Identification of Cigarette Smoking-Related Novel Biomarkers in Lung Adenocarcinoma

Yuan Zhang,¹ Qiong Wang,¹ Ting Zhu,¹ and Hui Chen²

¹Department of Respiratory Medicine, Zhenhai Hospital of Traditional Chinese Medicine, Ningbo, 315000 Zhejiang, China
²Department of Preventive Medicine of Traditional Chinese medicine, Jiangnan Hospital Affiliated to Zhejiang Chinese Medical University (Hangzhou Xiaoshan Hospital of Traditional Chinese Medicine), Hangzhou 310000, Zhejiang, China

Correspondence should be addressed to Hui Chen; huichen26@outlook.com

Received 26 April 2022; Revised 24 May 2022; Accepted 25 May 2022; Published 19 June 2022

Academic Editor: Yingbin Shen

Copyright © 2022 Yuan Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. The aims of this study were to screen the gene mutations that are able to predict the risk of cigarette smoking-related lung adenocarcinoma (LUAD) and to evaluate its prognostic significance.

Methods. Clinical data and genetic information were retrieved from the TCGA database, and the patients with LUAD were divided into three groups including never smoking, light smoking, and heavy smoking according to cigarette smoking dose. Differentially mutated genes (DMGs) of each group were analyzed. At the same time, the function of DMGs in three smoking groups was evaluated by GO function and KEGG pathway analysis. The driver genes and protein variation effect of DMGs were performed to further screen key genes. The survival characteristics of the gene expression and mutation of those genes were analyzed and plotted to visualize by the Kaplan-Meier model.

Result. The DMGs for different smoking doses were identified. The driver and deleterious mutation in the DMGs were screened and gene interaction network was constructed. The DMGs with driver mutations and deleterious mutations that were associated with the overall survival in the heavy smoking patients were considered as the candidate genes for novel markers of smoking-related LUAD. The final novel risk factor gene was identified as MYH7 and the high express of MYH7 in LUAD correlation with patients’ gender, lymph node metastasis, T stage, and clinical stage.

Conclusions. In summary, it can be concluded that MYH7 is a novel biomarker for heavy smoking-related LUAD and it is significantly correlated with the prognosis of lung cancer and is related to the clinical characteristics of lung cancer.

1. Introduction

Lung cancer is the most common malignancy in humans which leads to high cancer-related deaths worldwide. Lung adenocarcinoma (LUAD) is the main histological type, including more than 40% of lung cancer [1, 2]. The 5-year survival rate of patients with LUAD is less than 10%, and 90% of them die of complications related to tumor metastasis [3, 4]. Most patients with LUAD are diagnosed at advanced stages, thus miss best opportunities for surgical treatments. To make matters worse, LUAD is not sensitive to radiotherapy and chemotherapy, and the prognosis of patients with LUAD remains poor. In recent years, the incidence and mortality of lung cancer have been increasing year by year, which has caused serious negative effects on patients and society [5].

Many studies have shown that cigarette smoking is the main cause of lung cancer [6–8]. Tobacco smoke contains polycyclic aromatic hydrocarbons and the nicotine-derived nitrosamines, which induce gene mutations in known oncogenes such as KRAS and TP53 [9]. Moreover, it is reported that tobacco aldehydes inhibit the DNA repair [10]. Smoking increases the risk for development of the lung cancer via these mechanisms, and thus, smoking-associated LUAD has its specific gene mutations compared with general LUAD. In the current context of precision treatment of cancer, it is necessary to explore biomarkers or molecular targets for cigarette smoking-associated LUAD. Understanding the mechanism of the occurrence and development of cigarette smoking-associated LUAD contributes to identifying therapeutic targets and approaches for the prevention and management.
In this study, data of gene mutation for lung adenocarcinoma patients were downloaded from The Cancer Genome Atlas (TCGA), and the differentially mutated genes (DMGs) among three groups including never smoking, light smoking, and heavy smoking groups were screened. We analyzed the gene function enrichment of the specific DMGs for heavy smoking patients and identified the oncogenic drivers in them. We also analyzed gene-gene interaction of the specific DMGs and their association with prognosis for overall survival. Combining the above results, we found a novel biomarker, MYH7, with high occurrence of mutation in heavy smoking patients. There are to date few reports for MYH7 in lung cancer. Therefore, MYH7 can be used as a novel target for the diagnosis of smoking-associated lung cancer or for targeted precision therapy targeting MYH7.

2. Materials and Methods

2.1. Datasets. The clinical data and gene expression information of lung cancer patients were downloaded from the American Cancer Genome Atlas Database (TCGA), and lung adenocarcinoma (Broad, Cell 2012) dataset was used to obtain lung cancer patients’ information. A total of 184 samples were included in this study. A total of 65,768 somatic mutations were detected.

2.2. Identification of Differentially Mutated Genes. Differentially mutated gene analysis for the never smoker, light smoker, and heavy smoker groups in the LUAD dataset was performed by using the clinical enrichment function of the maftools package in R software. p value < 0.05 was defined as the significant difference.

2.3. Functional Annotation. As for the obtained different genes, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotation were performed with the R package (clusterProfiler). GO annotation was carried out from the aspects of biological process (BP), molecular function (MF), and cellular component (CC). Fisher’s test was used to calculate the p value of significance level, so as to screen the GO with significant enrichment of different genes. The p value < 0.01 was marked with red as the significant enrichment item and the blue as the nonsignificant item. KEGG database was used to explore the signal pathway of significantly differentially expressed gene enrichment, with p value < 0.05 as the threshold.

2.4. Driver Gene Analysis Based on Mutation Location Clustering. Oncogene mutations usually gather at specific locations of proteins (also known as mutation hot spots), and the mutations in these domains are beneficial to the growth or proliferation of cancer cells. We used OncodriveCLUST algorithm to cluster the mutation sites of gene bases to identify cancer genes. The key information calculated included the number of mutation hot spots, the number of mutations clustering in the hot spots, the length of amino acids corresponding to the protein, the proportion of clustering mutations in all mutations of the gene, and the p value and FDR values. The smaller the value, the stronger the driving force.

2.5. Mutation Damaging Was Assessed Based on PROVEAN and SIFT Software. Homologous proteins were found in the database, and protein sequences with high similarity and consistent function were selected for multisequence PSI-BLAST alignment to evaluate the conserved protein sites, and the risk was evaluated by PROVEAN/SIFT database score.

2.6. Interacting Network Analysis. The STRING database (https://string-db.org/) is used to explore the interactions between proteins and genes. The SRING database contains experimental data, direct interactions, and indirect functional correlations between proteins and obtains the PPI interaction network diagram.

2.7. Statistical Analysis. The gene expression information and overall survival (OS) data were obtained from TCGA database. The Kaplan-Meier analysis was used to calculate the hazard ratio (HR), and the survival curve was drawn. p < 0.05 was considered to be significantly related to the prognosis of lung cancer patients.

### Table 1: Clinical characteristics of patients with different smoking degrees.

| Clinical | Never smoker | Light smoker | Heavy smoker | p value |
|----------|--------------|--------------|--------------|---------|
| Smoking amount | 0 | 1-10 | 11-128 | 0.205 |
| Age | 67.59 (43-87) | 59.25 (42-78) | 65.69 (38-84) | 0.015 |
| Gender | 0.015 | 0.015 | 0.015 | 0.015 |
| Men | 7 | 7 | 66 | 0.567 |
| Women | 20 | 10 | 52 | |
| TNM stage | I | 13 | 60 | 0.023 |
| II | 5 | 1 | 24 | 0.023 |
| III | 4 | 3 | 15 | 0.023 |
| IV | 1 | 0 | 8 | 0.023 |
| Overall survival (OS) | 21.69 | 19.86 | 14.13 | 0.023 |

BioMed Research International
Figure 1: The profile of somatic mutations for never smokers (a), light smokers (b), and heavy smokers (c) in LUAD, respectively. From top to bottom, each row is the statistics of mutation types, the type and number of mutation bases (vertical axis is classified; horizontal axis scale is counted), and the count box diagram of mutation number and mutation species in each sample.
The top 10 mutant genes in the heavy smoking group were genes in light smoking group were
TTN, Scand3, EGFR, The top 10 mutant genes in the nonsmoking group were
Lung adenocarcinoma patients were divided into nonsmoking, light smoking, and heavy smoking patients, boxplots of different smoking levels.

3. Results

3.1. Screening of Differential Mutation Genes in Lung Cancer Patients with Different Smoking Levels. The somatic gene mutation profiles and clinical data were acquired from the TCGA database, which included 184 patients. The result of the survival analysis showed that the smoking situation significantly associated with patient’s OS (Table 1, p = 0.023). Lung adenocarcinoma patients were divided into nonsmoking group, light smoking group, and heavy smoking group based on their total amount of smoking (the product of the number of packs smoked and the number of years) up to the time of tumor diagnosis: heavy (>10), light (>0 and<10), and never (=0). The mutation status of patients in each group was statistically analyzed, and the results are shown in Figure 1. As can be seen from the result, the single nucleotide missense mutation was the dominant mutation in the three types of patients with different smoking levels. Patients in the nonsmoking group mutated the base type to replace thymine cytosine nucleotide with cytosine nucleotide (C>T), followed by cytosine nucleotide substitution for adenine nucleotide substitution (C>A), while cytosine nucleotide substitution for adenine nucleotide substitution in light and heavy smoking groups (C>A) is the most common, followed by cytosine nucleotide instead of thymine (C>T). The top 10 mutant genes in the nonsmoking group were EGFR, ZFHX3, CDC42BPA, TP53, SPHKAP, SLC17A6, Scand3, RelN, GRTN2A, and DIO1. The top 10 mutant genes in light smoking group were TTN, MUC16, HMCN1, FLG, TP53, EGFR, USH2A, CSMD3, NALCN, and HSPG2. The top 10 mutant genes in the heavy smoking group were TTN, MUC16, LRPIB, CSMD3, RyR2, USH2A, TP53, SPTA1, ZFHX4, and KRAS. The 1122 DMGs were identified in heavy smoking, 432 DMGs were identified in light smoking, and 327 DMGs were identified in never smoking, and the significant genes are shown in Figure 2.

3.2. GO and KEGG Pathway Analysis of Mutated Differential Genes. Using GO analysis, the difference of gene has been studied, and the results are shown in Figure 3; the difference of gene biological pathways is mainly related to cell adhesion, involving the main molecular function of the ion channels combining exercise, calcium ion, and extracellular matrix structure; these genes mainly located in the plasma membrane and organelle membrane, which are involved in cell information exchange, may be related to the spread of cancer cells to metastasize. KEGG pathway results are shown in Figure 3. These genes were significantly correlated with adhesion, ECM receptor interaction, olfaction transduction, and other signaling pathways.

3.3. Protein Variation Effect of Mutated Genes and Candidate Marker Genes. In order to validate the protein variation effect of mutation genes between never smoking, light smoking, and heavy smoking patients, boxplots of model genes were drawn, and both PROVEN and SIFT programs showed that the variation effect scores for the protein functions between never, light, and heavy smoking groups were significantly different (p<0.05, Figure 4), while PROVEN and SIFT scores were conflicting in light smoking group. Mutations in light smokers were more deleterious in the SIFT scores while contrary in the PROVEN scores.
Driver gene analysis was performed on the mutation data of lung cancer dataset based on mutation location clustering. The results of cancer driver genes with $p$ value less than 0.05 are shown in Table 2. The oncogenes significantly associated with lung cancer were KRAS, NR4A2, CDKN2A, EGFR, OR5AS1, OR5D14, DOCK11, TFEB, and ZNF335.

Based on the results of differential mutation, cancer driving gene analysis, and mutation harmfulness analysis, the genes were intersected. Differential mutations that may be cancer drivers in the never smoker, light smoker, and heavy smoker groups were obtained ($p$ value < 0.1), and damaging and deleterious genes are considered as key candidate genes.
in PROBEAN/SIFT prediction, which are arranged by mutation frequency, as shown in Table 2.

3.4. Interacting Networks of Important Differential Mutants. The interaction between proteins of cancer-driving genes was explored based on the STRING database, which included experimental data, results mined from PubMed abstracts and integrated data from other databases, as well as results predicted by bioinformatics methods. The PPI interaction network diagram is shown in Figure 5. It can be seen from the diagram that CDKN2A, KRAS, EGFR, TLR4, and TP53 with high-grade index are the core genes, followed by STK11, SPTA1, MYH8, MYH7, and MYO10, and most of the core genes have been reported. Literature mining was performed for searching the association of those genes to the smoking lung cancer. The results showed that only MYH7 and MYH8 genes had not been reported yet, and they were candidate genes related to lung cancer of new types of smoking. Although there is an enrichment of MYH7 mutation in heavy smoking patients, the mutation loci varied in the patients (Supplementary Table 1).

3.5. Novel Biomarkers of Smoking-Related LUAD. PubMed was used to search for papers related to the key node genes of result 2.4 and cancer caused by smoking. The results showed that only MYH7 and MYH8 genes had not been reported yet, and they were candidate genes related to lung cancer of new types of smoking. Although there is an enrichment of MYH7 mutation in heavy smoking patients, the mutation loci varied in the patients (Supplementary Table 1).

| Hugo_Symbol | Total mutated | Mutated samples | Hugo_Symbol | Total mutated | Mutated samples | Hugo_Symbol | Total mutated | Mutated samples |
|-------------|---------------|-----------------|-------------|---------------|-----------------|-------------|---------------|-----------------|
| TP53        | 67            | 63              | CNTNAP5     | 21            | 17              | SCN11A      | 11            | 9               |
| TTN         | 144           | 61              | CTNNND2     | 20            | 17              | ARAP2       | 10            | 9               |
| MUC16       | 95            | 51              | KIF2B       | 20            | 17              | MOY010      | 10            | 9               |
| LRP1B       | 85            | 48              | LRP2        | 20            | 17              | OR10AG1     | 10            | 9               |
| SPTA1       | 49            | 39              | MYO18B      | 20            | 17              | PCDH1A9      | 10            | 9               |
| ZFHX4       | 51            | 36              | EPFA5       | 20            | 16              | GABRA5      | 9             | 9               |
| KRAS        | 36            | 36              | SORCS3      | 18            | 15              | OR4M2       | 9             | 9               |
| PCLO        | 48            | 34              | POTEC       | 16            | 15              | WSCD1       | 11            | 8               |
| XIRP2       | 47            | 34              | KEAP1       | 15            | 15              | THBS2       | 10            | 8               |
| PCDH15      | 36            | 29              | PCDH10      | 18            | 14              | MKRN3       | 9             | 8               |
| CSDM1       | 40            | 26              | TRPS1       | 17            | 14              | AGBL1       | 8             | 8               |
| LPHN3       | 30            | 26              | CMYA5       | 15            | 14              | CDKN2A      | 8             | 8               |
| RP11L1      | 31            | 25              | LRRRC4C     | 15            | 14              | GP2         | 8             | 8               |
| RELN        | 29            | 24              | MYH7        | 15            | 14              | OR5D14      | 8             | 8               |
| DNAH9       | 34            | 23              | CDH7        | 13            | 13              | SLC6A2      | 8             | 8               |
| EGFR        | 26            | 23              | KNTC2       | 15            | 12              | LPA         | 8             | 7               |
| ZNF804A     | 26            | 22              | KIAA1211    | 15            | 12              | ADCY5       | 7             | 7               |
| ZNF536      | 30            | 21              | LRRTM4      | 13            | 12              | OR5B17      | 7             | 7               |
| CUBN        | 27            | 21              | MYH8        | 14            | 11              | PKP2        | 7             | 7               |
| FAM5C       | 27            | 21              | BRAF        | 12            | 11              | SAGE1       | 7             | 7               |
| STK11       | 21            | 21              | FAM71B      | 12            | 11              | TSHR        | 7             | 7               |
| BAI3        | 24            | 20              | TLR4        | 12            | 11              | VSTM2A      | 7             | 7               |
| MXRA5       | 22            | 20              | CNTN5       | 11            | 11              | ADAM21      | 7             | 6               |
| Tpte        | 22            | 20              | POM121L12   | 11            | 11              | FCRLA       | 7             | 6               |
| CDH10       | 24            | 19              | TGF2LX      | 11            | 11              | OR5A51      | 7             | 6               |
| FLG2        | 21            | 18              | SLC17A6     | 13            | 10              | C7orf10     | 6             | 6               |
| DNAH3       | 20            | 18              | MMP16       | 12            | 10              | CLCNAK      | 6             | 6               |
| EPHA3       | 20            | 18              | OR2M2       | 11            | 10              | NR4A2       | 6             | 6               |
| PRDM9       | 20            | 18              | TRHDE       | 11            | 10              | OR10A4      | 6             | 6               |
| CSMD2       | 19            | 18              | KCNJ3       | 10            | 10              | OR10Z1      | 6             | 6               |
| PKHDL1      | 26            | 17              | RAG1        | 10            | 10              | TRIM48      | 6             | 6               |
the prognostic survival rate of patients. The high expression of MYH8 gene had no significant effect on the prognostic survival rate. Therefore, MYH7 was screened as a new smoking-induced lung cancer target gene.

4. Discussion

In this study, we focused on the analysis of mutated genes associated with tobacco smoking in LUAD. We identified specific mutations in LUAD patients with heavy smoking that were distinct from the nonsmoking group. Among these mutations, we screened the genes with driver mutations and those with deleterious mutations. Considering that these mutated genes have regulatory relationships and affect the occurrence of LUAD through common pathways, we subsequently performed gene interaction analysis for these mutated genes and constructed a gene network for smoking-related LUAD centered on genes known to be high frequency mutated in LUAD, such as KRAS and TP53. Based on the results of the literature search, most of these smoking-related core genes (CDKN2A, EGFR, KRAS, TLR4, TP53, SPTA1, and STK11) we identified have been reported in many studies for their association with lung cancer. However, MYH7 has not been studied to elaborate its association with lung cancer. In LUAD, MYH7 has a high mutation frequency (11 of 90), so MYH7 can be used as a novel diagnostic biomarker. Meanwhile, the gene expression of MYH7 correlated with the overall survival of LUAD patients and the tumor stage and lymph node metastasis of patients, suggesting that MYH7 is associated with the progression of LUAD, and thus precise targeted therapies targeting MYH7 can be carried out in the future.

Current research on MYH7 has focused on studies in cardiomyopathies, as it is predominantly expressed in the normal human ventricle. Mutations in this gene are associated with familial hypertrophic cardiomyopathy, myosin storage myopathy, dilated cardiomyopathy, and Laing early-onset distal myopathy [11–14]. In our results, MYH7 was shown to be highly expressed in LUAD tumor tissue. In addition, only a small number of studies have shown that MYH7 is associated with tumorigenesis. Sun et al. reported that MYH7 is one of the top ten hub genes in PTEN mutation prostate cancer [15]. Huang et al. reported that mutations in MYH7 occur in Epstein-Barr virus-associated intrahepatic cholangiocarcinoma [16]. This paper is the first to propose that the lack of function of MYH7 is one of the causes of LUAD, especially for smoking-associated LUAD.
Expression of MYH7 in LUAD based on sample types

Expression of MYH7 in LUAD based on individual cancer stages

Expression of MYH7 in LUAD based on patient’s race

Expression of MYH7 in LUAD based on patient’s gender

Figure 6: Continued.
Although cigarette smoking is the main cause of lung cancer, the incidence of lung cancer is increasing among nonsmokers. It is estimated that about 25% of lung cancer cases are observed in nonsmokers, and some studies have observed that 40% of nonsmoking men and 31.2% of nonsmoking women have no known exposure history to major carcinogens [17, 18]. If lung cancer in nonsmokers were considered as a single cancer, it would be the seventh leading cancer death in the world [17]. If the current growth rate of nonsmoking lung cancer continues, it is predicted that nonsmoking lung cancer will be the main type of lung cancer in the next 10 years [19]. Current evidence shows that nonsmoking lung cancer shows a different pattern from smokers' lung cancer, and there are essential differences between nonsmoking lung cancer and smoking-related lung cancer in terms of gender, clinical characteristics, and molecular genetic changes [20, 21]. Heavy smokers were found to have many specific gene mutations in this study, while never smokers did not seem to have specific gene mutations, compared to other smoking patients. Therefore, the results of the present study cannot explain the etiology of non-smoking-related LUAD. Considering the high rate of non-smoking-related lung cancer as well, more studies are still needed for non-smoking-related LUAD, but we suggest that studies can be conducted at levels other than gene mutations.

**Data Availability**

The data used to support the findings of this study are included within the article.

**Conflicts of Interest**

The authors declare that they have no competing interest.

**Supplementary Materials**

Supplementary Table 1: the mutated sites of MYH7 in the heavy smoking patients. (Supplementary Materials)

**References**

[1] M. Noguchi, A. Morikawa, M. Kawasaki et al., "Small adenocarcinoma of the lung. Histologic characteristics and prognosis," Cancer, vol. 75, no. 12, pp. 2844–2852, 1995.

[2] C. Zappa and S. A. Mousa, "Non-small cell lung cancer: current treatment and future advances," Translational lung cancer research, vol. 5, no. 3, pp. 288–300, 2016.

[3] B. Y. Wang, J. Y. Huang, H. C. Chen et al., "The comparison between adenocarcinoma and squamous cell carcinoma in lung cancer patients," Journal of Cancer Research and Clinical Oncology, vol. 146, no. 1, pp. 43–52, 2020.

[4] D. Zhang, Q. Jiang, X. Ge et al., "RHOV promotes lung adenocarcinoma cell growth and metastasis through JNK/c-Jun pathway," International Journal of Biological Sciences, vol. 17, no. 10, pp. 2622–2632, 2021.

[5] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2021," CA: a Cancer Journal for Clinicians, vol. 71, no. 1, pp. 7–33, 2021.

[6] I. Petersen, "The morphological and molecular diagnosis of lung cancer," Deutsches Ärzteblatt International, vol. 108, no. 31-32, pp. 525–531, 2011.

[7] H. M. Schuller, "The impact of smoking and the influence of other factors on lung cancer," Expert Review of Respiratory Medicine, vol. 13, no. 8, pp. 761–769, 2019.

[8] G. Siasos, V. Tsigkou, E. Kokkou et al., "Smoking and atherosclerosis: mechanisms of disease and new therapeutic approaches," Current Medicinal Chemistry, vol. 21, no. 34, pp. 3936–3948, 2014.

[9] S. S. Hecht, "Progress and challenges in selected areas of tobacco carcinogenesis," Chemical Research in Toxicology, vol. 21, no. 1, pp. 160–171, 2008.

[10] M. W. Weng, H. W. Lee, S. H. Park et al., "Aldehydes are the predominant forces inducing DNA damage and inhibiting DNA repair in tobacco smoke carcinogenesis," Proceedings of the National Academy of Sciences of the United States of America, vol. 115, no. 27, pp. E6152–E6161, 2018.

[11] S. Klaassen, S. Probst, E. Oechslin et al., "Mutations in sarcomere protein genes in left ventricular noncompaction," Circulation, vol. 117, no. 22, pp. 2893–2901, 2008.
[12] H. Morita, H. L. Rehm, A. Menesses et al., “Shared genetic causes of cardiac hypertrophy in children and adults,” The New England Journal of Medicine, vol. 358, no. 18, pp. 1899–1908, 2008.

[13] L. Song, Y. Zou, J. Wang et al., “Mutations profile in Chinese patients with hypertrophic cardiomyopathy,” Clinica Chimica Acta, vol. 351, no. 1-2, pp. 209–216, 2005.

[14] E. Villard, L. Duboscq-Bidot, P. Charron et al., “Mutation screening in dilated cardiomyopathy: prominent role of the beta myosin heavy chain gene,” European Heart Journal, vol. 26, no. 8, pp. 794–803, 2005.

[15] J. Sun, S. Li, F. Wang, C. Fan, and J. Wang, “Identification of key pathways and genes in PTEN mutation prostate cancer by bioinformatics analysis,” BMC Medical Genetics, vol. 20, no. 1, p. 191, 2019.

[16] Y. H. Huang, C. Z. Zhang, Q. S. Huang et al., “Clinicopathologic features, tumor immune microenvironment and genomic landscape of Epstein-Barr virus-associated intrahepatic cholangiocarcinoma,” Journal of Hepatology, vol. 74, no. 4, pp. 838–849, 2021.

[17] S. Sun, J. H. Schiller, and A. F. Gazdar, “Lung cancer in never smokers – a different disease,” Nature Reviews. Cancer, vol. 7, no. 10, pp. 778–790, 2007.

[18] T. Zhang, P. Joubert, N. Ansari-Pour et al., “Genomic and evolutionary classification of lung cancer in never smokers,” Nature Genetics, vol. 53, no. 9, pp. 1348–1359, 2021.

[19] S. Dean, R. Lennox, C. Senko, and S. Parakh, “Lung cancer in non-smokers: a diagnosis of increasing importance,” The Medical Journal of Australia, vol. 216, no. 7, pp. 342–343, 2022.

[20] Y. J. Chen, T. I. Roumeliotis, Y. H. Chang et al., “Proteogenomics of non-smoking lung cancer in East Asia delineates molecular signatures of pathogenesis and progression,” Cell, vol. 182, no. 1, pp. 226–244.e17, 2020.

[21] L. Y. Gou, F. Y. Niu, Y. L. Wu, and W. Z. Zhong, “Differences in driver genes between smoking-related and non-smoking-related lung cancer in the Chinese population,” Cancer, vol. 121, Supplement 17, pp. 3069–3079, 2015.