**Name of Journal:** World Journal of Clinical Cases

**Manuscript NO:** 76061

**Manuscript Type:** ORIGINAL ARTICLE

**Retrospective Study**

Prognostic role of the abnormal multiple genes in non-small-cell lung cancer

Lu-Da Yan, Liu Yang, Na Li, Meng Wang, Yan-Hua Zhang, Wen Zhou, Zhi-Qiong Yu, Xiao-Chun Peng, Jun Cai
Abstract

BACKGROUND

Non-small-cell lung cancer (NSCLC) has the highest morbidity and mortality rates among all malignant tumor types. Although therapies targeting the mutated genes such as KRAS have been used in the clinic for many years, the prognosis remains poor. Therefore, it is necessary to further study the aberrant expression or mutation of non-target genes affecting the survival and prognosis.

AIM

To explore the impact of simultaneous abnormalities of multiple genes on the prognosis and survival of patients.

METHODS

We used the R package pheatmap to analyze gene expression data and clinical data downloaded from The Cancer Genome Atlas (TCGA) database. We also collected samples from 85 NSCLC patients from the First People’s Hospital of Jingzhou City and retrospectively followed up the patients. Multivariate Cox regression analysis and survival analysis were performed.

RESULTS

Analysis of gene expression data from TCGA revealed that the overexpression of the following single genes affected overall survival: TP53 \( (P = 0.79) \), PTEN \( (P = 0.94) \), RB1 \( (P = 0.49) \), CTNNB1 \( (P = 0.24) \), STK11 \( (P = 0.32) \), and PIK3CA \( (P = 0.013) \). However, the probability of multiple genes \( (TP53, PTEN, RB1, \text{ and } STK11) \) affecting survival was 0.025. Retrospective analysis of clinical data revealed that sex (hazard ratio \( \text{HR}=1.29 \ [95\% \text{CI}: 0.64-2.62] \)), age \( \text{HR}=1.05 \ [95\% \text{CI}: 1.02-1.07] \), smoking status \( \text{HR}=2.26 \ [95\% \text{CI}: 1.16-4.39] \), tumor histology \( \text{HR}=0.58 \ [95\% \text{CI}: 0.30-1.11] \), cancer stage \( \text{HR}=16.63 \ [95\% \text{CI}: 4.8-57.63] \), epidermal growth factor receptor (EGFR) mutation \( \text{HR}=1.82 \ [95\% \text{CI}: 1.05-3.16] \), abundance \( \text{HR}=4.95 \ [95\% \text{CI}: 0.78-31.36] \), and treatment with
tyrosine kinase inhibitors (TKIs) (HR=0.58 [95% CI: 0.43-0.78]) affected patient survival. Co-occurring mutation of TP53, PTEN, RB1, and STK11 did not significantly affect the overall survival of patients receiving chemotherapy ($P = 0.96$) but significantly affected the overall survival of patients receiving TKIs ($P = 0.045$).

CONCLUSION
Co-occurring mutation or overexpression of different genes has different effects on the overall survival and prognosis of patients. Combined with TKI treatment, the co-occurring mutation of some genes may have a synergistic effect on the survival and prognosis of NSCLC patients.

**Key Words:** non-small-cell lung cancer; gene mutation; tyrosine kinase inhibitor; overexpression; next-generation sequencing; epidermal growth factor receptor; KRAS


**Core Tip:** Non-small-cell lung cancer (NSCLC) has the highest morbidity and mortality rates among all malignant tumor To explore the impact of simultaneous abnormalities of multiple genes on the prognosis and survival of patients. We used the R package to analyze gene expression data and clinical data downloaded from The Cancer Genome Atlas (TCGA) database. We also collected samples from 85 NSCLC patients from the First People's Hospital of Jingzhou City and retrospectively followed up the patients for multivariate Cox regression analysis and survival analysis. Co-occurring mutation or overexpression of different genes has different effects on the overall survival and prognosis of patients. Combined with TKI treatment, the co-occurring mutation of some genes may have a synergistic effect on the survival and prognosis of NSCLC patients.
INTRODUCTION

The prognosis of lung cancer is poor, and the associated mortality rate is among the highest due to the tumor's highly invasive and metastatic nature [1]. Non-small-cell lung cancer (NSCLC) accounts for 85% of lung malignancy cases [2]. Despite advances in gene-targeted therapy and immunotherapy, the long-term survival benefits of patients are still limited [3]. Cancer patients receiving molecularly targeted therapies have clinically different survival prognoses mainly because currently used targeted therapies primarily target single-gene mutations, while tumor tissues are highly heterogeneous. Moreover, the complex tumor microenvironment plays a crucial role in the survival, ability to evade immune surveillance, and drug resistance of cancer cells [4].

Single-gene-targeted therapy has provided a survival benefit to patients over conventional chemotherapy. For example, patients with advanced epidermal growth factor receptor (EGFR) mutation-positive lung cancer can be treated with tyrosine kinase inhibitors (TKIs) such as gefitinib or erlotinib, and those with anaplastic lymphoma kinase-positive lung cancer can be treated with crizotinib [5-7]. Third-generation TKIs including osimertinib further improve the survival of patients with EGFR mutation-positive lung cancer that is also characterized by T790M mutation, in which cases first-generation TKIs are not effective [8]. However, not all patients with EGFR mutation-positive lung cancer respond well to TKIs, and some even develop drug resistance, which will lead to disease progression [9].

Next-generation sequencing analyses have revealed significant differences in gene mutation sites among patients, with the differences also being apparent between early and late stages of cancer and in the mutation frequency of each site [10]. The differences in gene mutations may also be responsible for differences in the risk of drug resistance and differences in individual treatment responses.

The co-occurrence of TP53 mutation and EGFR is often associated with a worse prognosis and accelerated proliferation and invasion of cancer cells [11, 12]. In this study, we collected clinical data from The Cancer Genome Atlas (TCGA) and analyzed
the impact of common mutations in NSCLC patients on targeted therapy and survival prognosis.

MATERIALS AND METHODS

Data collection and analysis

First, we downloaded the clinical and original gene expression data (counts) of lung adenocarcinoma (LUAD) patients from the TCGA public database. Next, we screened 21 target genes detected by second-generation sequencing during clinical treatment and observed the effect of multiple gene expression on the survival and prognosis of patients using the heatmap package of R software to plot a heat map for visual analysis.

Then, we used the patient mutation information provided in the database and divided all patient samples into EGFR and KRAS mutation groups to study the gene expression differences and explore the correlation between mutations at other loci and these two most common mutations.

Finally, we analyzed the overall survival (OS) time of patients whose samples showed different gene expression profiles. The significance of single-gene analysis was improved by considering an expression Z score of more than 1 as high expression and an expression Z score less than -1 as low expression. Then, considering 0 as the critical value, we analyzed the co-expression of multiple genes.

Collection of clinical data from patients

We collected the next-generation sequencing results of more than 300 NSCLC patients from the First People’s Hospital of Jingzhou City from 2017 to 2020. After follow-up, clear OS and detailed data of 85 patients were obtained. If some patients were examined many times during the treatment, we took the first detection results as the basis for analysis.

We collected the medical history and general clinical data of the subjects through the hospital information system and telephone return visit in the hospital’s oncology department. We collected the patient’s sex, age, smoking history, pathological type,
cancer stage, next-generation sequencing results, treatment with TKIs, OS, and other results and divided the patient population based on the gene mutation status for multivariate Cox regression analysis. Finally, the patients were divided into two groups based on whether they received targeted TKI therapy or chemotherapy to explore the effect of co-occurrence of gene mutations on the OS of patients. All clinical data were collected after being submitted to the ethics committee of Jingzhou First People’s Hospital for approval. All patients signed the informed consent before the next-generation sequencing analysis.

Statistical analysis

All patients were followed up until December 31, 2020. We screened patients with complete basic information and definite OS data. The results of all patients were obtained using the same high-throughput sequencing equipment. All statistical analyses were performed with R software (version 4.0.3) and used several R packages such as edgeR, DESeq, TCGAbiolinks, and ggplot2. We used the Kaplan-Meier method to analyze the differences in OS and compared the effects of gene mutations using the log-rank test. We also investigated the influence of various factors on the total survival of patients using multivariate Cox regression analysis. The Fisher exact test was used for comparing different groups. All P values are based on a two-tail hypothesis with statistical significance defined as P<0.05.

RESULTS

Differential gene expression

The volcano map (Figure 1) of gene expression data of 533 cancer tissues and 53 normal tissues from TCGA reveals a large number of upregulated (red) and downregulated (green) genes [13]. Differences in gene expression were also seen among different subtypes of cancer tissues. In this study, 21 genes detected by next-generation sequencing were selected to explore the effect of gene mutation and overexpression on OS. A 2-D heat map of various parameters was plotted to intuitively observe patients' basic indicators and gene expression (Figure 2). It showed apparent differences in the
expression of different genes, but it is necessary to clarify which indicators impact the OS of patients.

Clinical data regarding the EGFR and KRAS mutation status were also analyzed and compared with data on mutations of other genes [14, 15]. The National Comprehensive Cancer Network guidelines have pointed out that KRAS mutation can reduce sensitivity toward EGFR inhibitors. Clinically, the probability of simultaneous occurrence of KRAS mutation and EGFR mutation is very low, and a mutually exclusive relationship between them is also reported [16]. After grouping the samples based on the EGFR and KRAS mutation status with more than 20 samples in each group, the prepared bubble chart and box chart (Figure 3) showed that RET, KIT, and TERT exhibited significantly different expression levels between the two groups. These three genes were upregulated in patients with KRAS mutation, while EGFR and BRAF were downregulated (Figure 3A).

In the analysis of the survival of patients with single-gene mutations, to amplify the single-gene effect, we considered genes with a Z score greater than 1 to be highly expressed and those with a Z score less than -1 to have a low expression level. After calculating the P-value, it was found that except for PIK3CA (P<0.05), there was no statistical significance in the high expression of other single genes: TP53 (P = 0.79), PTEN (P = 0.94), RB1 (0.49), CTNNB1 (P = 0.24), STK11 (P = 0.32) (Figure 4). We speculated that significant PIK3CA overexpression indicates poor prognosis and OS [17-19]. Given that the expression of other single genes did not seem to affect prognosis significantly, we suspected the simultaneous overexpression of multiple genes, especially the four tumor suppressor genes (TP53, PTEN, RB1, and STK11), would have some clinical implication. Therefore, we used the Z score of 0 as the critical value and divided the four genes into two groups in which all had a high expression or a low expression level at the same time (Figure 5). A P-value of 0.025 showed that when TP53, PTEN, RB1, and STK11 are highly expressed simultaneously, the OS is significantly different from that when these genes show a low expression level. Moreover, the high expression group had a significantly shorter OS.
Differences in the co-occurrence of gene mutations between patients receiving chemotherapy and TKI therapy

We visualized data regarding gene mutations and basic clinical information collected from patients by plotting a heat map (Figure 6). First, we divided the patients based on whether they received chemotherapy or TKI therapy and then carried out Fisher’s exact test (Table 1). We found that sex (P = 0.0021), smoking history (P = 0.0302), and OS (P = 0.0022) differed significantly based on the treatment. To understand the impact of various factors on patients, we separately analyzed the impact of basic indicators and gene mutations by multivariate Cox regression (Figure 7). Sex (hazard ratio [HR]=1.29 [95%CI: 0.64-2.62; P = 0.475]), age (HR=1.05 [95%CI: 1.02-1.07]; P<0.001), smoking history (HR=2.26 [95%CI: 1.16-4.39]; P = 0.017), tumor histology (HR=0.58 [95%CI: 0.30-1.11]; P = 0.098), cancer stage (HR=16.63 [95%CI: 4.8-57.63]; P<0.001), EGFR mutation (HR=1.82 [95%CI: 1.05-3.16]; P = 0.034), abundance (HR=4.95 [95%CI: 0.78-31.36]; P = 0.09), and TKI treatment (HR=0.58 [95%CI: 0.43-0.78]; P<0.001).

Cox regression analysis (Figure 8) revealed a significant effect of all gene mutations except for BRAF (P = 0.02), which indicates that the OS of patients is under the combined influence of multiple-locus genes. The HR value of some gene mutations was positive, such as MAP2K1 (HR=0.0014), CTNNB1 (HR=0.1629), and RET (HR=0.1089), indicating that mutations of these genes may benefit the patients, but further confirmation using more samples is necessary. We selected TP53, PTEN, Rb1, and STK11, whose mutation frequency is high in NSCLC patients, to study if the co-occurrence of mutations of these genes has a similar superposition effect [20-22]. Patients with mutations in these tumor suppressor genes have a worse survival prognosis [23, 24]. These co-occurring mutations can allow cancer cells to escape immune surveillance, proliferate aberrantly to malignancy and develop resistance to targeted therapy. After identifying the effect of TP53 (HR=1.2602), PTEN (HR=1.4428), Rb1 (HR=2.2605), and STK11 (HR=3.5352), we sought to determine whether either of them could be used individually as an indicator of survival and
prognosis of patients. However, the effect of age (P<0.001) was significant even if its HR value was only 1.05 (Figure 7).

The patients were divided into two groups based on whether they received chemotherapy (n = 41) or TKI treatment (n = 44). According to the number of mutations in the four tumor suppressor genes (TP53, PTEN, Rb1, and STK11), we classified those with more than one mutation into the greater than (GT) 1 group (Figure 9). The number of mutations of these tumor suppressor genes in the chemotherapy group did not bear significance between the GT1 and non-GT1 groups (P = 0.96). Still, it was significant in the TKI treatment group (P = 0.045). Co-occurrence of mutations in these genes worsened the prognosis similarly in both groups. We found that some patients discontinued targeted therapy not because of disease progression but because of economic reasons; this could have affected the results. The complexity of the tumor genome determines that cancer treatment cannot target a single gene. Co-mutation is likely to completely change the biological characteristics of the original tumor through synergy, endow the tumor with new biological characteristics, and make the tumor tolerant to targeted therapy. This co-mutation may occur gradually in the process of targeted therapy.

DISCUSSION

In this study, we focused on the effect of the co-occurrence of gene mutations using TCGA database and clinical patient data. Our analyses revealed significant differences in the gene expression profiles between adenocarcinoma and normal tissues in NSCLC. The upregulated and downregulated genes also differed in different patients and between LUAD patients with EGFR and KRAS mutations. Most tumor-derived TP53 mutations occur in the region encoding the DNA binding domain of p53. The TP53 mutation significantly impacts the progression of various types of cancer [25,26]. While the overexpression of some genes was significantly associated with good OS and prognosis, such as PIK3CA, the single expression of most genes did not have a significant effect. The simultaneous overexpression of multiple tumor suppressor genes
(TP53, PTEN, RB1, and STK11) was associated with poor OS. Cox multivariate regression analysis revealed that for NSCLC patients, the most critical factor affecting OS is not the type of treatment or gene mutation but the disease stage, which underscores the importance of early diagnosis of solid tumors. The effects of recognized risk factors such as smoking history were also confirmed in the analysis. Finally, after grouping based on treatment, we found that in patients receiving traditional chemotherapy, mutations of TP53, PTEN, RB1, and STK11 had no significant influence on the OS; however, these mutations had a significant effect in patients receiving TKIs. Simultaneous mutations in multiple tumor suppressor genes resulted in a risk superposition effect[24].

These genes have been studied in the context of non-target therapy. While the frequency of mutations in these genes is high in NSCLC patients, the problem of tumor heterogeneity and the possibility of personalized medicine need to be further explored. The influence of these gene mutations on the OS and prognosis of patients receiving immunotherapy has also received attention [27-29]. In the future, the development of new molecular targeted drugs will help deal with the heterogeneity of different mutant subtypes.

CONCLUSION

In conclusion, this study summarizes the impact of the co-occurrence of mutations or overexpression of multiple genes on the OS and prognosis of NSCLC patients. The results indicate that the co-occurrence of mutations results in a risk superposition effect, and such genes must be studied further when predicting patients' disease progression.

ARTICLE HIGHLIGHTS

Research perspectives

In the future, the development of new molecular targeted drugs will help deal with the heterogeneity of different mutant subtypes.
Research conclusions
Co mutation or overexpression of different genes have different effects on the overall survival and prognosis of patients. Combined with TKI treatment, the co mutation of some genes may have a synergistic effect on the survival and prognosis of NSCLC patients.

Research results
The probability of multiple genes (TP53, PTEN, RB1, and STK11) affecting survival was 0.025. Retrospective analysis of clinical data revealed that sex (hazard ratio [HR]=1.29), age (HR=1.05), smoking status (HR=2.26), tumor histology (HR=0.58), cancer stage (HR=16.63), epidermal growth factor receptor (EGFR) mutation (HR=1.82), abundance (HR=4.95), and treatment with tyrosine kinase inhibitors (TKIs) (HR=0.58) affected patient survival. Co-occurring mutation of TP53, PTEN, RB1, and STK11 did not significantly affect the overall survival of patients receiving chemotherapy (P = 0.96) but significantly affected the overall survival of patients receiving TKIs (P = 0.045).

Research methods
We used the R package pheatmap to analyze gene expression data and clinical data downloaded from The Cancer Genome Atlas (TCGA) database. We also collected samples from 85 NSCLC patients from the First People’s Hospital of Jingzhou City and retrospectively followed up the patients. Multivariate Cox regression analysis and survival analysis were performed.

Research objectives
At the same time, the lack of target inhibition and tumor heterogeneity caused by gene mutation finally promote the re emergence of drug resistance in patients with acquired targeted therapy. Based on the above reasons, we carried out this project to study the effect of polygene mutation and abnormal expression on the prognosis and survival of patients with non-small cell lung cancer.
Research motivation
To investigate the effect of simultaneous polygenic abnormalities on the prognosis and survival of patients.

Research background
Among all malignant tumor types, non-small cell lung cancer (NSCLC) has the highest incidence rate and mortality.

ACKNOWLEDGEMENTS
Not applicable.
### Similarity Index

4%

### Primary Sources

<table>
<thead>
<tr>
<th>Rank</th>
<th>Source</th>
<th>Words</th>
<th>Similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><a href="http://www.science.gov">www.science.gov</a></td>
<td>25</td>
<td>1%</td>
</tr>
<tr>
<td>2</td>
<td>e-sciencecentral.org</td>
<td>23</td>
<td>1%</td>
</tr>
<tr>
<td>3</td>
<td><a href="http://www.spandidos-publications.com">www.spandidos-publications.com</a></td>
<td>19</td>
<td>1%</td>
</tr>
<tr>
<td>4</td>
<td>consumer.healthday.com</td>
<td>14</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>5</td>
<td>effiloop.com</td>
<td>13</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>6</td>
<td><a href="http://www.jthoracdis.com">www.jthoracdis.com</a></td>
<td>13</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>7</td>
<td>dev.physiciansweekly.com</td>
<td>12</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>8</td>
<td><a href="http://www.ncbi.nlm.nih.gov">www.ncbi.nlm.nih.gov</a></td>
<td>12</td>
<td>&lt; 1%</td>
</tr>
</tbody>
</table>

EXCLUDE QUOTES: ON
EXCLUDE BIBLIOGRAPHY: OFF
EXCLUDE SOURCES: < 12 WORDS
EXCLUDE MATCHES: < 12 WORDS