

Original Research

# Identification of Candidate mRNA Genes and Their Potential MicroRNA Targets in Lung Cancer Induced by Smoking Tobacco

Amresh Kumar Mishra<sup>1</sup>, Neha Mumtaz<sup>1</sup>, Maneesh Kumar Misra<sup>2,\*</sup> <sup>1</sup>Environmental Engineering, Department of Civil Engineering, Integral University, Lucknow, Uttar Pradesh 226026, India<sup>2</sup>Department of Pathology, The University of Chicago Medicine, Chicago, IL 60637, USA\*Correspondence: [mmisra@bsd.uchicago.edu](mailto:mmisra@bsd.uchicago.edu) (Maneesh Kumar Misra)

Academic Editor: Alessandro Poggi

Submitted: 6 July 2023 Revised: 1 September 2023 Accepted: 27 October 2023 Published: 27 November 2023

## Abstract

**Background:** Smoking is considered the single highest risk factor for lung cancer and has been suggested to be associated with accelerated somatic mutations in respiratory mucosa that lead to the development of lung cancer. MicroRNAs serve as modulators in smoking-induced mRNA gene expression changes in the human airway epithelium and are linked to the development of lung cancer. The thermodynamics in the microRNA (miRNA)–mRNA interactions may be affected in tobacco smokers, consequently, leading to phenotypic variations in lung cancer patients. Therefore, this study aimed to investigate the impact of smoking tobacco on somatic mutations in mRNA genes and assess their potential impact on miRNA–mRNA interactions in lung cancers. **Methods:** The clinically significant pathogenic variants in mRNA genes in the dataset in lung cancer cases linked to smoking tobacco (n = 330) were obtained from the Cancer Atlas database (TCGA, <http://cancergenome.nih.gov/>) and used to assess the potential role of tobacco consumption in driving the genetic alterations in proto-oncogenes associated with lung cancer. The analysis of the miRNA interaction with the top five altered mRNA proto-oncogenes in lung cancer cases due to tobacco consumption was performed using the target prediction function in the miRDP program (Database version 5.2.3.1, <https://mirdb.org/>). **Results:** We identified the top five mRNA proto-oncogenes enriched with simple somatic mutations (SSM) in lung cancer were *TP53*, *EGFR*, *KRAS*, *FAT4*, and *KMT2D*. Interestingly, we observed the highest incidence of SSM in the *Tumor Protein p53 (TP53)* gene at 63.64%. Similarly, the SSM incidence was 23.94% in the *Epidermal Growth Factor Receptor (EGFR)*, 22.12% in the *Kirsten Rat Sarcoma Viral Oncogene Homolog (KRAS)*, 18.48% in the *FAT Atypical Cadherin 4 (FAT4)*, and 14.24% in the *Lysine (K)-Specific Methyltransferase 2D (KMT2D)* genes. Subsequently, we used a bioinformatics approach to assess the effect of miRNA–mRNA interactions in lung cancer among the top five SSM-enriched mRNA proto-oncogenes. Among the top 20 identified and selected miRNAs, we observed 18 unique microRNAs that bind specifically to *TP53*, *KRAS*, and *FAT4* genes and 17 and 19 microRNAs that exclusively bind with the *EGFR* and *KMT2D* genes, respectively. **Conclusions:** Our study found that the top five SSM-enriched mRNA proto-oncogenes in lung cancers among tobacco smokers were *TP53*, *EGFR*, *KRAS*, *FAT4*, and *KMT2D*. Further, our results provide an important insight into the involvement of the intricate network of mRNA–miRNA interactions in the development of lung cancer.

**Keywords:** lung cancer; tobacco consumption; somatic mutation; mRNA; microRNA

## 1. Introduction

Lung cancer is considered the most commonly diagnosed cancer [1,2], and is the number one cause of cancer-related deaths worldwide [1], which primarily result from smoking tobacco [2,3]. The risk of causation of lung cancer in smokers is reported to be nearly ten-fold higher compared to non-smokers [3,4]. Tobacco smoke contains numerous toxic and carcinogenic compounds, such as nicotine, nitrogen oxides, and carbon monoxide or cadmium [5,6]. Tobacco consumption is linked to an alteration in the expression of oncogenes and tumor suppressor genes [7]. Previous reports indicate that alterations in gene expression may affect the incidence and development of tumors intimately induced by epigenetic modulations [7,8]. The epigenetic alterations could be in microRNA (miRNA), DNA methylation, histone modifications, and nucleosome remodeling [7]. These epigenetic alterations are standalone but may

interact with each other to systematically control overall gene expression [9]. Thus, the interruptions of epigenetic pathways might encourage the procurement of a cancerous phenotype, the development of lung cancer cells, and the modulation of the response to lung cancer therapy among individuals [10].

MiRNAs are small, single-stranded, and highly conserved non-coding RNAs comprising about 19–24 nucleotides and are considered an important gene regulator of post-transcriptional modifications in lung cancer [11]. A single miRNA can regulate numerous different transcripts in the human genome alongside multiple RNA-coding transcripts [12]. Previous studies from our group and others have shown that miRNAs are involved in numerous biological processes by binding with their putative mRNA targets and regulating their expression [13–15]. The dysregulation of miRNAs is linked to the progression and suppression of



cancer, thereby establishing miRNAs as important players in tumor suppressor genes or oncogenes [16].

The advent of high-throughput next-generation sequencing technology has reduced the costs and provided increased coverage to detect accurately the single nucleotide variants that occur in the coding and non-coding regions of human genomes during lung tumorigenesis. The precise analysis of these data detects novel variants and vital genes and biological processes related to the causation of tumors and provides an improved understanding of the mechanisms of occurrence, development, and prognosis associated with lung cancer [17,18]. The Clinical Proteomic Tumor Analysis Consortium (CPTAC) utilizes cutting-edge proteomic technologies, and workflows to clinical tumor samples with characterized genomic, and transcript profiles to fast-track and pinpoint the molecular basis of cancer under the auspices of the National Cancer Institute's Office of Cancer Clinical Proteomics Research [19]. The CPTAC Data Portal is hosted under the auspices of the Cancer Genome Atlas database (TCGA, <http://cancergenome.nih.gov/>), and has revealed clinically significant pathogenic variants in mRNA through molecular studies in lung squamous cell carcinoma (LUSC) [17] and adenocarcinoma (LUAD) [18]. The TCGA database comprises data on clinically significant somatic mutations in human genes among lung cancer patients with a history of tobacco consumption. However, a systematic analysis is lacking that assesses the interactions between top enriched mRNA genes and miRNAs relating to the causation of lung cancer from tobacco consumption. Therefore, the present study was designed to perform this analysis. The goal of this study was to evaluate the impact of smoking tobacco on the incidence of simple somatic mutations in lung cancer. Subsequently, we also investigated the associations between miRNAs and mRNA candidate genes in the susceptibility and prognosis of lung cancer in tobacco smokers.

## 2. Material and Methods

### 2.1 Lung Cancer Dataset of Somatic Mutations in mRNA Proto-Oncogenes in Tobacco Smokers

The clinically significant pathogenic variants in mRNA genes in the dataset from tobacco smokers with lung cancer ( $n = 330$ ) were obtained from the CPTAC Data Portal, hosted at the Cancer Atlas database (TCGA, <http://cancergenome.nih.gov/>), to assess the potential role of tobacco consumption in driving the genetic alterations in proto-oncogenes for the causation risk of lung cancer (Fig. 1). Among this dataset of 330 lung cancer cases (229 males, 101 females; 37–88 years old), we selected the top five enriched proto-oncogenes (mRNA) to determine their interactions with miRNAs. The workflow of the data extraction from the CPTAC data portal and smoking interaction with mRNA genes is shown in Fig. 2.

### 2.2 Identification of the Most Probable miRNA Targets of Simple Somatic Mutations (SSM)-Enriched mRNA Proto-Oncogenes

The analysis of the miRNA interaction with the top five altered proto-oncogenes in lung cancer cases resulting from tobacco consumption was performed using the target prediction function in the miRDP program (Database version 5.2.3.1) [20], which provides a more reliable prediction of the mRNA gene miRNA targets. The miRDP program uses a database of nearly 152 million human miRNA–target predictions, which have been collected across thirty different resources, and provides an integrative score that is statistically inferred from the obtained predictions and assigned to each unique microRNA–target interaction to give a unified measure of confidence [20]. This algorithm produces a more reliable prediction of miRNA targets to remove the selection bias, and the predictions are cross-verified using an experimentally validated miRNA–mRNA target interaction dataset [20]. The workflow relating to the miRNA–mRNA interactions is illustrated in Fig. 3.

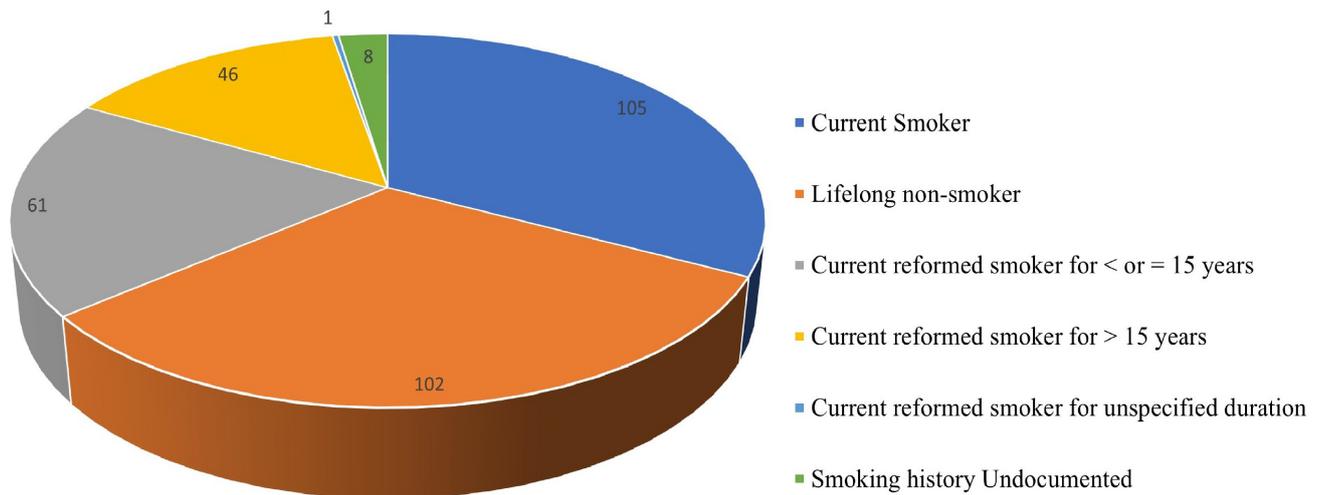
### 2.3 Statistical Analysis

Qualitative data are expressed as frequency and percentage. A two-tailed Fisher exact test was used to compute the differences in the incidence of simple somatic mutations in the proto-oncogenes of lung cancer patients who were either tobacco smokers or non-smokers. The magnitude of the effect was estimated by the odds ratio (OR) and its 95% confidence interval (CI). The statistical significance was considered only when the  $p$ -value was  $<0.05$ . Statistical tests were performed using the R language for statistical computing (<https://cran.r-project.org/bin/windows/base/old/4.2.2/>).

## 3. Results

### 3.1 Incidence of Simple Somatic Mutations in mRNA Proto-Oncogenes in Lung Cancer Cases among Tobacco Smokers and Non-Smokers

A summary of tobacco smoking status is summarized in **Supplementary Table 1** for each lung cancer case. We identified and selected the top five enriched proto-oncogenes containing simple somatic mutations in lung cancer cases ( $n = 330$ ) with a history of tobacco consumption, and the results are presented in Table 1. We observed the highest SSM incidence in the *TP53* gene at 63.64% (Table 1). Similarly, the SSM incidence was 23.94% in the *EGFR*, 22.12% in the *KRAS*, 18.48% in the *FAT4*, and 14.24% in the *KMT2D* genes (Table 1). Next, we examined whether the mutation frequency in the top five enriched proto-oncogenes with SSMs differed between lung cancer patients who were tobacco smokers and those who did not smoke (Table 1). Interestingly, we observed a three-fold and almost two-fold higher incidence of SSMs in two proto-oncogenes, namely, *EGFR* (OR = 3.33, 95% CI = 2.36–4.69,  $p = 4.67 \times 10^{-12}$ ) and *KRAS* (OR = 1.60, 95% CI =



**Fig. 1. Lung cancer cases (n = 330) according to tobacco smoking status.** Current smoker (n = 105) is represented in blue color; lifelong non-smoker (n = 102) in orange color; current reformed smoker for < or = 15 years (n = 61) in grey color; current reformed smoker for >15 years (n = 46) in gold color; current reformed smoker for the unspecified duration (n = 1) in light blue color; smoking history undocumented (n = 8) shown in green color.

**Table 1. Simple somatic mutation incidences in lung cancer cases among tobacco smokers and non-smokers.**

Gene symbol	Cytoband	#SSMs incidence in tobacco smokers	#SSMs incidence in non-smokers	OR	95% CI	p-value
<i>TP53</i>	17p13.1	210/330 (63.64%)	687/1041 (65.99%)	0.90	0.69–1.18	0.465
<i>EGFR</i>	7p11.2	79/330 (23.94%)	93/1041 (8.93%)	3.33	2.36–4.69	$4.67 \times 10^{-12}$ *
<i>KRAS</i>	12p12.1	73/330 (22.12%)	157/1041 (15.08%)	1.60	1.15–2.20	0.004*
<i>FAT4</i>	4q28.1	61/330 (18.48%)	184/1041 (17.68%)	1.06	0.75–1.47	0.742
<i>KMT2D</i>	12q13.12	47/330 (14.24%)	154/1041 (14.79%)	0.96	0.66–1.37	0.859

\*Statistically significant p-value.

Gene names, *TP53*, tumor protein p53; *EGFR*, epidermal growth factor receptor; *KRAS*, Kirsten rat sarcoma viral oncogene homolog; *FAT4*, FAT atypical cadherin 4; *KMT2D*, lysine (K)-specific methyltransferase 2D; #SSMs, simple somatic mutations; OR, Odds Ratio; CI, Confidence Interval.

1.15–2.20,  $p = 0.004$ ), respectively, among tobacco smoking lung cancer patients compared to non-smoker lung cancer patients (Table 1). However, no significant difference was found in the SSM incidence for *TP53* (OR = 0.90, 95% CI = 0.69–1.18,  $p = 0.465$ ), *FAT4* (OR = 1.06, 95% CI = 0.75–1.47,  $p = 0.742$ ), and *KMT2D* (OR = 0.96, 95% CI = 0.66–1.37,  $p = 0.859$ ) proto-oncogenes in lung cancer cases of tobacco smokers compared to non-smokers (Table 1).

### 3.2 Identification of Most Probable MicroRNA Targets of the Enriched mRNA Proto-Oncogenes in Lung Cancer

We used a bioinformatic approach to identify and select the top 20 most probable microRNA targets of the enriched mRNA proto-oncogenes in lung cancer, and the results are presented in Fig. 4 as a heatmap diagram. Interestingly, we identified 18 unique microRNAs that bind to the *TP53*, *KRAS*, and *FAT4* genes and 17 and 19 microRNAs that exclusively bind to the *EGFR* and *KMT2D* genes, respectively. Two microRNAs, namely, *has-miR-6812-5p* and *has-miR-6819-5p* both bind to *TP53* and *EGFR*. Simi-

larly, *has-miR-193a-3p* and *has-miR-193b-3p* both bind to the *KRAS* and *FAT4* genes, while *has-miR-608* binds to both *EGFR* and *KMT2D* (Table 2).

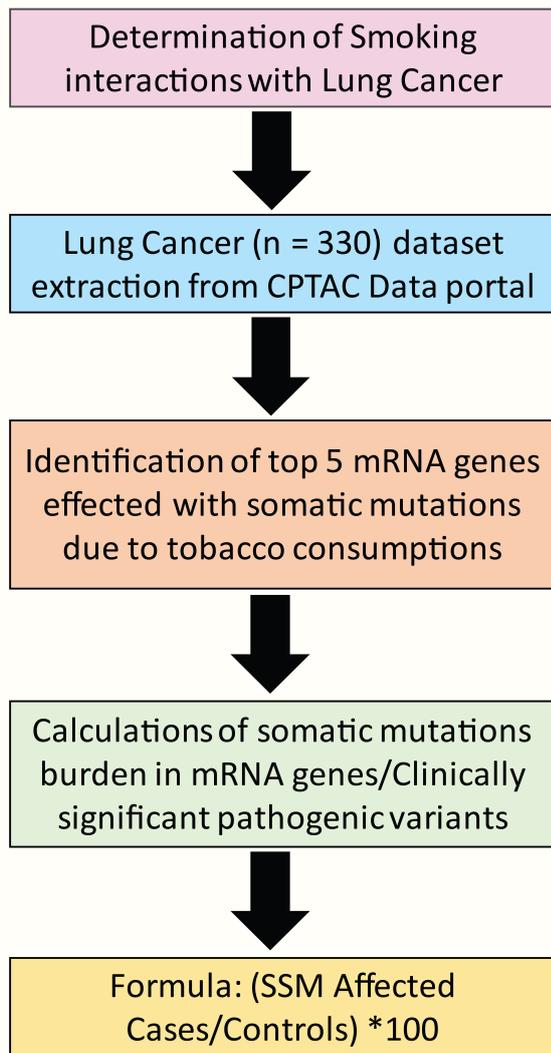
**Table 2. Common microRNAs that bind to the top five somatic mutations enriched in the mRNA proto-oncogenes found in lung cancer cases.**

MicroRNA	<i>TP53</i>	<i>KRAS</i>	<i>EGFR</i>	<i>FAT4</i>	<i>KMT2D</i>
<i>has-miR-6812-5p</i>	0.374 <sup>#</sup>		0.391 <sup>#</sup>		
<i>has-miR-6819-5p</i>	0.364 <sup>#</sup>		0.450 <sup>#</sup>		
<i>has-miR-193a-3p</i>		0.844*		0.751*	
<i>has-miR-193b-3p</i>		0.830*		0.723*	
<i>has-miR-608</i>			0.668 <sup>‡</sup>		0.667 <sup>‡</sup>

<sup>#</sup>Human microRNAs that bind both *TP53* and *EGFR*.

\*Human microRNAs that bind both *KRAS* and *FAT4*.

<sup>‡</sup>Human microRNAs that bind both *EGFR* and *KMT2D*.



**Fig. 2. Workflow of smoking interactions with mRNA genes in lung cancer.** CPTAC, Clinical Proteomic Tumor Analysis Consortium; SSM, Simple Somatic Mutations.

### 3.3 Construction of the miRNA–mRNA Regulatory Network

We constructed the miRNA–mRNA regulatory network using the Cytoscape software (version 3.9.1, <https://cytoscape.org/>) [21]. The top 20 most suitable miRNAs with the highest integrated score were selected for each of the top five enriched proto-oncogenes with simple somatic mutations in the lung cancer cases to construct a miRNA–mRNA regulatory network that consisted of interactions for a total of one hundred miRNAs with five SSM-enriched mRNA proto-oncogenes (Fig. 5).

### 3.4 Protein–Protein Interaction Analysis of the Top Five SSM-Enriched Proto-Oncogenes in Lung Cancer, and KEGG Pathway Enrichment Analysis

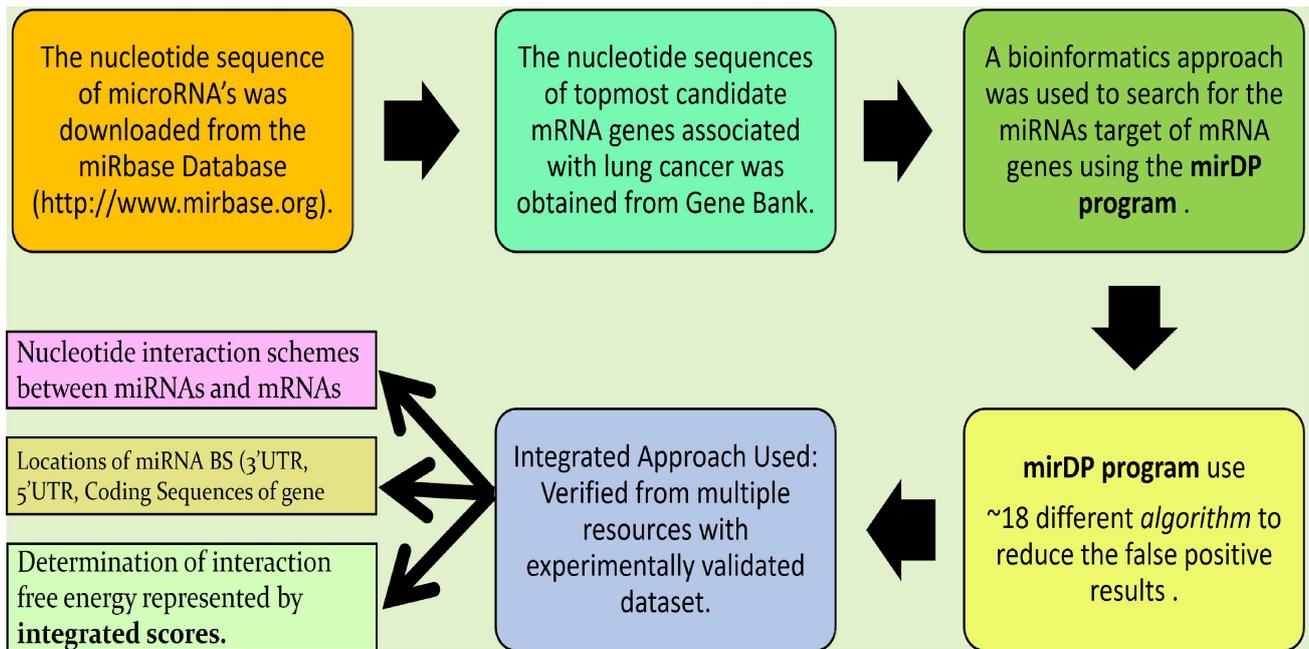
We performed protein–protein interaction (PPI) analysis on the top five enriched mRNA proto-oncogenes in

lung cancer and assessed their potential interactions with the top fifty closest neighboring genes in the network using the STRING database in Cytoscape software (version 3.9.1, <https://cytoscape.org/>) [21] and found a PPI network that consisted of 55 unique nodes with 1035 edges (Supplementary Fig. 1). The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed using the same STRING database to predict the potential biological functions of the overlapping genes that are overrepresented in the identified PPI network with the top five SSM-enriched mRNA proto-oncogenes in lung cancer. The results of the top thirty identified biological pathways are presented in Fig. 6.

## 4. Discussion

MicroRNAs serve as modulators of smoking-induced mRNA gene expression changes in the human airway epithelium [22] and have been implicated in the process of tumorigenesis through regulation of the cell cycle, metastasis, angiogenesis, metabolism, and apoptosis [23,24]. Smoking is considered the single highest risk factor for lung cancer [25,26] and has been suggested as being associated with accelerated somatic mutations in the respiratory mucosa that lead to the development of lung cancer [27]. The thermodynamics in the RNA–RNA interactions between miRNA and their mRNA target sites may be affected because of the presence of somatic mutations, resulting in the deregulation of target mRNA genes, and consequently, leading to phenotypic variations and lung cancer susceptibility. Considering these factors, this study examined the impact of smoking tobacco on the development of somatic mutations in mRNA genes and assessed their potential impact on the miRNA–mRNA interactions in lung cancers.

We found that the top five SSM-enriched mRNA proto-oncogenes in lung cancer were *TP53*, *EGFR*, *KRAS*, *FAT4*, and *KMT2D*. Interestingly, we observed the highest incidence of SSM in the *TP53* gene at 63.64%. Similarly, the incidence of SSM was 23.94% in *EGFR*, 22.12% in *KRAS*, 18.48% in *FAT4*, and 14.24% in *KMT2D* (Table 1). Our results are consistent with previous reports, which have shown that the *TP53* gene possesses the commonest somatic mutation in all human cancers, including lung cancers associated with smoking cigarettes [28–30]. Large-scale screening for somatic mutations in lung cancer studies has further strengthened our observations of the reported association between *TP53*, *EGFR*, and *KRAS* somatic mutations and lung cancer susceptibility [29,31,32]. The histone methyltransferase *KMT2D* is among the most highly inactivated epigenetic modifiers in lung cancer, and somatic mutations in this gene have been previously reported to be associated with vulnerability to lung cancer [33,34]. *FAT4* is an enormously large, atypical cadherin with critical roles in the regulation of planar cell polarity (PCP) and controlling the Hippo signaling pathway [35], while somatic mutations in this gene are also linked to lung cancer [36]. Similar to



**Fig. 3. Summary of the bioinformatics pipeline used to assess the microRNA (miRNA)–mRNA interactions.**

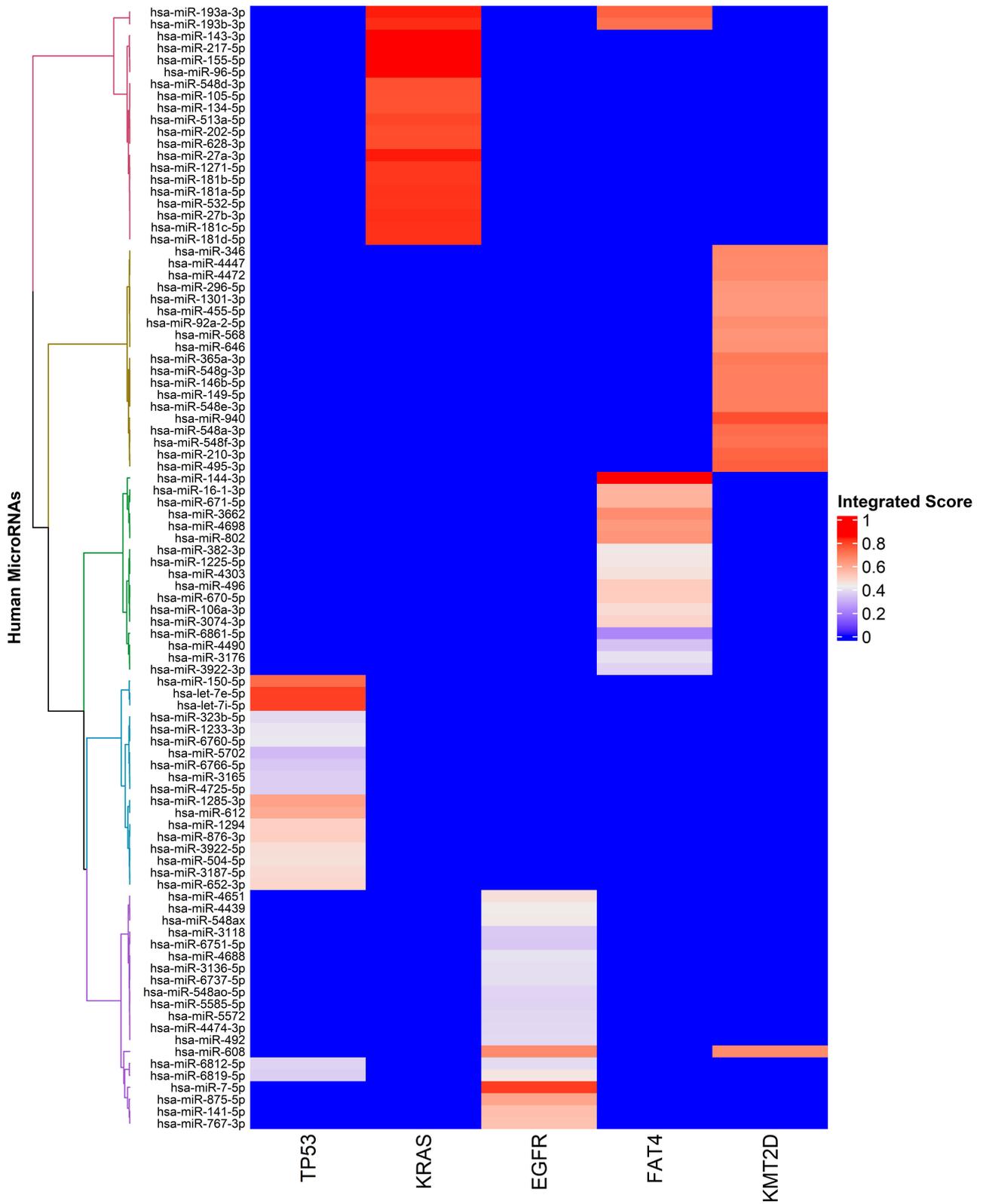
our observation in this study, the comprehensive genomic profiling in a previous lung cancer study revealed a significant association of high tumor burden with somatic mutations in the *TP53*, *KMT2D*, and *FAT1* genes [36].

We observed a three-fold and almost two-fold higher incidence of SSMs in two proto-oncogenes, namely, *EGFR* (OR = 3.33, 95% CI = 2.36–4.69,  $p = 4.67 \times 10^{-12}$ ) and *KRAS* (OR = 1.60, 95% CI = 1.15–2.20,  $p = 0.004$ ), respectively, in lung cancer cases among tobacco smokers compared to non-smokers (Table 1). Similarly, the *EGFR* and *KRAS* somatic mutations were also found to be more frequent in smokers compared to non-smokers [37,38]. Further, we found no significant difference in the SSM incidences for *TP53*, *FAT4*, and *KMT2D* proto-oncogenes in lung cancer cases related to tobacco smokers compared to non-smokers (Table 1). Our findings are consistent with a previous report that found no significant differences in the *TP53* somatic mutation frequencies in lung cancer cases of smokers compared to non-smokers [39]. To the best of our knowledge, this is the first study that assessed and showed the frequency differences in somatic mutations in *FAT4* and *KMT2D* proto-oncogenes between smokers and non-smokers in cases of lung cancer. Future studies are needed to confirm our findings regarding the differences in lung cancer cases between the frequency of somatic mutations in smokers and non-smokers.

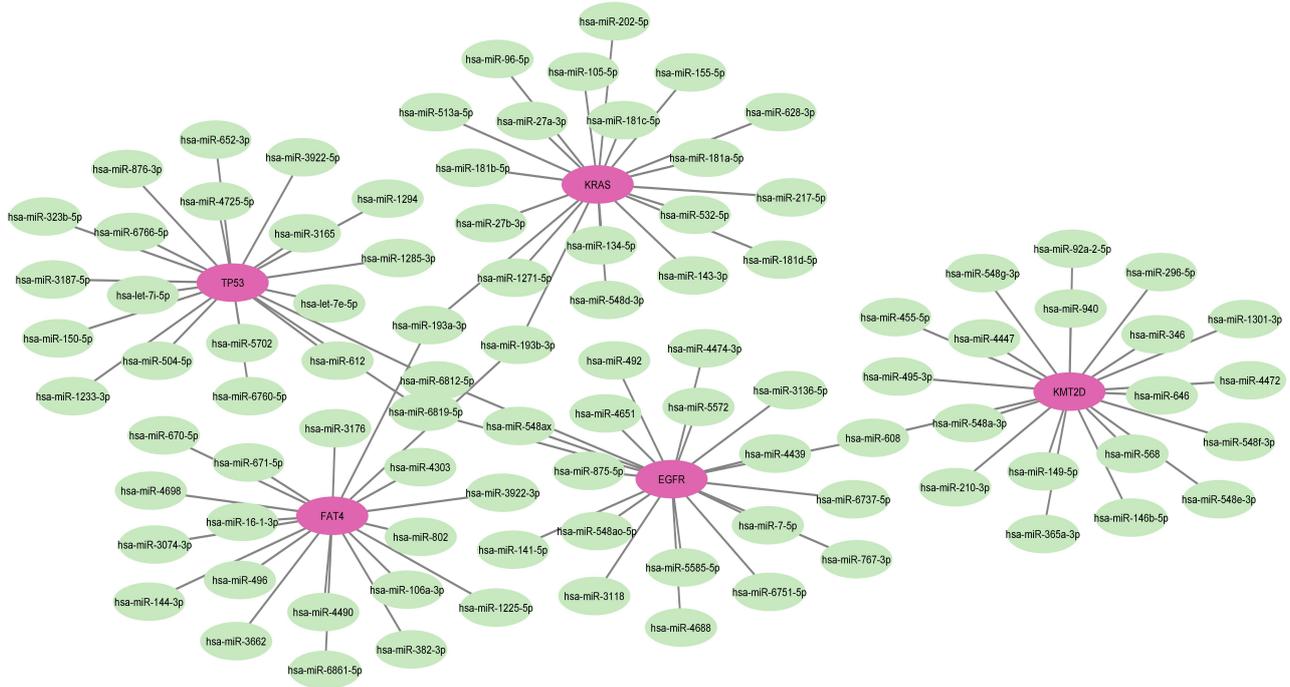
In the next step, we used a bioinformatics approach to evaluate the effect of miRNA and mRNA interactions in lung cancer among the top SSM-enriched mRNA proto-oncogenes, namely, *TP53*, *EGFR*, *KRAS*, *FAT4*, and *KMT2D*. We found the top 20 most probable miRNAs that target these five enriched mRNA proto-oncogenes in

lung cancer (Fig. 4), which were displayed in the miRNA–mRNA regulatory network (Fig. 5). Among these top 20 miRNAs, we observed 18 unique microRNAs that bind specifically to *TP53*, *KRAS*, and *FAT4* genes and 17 and 19 microRNAs that bind exclusively to the *EGFR* and *KMT2D* genes, respectively. Two microRNAs, namely, *has-miR-6812-5p* and *has-miR-6819-5p* both bind to *TP53* and *EGFR*. Similarly, *has-miR-193a-3p* and *has-miR-193b-3p* both bind to the *KRAS* and *FAT4* genes, while *hsa-miR-608* binds to both *EGFR* and *KMT2D*. Recent reports suggest that the role of miRNAs in targeting the *TP53* [40], *EGFR* [41], *KRAS* [42], *FAT4* [43], and *KMT2D* [44] genes, which play a pivotal role in tumorigenesis [40–44], are consistent with our findings, whereby miRNAs target these mRNA proto-oncogenes in lung cancer, namely, *TP53*, *EGFR*, *KRAS*, *FAT4*, and *KMT2D*. Further, the protein–protein interactions (PPIs) associated with the smoking-induced simple somatic mutations in the top five enriched mRNA proto-oncogenes found in lung cancer and the fifty closest neighboring genes were used to construct a network of 55 proteins with 1035 unique interactions (Supplementary Fig. 1), while the KEGG pathway enrichment analysis provided an important insight into the potential biological functions of these overlapping genes in lung cancer (Fig. 6).

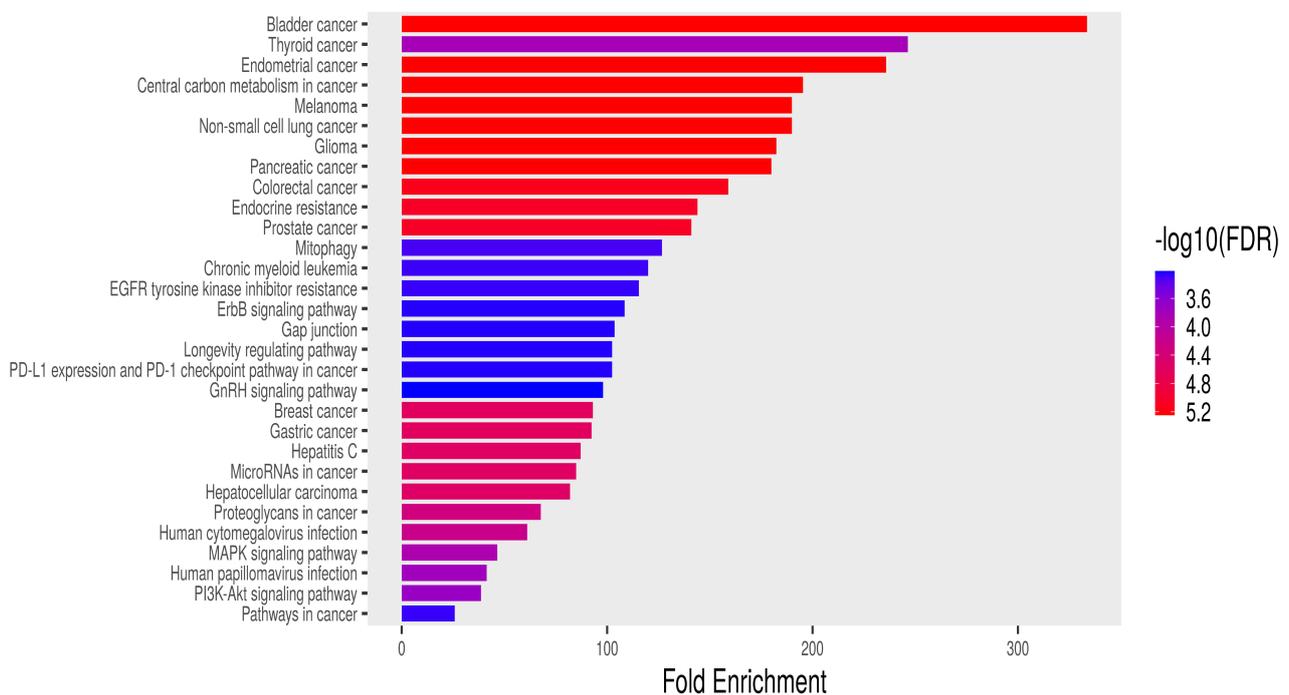
The tumor suppressor *TP53* gene is referred to as the guardian of the genome and plays a central role in tumor suppression [45,46]. Thus, interruption of the usual *TP53* function frequently leads to the onset and/or progression of cancer [45]. Previous studies have proven that miRNAs contribute to the regulation of *TP53* levels and function [40]. Recently, more than 20 miRNAs have been demonstrated to serve as direct negative regulators of the *TP53*



**Fig. 4. Heatmap of Integrated binding score for Human MicroRNA and mRNA gene interactions.** The human microRNAs are listed on the Y-axis and the integrated binding score of the mRNA proto-oncogenes in tobacco-smoking lung cancer patients is shown on the X-axis. Each cell represents the integrated binding score of the human miRNA–mRNA interactions on a heatmap color scale.



**Fig. 5. MicroRNAs and mRNA interaction network consisting of interactions for a total of one hundred miRNAs with five SSM-enriched mRNA proto-oncogenes.** Human microRNAs are represented in the light lime green color and mRNA genes are shown in purple.



**Fig. 6. Bar plot of the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis for the overrepresented genes in the protein–protein interaction network of the top five enriched mRNA proto-oncogenes in lung cancer with overlapping genes.** The KEGG pathway displays thirty enriched biological pathways from the top to bottom of the Y-axis, and the X-axis represents fold enrichment of the genes.

level, with several of them frequently showing increased expression in tumors [40]. Similarly, we have also found more than 20 miRNAs that target the *TP53* gene and may potentially downregulate *TP53* expression. Some of the miRNAs (for example: hsa-let-7) that were identified as a potential target for *TP53* have also been previously found in these studies [40,47], thereby further strengthening our findings.

The *epidermal growth factor receptor (EGFR)* gene was among the first molecules to be chosen for targeted gene therapy in lung cancer [48]. A recent report suggests that *EGFR* mutations can be controlled by miRNAs in cancer therapies [49]. Further reports demonstrated that specific miRNA/*EGFR* axes may play crucial roles in lung tumorigenesis [41]. In the present study, we selected the top 20 miRNAs that could potentially target *EGFR* expression, along with *KRAS*, *FAT4*, and *KMT2D*. Further, experimental validation of these observed mRNA–miRNA interactions in our study may provide an important insight into the role of these interactions in lung cancer and their consequences in lung carcinogenesis. The improved knowledge of miRNA biology will assist in establishing the basis for the development of novel approaches to alter miRNA expressions and functions. The development of a novel class of synthetic oligonucleotide molecules to act as an inhibitor or enhancer can work to regulate the expression of a specific miRNA. Anti-miRs serve as inhibitors, which is an important characteristic of an ideal therapeutic agent.

## 5. Conclusions

To conclude, our study identified the top five SSM-enriched mRNA proto-oncogenes in lung cancers among tobacco smokers, namely, *TP53*, *EGFR*, *KRAS*, *FAT4*, and *KMT2D*. Further, the results of our study provide an important insight into the intricate network of mRNA–miRNA interactions involved in the development of lung cancer. *In silico* analysis showed quite significant observations, which suggest that expression-based studies should be carried out to assess the impact of predicted miRNA targets of altered mRNA proto-oncogenes in lung cancers among tobacco smokers.

## Availability of Data and Materials

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Author Contributions

MKM, AKM, and NM designed the research study, and performed the research. AKM collected the data. NM, and MKM oversaw initial data collection. AKM, and MKM analyzed the data. AKM and MKM wrote the manuscript. All authors have participated sufficiently in the work to take

public responsibility for appropriate portions of the content and agreed to be accountable for all aspects of the work in ensuring that questions related to its accuracy or integrity.

## Ethics Approval and Consent to Participate

Not applicable.

## Acknowledgment

Not applicable.

## Funding

This research received no external funding.

## Conflict of Interest

The authors declare no conflict of interest.

## Supplementary Material

Supplementary material associated with this article can be found, in the online version, at <https://doi.org/10.31083/j.fbs1504013>.

## References

- [1] Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lung cancer. *Lancet* (London, England). 2021; 398: 535–554.
- [2] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a Cancer Journal for Clinicians*. 2018; 68: 394–424.
- [3] Shields PG. Molecular epidemiology of smoking and lung cancer. *Oncogene*. 2002; 21: 6870–6876.
- [4] Hecht SS. Tobacco smoke carcinogens and lung cancer. *Journal of the National Cancer Institute*. 1999; 91: 1194–1210.
- [5] Talhout R, Schulz T, Florek E, van Benthem J, Wester P, Opperhuizen A. Hazardous compounds in tobacco smoke. *International Journal of Environmental Research and Public Health*. 2011; 8: 613–628.
- [6] Rennard SI. Cigarette smoke in research. *American Journal of Respiratory Cell and Molecular Biology*. 2004; 31: 479–480.
- [7] Nebbioso A, Tambaro FP, Dell'Aversana C, Altucci L. Cancer epigenetics: Moving forward. *PLoS Genetics*. 2018; 14: e1007362.
- [8] Flavahan WA, Gaskell E, Bernstein BE. Epigenetic plasticity and the hallmarks of cancer. *Science* (New York, N.Y.). 2017; 357: eaal2380.
- [9] Portela A, Esteller M. Epigenetic modifications and human disease. *Nature Biotechnology*. 2010; 28: 1057–1068.
- [10] Duruisseau M, Esteller M. Lung cancer epigenetics: From knowledge to applications. *Seminars in Cancer Biology*. 2018; 51: 116–128.
- [11] Kang SM, Lee HJ. MicroRNAs in human lung cancer. *Experimental Biology and Medicine* (Maywood, N.J.). 2014; 239: 1505–1513.
- [12] Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell*. 2009; 136: 215–233.
- [13] Misra MK, Mishra A, Pandey SK, Kapoor R, Sharma RK, Agrawal S. Genetic variation in Micro-RNA genes of host genome affects clinical manifestation of symptomatic Human Cytomegalovirus infection. *Human Immunology*. 2015; 76: 765–769.

- [14] Misra MK, Pandey SK, Kapoor R, Sharma RK, Agrawal S. Genetic variants of MicroRNA-related genes in susceptibility and prognosis of end-stage renal disease and renal allograft outcome among north Indians. *Pharmacogenetics and Genomics*. 2014; 24: 442–450.
- [15] Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*. 2005; 433: 769–773.
- [16] Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. *Nature Reviews. Cancer*. 2006; 6: 259–269.
- [17] Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. 2012; 489: 519–525.
- [18] Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014; 511: 543–550.
- [19] Edwards NJ, Oberti M, Thangudu RR, Cai S, McGarvey PB, Jacob S, *et al.* The CPTAC Data Portal: A Resource for Cancer Proteomics Research. *Journal of Proteome Research*. 2015; 14: 2707–2713.
- [20] Tokar T, Pastrello C, Rossos AEM, Abovsky M, Hauschild AC, Tsay M, *et al.* mirDIP 4.1-integrative database of human microRNA target predictions. *Nucleic Acids Research*. 2018; 46: D360–D370.
- [21] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*. 2003; 13: 2498–2504.
- [22] Schembri F, Sridhar S, Perdomo C, Gustafson AM, Zhang X, Ergun A, *et al.* MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106: 2319–2324.
- [23] Yang W, Lee DY, Ben-David Y. The roles of microRNAs in tumorigenesis and angiogenesis. *International Journal of Physiology, Pathophysiology and Pharmacology*. 2011; 3: 140–155.
- [24] Iqbal MA, Arora S, Prakasam G, Calin GA, Syed MA. MicroRNA in lung cancer: role, mechanisms, pathways and therapeutic relevance. *Molecular Aspects of Medicine*. 2019; 70: 3–20.
- [25] Aredo JV, Luo SJ, Gardner RM, Sanyal N, Choi E, Hickey TP, *et al.* Tobacco Smoking and Risk of Second Primary Lung Cancer. *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer*. 2021; 16: 968–979.
- [26] Fearon ER. The smoking gun and the damage done: genetic alterations in the lungs of smokers. *Journal of the National Cancer Institute*. 1997; 89: 834–836.
- [27] Yoshida K, Gowers KHC, Lee-Six H, Chandrasekharan DP, Coorens T, Maughan EF, *et al.* Tobacco smoking and somatic mutations in human bronchial epithelium. *Nature*. 2020; 578: 266–272.
- [28] Van Egeren D, Kohli K, Warner JL, Bedard PL, Riely G, Lepisto E, *et al.* Genomic analysis of early-stage lung cancer reveals a role for TP53 mutations in distant metastasis. *Scientific Reports*. 2022; 12: 19055.
- [29] Frankell AM, Dietzen M, Al Bakir M, Lim EL, Karasaki T, Ward S, *et al.* The evolution of lung cancer and impact of subclonal selection in TRACERx. *Nature*. 2023; 616: 525–533.
- [30] Suzuki H, Takahashi T, Kuroishi T, Suyama M, Ariyoshi Y, Takahashi T, *et al.* p53 mutations in non-small cell lung cancer in Japan: association between mutations and smoking. *Cancer Research*. 1992; 52: 734–736.
- [31] Rosell R, Karachaliou N. Large-scale screening for somatic mutations in lung cancer. *Lancet (London, England)*. 2016; 387: 1354–1356.
- [32] van den Heuvel GRM, Kroeze LI, Ligtenberg MJL, Grünberg K, Jansen EAM, von Rhein D, *et al.* Mutational signature analysis in non-small cell lung cancer patients with a high tumor mutational burden. *Respiratory Research*. 2021; 22: 302.
- [33] Alam H, Tang M, Maitiuheti M, Dhar SS, Kumar M, Han CY, *et al.* KMT2D Deficiency Impairs Super-Enhancers to Confer a Glycolytic Vulnerability in Lung Cancer. *Cancer Cell*. 2020; 37: 599–617.e7.
- [34] Pan Y, Han H, Hu H, Wang H, Song Y, Hao Y, *et al.* KMT2D deficiency drives lung squamous cell carcinoma and hypersensitivity to RTK-RAS inhibition. *Cancer Cell*. 2023; 41: 88–105.e8.
- [35] Ning Y, Yang Y, Zheng H, Zhan Y, Zang H, Wen Q, *et al.* Increased expression of FAT4 suppress metastasis of lung adenocarcinoma through regulating MAPK pathway and associated with immune cells infiltration. *Cancer Medicine*. 2023; 12: 1616–1629.
- [36] Hu J, Wang Y, Zhang Y, Yu Y, Chen H, Liu K, *et al.* Comprehensive genomic profiling of small cell lung cancer in Chinese patients and the implications for therapeutic potential. *Cancer Medicine*. 2019; 8: 4338–4347.
- [37] Jiang AG, Lu HY. k-RAS mutations in non-small cell lung cancer patients treated with TKIs among smokers and non-smokers: a meta-analysis. *Contemporary Oncology (Poznan, Poland)*. 2016; 20: 124–129.
- [38] Ahrendt SA, Decker PA, Alawi EA, Zhu Yr YR, Sanchez-Cespedes M, Yang SC, *et al.* Cigarette smoking is strongly associated with mutation of the K-ras gene in patients with primary adenocarcinoma of the lung. *Cancer*. 2001; 92: 1525–1530.
- [39] Halvorsen AR, Silwal-Pandit L, Meza-Zepeda LA, Vodak D, Vu P, Sagerup C, *et al.* TP53 Mutation Spectrum in Smokers and Never Smoking Lung Cancer Patients. *Frontiers in Genetics*. 2016; 7: 85.
- [40] Hermeking H. MicroRNAs in the p53 network: micromanagement of tumour suppression. *Nature Reviews. Cancer*. 2012; 12: 613–626.
- [41] Li YL, Liu XM, Zhang CY, Zhou JB, Shao Y, Liang C, *et al.* MicroRNA-34a/EGFR axis plays pivotal roles in lung tumorigenesis. *Oncogenesis*. 2017; 6: e372.
- [42] Shi L, Middleton J, Jeon YJ, Magee P, Veneziano D, Laganà A, *et al.* KRAS induces lung tumorigenesis through microRNAs modulation. *Cell Death & Disease*. 2018; 9: 219.
- [43] Dhawan A, Scott JG, Harris AL, Buffa FM. Pan-cancer characterisation of microRNA across cancer hallmarks reveals microRNA-mediated downregulation of tumour suppressors. *Nature Communications*. 2018; 9: 5228.
- [44] Hou G, Xu W, Jin Y, Wu J, Pan Y, Zhou F. MiRNA-217 accelerates the proliferation and migration of bladder cancer via inhibiting KMT2D. *Biochemical and Biophysical Research Communications*. 2019; 519: 747–753.
- [45] Liu J, Zhang C, Hu W, Feng Z. Tumor suppressor p53 and its mutants in cancer metabolism. *Cancer Letters*. 2015; 356: 197–203.
- [46] Vousden KH, Prives C. Blinded by the Light: The Growing Complexity of p53. *Cell*. 2009; 137: 413–431.
- [47] Liu J, Zhang C, Zhao Y, Feng Z. MicroRNA Control of p53. *Journal of Cellular Biochemistry*. 2017; 118: 7–14.
- [48] Brambilla E, Gazdar A. Pathogenesis of lung cancer signalling pathways: roadmap for therapies. *The European Respiratory Journal*. 2009; 33: 1485–1497.
- [49] Yu HA, Riely GJ, Lovly CM. Therapeutic strategies utilized in the setting of acquired resistance to EGFR tyrosine kinase inhibitors. *Clinical Cancer Research: an Official Journal of the American Association for Cancer Research*. 2014; 20: 5898–5907.