Differentiation-related genes in tumor-associated macrophages as potential prognostic biomarkers in non-small cell lung cancer

Zhaoxun Li†, Bin Zhou†, Xinsheng Zhu†, Fujun Yang†, Kaiqi Jin, Jie Dai, Yuming Zhu, Xiao Song* and Gening Jiang*

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

Background: The purpose of this study was to evaluate the role of differentiation-related genes (DRGs) in tumor-associated macrophages (TAMs) in non-small cell lung cancer (NSCLC).

Methods: Single cell RNA-seq (scRNA-seq) data from GEO and bulk RNA-seq data from TCGA were analyzed to identify DRGs using trajectory method. Functional gene analysis was carried out by GO/KEGG enrichment analysis. The mRNA and protein expression in human tissue were analyzed by HPA and GEPIA databases. To investigate the prognostic value of these genes, three risk score (RS) models in different pathological types of NSCLC were generated and predicted NSCLC prognosis in datasets from TCGA, UCSC and GEO databases.

Results: 1,738 DRGs were identified through trajectory analysis. GO/KEGG analysis showed that these genes were predominantly related to myeloid leukocyte activation and leukocyte migration. 13 DRGs (*C1QB, CCL4, CD14, CD84, FGL2, MS4A6A, NLRP3, PLEK, RNASE6, SAMSN1, SPN, TMEM176B, ZEB2*) related to prognosis were obtained through univariate Cox analysis and Lasso regression. *C1QB, CD84, FGL2, MS4A6A, NLRP3, PLEK, SAMSN1, SPN*, and *ZEB2* were downregulated in NSCLC compared to non-cancer tissue. The mRNA of 13 genes were significantly expressed in pulmonary macrophages with strong cell specificity. Meanwhile, immunohistochemical staining showed that *C1QB, CCL4, SPN, CD14, NLRP3, SAMSN1, MS4A6A, TMEM176B* were expressed in different degrees in lung cancer tissues. *ZEB2* (HR=1.4, P<0.05) and *CD14* (HR=1.6, P<0.05) expression were associated with a worse prognosis in lung squamous cell carcinoma; *ZEB2* (HR=0.64, P<0.05), *CD84* (HR=0.65, P<0.05), *PLEK* (HR=0.71, P<0.05) and *FGL2* (HR=0.61, P<0.05) expression were associated with a better prognosis in lung adenocarcinoma. Three RS models based on 13 DRGs both showed that the high RS was significantly associated with poor prognosis in different pathological types of NSCLC.

Conclusions: This study highlights the prognostic value of DRGs in TAMs in NSCLC patients, providing novel insights for the development of therapeutic and prognostic targets based on TAM functional differences.

1 Introduction

Lung cancer is one of the main global causes of cancer-related deaths (1, 2). Non-small cell lung cancer (NSCLC) is the most common type of lung cancer (2, 3). Despite improvements in therapy and the use of comprehensive treatments consisting of a variety of approaches, the overall survival (OS) of NSCLC patients remains poor (2). Significantly, tumor heterogeneity is found to be associated with drug resistance, tumor metastasis, and poor prognosis (4, 5).

A large number of recent studies has focused on the tumor microenvironment (TME) (6, 7). The composition of the TME can be complex and heterogeneous and that includes macrophages, T cells, bone marrow derived inflammatory cells, NK cells, fibroblasts, B cells, extracellular matrix, and various signaling molecules (6, 8, 9). Cellular interactions in the TME are thought to be closely associated with tumor invasion, growth and metastasis (8–10), and components of the TME can represent biomarkers with important roles in the detection, treatment and prognosis of tumors (7, 11–13). This is also the case for NSCLC, where several potential TEM targets have been explored in relation to diagnosis, treatment and prognosis (6, 11–13).

Macrophages are major component of TEM (14, 15) and their functional diversity and phenotypic plasticity has attracted increasing research interest (10). Several studies have shown that tumor associated macrophages (TAMs) are related to tumor metastasis, invasion, angiogenesis, and immunosuppression (9, 14, 16). TAMs are heterogeneous and consist of several subtypes, which have traditionally been grouped into "M1" and "M2" types. M1 macrophages exhibit proinflammatory and anti-tumor properties, while M2 macrophages are associated with inflammation resistance, angiogenesis, and tumorigenesis. Importantly, macrophages can transition between M1 and M2 subtypes (9, 14, 16, 17). TAMs have shown great potential in therapy and prognosis prediction of lung cancer (15–18) but due to the complex cellular heterogeneity in the TME, it has been difficult to clearly define their biological function and clinical value.

Traditional gene sequencing methods obtain the average gene expression of different cell types in a sample, which renders it difficult to identify and describe distinct immune cell states and types, and may result in the loss of important cell subtype information (19). In contrast, single-cell sequencing is a novel method that enables to assess the gene expression at single cell level, which offers great advantages for the elucidation of cellular heterogeneity in different tumors and their TME (4, 5). Furthermore, single-cell sequencing allows for the simulation of cell fate or differentiation trajectories and identification of fate or differentiation related genes (DRGs), enabling in-depth exploration of cellular phenotypes and biological differences between various cells types in the TME.

In this study, we designed a data mining experiment using single-cell sequencing data of NSCLC. Single-cell transcriptomic analysis was applied to probe gene expression in tumor samples and identify DRGs in TAMs. Gene Ontology (GO: molecular function, cellular component, and biological process) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enrichment analyses were conducted to evaluate the function of TAM DRGs. Through univariate Cox analysis and Lasso regression using TCGA-NSCLC bulk RNA-seq data, we screened 13 DRGs significantly related to patient prognosis. Next, we analyzed the gene function, involving pathway and expression of 13 DRGs. By constructing RS models in multiple data sets, we further explored the prognostic value of 13 DRGs in different pathological types of NSCLC. Our research reveals a potential role for TAM DRGs in the prognosis of NSCLC, provides clues for illuminating the function of TAM DRGs in the NSCLC TME, and discovers potential future therapeutic and prognostic targets for NSCLC.

2 Methods

2.1 Data mining

Single-cell RNA sequencing (scRNA-seq) data from human NSCLC samples were download from the GSE116947 dataset in the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/). Bulk RNA-seq and clinical data of NSCLC patients were obtained from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/) to establish and verify the risk prediction models. In addition, multiple sequencing data (Lung cancer RAPONI 2006; GSE157009; GSE31210) of NSCLC from GEO and University of California Santa Cruz (USCS) databases (https://xena.ucsc.edu/public/) were used as external validation sets to verify the effectiveness of the prognosis models. All data used in this study are freely available from the respective databases.

2.2 Data processing

The 'Seurat' package in R 3.5.1 was used for data quality control and preliminary data exploration. Data filtering was conducted according to the following criteria: 1) Genes that were only detected in less than 3 cells were excluded; 2) Cells with less than 50 genes detected in total were excluded; 3) Cells with mitochondrial gene expression of equal to or more than 5% were excluded. Data were normalized using the Log Normalization algorithm and gene expression was subsequently normalized using a linear regression model. Significant and effective dimensions were determined by principal component analysis (PCA) with a P value <0.05. The t-Stochastic Neighbor Embedding (t-SNE) and Uniform Manifold Approximation and Projection for dimension reduction (UMAP) algorithm were used to reduce the dimension of the top 15 principal components (PCs) and obtain major cells clusters. For differential gene expression analysis and identification of marker genes for each cell cluster, we used the 'Seurat' package. |log2 (fold change) |>0.25 was the threshold for marker gene identification. Cell annotation was performed using the CellMarker database (20) and reports from the literature (21) based on the composition pattern of marker genes. Data were visualized using the 'ggplot2' package in R 3.5.1.

2.3 Single cell trajectory analysis

In many diseases, cellular state transitions are characterized by cascading changes in gene expression. In order to infer the gene regulatory events that drive the transition from one cellular state to another, we used the Monocle 2 algorithm (22) to construct a single-cell pseudo-time trajectory of scRNA-seq data. Cells in each branch show different fates and functions. Genes differentiated between branches were defined as differentiation-related genes (DRGs), which essentially reflect the different functions of cells in different states. Functional enrichment analysis (GO and KEGG pathways analysis) of DRGs was performed using Metascape (23) (http://metascape.org). According to membership similarities, terms with a P-value<0.01, minimum count of 3, and enrichment factor>1.5 were grouped into clusters. A network plot was rendered by a subset of selected enriched terms to show the relationships between terms, where terms with a similarity>0.3 were connected by edges.

2.4 Establishment and validation of a risk score

The relationship between patients' survival and expression of risk genes was evaluated by univariate Cox regression analysis in the TCGA training cohort. Prognostic genes significantly associated with survival (P<0.05) were further filtered using least absolute shrinkage and selection operator (LASSO) with five times cross validation and multivariate Cox regression methods. The risk score for each patient was then calculated as follows: Risk scores (RS) = Exp (GENE$_1$) × β$_1$ + Exp (GENE$_2$) × β$_2$ +... + Exp (GENE$_n$) × β$_n$,

where "Exp" represents the expression level of the corresponding gene (GENE$_n$) and "β$_n$" represents the regression coefficient in Cox analysis as obtained from multiple regression. Individuals in the TCGA cohort were then either classified as low (low RS) or high risk (high RS) based on median RS values. Kaplan-Meier survival analysis was used to assess the overall survival (OS) of the two groups, and differences in survival were assessed using the bilateral log-rank test. 'Survcomp' and 'SurvivalROC' packages in R were used to generate ROC and calibration curves for evaluating the predictive accuracy of the RS score. The AUC value could range from 0.5 and 1, with 1 representing complete discrimination, 0.5 representing no discrimination.

2.5 Analysis of differential expression and prognosis

Gene Expression Profiling Interactive Analysis (GEPIA) (24) is a Web-based tool that provides a variety of analytical capabilities based on TCGA and Genotype-Tissue Expression (GTEx) data. We used NSCLC data from TCGA and GTEx in GEPIA website to explore the differential expression of the 13 DRGs in tumor and non-tumor samples and assess their relationship with prognosis. OS was selected as the prognostic outcome. The relationship between gene expression and prognosis was evaluated by hazard ratio (HR).

2.6 Gene expression analysis

The protein expression data of the Human Protein Atlas (HPA) (25) database (http://proteinatlas.org) were used to analyze the expression of proteins encoded by DRGs in lung cancer tissues. The HPA mRNA expression data in single cell lines were used to analyze the mRNA expression of DRGs in different lung cell types, and transcripts per kilobase of exon model per million mapped reads (TPM) and Z-score were used to calculate the expression of mRNA in cell lines. Immunohistochemical staining assay was used to show the expression of proteins in lung cancer tissues.

2.7 Statistical analysis

All tests conducted in this study were two-tailed, with P<0.05 being considered statistically significant. We used Kaplan-Meier analysis and log-rank tests for survival analysis. Data visualization and statistical analysis were carried out using R version 3.5.1. The analysis pipeline is illustrated in Figure 1.

3 Results

3.1 Identification of differentially expressed genes and cell annotation

We obtained macrophages single cell data from GSE116947. Before filtering, there were 31,760 features for 11,713 cells in the

FIGURE 1

Analysis overview. TAMs, tumor-associated macrophages; NTAMs, non-tumor-associated macrophages; TCGA, The Cancer Genome Atlas; PCA, Principal Component Analysis; GEPIA, Gene Expression Profiling Interactive Analysis; LASSO, Least absolute shrinkage and selection operator; t-SNE, t-Stochastic Neighbor Embedding; UMAP, Uniform Manifold Approximation and Projection for Dimension Reduction; USCS, University of California Santa Cruz; GEO, Gene Expression Omnibus.

NSCLC tumor sample. After data standardization and quality control, we finally detected 15,194 genes for further analysis. The sequencing depth was significantly positively correlated with total intracellular sequences (R=0.95), but not with mitochondrial gene sequences (R=0.03, Figure 2A). Analysis of variance showed 2,000 highly variable genes (Figure 2B). Data dimensionality reduction was conducted by principal component analysis (PCA), revealing no obvious separation trend of cells (Supplementary Figure 1A). Finally, 15 principal components with significant differences were selected for further analysis (Supplementary Figures 1C, D). Using t-SNE and UMAP, cells were divided into 9 subgroups for which 1,521 marker genes were identified by differential expression analysis. Based on subgroup marker genes, the cell cluster highly expressing LGMN was annotated as M2 cells, the cell cluster with high expression of CXCL9 was annotated as M1 cells, and the cell cluster highly expressing FABP4 was annotated as non-tumor-associated macrophages (NTAMs; Figure 2C).

3.2 Differentiation trajectory analysis and identification of DRGs

According to the results of cell clustering and annotation above, we included the TAMs and NTAMs into the pseudo-time cell differentiation trajectory analysis (Figure 2D). We identified two branches with distinct differentiation types. Branch I contained 3,950 NTAM cells and branch II contained 314 M1/M2 cells, and a total of 1,738 DRGs were identified. GO enrichment revealed that DRGs were predominantly related to myeloid leukocyte activation and leukocyte migration, and lymphocyte activation, positive regulation of cytokine production, cell death, cell migration, apoptosis signal pathway, immune response regulation pathways

were negatively correlated with DRGs (Figure 3). Transcription factor enrichment analysis indicated that DRGs had several common TFs, including *PSMB5*, *GTF2A2*, *FOXE1*, *MAPK3*, and *MXD1*, amongst others (Supplementary Figure 2A). Upstream TF enrichment analysis showed that the expression of DRGs was regulated by *RELA*, *NFKB1*, *SP1*, *STAT3*, and *JUN*, amongst others (Supplementary Figure 2B). KEGG analysis showed that DRGs were highly expressed in pneumonitis, myocardial ischemia, lupus nephritis and lung diseases, amongst others (Supplementary Figure 2C). DRGs expression was found to be tissue and cell specific, with blood, spleen, bone marrow, lung tissue, CD33 positive myeloid, adipocyte and B lymphocyte being enriched for these DRGs (Supplementary Figure 2D).

3.3 Development of a differentiation-related gene prognostic risk score

Through univariate analysis and Lasso regression, we screened 13 DRGs (*C1QB, CCL4, CD14, CD84, FGL2, MS4A6A, NLRP3, PLEK, RNASE6, SAMSN1, SPN, TMEM176B, ZEB2*) related to prognosis in TCGA NSCLC dataset (Supplementary Figure 3). We next constructed a risk score (RS) model based on 13 DRGs by Multivariate Cox. The RS of each sample was calculated by the relative coefficient and expression (Exp) of each gene:

RS = 0.4605 * Exp (*C1QB*) - 0.1733 * Exp (*CCL4*) + 0.2277 * Exp (*CD14*) + 0.5458 * Exp (*CD84*) - 0.2402 * Exp (*FGL2*) - 0.4388 * Exp (*MS4A6A*) + 0.1058 * Exp (*NLRP3*) - 0.3260 * Exp (*PLEK*) + 0.5543 * Exp (*RNASE6*) - 0.1314 * Exp (*SAMSN1*) - 0.0766 * Exp (*SPN*) - 0.1092 * Exp (*TMEM176B*) + - 0.3229 * Exp (*ZEB2*).

Patients in the TCGA cohort were divided into two groups based to the median RS, resulting in a high and low RS group. The Kaplan-Meier survival curve suggested that the overall survival (OS) in high RS group was significantly lower than in the low RS group (Figure 4B), which indicated a relationship between RS and prognosis. Multivariate Cox analysis of the 13 genes found a significant association *C1QB1*, *CD84*, *PLEK*, *ZEB2*, and *RNAASE6* with survival (Table 1). A high expression of *C1QB* (hazard ratio (HR)=1.58, P=0.006), *CD84* (HR=1.73, P=0.005), and *RNAASE6* (HR=1.74, P=0.005) was associated with poor prognosis, while the high expression of *PLEK* (HR=0.72, P=0.036) or *ZEB2* (HR=0.72, P=0.021) was associated with a good prognosis. Receiver operating characteristic (ROC) curves and the c-index were used to validate our model, which identified an area under the ROC curve for prediction of 5-year OS of 0.654 (Figure 4C).

A hallmark enrichment analysis of the 13 DRGs in risk score showed that they associated with immune destruction, angiogenesis, resistance to cell death, apoptosis, growth signaling, tumor-promoting information, and oxidative stress (Figure 4A). This suggested that they may participate in tumorigenesis by promoting tumor formation, resistance to apoptosis, modifying genomic stability, and exhibiting anti-effects that promote tumor invasion and metastasis, which ultimately lead to a worse prognosis. It is worth noting that even if the genes *PLEK* and *ZEB2* were associated with a good prognosis in multivariate Cox survival analysis, their functions are related to tumor-promoting
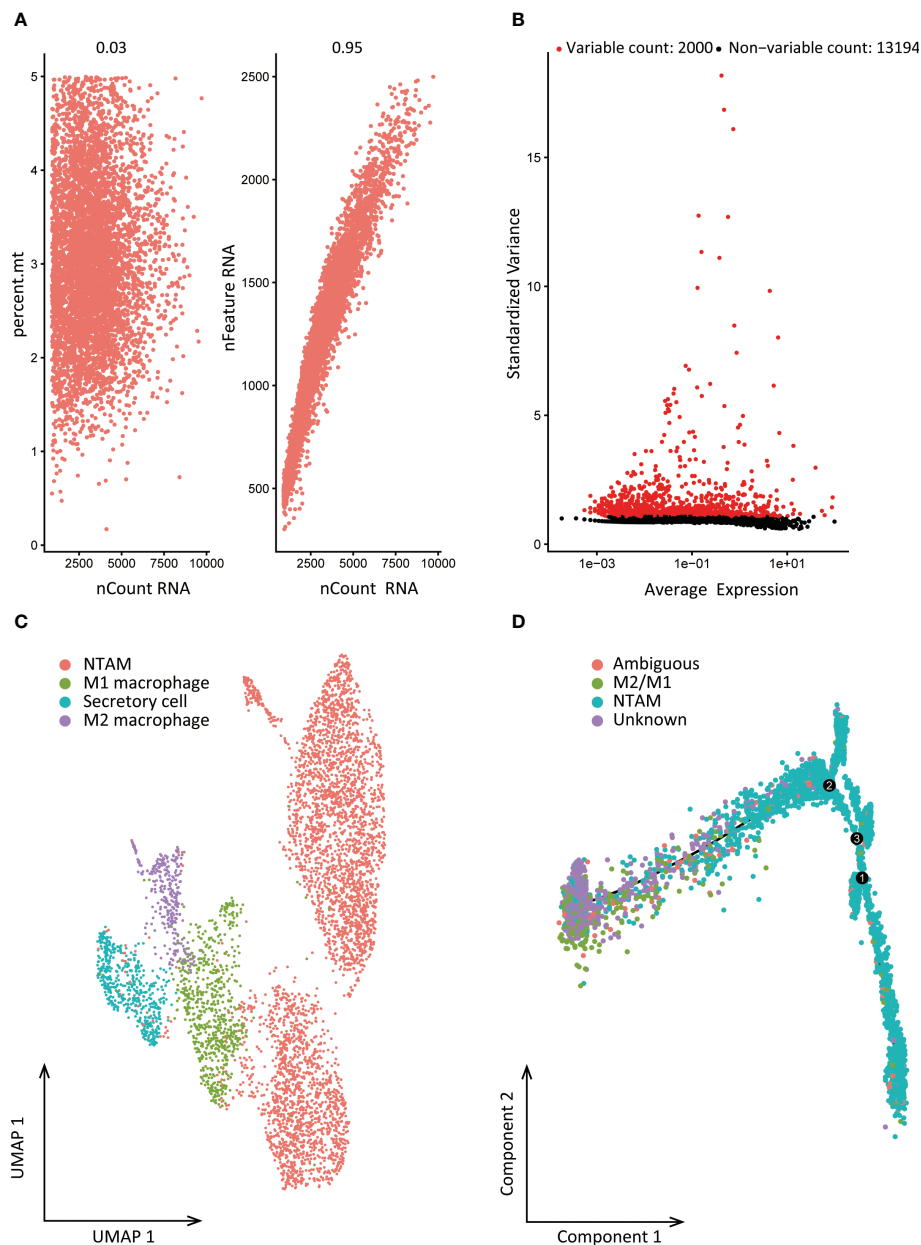
FIGURE 2
Cell clustering and differentiation trajectory analysis. **(A)** Correlation between sequencing depth and mitochondrial gene sequences or total intracellular sequences. **(B)** 15,194 genes were analyzed in total, of which 13,194 exhibited low intercellular variation and 2,000 had high intercellular variation. **(C)** Cell annotation based on gene markers. **(D)** Pseudo-time and trajectory analysis. PCA, principal component analysis; PCs, principal components; NTAMs, non-tumor-associated macrophages; UMAP, Uniform Manifold Approximation and Projection for Dimension Reduction.

inflammation (both *PLEK* and *ZEB2)*, immune destruction, resistance to cell death, invasion and metastasis (*ZEB2*), and therefore their role will need to be further evaluated.

3.4 Expression and prognosis analysis of 13 DRGs

We used NSCLC and normal lung tissue data from TCGA and GTEx in the GEPIA database to analyze expression levels of the 13 DRGs in lung cancer and normal samples. Compared with normal

lung tissue, *C1QB*, *NLRP3*, *SAMSN1*, *SNP*, and *ZEB2* were significantly downregulated in patients with lung adenocarcinoma (LUAD; Figure 5); *C1QB*, *FGL2*, *MS4A6A*, *NLRP3*, *PLEK*, *SAMSN1*, *SPN*, *CD84* and *ZEB2* were significantly downregulated in patients with lung squamous cell carcinoma (LUSC; Figure 5). Next, we constructed Kaplan-Meier curves of the 13 genes, dichotomizing by median expression. This revealed that high expression levels of *ZEB2* (HR = 0.64, P = 0.0034), *CD84* (HR = 0.65, P = 0.005), *PLEK* (HR = 0.71, P = 0.022) and *FGL2* (HR = 0.61, P = 0.0016) were significantly associated with a better prognosis in LUAD; high expression levels of *ZEB2* (HR = 1.4, P = 0.013) and *CD14* (HR =
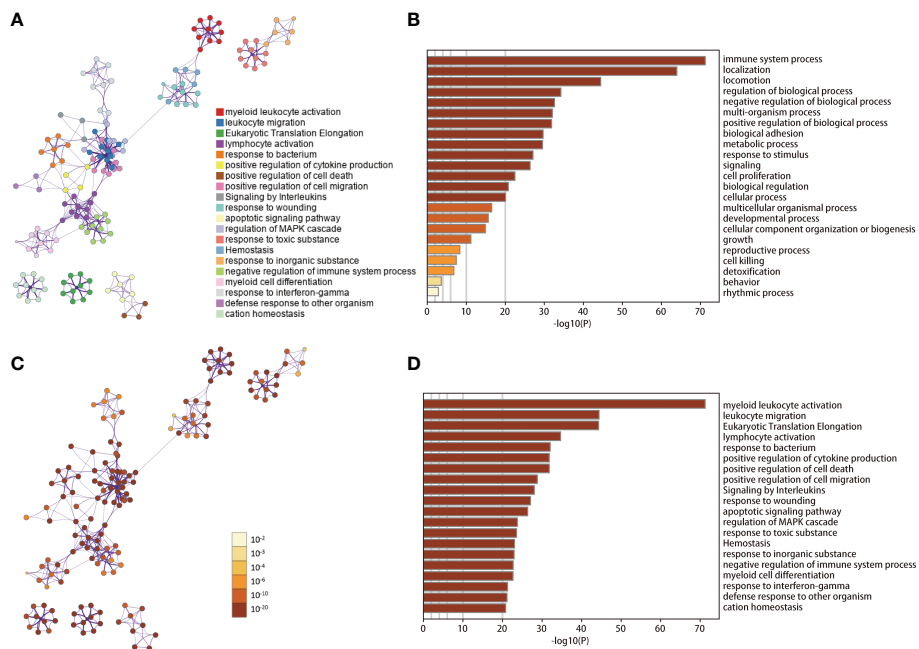
FIGURE 3
GO/KEGG enrichment analysis of DRGs. **(A, C)** Network of enriched terms. **(B, D)** Bar graph of GO/KEGG enrichment analysis across DRGs. Results in B-D are colored by p-values. DRGs, Differentiation Related Genes; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.
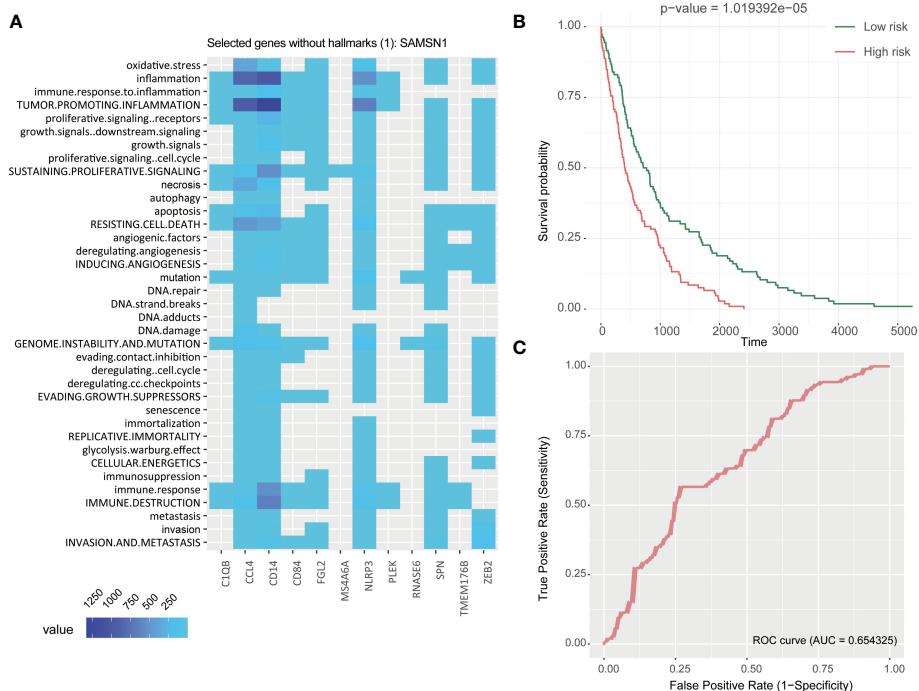
FIGURE 4
Establishment of 13 DRGs RS model. **(A)** Heatmap of hallmark analysis of 13 genes in the RS model. **(B)** Kaplan-Meier analysis of the RS in the TCGA NSCLC cohort. Patients were divided into high and low risk groups according to the median RS. **(C)** ROC curve for predicting 5-year OS in the TCGA NSCLC cohort. DRGs, Differentiation Related Genes; TCGA, The Cancer Genome Atlas; NSCLC, Non-Small Cell Lung cancer; RS, Risk Scores; ROC, Receiver Operating Characteristic.

TABLE 1   Multivariate Cox analysis of the 13 genes.

Gene Name	HR	P
C1QB	1.58 (1.14 - 2.20)	0.006
CD84	1.73 (1.18 - 2.53)	0.005
NLRP3	1.11 (0.81 - 1.52)	0.505
PLEK	0.72 (0.53 - 0.98)	0.036
ZEB2	0.72 (0.55 - 0.95)	0.021
CD14	1.26 (0.96 - 1.64)	0.096
FGL2	0.79 (0.53 - 1.16)	0.226
TMEM176B	0.90 (0.70 - 1.15)	0.395
MS4A6A	0.64 (0.41 - 1.02)	0.063
RNASE6	1.74 (1.18 - 2.56)	0.005
SPN	0.93 (0.71 - 1.21)	0.571
CCL4	0.84 (0.64 – 1.11)	0.233
SAMSN1	0.88 (0.70 – 1.10)	0.254

1.6, P < 0.05) were significantly associated with a worse prognosis in LUSC (Supplementary Figure 4).

We used the HPA database to further analyze the expression of 13 DRGs in human cells and tissues. The expression of 13 DRGs in lung cancer tissues was analyzed by immunohistochemical staining. Among them, CD84, FGL2, PLEK, RNASE6, ZEB2 were not detected in lung cancer tissues; C1QB, CCL4 and SPN showed low

staining and weak density in the lung cancer tissue; CD14, NLRP3, SAMSN1 showed moderate staining and moderate density; MS4A6A, TMEM176B showed high staining and strong density (Supplementary Figure 5). By analyzing the mRNA expression of each cell types in lung tissue, it was found that the mRNA of 13 DRGs was significantly expressed in lung macrophages, showing strong cell specificity (Supplementary Figures 6A, B). In addition, we analyzed the expression of these DRGs in different cell clusters (Supplementary Figures 6C, D). It can be seen that these genes were differentially expressed in TAMs and NTAMs, indicating that those differences in genes expression might be related to the tumorigenesis and anti-tumor function of TAMs.

3.5 Reconstruction of RS models under different pathological types

It should be noted that the expression and prognostic value of the same DRGs in LUAD and LUSC might not be consistent. Therefore, it is necessary to verify the prognostic value of DRGs in LUAD and LUSC respectively. We then selected the DRGs with Log-rank P < 0.05 in GEPIA prognosis analysis and reconstructed RS models in TCGA-LUAD and TCGA-LUSC datasets. In the TCGA-LUAD dataset, FGL2, CD84, PLEK, and ZEB2 were used to reconstruct the prognosis model (LUAD-RS = 0.1727* Exp (FGL2) - 0.1986* Exp (CD84) - 0.1155* Exp (PLEK) - 0.0195* Exp (ZEB2)). CD14 and ZEB2 were used to reconstruct a prognosis mode in the TCGA-LUSC dataset (LUSC-RS = 0.0374* Exp (CD14) + 0.1081* Exp (ZEB2)). The results showed that the prognosis of patients with
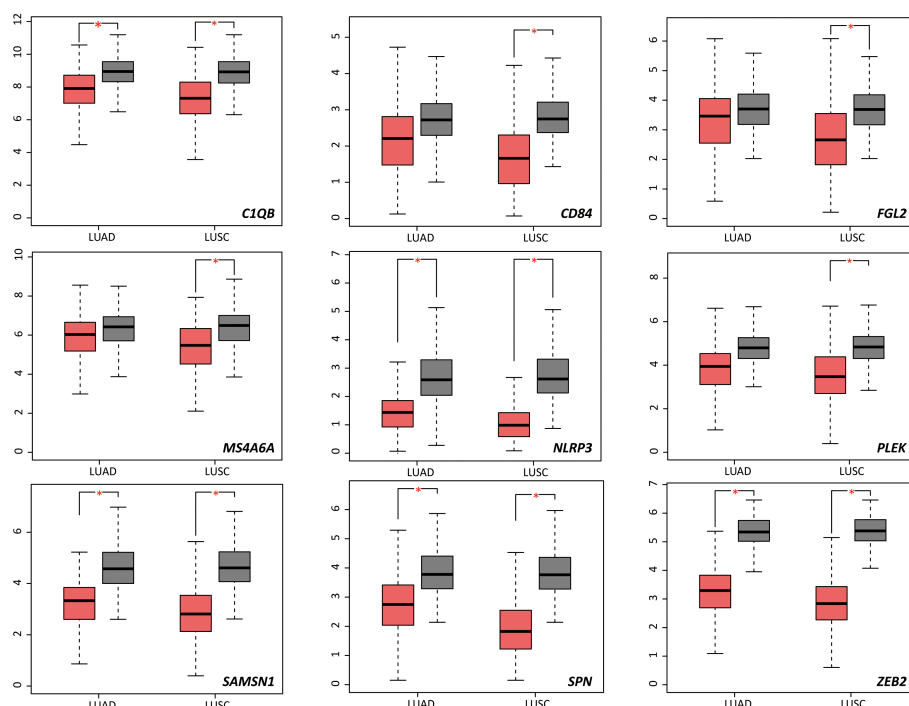

FIGURE 5
Differential expression analysis of 13 DRGs in lung squamous cell carcinoma and lung adenocarcinoma samples. DRGs, Differentiation Related Genes; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; T, tumor; N, non-tumor. *P < 0.05.

high RS was worse than that of patients with low RS (Figures 6A, D). In order to further validate two prognosis models, we used the lung cancer RAPONI 2006 data set and the GSE157009 data set to validate the results of LUSC-RS and LUAD-RS respectively. The results showed that patients with high RS had a worse prognosis than those with low RS (Figures 6B, E). In addition, we found that high LUAD-RS was associated with the tumor recurrence of LUAD patients with positive EGFR mutation in the GSE31210 data set (Figure 6C).

4 Discussion

Novel treatments involving the targeting of immune checkpoints or adoptive immune cell therapy provide promising strategies for cancer therapy. However, tumor heterogeneity (26), cell plasticity (27), and both primary and acquired drug resistance (28) remain significant problems faced by targeted therapies. By identifying and characterizing cancer subtypes with specific biological characteristics, personalized treatment plans and the implementation of precision medicine may be realized. The application of liquid biopsies and single-cell sequencing provide powerful methods for the development of personalized diagnosis and treatment schemes. In this study, we identified differentiation-related genes (DRGs) in tumor-associated macrophages through mining of NSCLC single-cell sequencing data from GEO database and trajectory analysis. Based on clinical data obtained from TCGA, we then screened 13 DRGs and constructed a RS model based on them. The RS was related to the prognosis of NSCLC patients. GO/KEGG analysis, differential expression analysis, and survival analysis further elucidated the role of these genes in the progression, metastasis, and drug resistance of lung cancer.

Besides, we reconstructed the RS models under different pathological types according to the difference in DRGs expression and prognosis analysis in LUAD and LUSC. Our results indicate a potential value of TAM DRGs in NSCLC and provide insights for the further exploration of their therapeutic, and prognostic roles.

Transcription factor enrichment analysis revealed several common TFs of DRGs (Supplementary Figure 2A). Several studies have previously described a role for TFs in the occurrence, development, metastasis, and chemoresistance of tumors. For instance, *EGR1*, *JUN*, *PPARG*, and *RELA* may have a beneficial effect in lung cancer, related to inhibition of tumor proliferation and metastasis, induction of apoptosis, and sensitization for chemotherapeutic drugs (29–37). Conversely, *NFB1*, *HDAC1*, *SP1*, *SPI1*, and *STAT3* have been found to be associated with the promotion of tumor proliferation, apoptosis resistance, cell migration, induction of angiogenesis, and drug resistance (38–49). It is worth noting that certain TFs such as *PPARA* (33, 34) and *STAT1* (50–53) exhibit both tumor-promoting and anti-tumor effects. Besides, upstream TF enrichment analysis indicated that the expression of DRGs might be regulated by *RELA*, *NFKB1*, *SP1*, *STAT3*, and *JUN*, amongst others. (Supplementary Figure 2B). These upstream genes have previously been shown to play a role in autophagy (54, 55), proliferation and metastasis of tumor cells (56), and are related to the drug resistance (57) and prognosis (58, 59). Moreover, these upstream TFs also play both anti-tumor (such as *PSMB5* (60)) and pro-tumor roles (such as *FOXE1* (55, 59), *GTF2A2* (58, 61), *MAPK3* (54, 62, 63), and *MXD1* (56)). These results suggested a multifaceted role of DRGs in NSCLC, which is in line with a dual role of TAMs in the TME.

CD14 is a classic monocyte marker (64), and a high prevalence of CD14-positive monocytes has previously been shown to be associated with better chemotherapeutic response and patient
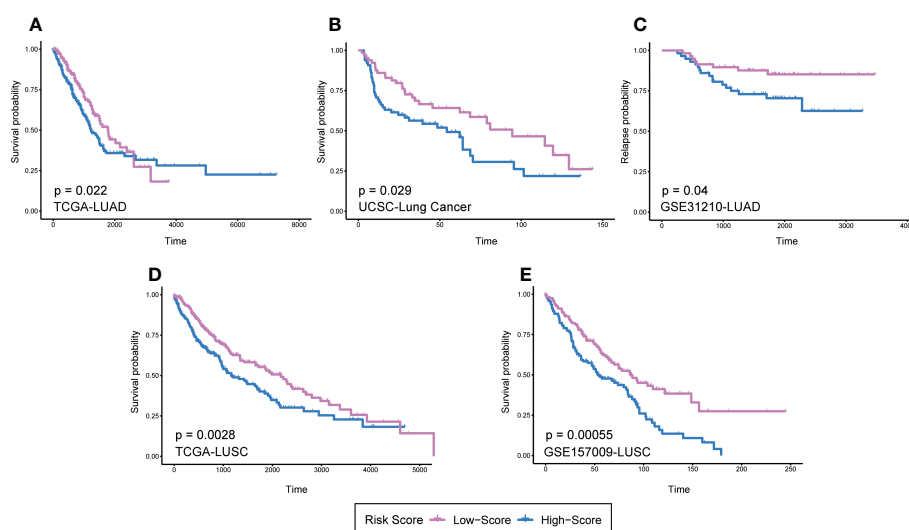
FIGURE 6
RS models of different pathological types based on 13 DRGs. **(A)** LUAD-RS model reconstruction in TCGA LUAD dataset. **(B)** LUAD-RS model validation in UCSC lung cancer dataset. **(C)** LUAD-RS model validation in GSE31210 dataset. **(D)** LUSC-RS model reconstruction in TCGA LUSC dataset. **(E)** LUSC-RS model validation in GSE157009 dataset. Patients were divided into high and low risk groups according to the median RS. DRGs, Differentiation Related Genes; RS, risk scores; TCGA, The Cancer Genome Atlas; USCS, University of California Santa Cruz; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma.

survival (64, 65). In this study, we found that patients with a high expression of *CD14* had a worse prognosis, which may indicate that *CD14* gene expression exhibits distinct effects in different cell types. In our study, CD14-positive cells in tumor samples were mainly macrophages and bone marrow-derived monocytes found to exhibit tumor-promoting and immunosuppressive effects. A previous study indicated that there is less infiltration of these cells in early lung cancer tissues which is consistent with our findings, suggesting that CD14-positive cells predominantly exist in advanced tumors and might related to a poor prognosis (66).

We found that *FGL2*, *MS4A6A*, and *SAMSN1* were downregulated in NSCLC tissues compared to normal lung. In previous studies, these genes were found to play a protective role in lung cancer. *FGL2* has previously been shown to be positively correlated with macrophage infiltration in lung adenocarcinoma and CD8-positive T cell activation, and is associated with a better prognosis (67). Anti-*MS4A1* therapy has achieved a promising results in non-Hodgkin's B cell lymphoma (68). *SAMSN1*, which is located in a common genomic deletion region in lung cancer and is associated with B cell differentiation (69), may act as a suppressor gene in lung cancer. Our results suggest that the above genes may have potential value in the treatment of lung cancer and their further exploration may be conducive to the development of new therapies.

A previous study found that overexpression of the C1QB protein was correlated with lymph node metastasis of lung cancer (70). In renal cell carcinoma (RCC), C1QB expression can influence CSF-1-induced macrophage migration and hamper their adhesion and chemotaxis (71). In our study, we found that C1QB was downregulated in NSCLC and related to poor prognosis in multivariate Cox analysis. Compared with macrophages in normal tissue, pancreatic cancer patients exhibited high expression of C1QB in TAMs and peripheral blood (72), indicating that it may be a suitable liquid biopsy biomarker to predict prognosis. Due to the lack of relevant research in lung cancer, the prognostic value of C1QB in lung cancer will need to be further explored.

CD84 promotes tumor cell survival in early chronic lymphocytic leukemia, and inhibition of *CD84* leads to cell death (73). However, the role of *CD84* in lung cancer is still not completely understood. One study showed that radiation-induced lung cell aging upregulates *CD84*, suggesting it may be related to radiation injury (74). In this study, we found differential expression of *CD84* in lung cancer and normal lung tissue and further uncovered an association of *CD84* expression with poor prognosis in multivariate Cox analysis. *CD84* is a cell surface receptor involved in leukocyte activation highly expressed on monocytes, macrophages, and granulocytes, and is related to TNF-alpha secretion induced by lipopolysaccharide (LPS) (75). A previous study found that certain substances can prevent or treat prostatic cancer by inhibition of *CD84* mRNA (76), suggesting that further exploration of the role of *CD84* in lung cancer may guide the identification of novel therapeutic targets.

PLEK gene expression was previously found to be associated with poor prognosis and chemoresistance in lung cancer patients (77). It is involved in the regulation NSCLC cell migration and vascular infiltration, and its expression is correlated with poor OS. Overexpression of *PLEK2* significantly promoted epidermal-mesenchymal transformation and tumor migration (78). Our results showed that *PLEK* was downregulated in LUSC tissues compared to normal lung, and multivariate Cox regression analysis as long as GEPIA survival analysis revealed that its expression was associated with a better prognosis. Therefore, our results indicated that a high expression of *PLEK* may play a protective role in NSCLC cancer, which is in contrast to previous findings. Interestingly, one previous study found that *PLEK* was negatively correlated with the purity of lung cancer tissue, and low expression of *PLEK* led to high tumor purity, low immune score, low CD8+ T lymphocyte content, and shorter 5-year survival (79). In multivariate Cox regression, the expression of *ZEB2* was associated with a better prognosis, while LUAD patients with high *ZEB2* expression had a poor prognosis in GEPIA survival analysis. *ZEB2* mutations were found to be related to immunologic ignorance and immune tolerance microenvironments and may predict response to checkpoint inhibitors, and tumors without *ZEB2* mutations are associated with lower risk of patient death (80). *ZEB2* is involved in epithelial-mesenchymal transformation and is related to cisplatin and paclitaxel resistance (81, 82). Based on the above findings, *ZEB2* may act as a tumor promoter in non-small cell lung cancer. However, some studies have shown that *ZEB2* can also promote the apoptosis of lung cancer cells (83) and inhibits their proliferation and invasion (83–85). Taken together, these seemingly contradicting results suggest that *PLEK* and *ZEB2* may show anti- or pro-tumor effects under different conditions which could be related to gene mutations and different cancer subtypes. Further exploration is needed to define their impact in lung cancer.

It is worth noting that there was inconsistency between the multivariate Cox results of 13 DRGs and the prognosis analysis in GEPIA database. In addition, the prognosis of DRGs was also affected by the pathological type of lung cancer. Since the expression of DRGs in various types of macrophages (M1/M2/NTAMs) is distinct, different types of macrophages can show anti-tumor or pro-tumor effects in lung cancer. We hypothesized that these genes might play different roles in different pathological types of lung cancer, which might be related to the differences in the expression of these genes in TAMs and NTAMs. By reconstructing new RS models in LUAD and LUSC, we proved that the prognostic value of DRGs was affected by pathological types. Therefore, it is necessary to further elucidate the role of differentially expressed genes of TAMs and NTAMs in different pathological types of lung cancer and explore their potential therapeutic value for crucial genes in the future.

We would like to point out the following limitations in this study: first, the sequencing data were mined in retrospective way and our results and proposed hypotheses need to be verified by further experiments. Second, genes included in our risk score may have distinct effects on promoting or inhibiting tumorigenesis which correspond to the distinct functions of different subtypes of macrophages in the TME; likely, not all will be directly related to the poor prognosis or malignant characteristics of lung cancer. Future single cell experiments may shed light on the distinct function of macrophage subpopulations and genes in our risk score. In the future, more effective prediction models based on DRGs may be developed.

5 Conclusion

Using single-cell sequencing data, the current research identifies a prognostic role of tumor-associated macrophage (TAM) DRGs and provides novel insights into the function of TAMs in the TME and potential therapeutic and prognostic targets for precision medicine in NSCLC patients.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: The data in this article are available in the TCGA, GEO, UCSC databases online.

Author contributions

Conception/Design: ZL, XS, BZ, XZ, FY. Provision of study material or patients: ZL, XS, BZ, XZ, FY. Collection and/or assembly of data: ZL, XS, BZ, XZ, FY. Data analysis and interpretation: ZL, XS, BZ, XZ. Manuscript writing: ZL, XS, BZ, XZ, FY, KJ, JD, YZ, GJ. Final approval of manuscript: ZL, XS, BZ, XZ, FY, KJ, JD, YZ, GJ. All authors contributed to the article and approved the submitted version.

Funding

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. The handling editor W-HW declared a shared parent affiliation with the authors at the time of the review.

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/fimmu.2023.1123840/full#supplementary-material

References

1. Barta JA, Powell CA, Wisnivesky JP. Global epidemiology of lung cancer. *Ann Glob Health* (2019) 85(1). doi: 10.5334/aogh.2419

2. Burotto M, Thomas A, Subramaniam D, Giaccone G, Rajan A. Biomarkers in early-stage non-small-cell lung cancer: current concepts and future directions. *Journal Of Thoracic Oncology* (2014) 9(11):1556–380. doi: 10.1097/JTO.0000000000000302

3. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. *CA Cancer J Clin* (2016) 66(2):115–32. doi: 10.3322/caac.21338

4. González-Silva L, Quevedo L, Varela I. Tumor functional heterogeneity unraveled by scRNA-seq technologies. *Trends Cancer* (2020) 6(1):13–9. doi: 10.1016/j.trecan.2019.11.010

5. Li L, Xiong F, Wang Y, Zhang S, Gong Z, Li X, et al. What are the applications of single-cell RNA sequencing in cancer research: a systematic review. *Journal Of Experimental & Clinical Cancer Research* (2020) 40(1). doi: 10.1186/S13046-021-01955-1

6. Arneth B. Tumor microenvironment.*LID* 15:1648–9144.

7. Wu T, Dai Y. Tumor microenvironment and therapeutic response. *Cancer Letters* (2017) 387:61–8. doi: 10.3390/Medicina56010015

8. Ocaña MC, Martínez-Poveda B, Quesada AR, Medina MÁA-O. Metabolism within the tumor microenvironment and its implication on cancer progression: An ongoing therapeutic target. 1098–128. doi: 10.1016/J.Canlet.2016.01.043

9. Liang WA-O, Huang X, Carlos CJJ, Lu X. Research progress of tumor microenvironment and tumor-associated macrophages. *Clinical & Translational Oncology* (2020) 22(12):2141–52. doi: 10.1007/S12094-020-02367-X

10. Greten FR, Grivennikov SI. Inflammation and cancer: Triggers, mechanisms, and consequences. *Immunity* (2019) 51(1):27–41. doi: 10.1016/J.Immuni.2019.06.025

11. Bader JE, Voss K, Rathmell JC. Targeting metabolism to improve the tumor microenvironment for cancer immunotherapy. *Molecular Cell* (2020) 78(6):1019–33.

12. Lei X, Lei Y, Li JK, Du WX, Li RG, Yang J, et al. Immune cells within the tumor microenvironment: Biological functions and roles in cancer immunotherapy. *Cancer Letters* (2020) 470:126–33. doi: 10.1016/j.canlet.2019.11.009

13. Shen M, Kang Y. Complex interplay between tumor microenvironment and cancer therapy. *Frontiers Of Medicine* (2018) 12(4):426–39. doi: 10.1007/S11684-018-0663-7

14. Shu Y, Cheng P. Targeting tumor-associated macrophages for cancer immunotherapy. *Biochim Biophys Acta (BBA) - Rev Cancer* (2020) 1874(2):188434. doi: 10.1016/j.bbcan.2020.188434

15. Vitale I, Manic G, Coussens LM, Kroemer G, Galluzzi L. Macrophages and metabolism in the tumor microenvironment. *Cell Metabolism* (2019) 30(1):36–50.

16. Ngambenjawong C, Gustafson HH, Pun SH. Progress in tumor-associated macrophage (TAM)-targeted therapeutics. *Advanced Drug Delivery Reviews* (2017) 114:1872–8294. doi: 10.1016/j.addr.2017.04.010

17. Brown JM, Recht L, Strober S. The promise of targeting macrophages in cancer therapy. *Clinical Cancer Research* (2017) 23(13):3241–50. doi: 10.1158/1078-0432.Ccr-16-3122

18. Choo YW, Kang M, Kim HA-O, Han J, Kang S, Lee JR, et al. M1 macrophage-derived nanovesicles potentiate the anticancer efficacy of immune checkpoint inhibitors. *Acs Nano* (2018) 12(9):8977–93.

19. Baslan T, Hicks J. Unravelling biology and shifting paradigms in cancer with single-cell sequencing. *Nature Reviews Cancer* (2017) 17(9):1474–768. doi: 10.1038/Nrc.2017.58

20. Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, et al. CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Research* (2019) 47 (D1):1362–4962. doi: 10.1093/Nar/Gky900

21. Garrido-Martin EM, Mellows TWP, Clarke J, Ganesan AP, Wood O, Cazaly A, et al. M1(hot) tumor-associated macrophages boost tissue-resident memory T cells

infiltration and survival in human lung cancer. *Journal For Immunotherapy Of Cancer* (2020) 8(2). doi: 10.1136/Jitc-2020-000778

22. Qiu XA-O, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. 1548–7105. (Electronic)/

23. Zhou Y, Zhou B, Pache L, Chang MA-OX, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. 2041–1723. (Electronic)

24. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* (2017) 45 (W1):W98–w102. doi: 10.1093/nar/gkx247

25. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. *Tissue-based map Hum proteome Sci* (2015) 347(6220):1260419.

26. Zhang H, Chen J. Current status and future directions of cancer immunotherapy. *J Cancer* (2018) 9(10):1773–81. doi: 10.7150/jca.24577

27. Boumahdi S, de Sauvage FJ. The great escape: tumour cell plasticity in resistance to targeted therapy., 1474–784. (Electronic)

28. Ward RA-O, Fawell S, Floc'h N, Flemington V, McKerrecher D, Smith PD. Challenges and opportunities in cancer drug resistance., 1520–6890. (Electronic)

29. Zhang H, Chen X, Wang J, Guang W, Han W, Zhang H, et al. EGR1 decreases the malignancy of human non-small cell lung carcinoma by regulating KRT18 expression. *Sci Rep* (2014) 4. doi: 10.1038/srep05416

30. Ryu J-w, Choe SS, Ryu S-H, Park E-Y, Lee BW, Kim TK, et al. Paradoxical induction of growth arrest and apoptosis by EGF *via* the up-regulation of PTEN by activating redox factor-1/Egr-1 in human lung cancer cells. *Oncotarget* (2017) 8 (3):4181–95. doi: 10.18632/oncotarget.13809

31. Ferraro B, Bepler G, Sharma S, Cantor A, Haura EB. EGR1 predicts PTEN and survival in patients with non-small-cell lung cancer. *J Clin Oncol* (2005) 23(9):1921–6. doi: 10.1200/JCO.2005.08.127

32. Sui X, Kong N, Ye L, Han W, Zhou J, Zhang Q, et al. p38 and JNK MAPK pathways control the balance of apoptosis and autophagy in response to chemotherapeutic agents. *Cancer Lett* (2014) 344(2):174–9. doi: 10.1016/j.canlet.2013.11.019

33. Wang M-s, Han Q-s, Jia Z-r, Chen C-s, Qiao C, Liu Q-q, et al. PPAR alpha agonist fenofibrate relieves acquired resistance to gefitinib in non-small cell lung cancer by promoting apoptosis via PPAR alpha/AMPK/AKT/FoxO1 pathway. *Acta Pharmacologica Sin* (2021) 43(1):167–76. doi: 10.1038/s41401-021-00638-z

34. Hann SS, Zheng F, Zhao S. Targeting 3-phosphoinositide-dependent protein kinase 1 by n-acetyl-cysteine through activation of peroxisome proliferators activated receptor alpha in human lung cancer cells, the role of p53 and p65. *J Exp Clin Cancer Res* (2013) 32. doi: 10.1186/1756-9966-32-43

35. Basseres DS, Ebbs A, Levantini E, Baldwin AS. Requirement of the NF-kappa b subunit p65/RelA for K-Ras-Induced lung tumorigenesis. *Cancer Res* (2010) 70 (9):3537–46. doi: 10.1158/0008-5472.CAN-09-4290

36. Stathopoulos GT, Sherrill TP, Han W, Sadikot RT, Polosukhin VV, Fingleton B, et al. Use of bioluminescent imaging to investigate the role of nuclear factor-kappa b in experimental non-small cell lung cancer metastasis. *Clin Exp Metastasis* (2008) 25 (1):43–51. doi: 10.1007/s10585-007-9100-z

37. Lomert E, Turoverova L, Kriger D, Aksenov ND, Nikotina AD, Petukhov A, et al. Co-Expression of RelA/p65 and ACTN4 induces apoptosis in non-small lung carcinoma cells. *Cell Cycle* (2018) 17(5):616–26.

38. Beaupre DM, Talpaz M, Marini FC 3rd, Cristiano RJ, Roth JA, Estrov Z, et al. Autocrine interleukin-1 beta production in leukemia: Evidence for the involvement of mutated RAS. *Cancer Res* (1999) 59(12):2971–80. (Print)

39. Zhang L, Bu L, Hu J, Xu Z, Ruan L, Fang Y, et al. HDAC1 knockdown inhibits invasion and induces apoptosis in non-small cell lung cancer cells. *Biol Chem* (2018) 399(6):603–10. doi: 10.1515/hsz-2017-0306

40. Wang L, Li H, Ren Y, Zou S, Fang W, Jiang X, et al. et al. targeting HDAC with a novel inhibitor effectively reverses paclitaxel resistance in non-small cell lung cancer. *via multiple mechanisms. Cell Death Dis* (2016) 7. doi: 10.1038/cddis.2015.328

41. Sun Y, Bao X, Ren Y, Jia L, Zou S, Han J, et al. et al. targeting HDAC/OAZ1 axis with a novel inhibitor effectively reverses cisplatin resistance in non-small cell lung cancer. *Cell Death Dis* (2019) 10. doi: 10.1038/s41419-019-1597-y

42. Tang Y-A, Chen C-H, Sun HS, Cheng C-P, Tseng VS, Hsu H-S, et al. Global Oct4 target gene analysis reveals novel downstream PTEN and TNC genes required for drug-resistance and metastasis in lung cancer. *Nucleic Acids Res* (2015) 43(3):1593–608. doi: 10.1093/nar/gkv024

43. Gorbacheva AM, Kuprash DV, Mitkin NA. Regulation of IL33 gene expression by SP1 and Foxa1 in breast and lung cancer cells. *Mol Biol* (2021) 55(1):92–101. doi: 10.1134/S0026893321010064

44. Cheng L, Tong Q. Interaction of FLNA and ANXA2 promotes gefitinib resistance by activating the wnt pathway in non-small-cell lung cancer. *Mol Cell Biochem* (2021) 476(10):3563–75. doi: 10.1007/s11010-021-04179-1

45. Kossenkov AV, Vachani A, Chang C, Nichols C, Billouin S, Horng W, et al. Resection of non-small cell lung cancers reverses tumor-induced gene expression changes in the peripheral immune system. *Clin Cancer Res* (2011) 17(18):5867–77. doi: 10.1158/1078-0432.CCR-11-0737

46. Gao N, Ye B. Original SPI1-induced upregulation of lncRNA SNHG6 promotes non-small cell lung cancer via miR-485-3p/VPS45 axis. *Biomed Pharmacother* (2020) 129. doi: 10.1016/J.Biopha.2020.110239

47. Vendramini-Costa DB, Carvalho JE. Molecular link mechanisms between inflammation and cancer. *Curr Pharm Design* (2012) 18(26):3831–52. doi: 10.2174/138161212802083707

48. Fujita Y, Yagishita S, Hagiwara K, Yoshioka Y, Kosaka N, Takeshita F, et al. The clinical relevance of the miR-197/CKS1B/STAT3-mediated PD-L1 network in chemoresistant non-small-cell lung cancer. *Mol Ther* (2015) 23(4):717–27. doi: 10.1038/mt.2015.10

49. Fang Z, Chen W, Yuan Z, Liu X, Jiang H. LncRNA-MALAT1 contributes to the cisplatin-resistance of lung cancer by upregulating MRP1 and MDR1 *via* STAT3 activation. *Biomed Pharmacother* (2018) 101:536–42. doi: 10.1016/j.biopha.2018.02.130

50. Zhao J, Sun X, Ji D, Chen J, Wang M, Jiang L, et al. Influence of STAT1 on proliferation and IFN-beta sensitivity of human nonsmall-cell lung cancer H1299 cells. *Chin J Pathophysiol* (2015) 31(5):852–6.

51. Yang L, Dong Y, Li Y, Wang D, Liu S, Wang D, et al. IL-10 derived from M2 macrophage promotes cancer stemness *via* JAK1/STAT1/NF-kappa B/Notch1 pathway in non-small cell lung cancer. *Int J Cancer* (2019) 145(4):1099–110. doi: 10.1002/ijc.32151

52. Huang Q, Duan L, Qian X, Fan J, Lv Z, Zhang X, et al. IL-17 promotes angiogenic factors IL-6, IL-8, and vegf production. *via Stat1 Lung Adenocarcinoma Sci Rep* (2016) 6. doi: 10.1038/srep36551

53. Chen Q, Boire A, Jin X, Valiente M, Er EE, Lopez-Soto A, et al. Carcinoma-astrocyte gap junctions promote brain metastasis by cGAMP transfer. *Nature* (2016) 533(7604):493–+. doi: 10.1038/nature18268

54. Gagliano T, Bellio M, Gentilin E, Mole D, Tagliati F, Schiavon M, et al. mTOR, p70S6K, AKT, and ERK1/2 levels predict sensitivity to mTOR and PI3K/mTOR inhibitors in human bronchial carcinoids. *Endocrine-Related Cancer* (2013) 20 (4):463–75. doi: 10.1530/ERC-13-0042

55. Ji GH, Cui Y, Yu H, Cui XB. Profiling analysis of FOX gene family members identified FOXE1 as potential regulator of NSCLC development. *Cell Mol Biol* (2016) 62(11):57–62.

56. Hu W, Jin P, Ding C, Liu W. miR-19a/b modulates lung cancer cells metastasis through suppression of MXD1 expression. *Oncol Lett* (2016) 12(3):1901–5. doi: 10.3892/ol.2016.4881

57. Ruiz Esparza-Garrido R, Eugenia Torres-Marquez M, Viedma-Rodriguez R, Claudia Velazquez-Wong A, Salamanca-Gomez F, Rosas-Vargas H, et al. Breast cancer cell line MDA-MB-231 miRNA profile expression after BIK interference: BIK involvement in autophagy. *Tumor Biol* (2016) 37(5):6749–59. doi: 10.1007/s13277-015-4494-8

58. Wang G, Zhang Z, Yang M, Xu B, Gao Q, Yang X. Comparative proteomics analysis of human osteosarcoma by 2D DIGE with MALDI-TOF/TOF MS. *J Bone Oncol* (2016) 5(4):147–52. doi: 10.1016/j.jbo.2016.05.002

59. Zhang D-L, Qu L-W, Ma L, Zhou Y-C, Wang G-Z, Zhao X-C, et al. Genome-wide identification of transcription factors that are critical to non-small cell lung cancer. *Cancer Lett* (2018) 434:132–43. doi: 10.1016/j.canlet.2018.07.020

60. Cho HY, Reddy SP, Kleeberger SR. Nrf2 defends the lung from oxidative stress. *Antioxidants Redox Signaling* (2006) 8(1-2):76–87. doi: 10.1089/ars.2006.8.76

61. Girardot M, Pecquet C, Chachoua I, Van Hees J, Guibert S, Ferrant A, et al. Persistent STAT5 activation in myeloid neoplasms recruits p53 into gene regulation. *Oncogene* (2015) 34(10):1323–32. doi: 10.1038/onc.2014.60

62. He T, Shen H, Wang S, Wang Y, He Z, Zhu L, et al. MicroRNA-3613-5p promotes lung adenocarcinoma cell proliferation through a RELA and AKT/MAPK positive feedback loop. *Mol Therapy-Nucleic Acids* (2020) 22:572–83. doi: 10.1016/j.omtn.2020.09.024

63. Buonato JM, Lazzara MJ. ERK1/2 blockade prevents epithelial-mesenchymal transition in lung cancer cells and promotes their sensitivity to EGFR inhibition. *Cancer Res* (2014) 74(1):309–19. doi: 10.1158/0008-5472.CAN-12-4721

64. Kapellos TS, Bonaguro L, Gemuend I, Reusch N, Saglam A, Hinkley ER, et al. Human monocyte subsets and phenotypes in major chronic inflammatory diseases. *Front Immunol* (2019) 10. doi: 10.3389/fimmu.2019.02035

65. Krieg C, Nowicka M, Guglietta S, Schindler S, Hartmann FJ, Weber LM, et al. High-dimensional single-cell analysis predicts response to anti-PD-1 immunotherapy. *Nat Med* (2018) 24(2):144–+. doi: 10.1038/nm.4466

66. Singhal SA-O, Stadanlick JA-O, Annunziata MJ, Rao AS, Bhojnagarwala PA-O, O'Brien SA-O, et al. Human tumor-associated monocytes/macrophages and their regulation of T cell responses in early-stage lung cancer., 1946–6242. (Electronic)

67. Yuan K, Feng Y, Wang H, Zhao L, Wang W, Wang T, et al. FGL2 is positively correlated with enhanced antitumor responses mediated by T cells in lung adenocarcinoma. *Peerj* (2020) 8. doi: 10.7717/peerj.8654

68. Cruse G, Beaven MA, Music SC, Bradding P, Gilfillan AM, Metcalfe DD. The CD20 homologue MS4A4 directs trafficking of KIT toward clathrin-independent endocytosis pathways and thus regulates receptor signaling and recycling. *Mol Biol Cell* (2015) 26(9):1711–27. doi: 10.1091/mbc.E14-07-1221

69. Yamada H, Yanagisawa K, Tokumaru S, Taguchi A, Nimura Y, Osada H, et al. Detailed characterization of a homozygously deleted region corresponding to a candidate tumor suppressor locus at 21q11-21 in human lung cancer. *Genes Chromosomes Cancer* (2008) 47(9):810–8. doi: 10.1002/gcc.20582

70. Li W, Zhang X, Wang W, Sun R, Liu B, Ma Y, et al. Quantitative proteomics analysis of mitochondrial proteins in lung adenocarcinomas and normal lung tissue using iTRAQ and tandem mass spectrometry. *Am J Trans Res* (2017) 9(9):3918–34.

71. Wang Y, Fu D, Su J, Chen Y, Qi C, Sun Y, et al. C1QBP suppresses cell adhesion and metastasis of renal carcinoma cells. *Sci Rep* (2017) 7. doi: 10.1038/s41598-017-01084-w

72. Kemp SB, Steele NG, Carpenter ES, Donahue KL, Bushnell GG, Morris AH, et al. Pancreatic cancer is marked by complement-high blood monocytes and tumor-associated macrophages. *LID*, 2575–1077. doi: 10.26508/Lsa.202000935

73. Tangye SG, Nichols KE, Hare NJ, van de Weerdt BC. Functional requirements for interactions between CD84 and src homology 2 domain-containing proteins and their contribution to human T cell activation., 0022–1767. (Print)

74. Xing Y, Zhang J, Lu L, Li D, Wang Y, Huang S, et al. Identification of hub genes of pneumocyte senescence induced by thoracic irradiation using weighted gene co-expression network analysis. *Mol Med Rep* (2016) 13(1):107–16. doi: 10.3892/mmr.2015.4566

75. Sintes J, Romero X, de Salort J, Terhorst C, Engel P. Mouse CD84 is a pan-leukocyte cell-surface molecule that modulates LPS-induced cytokine secretion by macrophages., 1938–3673. (Electronic)

76. Wang TTY, Pham Q, Kim YS. Elucidating the role of CD84 and AHR in modulation of LPS-induced cytokines production by cruciferous vegetable-derived compounds indole-3-Carbinol and 3,3'-diindolylmethane. *Int J Mol S* (2018) 19(2). doi: 10.3390/ijms19020339

77. Vuong H, Cheng F, Lin C-C, Zhao Z. Functional consequences of somatic mutations in cancer using protein pocket-based prioritization approach. *Genome Med* (2014) 6. doi: 10.1186/s13073-014-0081-7

78. Wu D-M, Deng S-H, Zhou J, Han R, Liu T, Zhang T, et al. PLEK2 mediates metastasis and vascular invasion *via* the ubiquitin-dependent degradation of SHIP2 in non-small cell lung cancer. *Int J Cancer* (2020) 146(9):2563–75. doi: 10.1002/ijc.32675

79. Bai M, Pan Q, Sun C. Tumor purity coexpressed genes related to immune microenvironment and clinical outcomes of lung adenocarcinoma. *J Oncol* (2021) 2021. doi: 10.1155/2021/9548648

80. Parra ER, Jiang M, Machado-Rugolo J, Yaegashi LB, Prieto T, Farhat C, et al. Variants in epithelial-mesenchymal transition and immune checkpoint genes are associated with immune cell profiles and predict survival in non-small cell lung cancer. *Arch Pathol Lab Med* (2020) 144(10):1234–44. doi: 10.5858/arpa.2019-0419-OA

81. Han M-l, Zhao Y-f, Tan C-h, Xiong Y-j, Wang W-j, Wu F, et al. Liang z-q: Cathepsin l upregulation-induced EMT phenotype is associated with the acquisition of cisplatin or paclitaxel resistance in A549 cells. *Acta Pharmacologica Sin* (2016) 37 (12):1606–22. doi: 10.1038/aps.2016.93

82. Duan X, Fu Z, Gao L, Zhou J, Deng X, Luo X, et al. Direct interaction between miR-203 and ZEB2 suppresses epithelial-mesenchymal transition signaling and reduces lung adenocarcinoma chemoresistance. *Acta Biochim Et Biophys Sin* (2016) 48(11):1042–9. doi: 10.1093/abbs/gmw099

83. Jiang M, Zhong T, Zhang W, Xiao Z, Hu G, Zhou H, et al. Reduced expression of miR-205-5p promotes apoptosis and inhibits proliferation and invasion in lung cancer A549 cells by upregulation of ZEB2 and downregulation of erbB3. *Mol Med Rep* (2017) 15(5):3231–8. doi: 10.3892/mmr.2017.6398

84. Liu Q, Chen J, Wang B, Zheng Y, Wan Y, Wang Y, et al. miR-145 modulates epithelial-mesenchymal transition and invasion by targeting ZEB2 in non-small cell lung cancer cell lines. *J Cell Biochem* (2019) 120(5):8409–18. doi: 10.1002/jcb.28126

85. Jin L, Jiang X, He G, Shi J, Su F, Zhu H. SNHG20 knockdown suppresses proliferation, migration and invasion, and promotes apoptosis in non-small cell lung cancer through acting as a miR-154 sponge. *Biomed Pharmacother* (2019) 112. doi: 10.1016/J.Biopha.2019.108648

Glossary

DRGs	Differentiation Related Genes
TAMs	Tumor Associated Macrophages
NTAMs	Non Tumor Associated Macrophages
NSCLC	Non-Small Cell Lung Cancer
scRNA-seq	Single Cell RNA Sequencing
RS	Risk score
OS	Overall Survival
TME	Tumor Microenvironment
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
GEO	the Gene Expression Omnibus
TCGA	the Cancer Genome Atlas
PCA	Principal Component Analysis
PCs	Principal Components
GEPIA	Gene Expression Profiling Interactive Analysis
GTEx	Genotype-Tissue Expression
USCS	University of California Santa Cruz
TFs	Transcription Factors
HR	Hazard Ratio
ROC	Receiver Operating Characteristic
K-M	Kaplan-Meier
LASSO	Least absolute shrinkage and selection operator
t-SNE	t-Stochastic Neighbor Embedding
UMAP	Uniform Manifold Approximation and Projection for Dimension Reduction