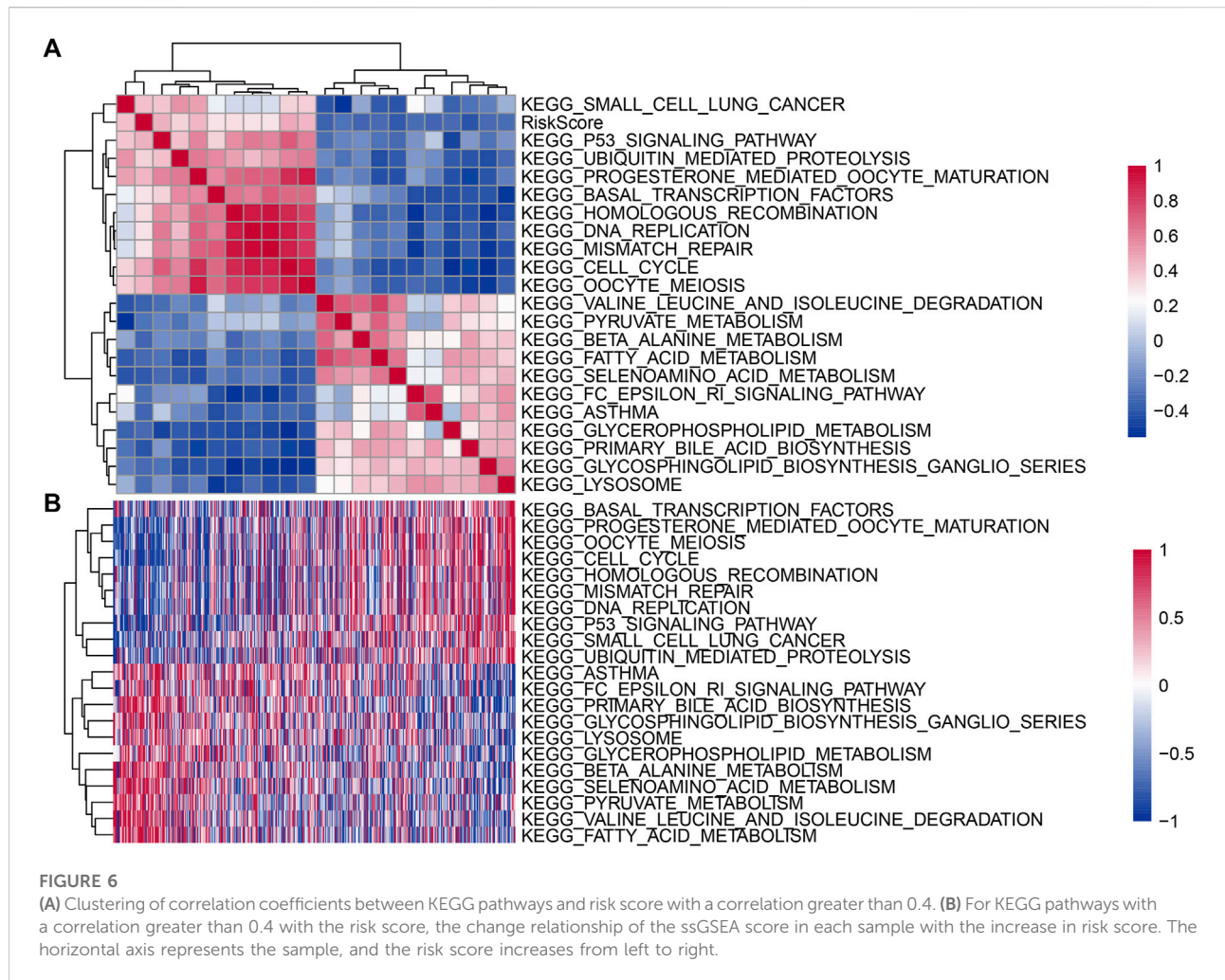
Relationship between risk score and channel

To further observe the relationship between the risk scores of different samples and biological functions, we used the R software package *GSVA* for single-sample GSEA of the expression profile corresponding to TCGA samples, calculated the scores of each sample on different functions, obtained the ssGSEA score of each sample corresponding to each function (Supplementary Table S8), further calculated the correlation between these functions and the risk score (Supplementary Table S9), and selected the function with a correlation no less than 0.3, as shown in Figure 6A. It can be seen that 11 are negatively correlated with the sample risk score, and 10 channels are positively correlated with the sample risk score. KEGG P53 SIGNALING PATHWAY, KEGG_CELL CYCLE and KEGG_OOCYTE MEIOSIS were positively correlated with the risk score. KEGG_VALINE LEUCINE AND ISOLEUCINE DEGRADATION and KEGG FATTY ACID METABOLISM were negatively correlated with the risk score. Twenty-two KEGG pathways with correlations no less than 0.3 were identified, and cluster analysis was performed according to their enrichment scores (Figure 6B).

Expression of the unreported signature genes HLA-DRB5 and CCDC50 in LUAD

CCL20, *CP*, *RHOV*, *CYP4B1*, *BASP1*, *ACSL4*, *GNG7* and *SPATS2* have been reported, and their dysregulation is associated with the prognosis of LUAD. However, the expression and function of *HLA-DRB5* and *CCDC50* have not yet been reported.

We applied immunohistochemistry and qRT–PCR to detect the differences in the expression of *HLA-DRB5* and *CCDC50* between paired tumor tissues and normal tissues. The qRT–PCR results revealed that the levels of *HLA-DRB5* were lower and the levels of *CCDC50* were higher in five high-risk tumor tissues (Figure 7A). The protein (Figure 7C) levels of *HLA-DRB5* were lower and the levels of *CCDC50* (Figures 7B,D) were higher in high-risk tumor tissues.



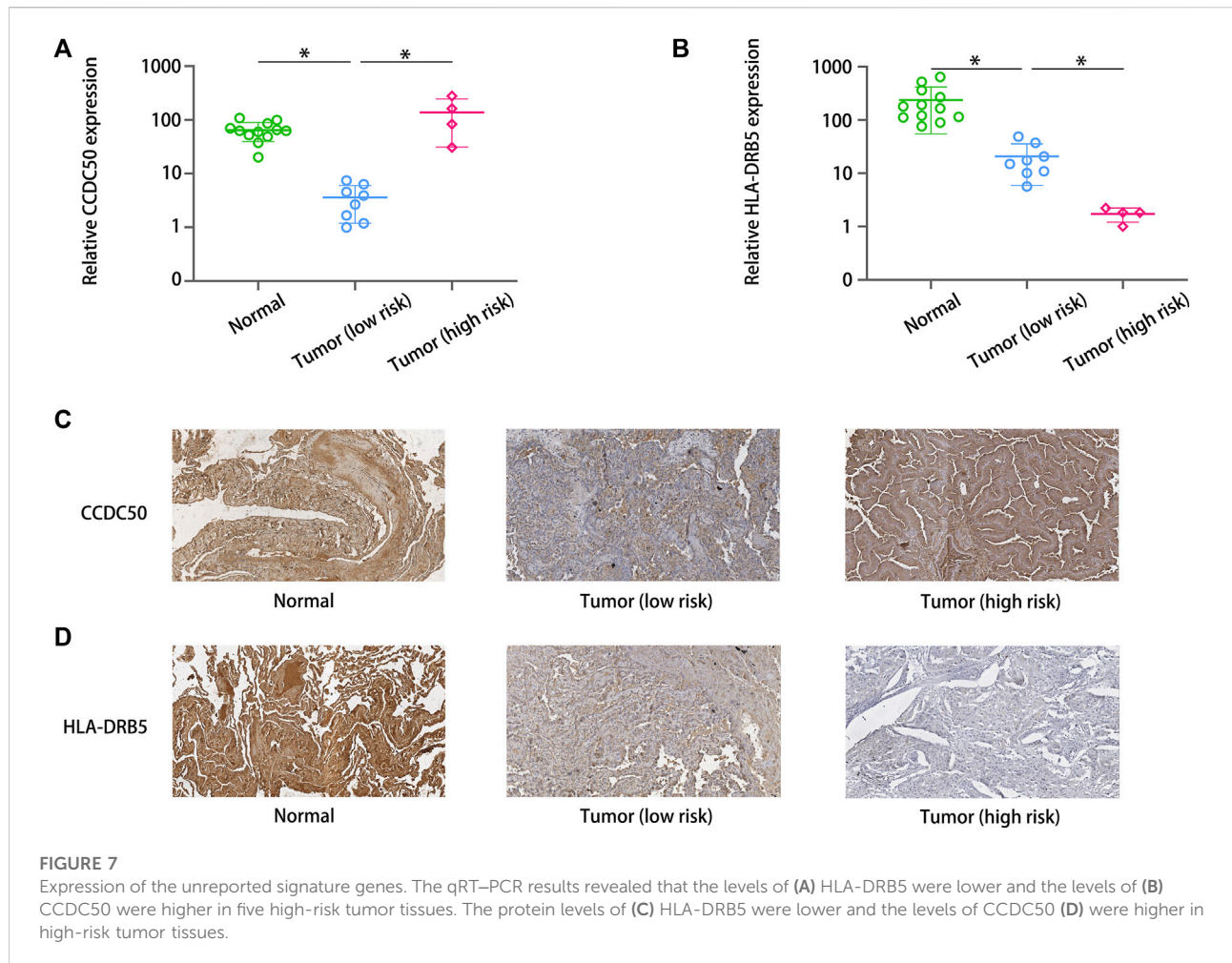
Discussion

At present, most studies on LUAD focus on the transcriptome level, and there are few studies on the single-cell level and tumor microenvironment (Guo et al., 2018; He et al., 2021; Spella and Stathopoulos, 2021). The heterogeneity between and within tumors is closely related to tumor progression and metastasis and will affect the response to targeted therapy and the final survival results (Chen Z. et al., 2021; Spella and Stathopoulos, 2021). Therefore, it is necessary to screen marker genes related to prognosis at the single-cell level of LUAD and construct a risk model accordingly.

First, we obtained bulkRNA-seq data for LUAD samples based on the TCGA database and downloaded the single-cell sequencing data GSE149655 from the GEO database. A total of four samples were detected, including two LUAD samples and two normal samples. The single-cell dataset GSE149655 was subjected to quality control, filtration and dimensionality reduction through the seraut package, and finally, 23 subsets

were screened. Next, 23 subpopulations were annotated by the marker gene of cellmarkers, and a total of 11-cell types were annotated by 23 subpopulations. Through statistical analysis of 11 subpopulations of LUAD and normal samples, five important subpopulations were selected, namely, lung epithelial cells, macrophages, neuroendocrine cells, secret cells and T cells. Next, we screened the differentially expressed genes in TCGA-LUAD, constructed a prognostic risk model based on key genes by using univariate risk analysis and multivariate risk analysis, and verified it with the independent GSE31210 dataset. Finally, we verified the genes in the model through experiments. The above situation shows that our model 10 gene signature has good prediction performance in clinical application value.

A growing number of studies have shown that cancer usually becomes more heterogeneous in the process of disease. Due to this heterogeneity, large tumors may include a variety of cell collections, which have different molecular characteristics and different sensitivities to treatment. This



heterogeneity may lead to an uneven distribution of tumor cell subsets with different genes between and within the disease site (spatial heterogeneity) or temporal changes in the molecular composition of cancer cells (temporal heterogeneity) (Dagogo-Jack and Shaw, 2018). ScRNA-seq can reveal the expression of all genes in the whole genome at the single-cell level and can study cell heterogeneity more intuitively (Navin et al., 2011; Francis et al., 2014). We screened the cell types with significant differences in subpopulation abundance between LUAD and normal tissues through single-cell analysis, screened the cell types with different subpopulation abundance during the occurrence and development of LUAD at the single-cell level, screened the marker genes of these key cell types, combined with TCGA data, screened the marker genes related to prognosis, and constructed the risk model accordingly.

The model we constructed includes 10 genes, including *CCL20*, *CP*, *HLA-DRB5*, *RHOV*, *CYP4B1*, *BASP1*, *ACSL4*, *GNG7*, *CCDC50*, and *SPATS2*. *CCL20* is a member of the chemokine family (Chen et al., 2020). Recent studies have

shown that high levels of *CCL20* are associated with malignancies of various cancers (Kapur et al., 2016; Zhang et al., 2016). *CCL20* can also recruit immune cells, such as DCs and Tregs, which further connect *CCL20* with the tumor microenvironment. *CCL20* upregulation can recruit CD8⁺ T cells to the immune microenvironment of LUAD, which is helpful for immunotherapy (Luo et al., 2017; Lyu et al., 2019; Ma et al., 2022). *CP* (ceruloplasmin) is a multi copper oxidase and a mammalian plasma ferrous oxidase (Hellman and Gitlin, 2002). Recent evidence suggests that ceruloplasmin is also associated with tumor development and progression. The expression of plasma ceruloplasmin in LUAD is significantly upregulated and significantly correlated with clinicopathological stage (Matsuoka et al., 2018). The expression of plasma ceruloplasmin was also significantly upregulated in high-grade clear cell renal cell carcinoma samples (Thibodeau et al., 2016). *HLA-DRB5*, whose expression products play a central role in the immune system by presenting peptides derived from extracellular proteins (Su et al., 2021). Studies have shown

that HLA-DRB5 is associated with risk factors for cervical cancer (Bao et al., 2018). HLA-DRB5 was expressed at low levels in all patients with multiple myeloma, in a subgroup of patients with ulcerative mucositis and in a control group (Marcussen et al., 2017). RHOV has been shown to promote cell differentiation and act as an important regulator of neural crest induction (Guémar et al., 2007; Song et al., 2015). RHOV is highly expressed in many lung cancer cell lines and promotes the growth and metastasis of LUAD cells (Zhang Y. et al., 2020; Chen H. et al., 2021; Zhang et al., 2021). CYP4B1 belongs to the mammalian CYP4 enzyme family and is mainly expressed in human lungs (Wiek et al., 2015). Studies have suggested that CYP4B1 is a prognostic biomarker and potential therapeutic target of LUAD and can also be used as a target of cancer treatment (Lim et al., 2020; Liu et al., 2021). BASP1 can regulate many types of cell biological behavior, including proliferation, apoptosis and differentiation (Sanchez-Niño et al., 2010; Tang et al., 2017). High expression of BASP1 is associated with poor prognosis of human LUAD and head and neck squamous cell carcinoma and promotes tumor progression (Jaikumarr Ram et al., 2020; Wang et al., 2021). ACSL4 is mainly located in mitochondria, peroxisomes and the endoplasmic reticulum and plays a crucial regulatory role in ferroptosis (Quan et al., 2021). In most cases, ACSL4 plays a carcinogenic role. The high expression of ACSL4 indicates that the prognosis of patients with ovarian cancer is poor. In LUAD, ACSL4 plays a tumor suppressor role by inhibiting tumor survival/invasion and promoting ferroptosis (Ye et al., 2016; Ma et al., 2021; Yang et al., 2021). GNG7 belongs to the large G protein γ family (Shibata et al., 1998). Many studies have shown that GNG7 is a tumor suppressor gene in squamous cell carcinoma, pancreatic cancer, esophageal cancer, gastrointestinal cancer and clear cell renal cell carcinoma. In LUAD, GNG7 is significantly downregulated in LUAD tissues and cell lines. Low expression of GNG7 is related to poor prognosis in LUAD patients, and GNG7 overexpression inhibits the proliferation and invasion of LUAD cells. (Shibata et al., 1999; Demokan et al., 2013; Liu et al., 2019; Xu et al., 2019; Fang et al., 2022). CCDC50 is a tyrosine phosphorylated protein that mediates apoptosis through the NF- κ B pathway (Bohgaki et al., 2008). However, research on CCDC50 in cancer is still insufficient. Some studies have shown that different splice variants of CCDC50 play opposite tumorigenic roles *in vitro* and *in vivo*. CCDC50-S promotes the metastasis of renal clear cell carcinoma, but CCDC50-FL and sh-CCDC50 inhibit the metastasis of renal clear cell carcinoma (Sun et al., 2020). SPATS2 is a cytoplasmic RNA-binding protein that plays an important role in spermatogenesis (Fang et al., 2022). In recent studies, the expression of SPATS2 was upregulated in HCC tissues. High expression of SPATS2 was associated with poor clinicopathological features and poor prognosis in HCC

patients. SPATS2 knockdown significantly inhibited the growth and invasion of HCC cells and promoted apoptosis and G1 arrest of HCC cells *in vitro* (Senoo et al., 2002). SPATS2 is also highly expressed in liver cancer and may be a new diagnostic and prognostic biomarker of liver cancer. In recent studies, SPATS2 has also been used as a diagnostic biomarker of LUAD (Dong et al., 2020; Xing et al., 2020). The above reports of gene dysregulation associated with LUAD are consistent with our risk gene prediction results, which showed that the 10-gene signature can be used as an effective prognostic tool for LUAD patients. However, there are still some deficiencies: 1. We need to use more clinical samples for further verification in the follow-up. The biological functions of newly discovered HLA-DRB5 and CCDC50 risk genes in lung cancer were further explored.

Conclusion

By analysing the single-cell sequencing data of LUAD, we established a 10-gene signature related to the prognosis of LUAD. This 10-gene signature has strong robustness and can achieve stable prediction efficiency in datasets from different platforms. We also performed qPCR and immunohistochemical sample verification on CCDC50 and HLA-DRB5, two genes that have not been verified in LUAD. The results are consistent with our prediction. These findings will contribute to a more accurate diagnosis of LUAD, which is very important for the precise treatment of LUAD.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/[Supplementary Material](#).

Ethics statement

The studies involving human participants were reviewed and approved by this study was approved by the Ethics Committee of Shanghai Pulmonary Hospital (K20-148Y). The patients/participants provided their written informed consent to participate in this study.

Author contributions

NS designed this study. YX drafted the manuscript. TX and YW performed the statistical analysis and contributed equally to the article. LL revised this manuscript.

Funding

This work was supported by Natural Science Foundation of Shanghai (22YF1437400).

Acknowledgments

We would like to thank the TCGA and GEO databases for the availability of the data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.975542/full#supplementary-material>

References

- Bao, X., Hanson, A. L., Madeleine, M. M., Wang, S. S., Schwartz, S. M., Newell, F., et al. (2018). HLA and KIR associations of cervical neoplasia. *J. Infect. Dis.* 218 (12), 2006–2015. doi:10.1093/infdis/jiy483
- Barta, J. A., Powell, C. A., and Wisnivesky, J. P. (2019). Global epidemiology of lung cancer. *Ann. Glob. Health* 85 (1), 8. doi:10.5334/aogh.2419
- Bohgaki, M., Tsukiyama, T., Nakajima, A., Maruyama, S., Watanabe, M., Koike, T., et al. (2008). Involvement of Ymer in suppression of NF-kappaB activation by regulated interaction with lysine-63-linked polyubiquitin chain. *Biochim. Biophys. Acta* 1783, 826–837. doi:10.1016/j.bbamcr.2007.09.006
- Chen, W., Qin, Y., and Liu, S. (2020). CCL20 signaling in the tumor microenvironment. *Adv. Exp. Med. Biol.* 1231, 53–65. doi:10.1007/978-3-030-36667-4_6
- Chen, H., Xia, R., Jiang, L., Zhou, Y., Xu, H., Peng, W., et al. (2021). Overexpression of RhoV promotes the progression and EGFR-TKI resistance of lung adenocarcinoma. *Front. Oncol.* 11, 619013. Published 2021 Mar 9. doi:10.3389/fgene.2021.619013
- Chen, Z., Huang, Y., Hu, Z., Zhao, M., Li, M., Bi, G., et al. (2021). Landscape and dynamics of single tumor and immune cells in early and advanced-stage lung adenocarcinoma. *Clin. Transl. Med.* 11 (3), e350. doi:10.1002/ctm2.350
- Choi, C. M., Kim, H. C., Jung, C. Y., Cho, D. G., Jeon, J. H., Lee, J. E., et al. (2019). Report of the Korean association of lung cancer registry (KALC-R), 2014. *Cancer Res. Treat.* 51 (4), 1400–1410. doi:10.4143/crt.2018.704
- Dagogo-Jack, I., and Shaw, A. T. (2018). Tumour heterogeneity and resistance to cancer therapies. *Nat. Rev. Clin. Oncol.* 15 (2), 81–94. doi:10.1038/nrclinonc.2017.166
- Demokan, S., Chuang, A. Y., Chang, X., Khan, T., Smith, I. M., Pattani, K. M., et al. (2013). Identification of guanine nucleotide-binding protein γ -7 as an epigenetically silenced gene in head and neck cancer by gene expression profiling. *Int. J. Oncol.* 42 (4), 1427–1436. doi:10.3892/ijo.2013.1808
- Dong, G., Zhang, S., Shen, S., Sun, L., Wang, X., Wang, H., et al. (2020). SPATS2, negatively regulated by miR-145-5p, promotes hepatocellular carcinoma progression through regulating cell cycle. *Cell Death Dis.* 11 (10), 837. doi:10.1038/s41419-020-03039-y
- Fang, C., Zhong, R., Qiu, C., and Zou, B. B. (2022). The prognostic value of GNG7 in colorectal cancer and its relationship with immune infiltration. *Front. Genet.* 13, 833013. PMID: 35281820; PMCID: PMC8906903. doi:10.3389/fgene.2022.833013
- Francis, J. M., Zhang, C. Z., Maire, C. L., Jung, J., Manzo, V. E., Adalsteinsson, V. A., et al. (2014). EGFR variant heterogeneity in glioblastoma resolved through single-nucleus sequencing. *Cancer Discov.* 4, 956–971. doi:10.1158/2159-8290.CD-13-0879
- Guémar, L., de Santa Barbara, P., Vignal, E., Maurel, B., Fort, P., and Faure, S. (2007). The small GTPase RhoV is an essential regulator of neural crest induction in *Xenopus*. *Dev. Biol.* 310, 113–128. doi:10.1016/j.ydbio.2007.07.031

SUPPLEMENTARY FIGURE S1

Quality control chart (before filtration) of LUAD single-cell subsets.

SUPPLEMENTARY FIGURE S2

Quality control chart (after filtration) of LUAD single-cell subsets.

SUPPLEMENTARY FIGURE S3

Number of cells in each sample before (A) and (B) after filtration. (C) PCA dimensionality reduction (left) shows the distribution of hypervariable genes and nonhypervariable genes, and the top 20 hypervariable genes are shown on the right.

SUPPLEMENTARY FIGURE S4

(A) PCA dimensionality reduction to find anchor points. In the trajectory diagram, 13 is mainly on the State2 branch. (B) Among the 2812 differentially expressed genes, 462 were marker genes of these four subgroups by overlap analysis.

SUPPLEMENTARY TABLE S1

The results of marker genes.

SUPPLEMENTARY TABLE S2

Cell marker genes.

SUPPLEMENTARY TABLE S3

Information on these 100 genes.

SUPPLEMENTARY TABLE S4

The enrichment results of KEGG enrichment.

SUPPLEMENTARY TABLE S5

The results of cell communication analysis.

SUPPLEMENTARY TABLE S6

The results of all gene difference analyses.

SUPPLEMENTARY TABLE S7

The univariate Cox analysis results.

SUPPLEMENTARY TABLE S8

The ssGSEA score of each corresponding sample.

SUPPLEMENTARY TABLE S9

The correlation between these functions and the risk score.

- Guo, X., Zhang, Y., Zheng, L., Zheng, C., Song, J., Zhang, Q., et al. (2018). Global characterization of T cells in non-small cell lung cancer by single-cell sequencing. *Nat. Med.* 24, 978–985. doi:10.1038/s41591-018-0045-3
- He, D., Wang, D., Lu, P., Yang, N., Xue, Z., Zhu, X., et al. (2021). Single-cell RNA sequencing reveals heterogeneous tumor and immune cell populations in early-stage lung adenocarcinomas harboring EGFR mutations. *Oncogene* 40 (2), 355–368. doi:10.1038/s41388-020-01528-0
- Hellman, N. E., and Gitlin, J. D. (2002). Ceruloplasmin metabolism and function. *Annu. Rev. Nutr.* 22, 439–458. doi:10.1146/annurev.nutr.22.012502.114457
- Houston, K. A., Henley, S. J., Li, J., White, M. C., and Richards, T. B. (2014). Patterns in lung cancer incidence rates and trends by histologic type in the United States, 2004–2009. *Lung Cancer* 86 (1), 22–28. doi:10.1016/j.lungcan.2014.08.001
- Jaikumarr Ram, A., Girija As, S., Jayaseelan, V. P., and Arumugam, P. (2020). Overexpression of BASP1 indicates a poor prognosis in head and neck squamous cell carcinoma. *Asian Pac. J. Cancer Prev.* 21 (11), 3435–3439. Published 2020 Nov 1. doi:10.31557/APJCP.2020.21.11.3435
- Kapur, N., Mir, H., Clark, C. E., Iii, Krishnamurti, U., Beech, D. J., Lillard, J. W., et al. (2016). CCR6 expression in colon cancer is associated with advanced disease and supports epithelial-to-mesenchymal transition. *Br. J. Cancer* 114 (12), 1343–1351. doi:10.1038/bjc.2016.113
- Kenfield, S. A., Wei, E. K., Stampfer, M. J., Rosner, B. A., and Colditz, G. A. (2008). Comparison of aspects of smoking among the four histological types of lung cancer. *Tob. Control* 17, 198–204. doi:10.1136/tc.2007.022582
- Lavin, Y., Kobayashi, S., Leader, A., Amir, E. D., Elefant, N., Bigenwald, C., et al. (2017). Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses. *Cell* 169, 750–765.e17. doi:10.1016/j.cell.2017.04.014
- Liang, L., Yu, J., Li, J., Li, N., Liu, J., Xiu, L., et al. (2021). Integration of scRNA-seq and bulk RNA-seq to analyse the heterogeneity of ovarian cancer immune cells and establish a molecular risk model. *Front. Oncol.* 11, 711020. doi:10.3389/fonc.2021.711020
- Lim, S., Alshagga, M., Ong, C. E., Chieng, J. Y., and Pan, Y. (2020). Cytochrome P450 4B1 (CYP4B1) as a target in cancer treatment. *Hum. Exp. Toxicol.* 39 (6), 785–796. doi:10.1177/0960327120905959
- Liu, L., Wang, S., Cen, C., Peng, S., Chen, Y., Li, X., et al. (2019). Identification of differentially expressed genes in pancreatic ductal adenocarcinoma and normal pancreatic tissues based on microarray datasets. *Mol. Med. Rep.* 20 (2), 1901–1914. doi:10.3892/mmr.2019.10414
- Liu, X., Jia, Y., Shi, C., Kong, D., Wu, Y., Zhang, T., et al. (2021). CYP4B1 is a prognostic biomarker and potential therapeutic target in lung adenocarcinoma. *PLoS One* 16 (2), e0247020. doi:10.1371/journal.pone.0247020
- Luo, K., Zavala, F., Gordy, J., Zhang, H., and Markham, R. B. (2017). Extended protection capabilities of an immature dendritic-cell targeting malaria sporozoite vaccine. *Vaccine* 35 (18), 2358–2364. doi:10.1016/j.vaccine.2017.03.052
- Lyu, M., Li, Y., Hao, Y., Lyu, C., Huang, Y., Sun, B., et al. (2019). CCR6 defines a subset of activated memory T cells of Th17 potential in immune thrombocytopenia. *Clin. Exp. Immunol.* 195 (3), 345–357. doi:10.1111/cei.13233
- Ma, L. L., Liang, L., Zhou, D., and Wang, S. W. (2021). Tumor suppressor miR-424-5p abrogates ferroptosis in ovarian cancer through targeting ACSL4. *Neoplasma* 68, 165–173. doi:10.4149/neo_2020_200707n705
- Ma, C., Li, F., Wang, Z., and Luo, H. (2022). A novel immune-related gene signature predicts prognosis of lung adenocarcinoma. *Biomed. Res. Int.* 2022, 4995874. PMID: 35437508; PMCID: PMC9013292. doi:10.1155/2022/4995874
- Marcussen, M., Bødker, J. S., Christensen, H. S., Johansen, P., Nielsen, S., Christiansen, I., et al. (2017). Molecular characteristics of high-dose melphalan associated oral mucositis in patients with multiple myeloma: A gene expression study on human mucosa. *PLoS One* 12 (1), e0169286. doi:10.1371/journal.pone.0169286
- Matsuoka, R., Shiba-Ishii, A., Nakano, N., Togayachi, A., Sakashita, S., Sato, Y., et al. (2018). Heterotopic production of ceruloplasmin by lung adenocarcinoma is significantly correlated with prognosis. *Lung Cancer* 118, 97–104. doi:10.1016/j.lungcan.2018.01.012
- Moncada, R., Barkley, D., Wagner, F., Chiodin, M., Devlin, J. C., Baron, M., et al. (2020). Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.* 38 (3), 333–342. doi:10.1038/s41587-019-0392-8
- Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., et al. (2011). Tumour evolution inferred by single-cell sequencing. *Nature* 472, 90–94. doi:10.1038/nature09807
- Quan, J., Bode, A. M., and Luo, X. (2021). ACSL family: The regulatory mechanisms and therapeutic implications in cancer. *Eur. J. Pharmacol.* 909, 174397. doi:10.1016/j.ejphar.2021.174397
- Remen, T., Pintos, J., Abrahamowicz, M., and Siemiatycki, J. (2018). Risk of lung cancer in relation to various metrics of smoking history: A case-control study in montreal. *BMC Cancer* 18 (1), 1275. doi:10.1186/s12885-018-5144-5
- Sanchez-Niño, M. D., Sanz, A. B., Lorz, C., Gnirke, A., Rastaldi, M. P., Nair, V., et al. (2010). BASP1 promotes apoptosis in diabetic nephropathy. *J. Am. Soc. Nephrol.* 21 (4), 610–621. doi:10.1681/ASN.2009020227
- Senoo, M., Hoshino, S., Mochida, N., Matsumura, Y., and Habu, S. (2002). Identification of a novel protein p59(scr), which is expressed at specific stages of mouse spermatogenesis. *Biochem. Biophys. Res. Commun.* 292, 992–998. doi:10.1006/bbrc.2002.6769
- Shibata, K., Mori, M., Tanaka, S., Kitano, S., and Akiyoshi, T. (1998). Identification and cloning of human G-protein gamma 7, downregulated in pancreatic cancer. *Biochem. Biophys. Res. Commun.* 246 (1), 205–209. doi:10.1006/bbrc.1998.8581
- Shibata, K., Tanaka, S., Shiraishi, T., Kitano, S., and Mori, M. (1999). G-protein gamma 7 is downregulated in cancers and associated with p 27kip1-induced growth arrest. *Cancer Res.* 59 (5), 1096–1101.
- Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer statistics, 2021. *Ca. Cancer J. Clin.* 71 (1), 7–33. doi:10.3322/caac.21654
- Song, R., Liu, X., Zhu, J., Gao, Q., Wang, Q., Zhang, J., et al. (2015). RhoV mediates apoptosis of RAW264.7 macrophages caused by osteoclast differentiation. *Mol. Med. Rep.* 11, 1153–1159. doi:10.3892/mmr.2014.2817
- Spella, M., and Stathopoulos, G. T. (2021). Immune resistance in lung adenocarcinoma. *Cancers (Basel)* 13 (3), 384. doi:10.3390/cancers13030384
- Su, W. M., Gu, X. J., Hou, Y. B., Zhang, L. Y., Bei Cao, B., Ou, R. W., et al. (2021). Association analysis of WNT3, HLA-DRB5 and IL1R2 polymorphisms in Chinese patients with Parkinson's disease and multiple system Atrophy. *Front. Genet.* 12, 765833. doi:10.3389/fgene.2021.765833
- Sun, G., Zhou, H., Chen, K., Zeng, J., Zhang, Y., Yan, L., et al. (2020). HnRNP A1-mediated alternative splicing of CCDC50 contributes to cancer progression of clear cell renal cell carcinoma via ZNF395. *J. Exp. Clin. Cancer Res.* 39 (1), 116. doi:10.1186/s13046-020-01606-x
- Tang, H., Wang, Y., Zhang, B., Xiong, S., Liu, L., Chen, W., et al. (2017). High brain acid soluble protein 1(BASP1) is a poor prognostic factor for cervical cancer and promotes tumor growth. *Cancer Cell Int.* 17 (1), 97. doi:10.1186/s12935-017-0452-4
- Thibodeau, B. J., Fulton, M., Fortier, L. E., Geddes, T. J., Pruetz, B. L., Ahmed, S., et al. (2016). Characterization of clear cell renal cell carcinoma by gene expression profiling. *Urol. Oncol.* 34, 168.e1–168.e9. doi:10.1016/j.urolonc.2015.11.001
- Tseng, C. H., Tsuang, B. J., Chiang, C. J., Ku, K. C., Tseng, J. S., Yang, T. Y., et al. (2019). The relationship between air pollution and lung cancer in nonsmokers in taiwan. *J. Thorac. Oncol.* 14 (5), 784–792. doi:10.1016/j.jtho.2018.12.033
- Wang, X., Cao, Y., BoPan, B., Meng, Q., and Yu, Y. (2021). High BASP1 expression is associated with poor prognosis and promotes tumor progression in human lung adenocarcinoma. *Cancer Invest.* 39 (5), 409–422. doi:10.1080/07357907.2021.1910290
- Wiek, C., Schmidt, E. M., Roellecke, K., Freund, M., Nakano, M., Kelly, E. J., et al. (2015). Identification of amino acid determinants in CYP4B1 for optimal catalytic processing of 4-ipomeanol. *Biochem. J.* 465 (1), 103–114. doi:10.1042/BJ20140813
- Xing, J., Tian, Y., Ji, W., and Wang, X. (2020). Comprehensive evaluation of SPATS2 expression and its prognostic potential in liver cancer. *Med. Baltim.* 99 (9), e19230. doi:10.1097/MD.00000000000019230
- Xu, S., Zhang, H., Liu, T., Chen, Y., He, D., and Li, L. (2019). G protein γ subunit 7 loss contributes to progression of clear cell renal cell carcinoma. *J. Cell. Physiol.* 234 (11), 20002–20012. doi:10.1002/jcp.28597
- Yang, H., Hu, Y., Weng, M., Liu, X., Wan, P., Hu, Y., et al. (2021). Hypoxia inducible lncRNA-CBSLR modulates ferroptosis through m6A-YTHDF2-dependent modulation of CBS in gastric cancer. *J. Adv. Res.* 37, 91–106. PMID: 35499052; PMCID: PMC9039740. doi:10.1016/j.jare.2021.10.001
- Ye, X., Zhang, Y., Wang, X., Li, Y., and Gao, Y. (2016). Tumor-suppressive functions of long-chain acyl-CoA synthetase 4 in gastric cancer. *IUBMB Life* 68, 320–327. doi:10.1002/iub.1486

Yu, J., Liu, T. T., Liang, L. L., Liu, J., Cai, H. Q., Zeng, J., et al. (2021). Identification and validation of a novel glycolysis-related gene signature for predicting the prognosis in ovarian cancer. *Cancer Cell Int.* 21 (1), 353. doi:10.1186/s12935-021-02045-0

Zhang, X. G., Song, B. T., Liu, F. J., Sun, D., Wang, K. X., and Qu, H. (2016). CCR6 overexpression predicted advanced biological behaviors and poor prognosis in patients with gastric cancer. *Clin. Transl. Oncol.* 18 (7), 700–707. doi:10.1007/s12094-015-1420-x

Zhang, D., Jiang, Q., Ge, X., Shi, Y., Ye, T., Mi, Y., et al. (2021). RHOV promotes lung adenocarcinoma cell growth and metastasis through JNK/c-Jun

pathway. *Int. J. Biol. Sci.* 17 (10), 2622–2632. Published 2021 Jun 22. doi:10.7150/ijbs.59939

Zhang, L., Li, Z., Skrzypczynska, K. M., Fang, Q., Zhang, W., O'Brien, S. A., et al. (2020). Single-cell analyses inform mechanisms of myeloid-targeted therapies in colon cancer. *Cell* 181 (2), 442–459.e29. doi:10.1016/j.cell.2020.03.048

Zhang, Y., Zhang, X., Lv, X., Zhang, M., Gao, X., Liu, J., et al. (2020). Development and validation of a seven-gene signature for predicting the prognosis of lung adenocarcinoma. *Biomed. Res. Int.* 2020, 1836542. doi:10.1155/2020/1836542