Discover Oncology

Research

City scale related to epidermal growth factor receptor mutations status in Chinese non-small cell lung cancer

Guang-Gui Ding[1] · Jie-Lun Wen[2] · Rui Gong[3] · Jiao-Jiao Yang[4] · Wen-Jian Cen[5] · Ling Deng[5] · Fang Wang[5] · Li-Yue Sun[6,7]

Received: 11 January 2025 / Accepted: 2 May 2025
Published online: 09 May 2025
© The Author(s) 2025 OPEN

Abstract

Purpose This study aimed to explore the relationship between city scale and epidermal growth factor receptor (*EGFR*) mutation status in Chinese patients with non-small cell lung cancer (NSCLC).

Methods A retrospective study enrolled NSCLC patients who underwent tissue *EGFR* mutation testing at Sun Yat-sen University Cancer Center from 2012 to 2017. City scale was categorized according to classifications made by the State Council of China in 2014. Multivariable logistic regression was utilized to determine independent predictors of *EGFR* mutation status in NSCLC patients.

Results A total of 4637 NSCLC patients were enrolled in this study. *EGFR* mutation related to gender, smoking statue, histological type, and city scale. Higher rate of *EGFR* mutations among patients of super large-sized and very large-sized cities compared to other cities. There was an inverse relationship between city scale and smoking status among patients. Multivariate analysis showed that city scale was not an independent predictor of *EGFR* mutation.

Conclusions Although there is a correlation between the size of a city and the rates of *EGFR* mutations in Chinese NSCLC patients, city scale does not independently predict these mutation rates. Instead, the variations in *EGFR* mutation rates could be indirectly related to different levels of urbanization, which may influence smoking behaviors among the populations.

Keywords City scale · *EGFR* mutation · NSCLC · Smoking · Chinese

Guang-Gui Ding, Jie-Lun Wen and Rui Gong have contributed equally.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12672-025-02558-4.

✉ Fang Wang, wangfang@sysucc.org.cn; ✉ Li-Yue Sun, sunly7@mail2.sysu.edu.cn | [1]Department of Thoracic Surgery, Shenzhen People's Hospital, The Second Clinical Medical College of Jinan University, The First Affiliated Hospital of South University of Science and Technology, Shenzhen, Guangdong, China. [2]School of Medicine, Jinan University, Guangzhou, Guangdong, China. [3]Health Management Center, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China. [4]Department of Oncology, The Seventh Affiliated Hospital of Sun Yat-Sen University, Shenzhen, Guangdong, China. [5]Department of Molecular Diagnostics, Sun Yat-Sen University Cancer Center, Guangzhou, Guangdong, China. [6]Department of Health Management Centre, Zhongshan Hospital, Fudan University, Shanghai, China. [7]Department of General Practice, Zhongshan Hospital, Fudan University, Shanghai, China.

1 Introduction

Over the past three decades, the prevalence of non-small cell lung cancer (NSCLC) has surged dramatically in China, making it the foremost cause of cancer-related deaths in the twenty-first century [1]. NSCLC typically manifests subtly and is often only detected at an advanced stage, significantly restricting surgical treatment options [2]. The advent of precision medicine has revolutionized the treatment landscape for NSCLC, enabling targeted therapeutic strategies based on genetic markers, thus broadening the treatment possibilities for patients in advanced stages of the disease [3]. The progressive identification of tumor driver genes in NSCLC patients has included epidermal growth factor receptor (*EGFR*), anaplastic lymphoma kinase (*ALK*), and receptor tyrosine kinase ROS proto-oncogene 1 (*ROS1*) mutants [4]. In clinical practice, tyrosine kinase inhibitors that target these driver genes have shown remarkable efficacy and safety. Among all identified therapeutic targets, mutations in the *EGFR* gene are the most prevalent in China [5]. Consequently, gaining a thorough understanding of the clinical characteristics of NSCLC patients with *EGFR* mutations is essential.

EGFR is a transmembrane protein localized on cell membranes, playing a pivotal role in signaling pathways that facilitate tumor growth and progression [6]. *EGFR* mutations, primarily found in exons 18–21, are common and account for approximately 50% of gene mutation-positive NSCLC cases in China [7]. Among these, deletions in exon 19 and the L858R substitution mutation in exon 21 account for about 90% and are the primary therapeutic targets for EGFR-TKIs such as osimertinib [8], gefitinib [9], and erlotinib [10]. Previous studies have shown that the incidence of *EGFR* mutations is higher among East Asian race, females, non-smokers, and histology of adenocarcinoma [11]. However, still 30–40% of East Asian non-smoking female NSCLC patients have wild-type *EGFR* genes [12], and a significant portion of male NSCLC patients who smoke exhibit *EGFR* mutations [13]. Therefore, it is necessary to identify more clinical characteristics of *EGFR* mutations.

EGFR mutations are influenced not only by unchangeable factors such as race and gender, but also by environmental factors like smoking. Previous studies on NSCLC patients from different regions in China have shown that the *EGFR* mutation rate is lower in less developed areas of the southwest, such as Yunnan Province [14] and Guangxi Province [15], while it is relatively higher among NSCLC patients in the north region of China [16], where the economic development level is generally higher than that in the southwest. Therefore, we speculate that the *EGFR* mutation rate may vary among NSCLC patients from cities with different levels of development.

To substantiate this hypothesis, we carried out a retrospective study aimed at investigating the relationship between the *EGFR* mutation rate in the Chinese NSCLC patients and the developmental level of the cities from which the patients originate, along with other potential influencing factors. Our aim is to discover new predictive markers for *EGFR* mutations.

2 Materials and methods

2.1 Patient selection

This retrospective study included NSCLC patients who underwent *EGFR* testing at the Sun Yat-sen University Cancer Center, from 2012 to 2017. Inclusion criteria were testing via the ARMS-PCR (Amplification Refractory Mutation System Polymerase Chain Reaction) method and a confirmed histological diagnosis of NSCLC. Exclusion criteria included the lack of detailed residential address information. The included clinical features include age, gender, histological type, smoking status, and address. To quantify the development levels of different cities, we adopted the city scale classification standards set by the State Council of the People's Republic of China in 2014 [17]. Cities with a permanent urban population of over 10 million are classified as super large cities (SLC), those with populations between 5 and 10 million as very large cities (VLC), and those with 1 million to 5 million as large cities (including Type I large cities (LI) with 3 to 5 million and Type II large cities (LII) with 1 to 3 million residents). Cities with populations between 500,000 and 1 million are considered medium-sized cities (MC), while those with fewer than 500,000 residents are classified as small cities (including Type I small cities (SI) with 200,000 to 500,000 and Type II small cities (SII) with fewer than 200,000 residents). Population statistics for the corresponding years were sourced from China's Sixth National Population Census of 2014 [18]. According to the Chinese Smoking Cessation Guidelines, patients' smoking index

(pack-years) was categorized as never smokers, light smokers (0 < index ≤ 10), moderate smokers (10 < index < 20), and heavy smokers (index ≥ 20). The clinical data collected encompassed patient gender, age, smoking history, and pathological type from medical records. The classification of city scale was determined based on both the ancestral home and current address of patients. This study received approval from the Ethics Committee of Sun Yat-sen University Cancer Center, with the approval number: B2020-344-01. Given the retrospective nature of this study, the requirement for informed consent was waived.

2.2 DNA extraction and *EGFR* mutation analysis

DNA extraction from NSCLC patient formalin-fixed paraffin-embedded (FFPE) samples was executed using the QIAamp DNA FFPE Tissue kit (QIAGEN), following the manufacturer's protocols. Before DNA extraction, the tumor areas were delineated and tumor tissue was enriched by two independent pathologists using hematoxylin and eosin (HE) staining. The purity and concentration of extracted DNA were assessed using the NanoDrop 2000 spectrophotometer (Thermo Fisher), ensuring an A260/A280 ratio within the specified range of 1.8–2.0. *EGFR* gene mutations within exons 18–21 were detected through the utilization of commercially available *EGFR* Mutation Detection Kit (AmoyDx) with commercial equipment (ABI7500 instrument). The determination of *EGFR* gene mutation status relies on the manufacturer's protocols.

2.3 Statistical analysis

Comparison of continuous data using t-test and one-way analysis of variance (ANOVA). The Chi-square test was employed to scrutinize the associations among diverse characteristics of patients with NSCLC, particularly evaluating the relationships between city scale, *EGFR* mutation status, and smoking status. Variables exhibiting a *p*-value less than 0.05 in univariate analysis were subsequently included in a multivariate logistic regression analysis to identify factors favorably associated with *EGFR* mutation status. A post-hoc power analysis was conducted using G*Power 3.1.9.7, based on the observed EGFR mutation rate (41%), $\alpha = 0.05$, and effect size (5% difference between city tiers). With the cohort size (N = 4367), the study achieved > 90% power to detect significant associations. Statistical significance was established at a *p*-value of < 0.05. All statistical analyses were performed using R version 4.2.3.

3 Results

3.1 Clinical characteristics of the NSCLC patient

A total of 4367 NSCLC patients were included in this study. The clinical characteristics of all NSCLC patients are shown in Table 1. Among these NSCLC patients, males, elderly patients, histological with lung adenocarcinoma, and non-smokers were predominant. Among NSCLC patients who smoke, the highest number of patients have a smoking index of 25.01–50 pack-years, accounting for approximately 20.49% (Fig. 1). In terms of the distribution of current address and ancestral home, most patients were from LII, followed by SLC and LI.

3.2 Relationship between clinical characteristics and *EGFR* mutation in patients with NSCLC

Table 2 shows the differences in clinical characteristics between NSCLC patients with *EGFR* mutation and *EGFR* wild-type. The results indicate that regardless of current address or ancestral home, *EGFR* mutation patients have significant differences compared to *EGFR* wild-type patients ($p < 0.001$). To further clarify the correlation between city development level and *EGFR* mutations, we combined city grades into a binary variable (Supplementary Table 1 and Supplementary Fig. 1) (Table 3 and Supplementary Fig. 2). The results indicate that the mutation rate of NSCLC patients in VLC is significantly higher than in other city patients, with statistical differences observed in both current address and ancestral home ($p < 0.001$). To address potential confounding from population size, we analyzed *EGFR* mutation rates (mutations/total tested) stratified by city scale. Patients from SLC and VLC exhibited a significantly higher *EGFR* mutation rate compared to those from others. After adjusting for population size and demographic factors via multivariable logistic regression, the odds of *EGFR* mutation remained higher in SLC and VLC compared to smaller cities, regardless of whether the analysis was based on current address or ancestral home (Supplementary Table 2).In addition, we also analyzed the correlation between different *EGFR* mutation types, including co-mutations and single mutations [19], hotspot mutations and rare

Table 1 Clinical characteristics of NSCLC patients

Characteristic	n (%)
Total	4367(100)
Gender	
Female	1637(37.47)
Male	2730(62.53)
Age (year)	
< 50	945(21.64)
> = 50	3422(78.36)
Histological type	
Adenocarcinoma	3546(81.20)
Squamous cell carcinoma	562(12.87)
Adenosquamous carcinoma	71(1.63)
Others	188(4.30)
Smoking history	
No	2412(55.24)
Yes	1954(44.76)
Quit smoking	
No	1318(67.45)
Yes	636(32.55)
Smoking habit	
Never smoker	2412(58.04)
Light smoker	180(4.33)
Moderate smoker	83(2.00)
Heavy smoker	1481(35.64)
City scale (address)	
Super large-sized city	985(22.56)
Very large-sized city	370(8.47)
Large-sized city type I	533(12.21)
Large-sized city type II	1671(38.27)
Medium-sized city	291(6.67)
Small-sized city type I	426(9.76)
Small-sized city type II	90(2.06)
City scale (ancestral home)	
Super large-sized city	871(20.11)
Very large-sized city	363(8.38)
Large-sized city type I	542(12.51)
Large-sized city type II	1738(40.12)
Medium-sized city	294(6.79)
Small-sized city type I	433(9.99)
Small-sized city type II	91(2.10)

mutations [20], with clinical characteristics in this study. The results showed no significant statistical differences between co-mutations and single mutations in terms of city scale, and the same was true for the comparison between hotspot mutations and rare mutations (Supplementary Table 2 ~ 5).

3.3 Relationship between clinical characteristics and city scale in patients with NSCLC

To further explore the reasons behind the differences in *EGFR* mutation across various city scale, we analyzed the clinical characteristics of patients from VLC and SLC compared to those from other cities (Table 4). The results revealed significant differences in age, histological type, smoking history, and smoking index among patients from various city scales. Patients residing in VLC and SLC were older on average compared to those from other cities ($p < 0.001$) and had a higher proportion

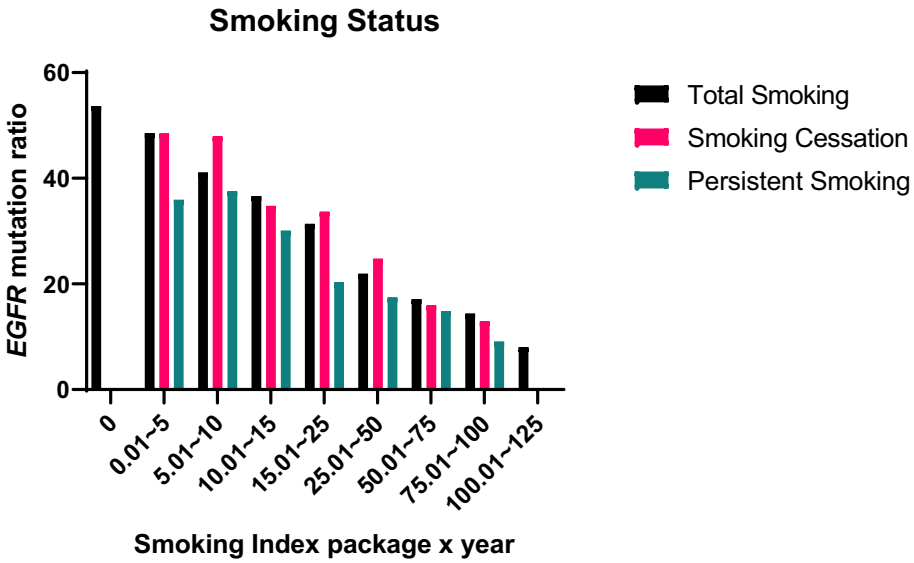

Fig. 1 EGFR mutation related to smoking index

of adenocarcinoma ($p < 0.001$). Notably, patients residing in VLC and SLC had significantly lower smoking rates (40.30% vs 46.74%, $p < 0.001$) and a lower smoking index among those who did smoke ($p = 0.001$). There was no significant correlation between gender or smoking cessation and city scale. These findings underscore the significance of integrating city-level factors into epidemiological studies. Consistent results were obtained whether patients were categorized based on their address or ancestral home.

3.4 Influence of city scale on *EGFR* mutations

To further investigate whether city scale can serve as a predictor for *EGFR* mutations, we assessed the impact of city scale on *EGFR* mutations using both univariate and multivariate analyses (Table 5 and Supplementary Table 6). Gender, histological type, smoking status, and city scale were included in the multivariate analysis. The results indicated that the *EGFR* mutation rate was lower in males (OR = 0.612, 95% CI 0.516–0.726, p = 0.000). Histological type was a significant factor, with a higher mutation rate in adenocarcinoma patients (OR = 1.835, 95% CI 1.113–3.026, p = 0.017) and a lower mutation rate in squamous cell carcinoma patients (OR = 0.168, 95% CI 0.117–0.242, p = 0.000). Additionally, smoking patients had a lower mutation rate (OR = 0.475, 95% CI 0.400–0.564, p = 0.000) (Supplementary Fig. 3). In the univariate analysis, both VLC and SLC showed statistical significance (p = 0.000). However, the multivariate analysis did not yield statistically significant results. We also performed a subgroup analysis on the smoking population (Supplementary Table 7 ~ 8). The results indicated that age, histological type, and smoking cessation status are independent predictors of *EGFR* mutation. However, city scale did not demonstrate significant statistical relevance in the univariate analysis.

4 Discussion

In the era of precision therapy, target therapy has revolutionized the management of NSCLC with driver gene mutation, rendering it a condition that can be effectively controlled [21]. Therefore, conducting genetic testing and molecular diagnostics is essential in the clinical management of NSCLC patients, enabling personalized treatment strategies tailored to individual needs [22]. Among the numerous driver gene mutations, *EGFR* mutation, as the most common therapeutic target, has demonstrated excellent clinical efficacy, allowing a significant portion of advanced NSCLC patients to experience a chronic progression of their treatment [23]. However, *EGFR* mutations are distributed variably among numerous NSCLC patients, predominantly seen in Asian populations, females, non-smokers, and adenocarcinoma patients [11]. Yet, these clinical factors are insufficient to effectively distinguish between patients with *EGFR* mutations and those with wild-type *EGFR*. Previous research has found significant differences in the rates of *EGFR* mutations across regions with different levels of development [14–16]. Therefore, this study, through a large-scale retrospective analysis, explores the

Table 2 Relationship between clinical characteristics and *EGFR* mutation status in patients with NSCLC

Characteristic	*EGFR*		
	Mutation	Wild-Type	*p*
Age (year)			
	57.74 ± 9.86	57.69 ± 10.93	0.877
Gender, n (%)			
Female	939(52.37)	697(27.09)	**0.000**
Male	854(47.63)	1876(72.91)	
Histological type, n (%)			
Adenocarcinoma	1722(95.99)	1824(70.89)	**0.000**
Squamous cell carcinoma	35(1.95)	527(20.48)	
Adenosquamous carcinoma	30(1.67)	41(1.59)	
Others	7(0.39)	181(7.03)	
City scale (address), n (%)			
Super large-sized city	429(23.93)	556(21.61)	**0.000**
Very large-sized city	184(10.26)	186(7.23)	
Large-sized city type I	202(11.27)	331(12.86)	
Large-sized city type II	669(37.31)	1002(38.94)	
Medium-sized city	96(5.35)	195(7.58)	
Small-sized city type I	172(9.59)	254(9.87)	
Small-sized city type II	41(2.29)	49(1.90)	
City scale (ancestral home), n (%)			
Super large-sized city	375(21.10)	496(19.41)	**0.000**
Very large-sized city	184(10.35)	179(7.01)	
Large-sized city type I	204(11.48)	338(13.23)	
Large-sized city type II	700(39.39)	1038(40.63)	
Medium-sized city	97(5.46)	197(7.71)	
Small-sized city type I	176(9.90)	257(10.06)	
Small-sized city type II	41(2.31)	50(1.96)	
Smoking history, n (%)			
No	1291(72.00)	1121(43.57)	**0.000**
Yes	502(28.00)	1452(56.43)	
Quit smoking, n (%)			
No	313(62.35)	1005(69.21)	**0.004**
Yes	189(37.65)	447(30.79)	
Smoking habit, n (%)			
Never smoker	1291(75.06)	1121(46.02)	**0.000**
Light smoker	79(4.59)	101(4.15)	
Moderate smoker	30(1.74)	53(2.18)	
Heavy smoker	320(18.60)	1161(47.66)	

Bold values indicate statistical significance at $p < 0.05$

Table 3 Relationship between other clinical characteristics and *EGFR* mutation in patients with NSCLC

Characteristic	*EGFR*		
	Mutation	Wild-Type	*p*
City scale (address), n (%)			
SLC and VLC	613(34.19)	742(28.84)	**0.000**
Others	1180(65.81)	1831(71.16)	
City scale (ancestral home), n (%)			
SLC and VLC	559(31.46)	675(26.42)	**0.000**
Others	1218(68.54)	1880(73.58)	

Bold values indicate statistical significance at $p < 0.05$

Table 4 Relationship between city scale and smoking status in patients with NSCLC

Characteristic	City scale (address)			City scale (ancestral home)		
	VLC and SLC	Others	*p*	VLC and SLC	Others	*p*
Age (year)						
	59.38 ± 10.79	56.96 ± 10.29	**0.000**	59.30 ± 10.73	57.05 ± 10.34	**0.000**
Gender, n (%)						
Male	824(60.81)	1905(63.29)	0.118	756(61.26)	1949(62.93)	0.306
Female	531(39.19)	1105(36.71)		478(38.74)	1148(37.07)	
Histological type, n (%)						
Adenocarcinoma	1149(84.80)	2396(79.57)	**0.000**	1042(84.44)	2475(79.89)	**0.000**
Squamous cell carcinoma	127(9.37)	435(14.45)		114(9.24)	443(14.30)	
Adenosquamous carcinoma	26(1.92)	45(1.49)		25(2.03)	45(1.45)	
Others	53(3.91)	135(4.48)		53(4.29)	135(4.36)	
Smoking history, n (%)						
No	809(59.70)	1603(53.26)	**0.000**	735(59.56)	1658(53.54)	**0.000**
Yes	546(40.30)	1407(46.74)		499(40.44)	1439(46.46)	
Quit smoking, n (%)						
No	373(68.32)	944(67.09)	0.605	344(68.94)	965(67.06)	0.440
Yes	173(31.68)	463(32.91)		155(31.06)	474(32.94)	
Smoking habit, n (%)						
Never smoker	809(62.04)	1603(56.23)	**0.003**	735(61.66)	1658(56.55)	**0.018**
Light smoker	49(3.76)	131(4.59)		44(3.69)	134(4.57)	
Moderate smoker	28(2.15)	55(1.93)		25(2.10)	56(1.91)	
Heavy smoker	418(32.06)	1062(37.25)		388(32.55)	1084(36.97)	

Bold values indicate statistical significance at $p < 0.05$

Table 5 Univariate and multivariate analyses of the NSCLC patients

Characteristic	*EGFR* Mutation			
	Univariate		Multivariate	
	OR (95% CI)	*p*	OR (95% CI)	*p*
Male	0.336 (0.296–0.382)	**0.000**	0.612 (0.516–0.726)	**0.000**
Age	1.000 (0.995–1.006)	0.895		
Adenocarcinoma	13.172 (9.257–18.743)	**0.000**	1.835 (1.113–3.026)	**0.017**
Squamous cell carcinoma	0.101 (0.079–0.131)	**0.000**	0.168 (0.117–0.242)	**0.000**
Adenosquamous carcinoma	0.925 (0.574–1.490)	0.747		
Smoking	0.299 (0.262–0.340)	**0.000**	0.475 (0.400–0.564)	**0.000**
VLC and SLC (address)	0.778 (0.683–0.886)	**0.000**	1.029 (0.693–1.529)	0.887
VLC and SLC (ancestral home)	0.780 (0.683–0.892)	**0.000**	0.830 (0.554–1.244)	0.367

Bold values indicate statistical significance at $p < 0.05$

correlation between *EGFR* mutations and different urban tiers, aiming to identify new predictive markers and management strategies for *EGFR* mutations.

Our results show that *EGFR* mutations are primarily found in female, non-smoking populations with lung adenocarcinoma or adenosquamous carcinoma. Additionally, the rate of *EGFR* mutations decreases with an increase in the smoking index. The mutation rate in patients who have quit smoking is also higher than in those who have not quit. Regarding the correlation between *EGFR* mutations and city scale, our results have demonstrated a correlation between the scale of Chinese cities and *EGFR* mutations; the more developed the city scale, the higher the rate of *EGFR* mutations. Conversely, the lower the city scale, the lower the rate of *EGFR* mutations. This indicates to some extent that surrounding environmental factors, especially the city scale, influence the rate of *EGFR* mutations. However, there is currently a lack of research on the correlation between *EGFR* mutation rates and different city scale among NSCLC patients. In addition

to uncontrollable factors such as race and gender, environmental factors like smoking also influence the *EGFR* mutations in NSCLC patients. In China, variations caused by different development rates in cities extend beyond mere differences in population size, they encompass a spectrum of disparities in environmental factors, healthcare provisions, medical infrastructure, and population demographics [24]. These multifaceted distinctions have garnered substantial attention, particularly in the context of NSCLC, as they may be correlated with an elevated incidence of *EGFR* mutations. The level of urbanization often correlates with heightened pollution levels [25], air quality deterioration, typified by the surge in PM2.5 pollution, accompanies urban development. Research has revealed a compelling link between PM2.5 pollution and the incidence of local lung adenocarcinoma, further extending its influence to the prevalence of *EGFR* mutations within affected populations. This interplay among urbanization, pollution, and *EGFR* mutations has been explored in pertinent studies [26]. In addition, the level of urban development significantly influences the healthcare provisions and distribution of medical resources. Disparities exist in the diagnostic capabilities of cities of varying sizes, with smaller cities often facing limitations in medical resources. This resource deficit can potentially impede the timely diagnosis and management of *EGFR* mutations. However, it is important to note that current research has not conclusively established a direct correlation between city scale and *EGFR* mutations. While our study highlights the association between city scale and *EGFR* mutation rates, it does not account for potential confounders such as city-specific pollution levels or smoking prevalence. Future prospective studies should incorporate these variables, including air quality indices (e.g., PM2.5 levels) and regional smoking patterns, to better understand their role in shaping *EGFR* mutation epidemiology. Leveraging publicly available datasets, such as China's National Environmental Monitoring Center for pollution indices and national health surveys for smoking prevalence, could provide a more comprehensive understanding of the environmental and behavioral factors influencing *EGFR* mutations in NSCLC patients. In summary, the diverse urban landscapes in China encompassing pollution levels, healthcare infrastructure, and population demographics may collectively contribute to the varying incidence of *EGFR* mutations in NSCLC patients. This multifaceted relationship underscores the need for comprehensive research to unravel the intricate interplay between urbanization and genetic alterations in lung cancer.

Although we found that the *EGFR* mutation rates in NSCLC patients from SLC and VLC are significantly higher than those in other cities, both univariate and multivariate analyses show that city tier is not an independent predictive factor for *EGFR* mutations. Further analysis revealed significant differences in smoking status among NSCLC patients from different city tiers, with lower smoking rates and smoking indices in SLC and VLC compared to other cities. This suggests that the differences in *EGFR* mutation rates between city tiers may be due to differences in smoking status. The association between smoking and *EGFR* mutations has been a subject of significant interest, particularly within the realm of non-small cell lung cancer research. Multiple investigations have consistently revealed an inverse relationship between smoking and *EGFR* mutations [27, 28]. While this inverse correlation is well-established, the precise underlying biological mechanism remains a topic of ongoing investigation. Taking a broader perspective, China has been grappling with a substantial tobacco consumption issue since the 1980s. Despite substantial efforts, the tobacco situation in the country remains a challenge [29]. The extensive rural-to-urban population migration has further complicated this landscape, with the migrant population exhibiting higher rates of tobacco and alcohol usage [30]. Consequently, smoking prevalence varies among individuals in different cities. Meanwhile, previous studies have also shown that differing levels of smoking across different regions in China lead to varying incidences of lung cancer [31]. Globally, significant disparities in *EGFR* mutation rates exist across various regions. Asian patients typically exhibit mutation rates ranging from 30 to 50%, whereas their American counterparts tend to have lower mutation rates [32].

Our study uncovered an association between city size and *EGFR* mutation rates. However, it is crucial to acknowledge the inherent limitations in our research. Employing a retrospective design and relying on real-world medical records offers advantages in terms of data availability and a substantial sample size, yet introduces the complexity of mixed information, with unrecorded variables potentially influencing *EGFR* mutation rates, unfortunately not accounted for in our analysis. Furthermore, the concentration of our study population in South China limits the extrapolation of our findings to a broader context, considering the significant geographical variations in China's cities. Smaller cities may have lower rates of EGFR testing due to limited access to molecular diagnostics, leading to underdiagnosis of mutations in these regions. Although our data were sourced from a major cancer center (mitigating referral bias), disparities in testing accessibility across city tiers could skew results. Future research endeavors should span multiple geographic regions to offer a more comprehensive understanding of the relationship between city scale and *EGFR* mutations. While our study provides preliminary evidence supporting the link between city scale and *EGFR* mutations, additional prospective investigations are imperative to validate this association and delve deeper into potential causal relationships. Such endeavors will establish a more robust foundation for the personalized treatment of NSCLC patients. In summary, despite inherent limitations, our study offers valuable preliminary insights into the correlation between city scale and *EGFR* mutations,

and future studies should strive to overcome these constraints, conducting in-depth investigations into the underlying mechanisms and influencing factors to advance the precision treatment of NSCLC patients.

5 Conclusion

In conclusion, city scale demonstrates a significant association with *EGFR* mutation, with smoking status among patients playing a pivotal role. This discovery enhances our understanding of the epidemiology of *EGFR* mutation and lays a foundation for personalized treatment strategies. Additional research is warranted to comprehensively elucidate its clinical implications and applications, thereby facilitating the delivery of more effective healthcare services and the allocation of medical resources.

Acknowledgements We are very grateful to Li Minrui for his contribution to the definition of city tiers in this study.

Author contributions Li-Yue Sun, Fang Wang and Guang-Gui Ding took part in article conception and design. Jie-Lun Wen acquired the data. Rui Gong, Jiao-Jiao Yang, Wen-Jian Cen, and Ling Deng analyzed and interpreted the data. Guang-Gui Ding, Li-Yue Sun and Jie-Lun Wen wrote the manuscript. Li-Yue Sun and Fang Wang supervised the whole research. All authors contributed to the article and approved the submitted version.

Funding This study was supported by National Natural Science Foundation of China (No. 82302640), Guangzhou Science and Technology Plan Project (Grant No. 2023A04J1129), Guangdong Medical Scientific Research (Grant No. B2023038), and Shanghai Rehabilitation Medicine Association Health Management Special Fund Project (Grant No. 2024JGYQ11).

Data availability The data that support the findings of this study are not openly available due to reasons of sensitivity and are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate This study received approval from the Ethics Committee of Sun Yat-sen University Cancer Center, with the approval number: B2020-344–01. Given the retrospective nature of this study, the requirement for informed consent was waived. The study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. Cao W, et al. Changing profiles of cancer burden worldwide and in China: a secondary analysis of the global cancer statistics 2020. Chin Med J (Engl). 2021;134(7):783–91.
2. Jie C, et al. Prospects and feasibility of synergistic therapy with radiotherapy, immunotherapy, and DNA methyltransferase inhibitors in non-small cell lung cancer. Front Immunol. 2023;14:1122352.
3. Wu J, Lin Z. Non-small cell lung cancer targeted therapy: drugs and mechanisms of drug resistance. Int J Mol Sci. 2022.1;23(23):15056.
4. Imyanitov EN, Iyevleva AG, Levchenko EV. Molecular testing and targeted therapy for non-small cell lung cancer: current status and perspectives. Crit Rev Oncol Hematol. 2021;157: 103194.
5. Lu HL, Jie GL, Wu YL. Epidermal growth factor receptor-targeted therapy for the treatment of non-small cell lung cancer: a review of phase II and III trials. Expert Opin Emerg Drugs. 2022;27(2):111–26.
6. Tomas A, Futter CE, Eden ER. EGF receptor trafficking: consequences for signaling and cancer. Trends Cell Biol. 2014;24(1):26–34.
7. Gristina V, et al. The significance of epidermal growth factor receptor uncommon mutations in non-small cell lung cancer: a systematic review and critical appraisal. Cancer Treat Rev. 2020;85: 101994.
8. Mok TS, et al. Osimertinib or platinum-pemetrexed in EGFR T790M-positive lung cancer. N Engl J Med. 2017;376(7):629–40.

9. Fukuoka M, et al. Biomarker analyses and final overall survival results from a phase III, randomized, open-label, first-line study of gefitinib versus carboplatin/paclitaxel in clinically selected patients with advanced non-small-cell lung cancer in Asia (IPASS). J Clin Oncol. 2011;29(21):2866–74.
10. Wheatley-Price P, et al. Erlotinib for advanced non-small-cell lung cancer in the elderly: an analysis of the National Cancer Institute of Canada Clinical Trials Group Study BR.21. J Clin Oncol. 2008;26(14):2350–7.
11. Castellanos E, Feld E, Horn L. Driven by mutations: the predictive value of mutation subtype in EGFR-mutated non-small cell lung cancer. J Thorac Oncol. 2017;12(4):612–23.
12. Saito S, et al. Current status of research and treatment for non-small cell lung cancer in never-smoking females. Cancer Biol Ther. 2017;18(6):359–68.
13. Kauffmann-Guerrero D, et al. NSCLC patients harbouring rare or complex EGFR mutations are more often smokers and might not benefit from first-line tyrosine kinase inhibitor therapy. Respiration. 2018;95(3):169–76.
14. Zhou Y, et al. Epidermal growth factor receptor T790M mutations in non-small cell lung cancer (NSCLC) of Yunnan in southwestern China. Sci Rep. 2018;8(1):15426.
15. Wei WE, et al. An analysis of EGFR mutations among 1506 cases of non-small cell lung cancer patients in Guangxi, China. PLoS ONE. 2016;11(12): e0168795.
16. Cheng Y, et al. Real-world EGFR testing in patients with stage IIIB/IV non-small-cell lung cancer in North China: a multicenter, non-interventional study. Thorac Cancer. 2018;9(11):1461–9.
17. Government, C., *China to apply new city classification standards.* 25-11-2014
18. Statistics, N.B.o., *China's Sixth National Population Census of 2014.* 29-4-2011
19. Barnet MB, et al. EGFR-co-mutated advanced NSCLC and response to EGFR tyrosine kinase inhibitors. J Thorac Oncol. 2017;12(3):585–90.
20. Ikemura S, et al. Molecular dynamics simulation-guided drug sensitivity prediction for lung cancer with rare EGFR mutations. Proc Natl Acad Sci U S A. 2019;116(20):10025–30.
21. Stine ZE, et al. Targeting cancer metabolism in the era of precision oncology. Nat Rev Drug Discov. 2022;21(2):141–62.
22. Vargas AJ, Harris CC. Biomarker development in the precision medicine era: lung cancer as a case study. Nat Rev Cancer. 2016;16(8):525–37.
23. Dalurzo ML, et al. Testing for EGFR mutations and ALK rearrangements in advanced non-small-cell lung cancer: considerations for countries in emerging markets. Onco Targets Ther. 2021;14:4671–92.
24. Zhao S, et al. Contemporary evolution and scaling of 32 major cities in China. Ecol Appl. 2018;28(6):1655–68.
25. Liang L, Gong P. Urban and air pollution: a multi-city study of long-term effects of urban landscape patterns on air quality trends. Sci Rep. 2020;10(1):18618.
26. Han SC, Wang GZ, Zhou GB. Air pollution, EGFR mutation, and cancer initiation. Cell Rep Med. 2023;4(5): 101046.
27. Barta JA, Powell CA, Wisnivesky JP. Global Epidemiology of Lung Cancer. Ann Glob Health. 2019. 22;85(1):8.
28. Sun LY, et al. Smoking status combined with tumor mutational burden as a prognosis predictor for combination immune checkpoint inhibitor therapy in non-small cell lung cancer. Cancer Med. 2021;10(19):6610–7.
29. Parascandola M, Xiao L. Tobacco and the lung cancer epidemic in China. Transl Lung Cancer Res. 2019;8(Suppl 1):S21-s30.
30. Zhang M, et al. Trends in smoking prevalence in urban and rural China, 2007 to 2018: Findings from 5 consecutive nationally representative cross-sectional surveys. PLoS Med. 2022;19(8): e1004064.
31. Sun D, et al. Cancer burden in China: trends, risk factors and prevention. Cancer Biol Med. 2020;17(4):879–95.
32. Shi Y, et al. A prospective, molecular epidemiology study of EGFR mutations in Asian patients with advanced non-small-cell lung cancer of adenocarcinoma histology (PIONEER). J Thorac Oncol. 2014;9(2):154–62.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.