Original Article
Differences in key genes in human alveolar macrophages between phenotypically normal smokers and nonsmokers: diagnostic and prognostic value in lung cancer

Yi-De Wang\textsuperscript{1,}\textsuperscript{*}, Zheng Li\textsuperscript{2,}\textsuperscript{*}, Feng-Sen Li\textsuperscript{2}

\textsuperscript{1}Department of Integrated Pulmonology, Fourth Affiliated Hospital of Xinjiang Medical University, Urumqi 830000, China; \textsuperscript{2}Xinjiang National Clinical Research Base of Traditional Chinese Medicine, Xinjiang Medical University, Urumqi 830000, China. \textsuperscript{*}Equal contributors

Received June 29, 2020; Accepted September 2, 2020; Epub November 1, 2020; Published November 15, 2020

Abstract: Objective: To explore the effect of smoking on gene expression in human alveolar macrophages and the value of identified key genes in the early diagnosis and prognosis of lung cancers. Methods: We downloaded three data sets (GSE8823, GSE2125, and GSE3212) from the Gene Expression Omnibus (GEO) database, including 31 non-smoking and 33 smoking human alveolar macrophage samples. We identified common differentially expressed genes (DEGs), from which we obtained module genes and hub genes by using STRING and Cytoscape. Then we analyzed the protein-protein interaction (PPI) network of DEGs, hub genes, and module genes and used David online analysis tool to carry out functional enrichment analysis of DEGs and module genes. Results: A total of 85 differentially expressed genes was obtained, including 42 up-regulated genes and 43 down-regulated genes. The Human Protein Atlas and Survival analysis showed that GBP1, ITGAM, CSF1, SPP1, COL1A1, LAMB1 and THBS1 may be closely associated with the carcinogenesis and prognosis of lung cancer. Conclusion: DEGs, module, and hub genes identified in the present study help explain the effects of smoking on human alveolar macrophages and provide candidate targets for diagnosis and treatment of smoking-related lung cancer.

Keywords: Smoking, alveolar macrophages, smoking-related lung cancers, diagnosis, prognosis

Introduction

Cigarette smoke is so harmful that a retrospective study found that secondhand smoke alone killed more than 600,000 people worldwide [1]. Smoking is also a significant risk factor for lung cancer and other lung diseases. It mainly causes damage to the lungs and other organs through the inflammatory response, oxidative stress, genetic changes, cell aging and different pathways [2]. Studies have found that smoking causes changes in alveolar macrophage genes involved in the occurrence and development of lung cancer [3]. For example, tumor-associated macrophages (TAM) can polarize into both M1 and M2 phenotypes. The M1 type is involved in the inflammatory and toxic reactions of tumor cells [4]. Transforming growth factor secreted by M2-type macrophages is associated with tumor progression [5]. In the face of intractable lung cancer, many studies have shown that intervention in the tumor microenvironment from the perspective of alveolar macrophages reduces cancer recurrence and development. In recent years, an anti-tumor strategy centering on macrophages by limiting the infiltration of cancer cells has been suggested as a cancer treatment method [6]. However, the study of the effects of smoking on gene expression in human alveolar macrophages is still immature. Therefore, it is necessary to strengthen the research on the molecular mechanism of the occurrence and development of lung cancer mediated by changes in alveolar macrophages caused by tobacco smoke exposure.

With rapid development in recent years, microarray technology has become more convenient and economical, and it can screen different genes in high throughput for research on the molecular mechanism of related diseases [7].
Identification of key genes in alveolar macrophages

The limitation of false positives, and the limited number of included samples reduce reliability of the analysis results. Based on microarray technology’s application to generate a large number of data sets, bioinformatics can efficiently utilize the above data to obtain more reliable results [8]. Therefore, we used bioinformatic tools to get three expression profile data sets, including GSE8823, GSE2125, and GSE3212. The DEGs of alveolar macrophages in smoke-exposed and non-smoking individuals were analyzed. Secondly, we screened out the most significant module gene and Hub gene and constructed the protein interaction network diagram of DEGs, module genes, and Hub genes. Then we conducted a functional enrichment analysis of the identified DEGs and module genes. Third, we analyzed the differential expression of Hub genes in lung cancer patients and healthy control groups and further selected 7 Hub genes to explain their value in predicting the prognosis in lung cancer. Finally, we tried to explore the influence of Hub gene mutation on the survival of lung cancer patients. In summary, the differentially identified genes may be therapeutic targets and biomarkers for smoking-related lung disease.

Materials and methods

Microarray data

Data came from GEO (http://www.ncbi.nlm.nih.gov/geo) [9], an online public functional genome data warehouse. We downloaded three gene expression datasets (GSE8823 [10], GSE2125 [11], and GSE3212 [12]). GSE8823 included 11 non-smoking alveolar macrophages samples and 13 smoking alveolar macrophages samples, based on platform GPL570[hg-u133_plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. GSE3212 included five non-smoking alveolar macrophage samples and five smoking alveolar macrophage samples, based on platform GPL80[Hu6800] Affymetrix Human full-length HuGeneFL Array. GSE2125 included 15 non-smoking alveolar macrophage samples and 15 smoking alveolar macrophage samples, based on platform GPL570[hg-u133_plus_2] Affymetrix Human Genome U133 Plus 2.0 Array.

Identification of DEGs

To further analyze the DEGs between smoking alveolar macrophages and non-smoking alveolar macrophages, we used an interactive online analysis tool--GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r). The deletion occurs when one gene corresponds to multiple probes, or one probe does not have a corresponding homologous gene. \(|\log FC\) (fold change) \(\geq 1\) and adj. \(p\)-value <0.01 were considered statistically significant.

PPI network construction, module analysis, and Hub genes selection

To explore the interactions between proteins and further explore the mechanism of disease occurrence or development, we first constructed a PPI network for screened DEGs based on an online protein interaction prediction tool -STRING [13]. Then we imported the resulting data into Cytoscape (version 3.7.2) [14] for analysis. Cytoscape is a biological information platform for visualization of molecular interactions. Third, we use Cytoscape’s plug-in MCODE (version 1.5.1) [15] to filter essential modules in the network diagram. The detailed criteria were as follows: Max depth =100, k-score =2, node score cutoff =0.2, degree cutoff =2, and find clusters in the whole network. Finally, we used cytoHubba [16], a plug-in of Cytoscape, to screen the Hub genes in the network graph, and obtained the top 20 genes with the highest score based on the MCC algorithm.

Functional enrichment analysis of DEGs

The Database for Annotation, Visualization, and Integrated Discovery (DAVID; http://david.ncifcrf.gov) (version 6.8) [17] is an online Database for feature enrichment. We performed GO enrichment analyses on screened DEGs, including biological process (BP), cellular component (CC), molecular function (MF), and KEGG analysis [18]. The visual network diagram presented the GO enrichment analysis results, and Cytoscape plugin, BiNGO (version 3.0.3) [19] drew the network diagram. The setting criteria were as follows: Organism selects Homo sapiens, Ontology file selects GO_full, P<0.05 was considered statistically significant.

Hub genes analysis

First, we conducted KEGG and GO enrichment analysis on Hub genes, and the specific operation was the same as DEGs enrichment analysis. Then the hierarchical clustering of Hub genes is constructed by using UCSC online
tools (https://xenabrowser.net/heatmap/). The Gene Expression Profiling Interactive Analysis (GEPIA, http://gepia.cancer-pku.cn/), UALCAN [20], and The Human Protein Atlas (https://www.proteinatlas.org) were used to analyze the differential expression of Hub genes in lung adenocarcinoma (LUAD), Lung squamous cell carcinoma (LUSC), and standard tissue samples. Among them, immunohistochemical (IHC) results from the Human Protein Atlas database were used to verify the translation level of hub genes. Fourth, we used UALCAN--an online platform for analyzing cancer omics, screening Genes correlated with the Hub Genes in LUSC. Finally, we used Kaplan-Meier plotter to draw the Kaplan-Meier survival graph [21]. Kaplan-Meier plotter is capable of detecting the effect of any gene or gene combination on survival in a variety of tumors. All forms of gene changes such as Missense Mutation, Truncating Mutation, and Amplification of Hub genes in patients with LUSC were analyzed using the cBioPortal database [22]. Furthermore, whether the genetic changes affect the overall and disease-free survival of LUSC patients was further investigated.

Figure 1. Identification of differentially expressed genes in human alveolar macrophages between smokers and non-smokers. A, B, abd C: GEO2R was used to analyze the expression of DEGs in three data sets (GSE8823, GSE2125, and GSE3212 respectively, which are presented as a volcano plots. The red and green dots represent up and down, respectively. D: Venn diagram shows that there are 85 DEGs in the three databases.
Identification of key genes in alveolar macrophages

Results

Identification of 85 DEGs shared by three gene expression datasets

Three microarray datasets (GSE8823, GSE2125, and GSE3212) were obtained through strict screening, including 1205,974 and 1312 DEGs, respectively. The analysis results were presented by a volcano diagram (Figure 1A-C). There were 85 overlapping genes in the three datasets, including 42 up-regulated genes and 43 down-regulated genes, as shown by the Venn diagram (Figure 1D).

PPI network construction, module analysis, and Hub gene selection

First, we constructed the PPI network of DEGs with STRING (Figure 2A). Then, we used Cytoscape’s plug-in MCODE to screen the most significant module genes in the system (Figure 2B). BiNGO, a plug-in of Cytoscape, was used to obtain 20 Hub genes in PPI network (Figure 2C). All Hub genes are shown in Table 1, including gene symbol, full name and primary function.

Functional enrichment analysis of DEGs and the most significant module

We used the DAVID tool to carry out GO and KEGG enrichment analysis of DEGs. The results of the GO analysis showed that the biological processes (BP) of DEGs were significantly enriched in cell adhesion, fatty acid metabolic process, immune response, inflammatory response, and positive regulation of angiogenesis (Figure 3A). Cellular component (CC) of DEGs is
Identification of key genes in alveolar macrophages

Table 1. Functional roles of 20 hub genes

<table>
<thead>
<tr>
<th>No.</th>
<th>Gene symbol</th>
<th>Full name</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GBP1</td>
<td>Guanylate binding protein 1</td>
<td>Overexpression of GBP1 promotes the progress of cervical cancer</td>
</tr>
<tr>
<td>2</td>
<td>IFIT2</td>
<td>Interferon induced protein with tetratrico- peptide repeats 2</td>
<td>Low IFIT2 expression is associated with poor survival in non-small cell lung cancer.</td>
</tr>
<tr>
<td>3</td>
<td>IFIT1</td>
<td>Interferon induced protein with tetratrico-peptide repeats 1</td>
<td>High expression of IFIT1 in head and neck squamous cell carcinoma is associated with poor prognosis</td>
</tr>
<tr>
<td>4</td>
<td>IFI44L</td>
<td>Interferon induced protein 44 like</td>
<td>IFI44L is associated with metastasis and drug resistance of liver cancer</td>
</tr>
<tr>
<td>5</td>
<td>MX2</td>
<td>MX dynamin like GTPase 2</td>
<td>MX2 is a restriction factor of hepatitis B virus replication.</td>
</tr>
<tr>
<td>6</td>
<td>IFI27</td>
<td>Interferon alpha inducible protein 27</td>
<td>Down-regulation of the IFI27 gene inhibited the proliferation of oral squamous cell carcinoma</td>
</tr>
<tr>
<td>7</td>
<td>IGF1</td>
<td>Insulin like growth factor 1</td>
<td>Association of serum IGF1 and adiponectin are involved in the metastasis of breast cancer.</td>
</tr>
<tr>
<td>8</td>
<td>ITGAM</td>
<td>Integrin subunit alpha M</td>
<td>ITGAM is a risk factor for systemic lupus erythematosus and may also be a protective factor for rheumatoid arthritis</td>
</tr>
<tr>
<td>9</td>
<td>ITPKB</td>
<td>Inositol-trisphosphate 3-kinase B</td>
<td>ITPKB may be associated with lung cancer cell metastasis</td>
</tr>
<tr>
<td>10</td>
<td>CSF1</td>
<td>Colony stimulating factor 1</td>
<td>CSF1 is involved in regulating macrophage differentiation and promoting the release of pro-inflammatory factors</td>
</tr>
<tr>
<td>11</td>
<td>SPP1</td>
<td>Secreted phosphoprotein 1</td>
<td>SPP1 is closely associated with non-small cell lung cancer, breast cancer, colorectal cancer and other cancers.</td>
</tr>
<tr>
<td>12</td>
<td>IL7</td>
<td>Interleukin 7</td>
<td>IL7 may have anti-tumor effects</td>
</tr>
<tr>
<td>13</td>
<td>TNFSF10</td>
<td>TNF superfamily member 10</td>
<td>Viral therapy using TNFSF10 combined with plasminogen k5 may treat cancer</td>
</tr>
<tr>
<td>14</td>
<td>COL1A1</td>
<td>Collagen type I alpha 1 chain</td>
<td>Down-regulation of COL1A1 expression inhibits the growth and metastasis of breast cancer cells</td>
</tr>
<tr>
<td>15</td>
<td>LAMB1</td>
<td>Laminin subunit beta 1</td>
<td>LAMB1 may be a biomarker for colorectal cancer</td>
</tr>
<tr>
<td>16</td>
<td>CD36</td>
<td>CD36 molecule</td>
<td>CD36 mediates inflammation, molecular adhesion, and apoptosis</td>
</tr>
<tr>
<td>17</td>
<td>MDM2</td>
<td>MDM2 proto-oncogene</td>
<td>MDM2 inhibitors may have anti-cancer effects.</td>
</tr>
<tr>
<td>18</td>
<td>THBS1</td>
<td>Thrombospondin 1</td>
<td>Activation of THBS1 promotes metastasis of oral squamous cell carcinoma.</td>
</tr>
<tr>
<td>19</td>
<td>AR</td>
<td>Androgen receptor</td>
<td>AR antagonists may be used in the treatment of breast cancer</td>
</tr>
<tr>
<td>20</td>
<td>FABP4</td>
<td>Fatty acid binding protein 4</td>
<td>FABP4 inhibitor inhibited oxidative stress in mice with acute lung injury</td>
</tr>
</tbody>
</table>

mainly enriched in the cell fraction, plasma membrane part, plasma membrane, extracellular region, and extracellular region (Figure 3B). The molecular functions (MF) of DEGs were significantly enriched in zinc ion binding, ribonuclease A activity, prostaglandin E receptor activity, 3',5'-cyclic-amp phosphodiesterase activity, and camp-dependent protein kinase inhibitor activity (Figure 3C). KEGG analysis showed that DEGs mainly concentrated in Hematopoietic cell lineage, Transcriptional misregulation in cancer, pi3k-Akt signaling pathway, focal adhesion, and ECM-receptor interaction (Figure 3D). The results of GO enrichment analysis are presented by the visual network diagram (Figure 3E).

Functional enrichment analysis of the most significant module

DAVID was used for the GO and KEGG enrichment analysis of the most significant module genes. The GO analysis results showed that the BP of module genes were mainly enriched in cell adhesion, fatty acid metabolic process, inflammatory response, immune response, and receptor internalization (Figure 4A). CC changes of module genes are mainly concentrated in extracellular space, extracellular exosome, extracellular region, external side of the plasma membrane, and cytosol (Figure 4B). The variations of MF of module genes were mainly concentrated in zinc ion binding, ribonuclease A activity, prostaglandin E receptor activity, cAMP-dependent protein kinase inhibitor activity, and 3',5'-cyclic nucleotide phosphodiesterase activity (Figure 4C). KEGG analysis showed that module genes were mainly enriched in Hematopoietic cell lineage, pi3k-Akt signaling pathway, Focal adhesion, ECM-receptor interaction, and Transcriptional misregulation in cancer (Figure 4D).

Hub gene analysis

First, we performed hierarchical clustering analysis on the 20 previously screened Hub genes, based on the UCSC online tool. The analysis results showed that there was a difference between the gene expression of the Hub
Identification of key genes in alveolar macrophages

Figure 3. GO and KEGG enrichment analyses of DEGs. Enrichment analysis of BP (A), CC (B), MF (C), and KEGG (D), was carried out by DAVID. (E) The results of GO enrichment analysis are presented as a visual network diagram.

genes in lung cancer tissue samples and standard tissue samples (Figure 5). Then, we used the GEPIA database and UALCAN cancer database to analyze the differential expression of 20 Hub genes in LUAD, LUSC, and standard tissue samples. The results showed that 7 Hub genes were found to be different, including GBP1, COL1A1, SPP1, ITGAM, CSF1, LAMB1, and THBS1. The expression of GBP1 and LAMB1 in LUSC tissues was lower than that in healthy tissues (Figure 6A, 6B, 6K and 6L). Compared with standard tissue samples, the two Hub genes SPP1 and COL1A1 were highly expressed in LUSC patients, and the three Hub genes ITGAM, CSF1 and THBS1 in lung tissues of LUSC patients were down-regulated, with
Identification of key genes in alveolar macrophages

Figure 4. GO and KEGG enrichment analyses of the most significant module. Enrichment analysis of BP (A), CC (B), MF (C), and KEGG (D) was carried out by DAVID.

Figure 5. Heatmap of hub genes. Blue is the Primary Tumor specimen, and dark red is Solid Tissue Normal. Red indicates high gene expression, and blue indicates low gene expression.
Identification of key genes in alveolar macrophages
Identification of key genes in alveolar macrophages

Figure 6. Differential expression of 7 Hub genes in lung cancer and healthy lung tissues. Compared with normal tissues, the expressions of GBP1, ITGAM, CSF1, LAMB1, THBS1 in NSCLC patients were down-regulated, and the expressions of SPP1 and COL1A1 were increased. In addition to GBP1, the difference was statistically significant (*P<0.05). Lung adenocarcinoma (left column); squamous cancer (right column).

statistically significant differences (P<0.05; Figure 6B-J and 6M-N). Subsequently, IHC staining results obtained from the Human Protein Atlas database were used to verify the high expression of COL1A1 and SPP1 Hub genes, as shown in Figure 7A-L, which showed that the expression levels of these two proteins were higher in LUSC patients than in the healthy control group. We analyzed the top 20 genes most closely related to the seven hub genes in Table 2. Further, we used the KM-plotter to plot the Kaplan-Meier curve of 7 Hub genes. High expression of SPP1 and COL1A1 genes were associated with a better survival rate in lung cancer patients. In contrast, high expression of the other five hub genes SPP1, ITGAM, CSF1, LAMB1, and THBS1 was associated with lower OS, all of which had statistical significance (P<0.05; Figure 8A-G). Finally, we analyzed the gene mutation of 7 Hub genes, and the results were as follows (Figure 9A-G): in 167 cases, 6 percent of GBP1 had gene change: 1 case Truncating Mutation, 1 case Amplification, and eight instances mRNA High. In 183 cases, 6 percent of the ITGAM gene from genetically altered: 6 cases Missense Mutation, 1 case Amplification, and three instances mRNA High. Among the 200 samples, 3 percent of CSF1 had gene changes: 2. Missense Mutation, 3 Deep Deletion, and one mRNA High. Among 182 cases, 11 percent of SPP1 had gene change: 1 case Missense Mutation, 1 case Amplification, 1 case Deep Deletion, and 17 instances mRNA High. Among the 175 samples, 12 percent of COL1A1 had gene changes. Among 278 samples, 9 percent of LAMB1 had gene changes: 1 Truncating Mutation, 3 Amplification, and mRNA High mixture. 4 Missense Mutation, 2 Truncating Mutation, 4 Amplification, and six mRNA High. Among the 200 samples, 3 percent of THBS1 had gene changes: 1 case of Missense Mutation, five instances of mRNA High. Mutations in all of the above genes were not associated with overall survival or disease-free survival of LUSC patients (Figure 10A-N).

Discussion

Cigarette smoke remains a significant risk factor for lung diseases such as COPD and lung
Identification of key genes in alveolar macrophages

Complex. A growing number of fundamental studies suggest that smoking-induced changes in alveolar macrophages may play an important role. In addition to increasing the number and activation of alveolar macrophages, smoke also induces their secretion of inflammatory mediators, adhesion molecules, and other mediators to increase expression, thus causing lung injury [26]. Therefore, it is urgent to find new candidate genes for smoking-induced lung cancer and screen out the.

Figure 7. IHC experiment showed that the expression of SPP1 and COL1A1 in NSCLC was higher than that in healthy tissues. (A-C) SPP1 was low in normal tissues, and (D-F) was moderately expressed in COL1A1 tissues. (G-I) COL1A1 expression was low in normal tissues, and (J-L) expression was up-regulated in NSCLC.

A recent study based on 2770 participants showed that smoking had an odds ratio (OR) of 2.63 (95% CI 1.86-3.73) for COPD [23]. In a systematic review of 12 lung cancer studies conducted by [24], the lung cancer prevalence of passive smokers was RR 1.29 (95% CI 1.10-1.39, P<0.001). In a study by [25], lung cancer mortality in male and female smokers was 4.46 (95% CI 3.10-6.41) and 3.58 (95% CI 2.24-5.73), respectively. The molecular mechanisms of smoking-induced lung cancer are very complex. A growing number of fundamental studies suggest that smoking-induced changes in alveolar macrophages may play an important role. In addition to increasing the number and activation of alveolar macrophages, smoke also induces their secretion of inflammatory mediators, adhesion molecules, and other mediators to increase expression, thus causing lung injury [26]. Therefore, it is urgent to find new candidate genes for smoking-induced lung cancer and screen out the.
Identification of key genes in alveolar macrophages

<table>
<thead>
<tr>
<th>No.</th>
<th>GBP1 genes</th>
<th>ITGAM genes</th>
<th>CSF1 genes</th>
<th>SPP1 genes</th>
<th>COL1A1 genes</th>
<th>LAMB1 genes</th>
<th>THBS1 genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GBP2</td>
<td>SH2B3</td>
<td>GPR84</td>
<td>CBR3</td>
<td>COL5A1</td>
<td>FSTL1</td>
<td>BICC1</td>
</tr>
<tr>
<td>2</td>
<td>BATF2</td>
<td>NFAM1</td>
<td>CSF1R</td>
<td>SRXN1</td>
<td>COL3A1</td>
<td>LEPRE1</td>
<td>CRISPLD2</td>
</tr>
<tr>
<td>3</td>
<td>STAT1</td>
<td>ITGB2</td>
<td>CACNA2D4</td>
<td>GCLM</td>
<td>COL6A3</td>
<td>FN1</td>
<td>ITGA1</td>
</tr>
<tr>
<td>4</td>
<td>TAP1</td>
<td>DOCK10</td>
<td>CLEC5A</td>
<td>EPHX1</td>
<td>COL1A2</td>
<td>LTBP1</td>
<td>CTGF</td>
</tr>
<tr>
<td>5</td>
<td>UBE2L6</td>
<td>LGALS9</td>
<td>HIC1</td>
<td>PRDX1</td>
<td>COL5A2</td>
<td>LAMC1</td>
<td>CYR61</td>
</tr>
<tr>
<td>6</td>
<td>IRF1</td>
<td>PILRA</td>
<td>CD14</td>
<td>GSR</td>
<td>PDGFRB</td>
<td>TCF7L1</td>
<td>SPON1</td>
</tr>
<tr>
<td>7</td>
<td>GBP5</td>
<td>SIGLEC9</td>
<td>RNASE2</td>
<td>OSGIN1</td>
<td>ADAMTS2</td>
<td>CRIM1</td>
<td>IL1R1</td>
</tr>
<tr>
<td>8</td>
<td>GBP4</td>
<td>NCKAP1L</td>
<td>AGAP2</td>
<td>AKR1B10</td>
<td>SPARC</td>
<td>LAMA5</td>
<td>LAMA4</td>
</tr>
<tr>
<td>9</td>
<td>APOL6</td>
<td>IL21R</td>
<td>STAB1</td>
<td>G6PD</td>
<td>NID2</td>
<td>PXDN</td>
<td>GREM1</td>
</tr>
<tr>
<td>10</td>
<td>PRF1</td>
<td>HK3</td>
<td>ABI3</td>
<td>NQO1</td>
<td>BGN</td>
<td>AFAP1</td>
<td>LRR3C2</td>
</tr>
<tr>
<td>11</td>
<td>LAP3</td>
<td>FCGR2C</td>
<td>ADAP2</td>
<td>AKR1C3</td>
<td>COL11A1</td>
<td>RBM9</td>
<td>MRV1</td>
</tr>
<tr>
<td>12</td>
<td>PSMB9</td>
<td>TNFRSF1B</td>
<td>CD300C</td>
<td>MAP2</td>
<td>COL6A2</td>
<td>C1TGNF6</td>
<td>FGF</td>
</tr>
<tr>
<td>13</td>
<td>CXCL9</td>
<td>LILRB4</td>
<td>ENG</td>
<td>CBR1</td>
<td>MMP2</td>
<td>LATS2</td>
<td>COL6A3</td>
</tr>
<tr>
<td>14</td>
<td>PSME2</td>
<td>LRRC25</td>
<td>LRRC25</td>
<td>CYP4F3</td>
<td>LRRC15</td>
<td>FBLN1</td>
<td>ACTA2</td>
</tr>
<tr>
<td>15</td>
<td>CCL4</td>
<td>SPI1</td>
<td>TNSFRS1B</td>
<td>CES1</td>
<td>AEBP1</td>
<td>ARHGEF17</td>
<td>ZEB2</td>
</tr>
<tr>
<td>16</td>
<td>IL18BP</td>
<td>STAT5A</td>
<td>COL8A1</td>
<td>ME1</td>
<td>THBS2</td>
<td>CDK14</td>
<td>C1orf33</td>
</tr>
<tr>
<td>17</td>
<td>B2M</td>
<td>CD4</td>
<td>NLRP3</td>
<td>TXNRD1</td>
<td>TIMP2</td>
<td>GLI3</td>
<td>NID2</td>
</tr>
<tr>
<td>18</td>
<td>LAG3</td>
<td>CD74</td>
<td>SIGLEC9</td>
<td>ASPH</td>
<td>ZNF469</td>
<td>IGDC4</td>
<td>CD93</td>
</tr>
<tr>
<td>19</td>
<td>WARS</td>
<td>HLA-DOA</td>
<td>ITGB7</td>
<td>AKR1C1</td>
<td>FNDC1</td>
<td>KIAA1462</td>
<td>PMP22</td>
</tr>
<tr>
<td>20</td>
<td>HLA-E</td>
<td>CLEC5A</td>
<td>SLC7A7</td>
<td>TDP2</td>
<td>POSTN</td>
<td>TGFB3</td>
<td>COL3A1</td>
</tr>
</tbody>
</table>

Note: CC correlation coefficient.
Identification of key genes in alveolar macrophages

module genes include inflammatory response, immune response, fatty acid lean process, and cell ledges. GG pathway analysis revealed that the most significant module was primarily enriched in the following aspects of the FoxO signaling pathway, Amoebiasis, Transcriptional misregulation in cancer, the ECM-receptor interaction, Focal adhesion, pi3k-Akt signaling pathway, and Hematopoietic cell lineage. Parts of the repeated KEGG pathway of the DEGs and the most significant module include Transcriptional misregulation in cancer, pi3k-Akt signaling pathway, Hematopoietic cell lineage, ECM-receptor interaction, and Focal migration. Many studies have confirmed the results of our analysis. Studies have found that cell markers with diagnostic and therapeutic potential for lung cancer.

To study the effects of smoking on gene expression in human alveolar macrophages, we included a total of three microarray datasets. Finally, 85 DEGs genes were selected, including 42 up-regulated genes and 43 low-expressed genes. We mapped the interaction network of DEGs and module gene and carried out functional enrichment analysis. The changes in BP term of the DEGs enrichment were mainly in the following aspects, including signal transduction, positive regulation of angiogenesis, inflammatory response, immune response, fatty acid metabolic process, and cell adhesion. EGs and module genes include inflammatory response, immune response, fatty acid lean process, and cell ledges. GG pathway analysis revealed that the most significant module was primarily enriched in the following aspects of the FoxO signaling pathway, Amoebiasis, Transcriptional misregulation in cancer, the ECM-receptor interaction, Focal adhesion, pi3k-Akt signaling pathway, and Hematopoietic cell lineage. Parts of the repeated KEGG pathway of the DEGs and the most significant module include Transcriptional misregulation in cancer, pi3k-Akt signaling pathway, Hematopoietic cell lineage, ECM-receptor interaction, and Focal migration. Many studies have confirmed the results of our analysis. Studies have found that cell markers with diagnostic and therapeutic potential for lung cancer.

Figure 8. Overall survival analysis of 7 hub genes was conducted using Kaplan-Meier plotter. The down-regulated expressions of GBP1 (A), ITGAM (B), CSF1 (C), LAMB1 (F), and THBS1 (G) in lung cancer patients were associated with poor survival. However, the low expression of SPP1 (D) and COL1A1 (E) in lung cancer patients was associated with a better prognosis (P<0.05).
adhesion inhibitors may play an important therapeutic role in the prevention and treatment of COPD [27]. Inflammatory response and immune response also play an essential role in the occurrence and development of diseases such as COPD and lung cancer [28]. Also, in terms of pathway enrichment, the FoxO signaling pathway is most closely related to the pi3k-Akt signaling pathway [29], especially the FoxOs protein in the FoxO signaling pathway, which plays an essential role in inhibiting tumor cell proliferation and inducing apoptosis. FoxO’s dysfunction is involved in the disease progression of breast cancer, prostate cancer, thymic tumors, and many other tumors [30]. Therefore, an in-depth study of the star molecules in the FoxO signaling pathway may have broad prospects for the discovery and development of anticancer drugs. For lung diseases and some tumors closely related to smoking, the conclusions of these studies are consistent with our results.

We used Cytoscape to screen the critical genes in the DEGs network map, identified the top 20 genes with the strongest correlation, and constructed PPI network, including GBP1, IFIT2, IFIT1, IFI44L, MX2IIF27, IGF1, ITGAM, ITPKB, CSF1, SPP1, IL7, TNFSF10, COL1A1, LAMB1, CD36, MDM2, THBS1, AR, FABP4. Furthermore, we conducted a hierarchical clustering analysis of the above genes, and the results showed that these Hub genes could be clearly distinguished between Primary Tumor and Solid Tissue Normal. Then, we compared the differential expression of 20 Hub genes in healthy...
Identification of key genes in alveolar macrophages

A

B

C

D

E

F

G

H

I

J

Overall Survival Kaplan-Meier Estimate
Altered group
Unaltered group
Logrank Test P-Value: 0.842

Overall Survival Kaplan-Meier Estimate
Altered group
Unaltered group
Logrank Test P-Value: 0.731

Overall Survival Kaplan-Meier Estimate
Altered group
Unaltered group
Logrank Test P-Value: 0.842

Overall Survival Kaplan-Meier Estimate
Altered group
Unaltered group
Logrank Test P-Value: 0.861

Overall Survival Kaplan-Meier Estimate
Altered group
Unaltered group
Logrank Test P-Value: 0.629

Overall Survival Kaplan-Meier Estimate
Altered group
Unaltered group
Logrank Test P-Value: 0.897

Overall Survival Kaplan-Meier Estimate
Altered group
Unaltered group
Logrank Test P-Value: 0.467

Overall Survival Kaplan-Meier Estimate
Altered group
Unaltered group
Logrank Test P-Value: 0.838

Overall Survival Kaplan-Meier Estimate
Altered group
Unaltered group
Logrank Test P-Value: 0.269

Identification of key genes in alveolar macrophages

Figure 10. Correlation analysis between Mutations in 7 hub genes and overall survival or disease-free survival. Mutations in GBP1 (A and B), ITGAM (C and D), CSF1 (E and F), SPP1 (G and H), COL1A1 (I and J), LAMB1 (K and L), and THBS1 (M and N) were not associated with overall survival or disease-free survival of LUSC patients.

tissues and LUAD, LUSC patients in the GEPIA database and screened out seven genes with significant differences, including GBP1, ITGAM, CSF1, SPP1, COL1A1, LAMB1, and THBS1. Then, we verified two genes-SPP1 and COL1A1, which were significantly highly expressed in the samples of patients with lung cancer in The Human Protein Atlas, and the results further confirmed that the expression levels of SPP1 and COL1A1 in lung cancer tissues were significantly higher than that in healthy tissues.

CSF1 is a growth factor secreted by many types of cells, such as macrophages and cancer cells, which has autocrine and parocrine effects on the CSF1 receptor (CSF1R). Now studies have found that mir-1207-5p activates the expression of CSF1 and is involved in the proliferation, migration, and invasion of lung cancer cells, thus affecting the survival and metastasis of lung cancer patients [31]. In addition to its involvement in the progression of lung cancer, CSF1 is also significantly expressed in breast, ovarian, prostate, and other diseases [32]. Our study found that SPP1 was significantly highly expressed in both LUAD and LUSC patients and that the up-regulation of SPP1 expression in lung cancer patients was closely associated with a low survival rate. Mutations in the SPP1 gene did not affect the overall or disease-free survival of LUSC patients. Our conclusion is consistent with the results of Zhang and Li [33, 34]. The latest research also found that SPP1 may be an essential mechanism of afatinib resistance. As we know, afatinib is applied as second-generation EGFR Tyrosine Kinase Inhibitor in the treatment of non-small cell lung cancer, and the high expression of SPP1 can enhance the resistance of afatinib and enhance the invasion ability of lung cancer cells. SPP1 may also be involved in disease progression by mediating macrophage polarization in lung adenocarcinoma [35], and the specific mechanism remains to be further studied. Compared with healthy tissues, LAMB1 expression in lung tissues of patients with pulmonary fibrosis was increased [36]. The up-regulation of LAMB1 expression was associated with the occurrence and development of multiple tumors in colorectal cancer, glioblastoma multiforme (GBM), and others. In our study, it was observed that the expression of THBS1 was down-regulated in lung cancer, and survival analysis found that the high expression of THBS1 was associated with a good prognosis. The above analysis results all suggested that
THBS1 plays a protective role in lung tumors, consistent with a prior study [37]. We further analyzed the relationship between THBS1 gene mutation and prognosis in patients with lung cancer and found that disease-free survival rate and overall survival rate of patients with lung cancer were not related to THBS1 gene mutation.

Through literature review, it was found that the correlation between GBP1, ITGAM, COL1A1, and pulmonary diseases has not been widely reported. Our study found that the expression of GBP1 in LUAD and LUSC patients was higher than that in controls, but the difference was not statistically significant. Down-regulation of the GBP1 gene in lung cancer patients was associated with lower survival. Some studies have observed that cytokines induce human ovarian cancer cells to secrete GBP1, and the up-regulation of the GBP1 gene in EOC patients is associated with a good prognosis [38]. However, other studies have shown that GBP1 is involved in cell proliferation. High expression of GBP1 promotes the invasion and migration of glioblastoma cells and predicts a poor prognosis. Moreover, the up-regulation of long non-coding RNA LINCO1783 in cervical cancer can regulate the expression of GBP1 and promote disease progression [39]. At present, the research on ITGAM mainly focuses on systemic lupus erythematosus (SLE), and no related research on ITGAM and lung diseases was found. Our bioinformatic analysis showed that ITGAM was protective against lung cancer. COL1A1 is involved in the development of gastric, breast, colon, and pancreatic cancers by promoting the proliferation, invasion, and metastasis of tumor cells. At present, there are still few studies on the relationship between COL1A1 and lung tumors. Oleksiewicz found that the expression of COL1A1 was increased in patients with non-small cell lung cancer [40]. Our research found that COL1A1 was overexpressed in LUAD and LUSC patients and was associated with low survival. Therefore, COL1A1 may be a marker for the diagnosis, treatment, and prognosis of lung cancer.

Conclusion

We used bioinformatic techniques to identify candidate genes for smoking-induced changes in the expression of genes in pulmonary macrophages that are associated with susceptibility to lung disease. The study identified 85 DEGs and 20 Hub genes that may be involved in the development and progression of smoking-related lung cancer. However, the biologic function and clinical diagnostic and prognostic value of these genes in lung cancer need more study.

Data sharing

Figure 10 and Table 2 are available from author on reasonable request. Figure 10: Correlation analysis between survival of the Hub genes and gene alteration. Table 2: The top 20 genes most closely related to the seven hub genes in LUSC.

Acknowledgements

The study was funded by the Tianshan youth program of Xinjiang Uygur autonomous region, China (2018Q090).

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Feng-Sen Li, Xinjiang National Clinical Research Base of Traditional Chinese Medicine, Xinjiang Medical University, 116 Huanghe Road, Urumqi, Xinjiang 830000, P. R. China. Tel: +86-13999980996; Fax: +86-0991-5810542; E-mail: fengsen602@163.com

References

[6] Pappas K, Papaioannou AI, Kostikas K and Tzanakis N. The role of macrophages in ob-
Identification of key genes in alveolar macrophages


[29] Sunters A, Madureira PA, Pomeranz KM, Aubert M, Brosens JJ, Cook SJ, Burgering BMT, Coombes RC and Lam EW. Paclitaxel-induced
Identification of key genes in alveolar macrophages


