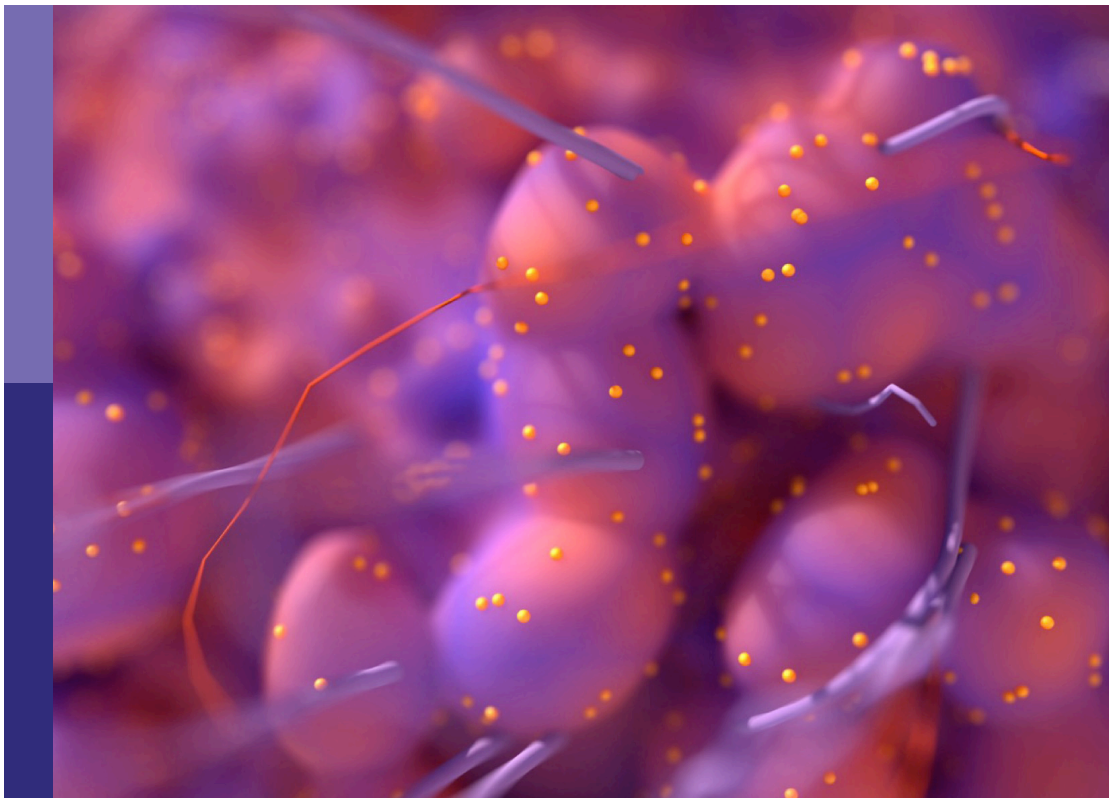
frontiers | Research Topics

Epidemiology, screening and diagnosis of lung cancer

Edited by
Yutong He, Xue Qin Yu, Chen Wanqing, Jackilen Shannon and Yingsong Lin

Published in
Frontiers in Oncology



ISSN 1664-8714
ISBN 978-2-8325-2287-5
DOI 10.3389/978-2-8325-2287-5

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Epidemiology, screening and diagnosis of lung cancer

Topic editors

Yutong He — Fourth Hospital of Hebei Medical University, China
Xue Qin Yu — The University of Sydney, Australia
Chen Wanqing — National Cancer Center of China, China
Jackilen Shannon — Oregon Health and Science University, United States
Yingsong Lin — Aichi Medical University, Japan

Citation

He, Y., Yu, X. Q., Wanqing, C., Shannon, J., Lin, Y., eds. (2023). *Epidemiology, screening and diagnosis of lung cancer*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2287-5

Table of contents

frontiers in Oncology

ORIGINAL RESEARCH
published: 17 November 2021
doi: 10.3389/fonc.2021.774937

Triglyceride-Glucose Index Is Not Associated With Lung Cancer Risk: A Prospective Cohort Study in the UK Biobank

Lijie Wang[1,2], *Shucheng Si*[1,2], *Jiqing Li*[1,2], *Yunxia Li*[1,2], *Xiaolu Chen*[1,2], *Fuzhong Xue*[1,2]* *and Wangang Ren*[3]*

[1] Department of Epidemiology and Health Statistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, China, [2] Institute for Medical Dataology, Shandong University, Jinan, China, [3] Department of Thoracic Surgery, Shandong Provincial Hospital, Cheeloo College of Medicine, Shandong University, Jinan, China

OPEN ACCESS

Edited by:
Jackilen Shannon, Oregon Health and Science University, United States

Reviewed by:
Xiaopan Li, Shanghai Pudong Disease Prevention and Control Center, China
Junmei Miao Jonasson, University of Gothenburg, Sweden

***Correspondence:**
Fuzhong Xue
xuefzh@sdu.edu.cn
Wangang Ren
renrwg@163.com

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *13 September 2021*
Accepted: *29 October 2021*
Published: *17 November 2021*

Citation:
Wang L, Si S, Li J, Li Y, Chen X, Xue F and Ren W (2021) Triglyceride-Glucose Index Is Not Associated With Lung Cancer Risk: A Prospective Cohort Study in the UK Biobank. Front. Oncol. 11:774937. doi: 10.3389/fonc.2021.774937

Background: The triglyceride-glucose (TyG) index is a practical substitute measure for insulin resistance (IR). The relationship between IR and lung cancer has been examined in previous studies; however, the findings have been controversial. In addition, previous studies had small sample sizes. Thus, we systematically examined the association between IR and lung cancer risk based on the UK Biobank with IR measured by the TyG index and further examined the interactions and joint effects for lung cancer.

Methods: A total of 324,334 individuals free from any type of cancer at recruitment from the UK Biobank prospective cohort were included. The participants were predominantly between 40 and 70 years old. After adjusting for relevant confounders, multivariable Cox regression models were constructed to examine the relationship between the TyG index and the risk of lung cancer. We also checked the interactions and joint effects using a polygenic risk score (PRS) for lung cancer.

Results: During a median follow-up of 9 years, 1,593 individuals were diagnosed with lung cancer. No association was found between the TyG index and lung cancer risk after multivariate Cox regression analysis adjusted for risk factors (hazard ratio: 0.91; 95% confidence interval: 0.64–1.18). No interaction or joint effects for genetic risk and the TyG index were observed.

Conclusion: The TyG index was not associated with the risk of lung cancer. Our results provide limited evidence that IR is not correlated with the risk of lung cancer.

Keywords: lung cancer, insulin resistance, UK Biobank, longitudinal study, triglyceride-glucose index

Abbreviations: BMI, body mass index; CI, confidence interval; HbA1c, glycated hemoglobin; HDL, high-density lipoprotein; HR, hazard ratio; ICD-10, International Classification of Diseases, 10th revision; IR, insulin resistance; LDL, low-density lipoprotein; PRS, polygenic risk score; SD, standard deviation; SNP, single nucleotide polymorphism; TC, total cholesterol; TG, triglyceride; TyG, triglyceride-glucose index; WHR, waist-hip ratio.

INTRODUCTION

Lung cancer is one of the most commonly diagnosed cancers causing many deaths each year (representing approximately one in 10 cancers diagnosed and one in five deaths in 2020) (1). Despite improved treatment, the diagnosis of lung cancer is associated with relatively poor survival (2). Identifying the population at high risk of lung cancer remains an arduous task (3).

Insulin resistance (IR) is one of the most common metabolic disorders (4, 5). The triglyceride-glucose index (TyG index), a surrogate indicator of combined triglycerides (TGs) and glucose, is considered to be a practical and effective measurement for IR (6–8). Some studies have shown that the TyG index plays an important role as a potential risk factor for some diseases, such as metabolic syndrome (9), acute pancreatitis (10), cardio-cerebrovascular diseases (11), and cancers of the digestive system (12). Although some previous studies have preliminarily examined the relationship between IR and lung cancer, the evidence is unconvincing and somewhat controversial. Several studies have shown that IR is positively related to lung cancer, whereas some studies found an invalid association (13–17). In addition, most of the literature had a small sample size and insufficient estimates of genetic predisposition. Whether IR can assist in predicting and diagnosing lung cancer remains unclear. Although genetic susceptibility alleles could explain approximately 12% of heritability for lung cancer (18), and more than 50 genetic susceptibility loci have been identified in different ethnic groups (19), no work has been done to investigate the joint effects or interactions between the TyG index and the genetic susceptibility for lung cancer.

In this study, we sought to systematically examine the association between IR and lung cancer risk based on the UK Biobank, with IR measured by the TyG index, and further examined the interactions and joint effects using a polygenic risk score (PRS) for lung cancer.

MATERIALS AND METHODS

Study Cohort

We used data from the UK Biobank, which is an ongoing population-based national prospective cohort study. Approximately half a million people between 37 and 73 years of age were recruited for the study across the United Kingdom from 2006 to 2010 (20, 21). At baseline assessment, participants completed a standardized questionnaire that included detailed information on socioeconomic and demographic characteristics, general health and medical information, lifestyle, and diet. Physical assessments, laboratory investigations, and genome-wide genotyping of all participants were also performed at baseline. Written informed consent was obtained from all participants, and the study protocol was approved by the North West Multi-Center Research Ethics Committee, the Patient Information Advisory Group in England and Wales, and the Community Health Index Advisory Group in Scotland.

Ascertainment of Exposures

Demographic data on age, sex, region, socioeconomic status (Townsend deprivation score), smoking status and alcohol intake frequency were obtained by administering a standardized questionnaire. Region refers to the UK Biobank Assessment Center at which the participant consented (22). Material deprivation was measured using the Townsend Index. Each participant was assigned a score corresponding to the output area in which their postcode was located. Smoking status was categorized as never, former, or current smoking. Alcohol intake frequency was categorized as never, special occasions only, 1-3 times/month, 1-2 times/week, 3 or 4 times/week, or daily or almost daily. Body mass index (BMI) was calculated by dividing body weight (kg) by height in meters squared (m^2). For analyses, BMI was categorized into normal (<25 kg/m^2), overweight (25–29.9 kg/m^2), and obese (≥30 kg/m^2). The waist-hip ratio (WHR) was calculated by dividing waist circumference by hip circumference.

Blood samples were collected at recruitment of participants into UK Biobank, and taken at random, and the time (hours) since the last meal and the fasting time were recorded during collection (23). Data on TG, glucose, total cholesterol (TC), high-density lipoprotein cholesterol (HDL), low-density lipoprotein cholesterol (LDL), and glycated hemoglobin (HbA1c), including fasting time before sampling, were collected during health examinations. The TyG index was calculated using the formula: ln [$triglyceride$ (mg/dL) × $glucose$ (mg/dL)/2]. (24) The HbA1c value was categorized into normal and high with a clinical cutoff of ≥42 mmol/mol (25, 26). Diabetes (International Classification of Diseases, 10th Revision [ICD-10]: E10-E14), dyslipidemia (ICD-10: E78), and hypertension (ICD-10: I10-I15) were defined from the health records of the UK Biobank. Further details on the definition of the selected variables are provided in **Table S1**.

Selection Criteria

We included participants who met the following criteria: 1) Caucasian; 2) no previous history of any type of cancer before enrollment; 3) without diabetes or dyslipidemia; and 4) no missing data. Further, we excluded participants with a follow-up time of less than 1 year from the study to minimize the possibility of reverse causality (i.e., parameters of interest affected by undiagnosed cancer) (24). **Figure 1** presents a flowchart of the study selection process.

Polygenic Risk Score Construction

Details of the genotyping and imputation procedures for single nucleotide polymorphisms (SNPs) in the UK Biobank can be found elsewhere (27). Eighteen SNPs that were significantly associated with lung cancer risk ($P < 5\times10^{-8}$) in Caucasians in the study by McKay et al. were identified from a genome-wide association study (**Table S2**) (18). The PRS for lung cancer was constructed for each participant by summing the risk allele numbers (i.e., participants have 0, 1, or 2 risk alleles) weighted by their respective effect sizes [β coefficients, as natural log-transformed values (odds ratios)] (28). The effect of missing SNP observations was set to a value of zero.

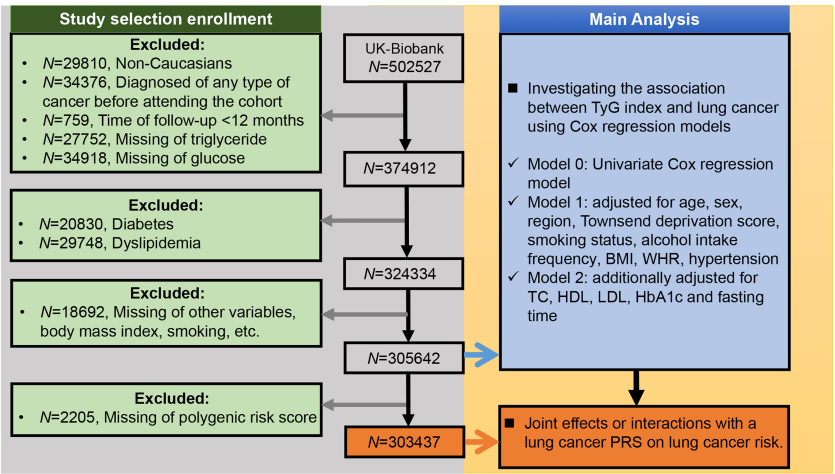


FIGURE 1 | Selection criteria for building the lung cancer cohort and the main analysis of this study. BMI, body mass index; WHR, waist-hip ratio; TC, total cholesterol; TyG index, triglyceride-glucose index; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; HbA1c, glycated hemoglobin; PRS, polygenic risk score.

Ascertainment of Outcome

The study outcome was incident diagnosis of lung cancer of any of the topographic subcategories, recorded *via* linkages to national cancer and death registries according to the ICD-10 code C34. Participants were followed up from the date of recruitment until the date of lung cancer diagnosis, the date of death, or 15 February 2018, whichever came first.

Statistical Analysis

Means ± standard deviations (SD) or medians (interquartile ranges) were calculated for continuous variables, whereas frequencies and proportions (*N*, %) were reported for categorical variables. Quantitative baseline variables were compared using the t-test. Categorical variables were compared between groups using the chi-square test.

Cox regression models were used to estimate hazard ratios (HRs) and 95% confidence intervals (CIs) for the risk of lung cancer according to the TyG index increment. A univariate Cox regression model (Model 0) was constructed in the first step. Subsequently, a multivariate Cox regression model (Model 1) was adjusted for age at recruitment (continuous), sex, region (10 categories: London, North-West, North-East, East Midlands, West Midlands, South-East, South-West, Wales, Yorkshire and Humber, Scotland), Townsend deprivation score (quartile), smoking status (never, former, current), alcohol intake frequency (6 categories: never, special occasions only, 1-3 times/month, 1-2 times/week, 3 or 4 times/week, and daily or almost daily), BMI (normal, overweight, obese), WHR (quartile), and hypertension (yes, no). Furthermore, Model 2 was additionally adjusted for TC (continuous), HDL (continuous), LDL (continuous), HbA1c (normal, high), and fasting time (continuous). Covariates were selected based on scientific plausibility and previous studies. Restricted cubic spline models were fitted to Cox proportional hazards models (29). Information about the data collection of covariates is provided in **Table S1**. The main analysis of this study is presented in **Figure 1**.

Furthermore, the continuous variable TyG index was divided into categories, and the above three Cox models were applied as described previously. First, the TyG index was divided by deciles, and the three Cox models were used to determine the overall trend of HRs for the TyG index with lung cancer risk. In the subsequent analysis, we categorized participants as having low-level and high-level TyG indices according to the median. Stratum-specific analyses were conducted to assess the potential effects of age group at recruitment (<55, 55-64, or ≥65 years), sex, Townsend deprivation score (tertiles), smoking status (never, former, or current), alcohol intake frequency (never, less than daily, daily or almost daily), BMI (normal, overweight, obese), WHR (tertiles), TC (tertiles), HDL (tertiles), LDL (tertiles), HbA1c (tertiles), fasting time (<8, ≥ 8 hours), and hypertension (yes, no).

Analyses of the genetic susceptibility of lung cancer were restricted to participants with a complete lung cancer PRS (**Figure 1**). We assessed the statistical significance of the potential effect modifications by interaction testing using likelihood ratio tests. Participants were divided into categories of low (quintile 1), intermediate (quintiles 2–4), and high (quintile 5) PRS (28). The effects of the TyG index were detected at each genetic level, and the joint effects were assessed in different genetic risk groups.

We conducted a sensitivity analysis by removing participants whose TyG index was beyond the range of mean ± 3SD. We also did the sensitivity analysis to examine the robustness of our results: removing participants with a follow-up time of less than 2 years. All analyses were performed using R version 4.0.3 (http://www.rproject.org/), and a two-tailed *P*-value < 0.05 was considered statistically significant.

RESULTS

Among the 502,527 UK Biobank participants, 29,810 were non-Caucasians, 34,376 had a pre-baseline diagnosis of any cancer, and 759 had less than 12 months of follow-up. Meanwhile, 27,752 and 34,918 participants were excluded because they were missing TG and glucose values, respectively. We also excluded 20,830 participants with diabetes and 29,748 participants with dyslipidemia (**Figure 1**). Among the remaining 324,334 UK Biobank participants, there were 1,593 incident lung cancer diagnoses during a median follow-up of 9.07 years (interquartile range: 8.34–9.74 years). The median age was 57 (range: 38-73) years old.

TyG Index and Baseline Characteristics

Baseline demographic and clinical characteristics are listed in **Table 1**. There were significant differences between the two groups in terms of age at recruitment, region, sex, smoking status, alcohol intake frequency, and hypertension (all $P < 0.01$). The TyG index was higher in the lung cancer group than in the group without lung cancer ($P < 0.01$). A significant excess risk of lung cancer was observed for Townsend deprivation score, WHR, fasting time, HbA1c and TG levels (all $P < 0.01$). However, TC, HDL-C, and LDL-C levels were significantly lower in patients with lung cancer than in those without lung cancer (all $P < 0.01$). The BMI and glucose levels were similar between the two groups ($P = 0.817$ and 0.296, respectively).

TyG Index and Risk of Lung Cancer

When applying the univariate Cox model to the TyG index, a significant excess risk of lung cancer was observed (HR: 1.757, 95% CI: 1.518–1.995, $P < 0.01$). There was no evidence of an elevated risk of lung cancer linked to the TyG index in Model 1 (HR: 0.895, 95% CI: 0.644–1.145, $P = 0.385$) after adjusting for age, sex, region, Townsend deprivation score, smoking status, alcohol intake frequency, BMI, WHR, and hypertension. Similar estimates were obtained when fasting time, TC, LDL-C, HDL-C, and HbA1c were included in Model 2 (HR: 0.911, 95% CI: 0.640–1.182, $P = 0.499$) (**Table 2**). Furthermore, we did not observe significant associations between the TyG index grouped by deciles and lung cancer risk since almost all CIs crossed the line with the HR equal to 1 in both Models 1 and 2, except for the last group of TyG index (**Supplementary Figure 1**). When dividing the TyG index by median level (TyG index = 8.639), the results were the same as before with no significant associations in Model 1 (HR: 0.974, 95% CI: 0.876-1.083, $P = 0.629$) and Model 2 (HR: 0.966, 95% CI: 0.850–1.097, $P = 0.589$) (**Table 2**).

Subgroup Analysis and Sensitivity Analyses

In addition, we restricted our analyses to participants whose fasting time was ≥ 8 h, and the results were similar to the results for the entire study population (**Supplementary Figure 2**). Meanwhile, the subgroup analysis also showed consistent results in that no evidence of effect modification was observed for the TyG index with lung cancer risk in different groups since all CIs contained 1 in both Model 1 and Model 2 (**Supplementary Figure 2**). Sensitivity analyses showed that the multivariable-adjusted associations remained unchanged after excluding individuals whose TyG index was beyond the range of mean ± 3SD (**Supplementary Figure 3**) or whose follow-up time less than 2 years (**Supplementary Figure 4**).

Joint Effects and Interactions for Lung Cancer According to PRS

We observed an increased risk of lung cancer in participants with higher lung cancer PRS ($N = 320820$, cases = 1579; adjusted HR per SD increase: 1.481; 95% CI: 1.300–1.687; **Supplementary Figure 5**), which is consistent with previous studies. The association of the TyG index with lung cancer was not significant after adjusting for PRS (HR: 1.193, 95% CI: 0.901–1.580, $P = 0.218$), and we did not identify a significant interaction between the TyG index and lung cancer (genetic score group [intermediate]: TyG index group [high], HR interaction: 0.767, 95% CI: 0.568–1.035, $P = 0.083$; and genetic score group [high]: TyG index group [high], HR interaction: 0.835, 95% CI: 0.592-1.178, $P = 0.304$) (also see **Figure 2** and **Supplementary Figure 6**).

DISCUSSION

Based on a large-scale prospective cohort study, we systematically examined the association between IR quantified using the TyG index and incident lung cancer risk. Although a significant excess risk of lung cancer was observed for the TyG index in the lung cancer group or in the univariate Cox model, no association was found between the TyG index and lung cancer risk after multivariate Cox regression analysis adjusted for risk factors. Subgroup and sensitivity analyses confirmed the results. Since both genetic and environmental factors could collectively contribute to lung cancer risk, we studied the joint effects of the TyG index and PRS on lung cancer risk. The studied PRS comprised 18 significant SNPs obtained from a published genome-wide association study (18). Our study found no statistically significant interaction between the TyG index and the genetic propensity for lung cancer.

IR refers to a condition of impaired insulin action in promoting glucose uptake and use. A decline in insulin sensitivity can lead to a series of disorders (4, 5, 30). In the 1960s, it was observed that diabetes, obesity, lipid metabolism disorders, and hypertension often occur simultaneously in the same individual. In 1995, Stern proposed the "common soil" theory, which states that IR is the common basis for the above-mentioned diseases (31). IR can be assessed using various methods. The gold standard test for measuring insulin insensitivity is the euglycemic-hyperinsulinemic clamp in which peripheral glucose uptake is measured under conditions of elevated insulin concentrations. However, it is difficult to apply the glucose clamp in larger population studies and clinical settings because it is expensive and time-consuming (32). In recent years, studies have shown that the TyG index, calculated using TGs and glucose, has high sensitivity and specificity for

TABLE 1 | Baseline demographic and clinical characteristics in the study.

Characteristics	Level	No Lung Cancer (n = 322741)	Lung Cancer (n = 1593)	Total (n = 324334)	*P* value
Age	mean (sd) median	55.805 (8.051) 57	61.08 (6.121) 62	55.831 (8.051) 57	<0.001
Sex	Female	180,245 (55.848)	812 (50.973)	181,057 (55.824)	<0.001
Sex	Male	142496 (44.152)	781 (49.027)	143,277 (44.176)	
Region	London	38,499 (11.929)	144 (9.040)	38,643 (11.915)	<0.001
Region	Wales	14300 (4.431)	64 (4.018)	14,364 (4.429)	
Region	North West	46327 (14.354)	279 (17.514)	46,606 (14.370)	
Region	North East	37438 (11.600)	196 (12.304)	37,634 (11.603)	
Region	Yorkshier and Humber	49346 (15.290)	218 (13.685)	49,564 (15.282)	
Region	West Midlands	28198 (8.737)	118 (7.407)	28,316 (8.731)	
Region	East Midlands	22512 (6.975)	98 (6.152)	22,610 (6.971)	
Region	South East	29875 (9.257)	131 (8.223)	30,006 (9.252)	
Region	South West	29594 (9.170)	135 (8.475)	29,729 (9.166)	
Region	Scotland	26652 (8.258)	210 (13.183)	26,862 (8.282)	
Townsend deprivation index	[-6.26,-3.73]	80,731 (25.044)	281 (17.640)	81,012 (25.008)	<0.001
Townsend deprivation index	(-3.73,-2.32]	80663 (25.023)	300 (18.832)	80,963 (24.993)	
Townsend deprivation index	(-2.32,0.128]	80614 (25.008)	375 (23.540)	80,989 (25.001)	
Townsend deprivation index	(0.128,11]	80347 (24.925)	637 (39.987)	80,984 (24.999)	
Townsend deprivation index	missing	386	0	386	
BMI	Normal	114,815 (35.673)	575 (36.369)	115,390 (35.676)	0.817
BMI	Overweight	137889 (42.842)	674 (42.631)	138,563 (42.841)	
BMI	Obese	69154 (21.486)	332 (20.999)	69,486 (21.483)	
BMI	missing	883	12	895	
WHR	<=0.796	80,958 (25.126)	251 (15.816)	81,209 (25.080)	<0.001
WHR	(0.796,0.864]	80432 (24.963)	365 (22.999)	80,797 (24.953)	
WHR	(0.864,0.927]	80633 (25.025)	430 (27.095)	81,063 (25.035)	
WHR	>0.927	80185 (24.886)	541 (34.089)	80,726 (24.931)	
WHR	missing	533	6	539	
Smoking status	Never	180,665 (56.154)	269 (16.993)	180,934 (55.962)	<0.001
Smoking status	Previous	108576 (33.747)	649 (40.998)	109,225 (33.783)	
Smoking status	Current	32492 (10.099)	665 (42.009)	33,157 (10.255)	
Smoking status	missing	1008	10	1018	
Alcohol intake frequency	Never	19,558 (6.064)	156 (9.805)	19,714 (6.082)	<0.001
Alcohol intake frequency	Special occasions only	32815 (10.174)	191 (12.005)	33,006 (10.183)	
Alcohol intake frequency	1-3 Times/month	36153 (11.209)	136 (8.548)	36,289 (11.196)	
Alcohol intake frequency	1-2 Times/week	86049 (26.680)	372 (23.382)	86,421 (26.664)	
Alcohol intake frequency	3 or 4 Times/week	79493 (24.647)	312 (19.610)	79,805 (24.623)	
Alcohol intake frequency	Daily or almost daily	68455 (21.225)	424 (26.650)	68,879 (21.251)	
Alcohol intake frequency	missing	218	2	220	
Hypertension	No	277,358 (85.938)	1,196 (75.078)	278,554 (85.885)	<0.001
Hypertension	Yes	45383 (14.062)	397 (24.922)	45,780 (14.115)	
Fasting time	mean (sd) median	3.754 (2.395) 3	4.11 (2.918) 3	3.755 (2.398) 3	<0.001
Fasting time	missing	8	0	8	
TG	mean (sd) median	1.692 (0.982) 1.437	1.812 (1.028) 1.574	1.692 (0.982) 1.438	<0.001
Glucose	mean (sd) median	4.987 (0.823) 4.9	4.965 (0.77) 4.901	4.987 (0.823) 4.9	0.296
TyG index	mean (sd) median	8.667 (0.541) 8.638	8.739 (0.529) 8.722	8.668 (0.541) 8.639	<0.001
TC	mean (sd) median	5.786 (1.087) 5.737	5.675 (1.124) 5.65	5.786 (1.087) 5.737	<0.001
TC	missing	79	0	79	
HDL	mean (sd) median	1.477 (0.38) 1.429	1.423 (0.386) 1.368	1.476 (0.38) 1.428	<0.001
HDL	missing	68	0	68	
LDL	mean (sd) median	3.624 (0.832) 3.582	3.55 (0.869) 3.534	3.623 (0.832) 3.582	<0.001
LDL	missing	455	1	456	
HbA1c	Normal	296,039 (96.406)	1,406 (92.744)	297,445 (96.388)	<0.001
HbA1c	High	11037 (3.594)	110 (7.256)	11,147 (3.612)	
HbA1c	missing	15665	77	15742	

sd, standard deviation; BMI, Body mass index; WHR, waist-hip ratio; TC, total cholesterol; TG, triglyceride; TyG index, Triglyceride-glucose index; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; HbA1c, glycated hemoglobin; TyG index, triglyceride-glucose index.

identifying IR. Moreover, it is fast, inexpensive, and easy to use (6, 7).

It has been recognized that IR is closely related to cardio-cerebrovascular diseases. Our previous studies have confirmed that the TyG index is a sensitive pre-diagnostic indicator for cardio-cerebrovascular diseases (11). Other studies have also shown that the TyG index is closely related to cancers of the digestive organs and kidneys and that an increased BMI has a substantial effect on the risk of these cancers (24). However, very few studies have examined the association between the TyG index and lung cancer. A recent study reported that the TyG index is remarkably higher in patients with non-small cell lung cancer than

TABLE 2 | Hazard ratios and 95% confidence intervals for lung cancer according to the TyG index.

Model	Beta	SE	HR (95%CI)	Wald	P
Model 0[a]	0.564	0.122	1.757 (1.518-1.995)	4.63	<0.001
Model 0[b]	0.259	0.051	1.295 (1.173-1.430)	5.123	<0.001
Model 1[a]	-0.111	0.128	0.895 (0.644–1.145)	-0.87	0.385
Model 1[b]	-0.026	0.054	0.974 (0.876-1.083)	-0.484	0.629
Model 2[a]	-0.093	0.138	0.911 (0.640-1.182)	-0.68	0.499
Model 2[b]	-0.035	0.065	0.966 (0.850-1.097)	-0.540	0.589

[a]and [b] represent the continuous type and categorial type (divided by median) of the TyG index respectively when we performed the Cox models. Model 0: univariate Cox models. Model 1: adjusted for age, sex, region, Townsend deprivation score, smoking status, alcohol intake frequency, body mass index, waist hip rate, and hypertension. Model 2: adjusted for model 1 plus fasting time, total cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, and glycated hemoglobin.

in controls in a Chinese population (16). Consequently, the TyG index may be a suitable tumor marker for non-small cell lung cancer. The main purpose of our study was to explore whether the TyG index is a suitable predictor of lung cancer.

Previous studies investigating the association between lung cancer and IR have yielded inconsistent results. Some studies have shown that IR increases the risk of lung cancer (15, 33). Many mechanisms have been proposed to explain this observation. For example, elevated insulin levels may potentiate the activity of insulin-like growth factor-I, which represents a potent growth-promoting factor for lung cancer (34), and insulin may stimulate the Ras signaling pathway to promote lung carcinogenesis (35, 36). Moreover, additional mechanisms, such as stimulation of local angiogenesis or direct growth promotion *via* insulin receptors available on lung cancer cells, cannot be excluded. However, some studies, especially observational studies, have shown that IR is not associated with lung cancer, which is consistent with our results. The exact mechanism underlying this phenomenon is not yet clear. One possible view suggests that it is related to abnormal lipid metabolism in patients with tumors. Many cancers can cause fatty acid oxidation and/or adipose tissue lipolysis, and excessive fatty acids in the circulation may lead to IR. Abnormal lipid metabolism has been proposed as a cause of IR in obesity and type 2 diabetes (17, 30, 31). Han et al. reported an association between the blockage of whole-body fatty acid oxidation or adipose tissue lipolysis, whole-body IR, and glucose intolerance (37). IR may not be a tumor inducer but an extrapulmonary symptom in the development of lung cancer. Furthermore, several factors including obesity can independently modify both cancer risk and insulin resistance (38, 39). 1) Obesity is a worldwide health problem that is closely associated with IR and hyperinsulinemia (40). Many studies have found a strong correlation between obesity and IR (5, 41–43). 2) Obesity is a major risk factor for several common cancers (24, 44). However, a high BMI has been correlated with a reduced risk of lung cancer (45–47). 3) Hence, obesity may represent a confounding factor in the analyses of studies linking IR and lung cancer. In this study, consistent conclusions were reached after multivariate analysis adjusting the BMI and other relevant factors (**Table 2**), and in

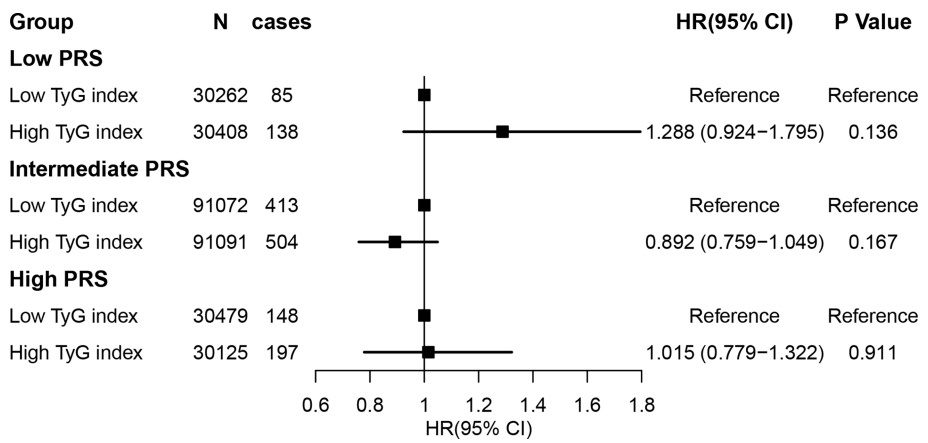


FIGURE 2 | Risk of incident lung cancer according to genetic risk and the TyG index. Participants were divided into categories of low (quintile 1), intermediate (quintiles 2–4), and high (quintile 5) genetic risk strata. The TyG index was defined as low and high according to median level (TyG index = 8.639). The hazard ratios were estimated using Cox regression models after adjusting for age, sex, region, Townsend deprivation score, smoking status, alcohol intake frequency, body mass index, waist hip ratio, and hypertension, fasting time, total cholesterol, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, and glycated hemoglobin as well as the top 10 principal components of the ancestry and genotyping array. TyG index, triglyceride-glucose index.

the subsequent stratification analysis by different levels of BMI (**Supplementary Figure 2**). Smoking is recognized as a major risk factor for lung cancer (3, 48). Smoking can suppress β-cell function, thereby reducing insulin secretion. A negative correlation between smoking and insulin has been observed in many epidemiological studies. This inverse relationship may also inhibit the tumor-promoting effects of hyperinsulinemia in lung cancer (49).

The exact mechanism between IR and lung cancer have not yet been fully understood, and more in-depth preclinical and clinical studies are needed to have a more detailed molecules and mechanisms understanding.

Our study is advantageous because of its design. First, this study had a large sample size and was prospectively designed. The longitudinal design could help reduce chance findings. The results from the multivariate Cox models at the continuous and categorical levels of the TyG index replicated each other to some degree and were further verified by incorporating the PRS of lung cancer. Second, to provide more stable and reliable results, subgroup analysis and sensitivity analyses were performed; however, the results remained the same. This somewhat reconfirmed that the analysis was stable and reliable.

Nevertheless, our study has some limitations. First, some information, such as physical activities, and female reproductive factors, was not included in the assessment, which could have affected the results. Second, we did not consider the confounding effect of inflammation, which could have overestimated the estimated indirect effect based on insulin. It has been demonstrated that the inflammatory milieu is associated with IR, obesity, dyslipidemia, and tumorigenesis (50). Third, the participants were all Caucasians from the UK Biobank, which may have contributed to selection bias while also affecting the generalizability of the results.

In conclusion, the present study showed that the TyG index level was not associated with the incidence of lung cancer. These results indicate that IR cannot identify the population at high risk of lung cancer, although the underlying mechanisms require further clarification.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: This research has been conducted using the UK Biobank Resource (Application ID: 51470). Researchers may have access to the data by submitting an application to the UK Biobank (https://www.ukbiobank.ac.uk/) through the UK Biobank Access Management System (https://bbams.ndph.ox.ac.uk/ams/).

ETHICS STATEMENT

The study protocol was approved by the North West Multicenter Research Ethics Committee, the Patient Information Advisory Group in England and Wales, and the Community Health Index Advisory Group in Scotland. All participants in the surveys have given informed consent. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LW and WR drafted this manuscript. LW, SS, JL, YL, and XC performed the data analysis. FX and WR conceived of the study and participated in its design and coordination. WR contributed to the interpretation. All authors discussed the results and commented on the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Key Research and Development Program of China (2020YFC2003500), the National Natural Science Foundation of China (81773547), the Natural Science Foundation of Shandong Province (ZR2019ZD02), and Shandong Province Major Science and Technology Innovation Project (2018CXGC1210).

ACKNOWLEDGMENTS

We would like to thank all cohort participants and team members who contributed to these studies.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.774937/full#supplementary-material

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
2. Surveillance, Epidemiology, and End Results (SEER) Program. *Cancer of the Lung and Bronchus - Cancer Stat Facts* (2021). Available from: https://seer.cancer.gov/statfacts/html/lungb.html.
3. Toumazis I, Bastani M, Han SS, Plevritis SK. Risk-Based Lung Cancer Screening: A Systematic Review. *Lung Cancer* (2020) 147:154–86. doi: 10.1016/j.lungcan.2020.07.007
4. Lebovitz HE. Insulin Resistance: Definition and Consequences. *Exp Clin Endocrinol Diabetes* (2001) 109 Suppl 2:S135–148. doi: 10.1055/s-2001-18576
5. Gluvic Z, Zaric B, Resanovic I, Obradovic M, Mitrovic A, Radak D, et al. Link Between Metabolic Syndrome and Insulin Resistance. *Curr Vasc Pharmacol* (2017) 15(1):30–9. doi: 10.2174/1570161114666161007164510

6. Guerrero-Romero F, Simental-Mendía LE, González-Ortiz M, Martínez-Abundis E, Ramos-Zavala MG, Hernández-González SO, et al. The Product of Triglycerides and Glucose, a Simple Measure of Insulin Sensitivity. Comparison With the Euglycemic-Hyperinsulinemic Clamp. *J Clin Endocrinol Metab* (2010) 95(7):3347–51. doi: 10.1210/jc.2010-0288
7. Simental-Mendía LE, Rodríguez-Morán M, Guerrero-Romero F. The Product of Fasting Glucose and Triglycerides as Surrogate for Identifying Insulin Resistance in Apparently Healthy Subjects. *Metab Syndr Relat Disord* (2008) 6 (4):299–304. doi: 10.1089/met.2008.0034
8. Navarro-González D, Sánchez-Íñigo L, Pastrana-Delgado J, Fernández-Montero A, Martinez JA. Triglyceride-Glucose Index (TyG Index) in Comparison With Fasting Plasma Glucose Improved Diabetes Prediction in Patients With Normal Fasting Glucose: The Vascular-Metabolic CUN Cohort. *Prev Med* (2016) 86:99–105. doi: 10.1016/j.ypmed.2016.01.022
9. Khan SH, Sobia F, Niazi NK, Manzoor SM, Fazal N, Ahmad F. Metabolic Clustering of Risk Factors: Evaluation of Triglyceride-Glucose Index (TyG Index) for Evaluation of Insulin Resistance. *Diabetol Metab Syndr* (2018) 10:74. doi: 10.1186/s13098-018-0376-8
10. Park JM, Shin SP, Cho SK, Lee JH, Kim JW, Kang CD, et al. Triglyceride and Glucose (TyG) Index Is an Effective Biomarker to Identify Severe Acute Pancreatitis. *Pancreatology* (2020) 20(8):1587–91. doi: 10.1016/j.pan.2020.09.018
11. Si S, Li J, Li Y, Li W, Chen X, Yuan T, et al. Causal Effect of the Triglyceride-Glucose Index and the Joint Exposure of Higher Glucose and Triglyceride With Extensive Cardio-Cerebrovascular Metabolic Outcomes in the UK Biobank: A Mendelian Randomization Study. *Front Cardiovasc Med* (2020) 7:583473. doi: 10.3389/fcvm.2020.583473
12. Okamura T, Hashimoto Y, Hamaguchi M, Obora A, Kojima T, Fukui M. Triglyceride-Glucose Index (TyG Index) Is a Predictor of Incident Colorectal Cancer: A Population-Based Longitudinal Study. *BMC Endocr Disord* (2020) 20(1):113. doi: 10.1186/s12902-020-00581-w
13. Loh WJ, North BV, Johnston DG, Godsland IF. Insulin Resistance-Related Biomarker Clustering and Subclinical Inflammation as Predictors of Cancer Mortality During 21.5 Years of Follow-Up. *Cancer Causes Control* (2010) 21 (5):709–18. doi: 10.1007/s10552-009-9499-4
14. Parekh N, Lin Y, Hayes RB, Albu JB, Lu-Yao GL. Longitudinal Associations of Blood Markers of Insulin and Glucose Metabolism and Cancer Mortality in the Third National Health and Nutrition Examination Survey. *Cancer Causes Control* (2010) 21(4):631–42. doi: 10.1007/s10552-009-9492-y
15. Petridou ET, Sergentanis TN, Antonopoulos CN, Dessypris N, Matsoukis IL, Aronis K, et al. Insulin Resistance: An Independent Risk Factor for Lung Cancer? *Metabolism* (2011) 60(8):1100–6. doi: 10.1016/j.metabol.2010.12.002
16. Yan X, Gao Y, Tong J, Tian M, Dai J, Zhuang Y. Association Between Triglyceride Glucose Index and Non-Small Cell Lung Cancer Risk in Chinese Population. *Front Oncol* (2021) 11:585388. doi: 10.3389/fonc.2021.585388
17. Argirion I, Weinstein SJ, Männistö S, Albanes D, Mondul AM. Serum Insulin, Glucose, Indices of Insulin Resistance, and Risk of Lung Cancer. *Cancer Epidemiol Biomarkers Prev* (2017) 26(10):1519–24. doi: 10.1158/1055-9965.EPI-17-0293
18. McKay JD, Hung RJ, Han Y, Zong X, Carreras-Torres R, Christiani DC, et al. Large-Scale Association Analysis Identifies New Lung Cancer Susceptibility Loci and Heterogeneity in Genetic Susceptibility Across Histological Subtypes. *Nat Genet* (2017) 49(7):1126–32. doi: 10.1038/ng.3892
19. Dai J, Shen W, Wen W, Chang J, Wang T, Chen H, et al. Estimation of Heritability for Nine Common Cancers Using Data From Genome-Wide Association Studies in Chinese Population. *Int J Cancer* (2017) 140(2):329–36. doi: 10.1002/ijc.30447
20. Collins R. What Makes UK Biobank Special? *Lancet* (2012) 379(9822):1173–4. doi: 10.1016/S0140-6736(12)60404-8
21. Palmer LJ. UK Biobank: Bank on it. *Lancet* (2007) 369(9578):1980–2. doi: 10.1016/S0140-6736(07)60924-6
22. Knuppel A, Papier K, Fensom GK, Appleby PN, Schmidt JA, Tong TYN, et al. Meat Intake and Cancer Risk: Prospective Analyses in UK Biobank. *Int J Epidemiol* (2020) 49(5):1540–52. doi: 10.1093/ije/dyaa142
23. Elliott P, Peakman TC on behalf of UK Biobank. The UK Biobank Sample Handling and Storage Protocol for the Collection, Processing and Archiving of Human Blood and Urine. *Int J Epidemiol* (2008) 37(2):234–44. doi: 10.1093/ije/dym276
24. Fritz J, Bjørge T, Nagel G, Manjer J, Engeland A, Häggström C, et al. The Triglyceride-Glucose Index as a Measure of Insulin Resistance and Risk of Obesity-Related Cancers. *Int J Epidemiol* (2020) 49(1):193–204. doi: 10.1093/ije/dyz053
25. World Health Organization. Use of Glycated Haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus. *Diabetes Res Clin Pract* (2011) 93(3):299–309. doi: 10.1016/j.diabres.2011.03.012
26. Kilpatrick ES, Bloomgarden ZT, Zimmet PZ. International Expert Committee Report on the Role of the A1C Assay in the Diagnosis of Diabetes: Response to the International Expert Committee. *Diabetes Care* (2009) 32(12):e159. author reply e160. doi: 10.2337/dc09-1231
27. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank Resource With Deep Phenotyping and Genomic Data. *Nature* (2018) 562(7726):203–9. doi: 10.1038/s41586-018-0579-z
28. Jin G, Lv J, Yang M, Wang M, Zhu M, Wang T, et al. Genetic Risk, Incident Gastric Cancer, and Healthy Lifestyle: A Meta-Analysis of Genome-Wide Association Studies and Prospective Cohort Study. *Lancet Oncol* (2020) 21 (10):1378–86. doi: 10.1016/S1470-2045(20)30460-5
29. Greenland S. Avoiding Power Loss Associated With Categorization and Ordinal Scores in Dose-Response and Trend Analysis. *Epidemiology* (1995) 6(4):450–4. doi: 10.1097/00001648-199507000-00025
30. Ighbariya A, Weiss R. Insulin Resistance, Prediabetes, Metabolic Syndrome: What Should Every Pediatrician Know? *J Clin Res Pediatr Endocrinol* (2017) 9 (Suppl 2):49–57. doi: 10.4274/jcrpe.2017.S005
31. Stern MP. Diabetes and Cardiovascular Disease. The "Common Soil" Hypothesis. *Diabetes* (1995) 44(4):369–74. doi: 10.2337/diab.44.4.369
32. Bonora E, Targher G, Alberiche M, Bonadonna RC, Saggiani F, Zenere MB, et al. Homeostasis Model Assessment Closely Mirrors the Glucose Clamp Technique in the Assessment of Insulin Sensitivity: Studies in Subjects With Various Degrees of Glucose Tolerance and Insulin Sensitivity. *Diabetes Care* (2000) 23(1):57–63. doi: 10.2337/diacare.23.1.57
33. Ho GYF, Zheng SL, Cushman M, Perez-Soler R, Kim M, Xue X, et al. Associations of Insulin and IGFBP-3 With Lung Cancer Susceptibility in Current Smokers. *J Natl Cancer Inst* (2016) 108(7):djw012. doi: 10.1093/jnci/djw012
34. Zhao S, Qiu Z, He J, Li L, Li W. Insulin-Like Growth Factor Receptor 1 (IGF1R) Expression and Survival in non-Small Cell Lung Cancer Patients: A Meta-Analysis. *Int J Clin Exp Pathol* (2014) 7(10):6694–704.
35. Ariga M, Nedachi T, Akahori M, Sakamoto H, Ito Y, Hakuno F, et al. Signalling Pathways of Insulin-Like Growth Factor-I That Are Augmented by cAMP in FRTL-5 Cells. *Biochem J* (2000) 348(Pt 2):409–16. doi: 10.1042/bj3480409
36. Gallardo A, Lerma E, Escuin D, Tibau A, Muñoz J, Ojeda B, et al. Increased Signalling of EGFR and IGF1R, and Deregulation of PTEN/PI3K/Akt Pathway Are Related With Trastuzumab Resistance in HER2 Breast Carcinomas. *Br J Cancer* (2012) 106(8):1367–73. doi: 10.1038/bjc.2012.85
37. Han X, Raun SH, Carlsson M, Sjøberg KA, Henriquez-Olguín C, Ali M, et al. Cancer Causes Metabolic Perturbations Associated With Reduced Insulin-Stimulated Glucose Uptake in Peripheral Tissues and Impaired Muscle Microvascular Perfusion. *Metabolism* (2020) 105:154169. doi: 10.1016/j.metabol.2020.154169
38. Godsland IF. Insulin Resistance and Hyperinsulinaemia in the Development and Progression of Cancer. *Clin Sci (Lond)* (2009) 118(5):315–32. doi: 10.1042/CS20090399
39. Amin MN, Hussain MS, Sarwar MS, Rahman Moghal MM, Das A, Hossain MZ, et al. How the Association Between Obesity and Inflammation may Lead to Insulin Resistance and Cancer. *Diabetes Metab Syndr* (2019) 13(2):1213–24. doi: 10.1016/j.dsx.2019.01.041
40. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, Regional, and National Prevalence of Overweight and Obesity in Children and Adults During 1980-2013: A Systematic Analysis for the Global Burden of Disease Study 2013. *Lancet* (2014) 384(9945):766–81. doi: 10.1016/S0140-6736(14)60460-8
41. Gobato AO, Vasques ACJ, Zambon MP, Barros Filho A de A, Hessel G. Metabolic Syndrome and Insulin Resistance in Obese Adolescents. *Rev Paul Pediatr* (2014) 32(1):55–62. doi: 10.1590/S0103-05822014000100010
42. Qatanani M, Lazar MA. Mechanisms of Obesity-Associated Insulin Resistance: Many Choices on the Menu. *Genes Dev* (2007) 21(12):1443–55. doi: 10.1101/gad.1550907
43. Kalupahana NS, Moustaid-Moussa N, Claycombe KJ. Immunity as a Link Between Obesity and Insulin Resistance. *Mol Aspects Med* (2012) 33(1):26–34. doi: 10.1016/j.mam.2011.10.011

44. Wang F, Liu L, Cui S, Tian F, Fan Z, Geng C, et al. Distinct Effects of Body Mass Index and Waist/Hip Ratio on Risk of Breast Cancer by Joint Estrogen and Progestogen Receptor Status: Results From a Case-Control Study in Northern and Eastern China and Implications for Chemoprevention. *Oncologist* (2017) 22(12):1431–43. doi: 10.1634/theoncologist.2017-0148
45. Rivera C, Pecuchet N, Wermert D, Pricopi C, Le Pimpec-Barthes F, Riquet M, et al. Obesity and Lung Cancer: Incidence and Repercussions on Epidemiology, Pathology and Treatments. *Rev Pneumol Clin* (2015) 71 (1):37–43. doi: 10.1016/j.pneumo.2014.11.006
46. Zhang X, Liu Y, Shao H, Zheng X. Obesity Paradox in Lung Cancer Prognosis: Evolving Biological Insights and Clinical Implications. *J Thorac Oncol* (2017) 12(10):1478–88. doi: 10.1016/j.jtho.2017.07.022
47. Yu D, Zheng W, Johansson M, Lan Q, Park Y, White E, et al. Overall and Central Obesity and Risk of Lung Cancer: A Pooled Analysis. *J Natl Cancer Inst* (2018) 110(8):831–42. doi: 10.1093/jnci/djx286
48. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global Cancer Statistics, 2012. *CA Cancer J Clin* (2015) 65(2):87–108. doi: 10.3322/caac.21262
49. Li Y, Zhang T, Han T, Li S, Bazzano L, He J, et al. Impact of Cigarette Smoking on the Relationship Between Body Mass Index and Insulin: Longitudinal Observation From the Bogalusa Heart Study. *Diabetes Obes Metab* (2018) 20 (7):1578–84. doi: 10.1111/dom.13259
50. Tilg H, Moschen AR. Inflammatory Mechanisms in the Regulation of Insulin Resistance. *Mol Med* (2008) 14(3):222–31. doi: 10.2119/2007-00119.Tilg

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Oncology

ORIGINAL RESEARCH
published: 10 December 2021
doi: 10.3389/fonc.2021.773958

Survival of Black and White Patients With Stage IV Small Cell Lung Cancer

*Huashan Shi[1], Kexun Zhou[1], Jordan Cochuyt[2], David Hodge[2], Hong Qin[3], Rami Manochakian[3], Yujie Zhao[3], Sikander Ailawadhi[3], Alex A. Adjei[4] and Yanyan Lou[3]**

[1] Department of Cancer Biology, Mayo Clinic, Jacksonville, FL, United States, [2] Department of Health Sciences Research/ Biomedical Statistics and Informatics, Mayo Clinic, Jacksonville, FL, United States, [3] Division of Hematology and Medical Oncology, Mayo Clinic, Jacksonville, FL, United States, [4] Division of Medical Oncology, Mayo Clinic, Rochester, MN, United States

Background: Small cell lung cancer (SCLC) is associated with aggressive biology and limited treatment options, making this disease a historical challenge. The influence of race and socioeconomic status on the survival of stage IV SCLC remains mostly unknown. Our study is designed to investigate the clinical survival outcomes in Black and White patients with stage IV SCLC and study the demographic, socioeconomic, clinical features, and treatment patterns of the disease and their impact on survival in Blacks and Whites.

Methods and Results: Stage IV SCLC cases from the National Cancer Database (NCDB) diagnosed between 2004 and 2014 were obtained. The follow-up endpoint is defined as death or the date of the last contact. Patients were divided into two groups by white and black. Features including demographic, socioeconomic, clinical, treatments and survival outcomes in Blacks and Whites were collected. Mortality hazard ratios of Blacks and Whites stage IV SCLC patients were analyzed. Survival of stage IV SCLC Black and White patients was also analyzed. Adjusted hazard ratios were analyzed by Cox proportional hazards regression models. Patients' median follow-up time was 8.18 (2.37-15.84) months. Overall survival at 6, 12, 18 and 24 months were 52.4%, 25.7%, 13.2% and 7.9% in Blacks in compared to 51.0%, 23.6%, 11.5% and 6.9% in Whites. White patients had significantly higher socioeconomic status than Black patients. By contrast, Blacks were found associated with younger age at diagnosis, a significantly higher chance of receiving radiation therapy and treatments at an academic/research program. Compared to Whites, Blacks had a 9% decreased risk of death.

Conclusion: Our study demonstrated that Blacks have significant socioeconomic disadvantages compared to Whites. However, despite these unfavorable factors, survival for Blacks was significantly improved compared to Whites after covariable adjustment. This may be due to Blacks with Stage IV SCLC having a higher chance of receiving radiation therapy and treatments at an academic/research program. Identifying and removing the barriers to obtaining treatments at academic/research programs or

OPEN ACCESS

Edited by:
Yutong He,
Fourth Hospital of Hebei Medical University, China

Reviewed by:
Xiaopan Li,
Shanghai Pudong Disease Prevention and Control Center, China
Naseer Ahmed,
CancerCare Manitoba, Canada

***Correspondence:**
Yanyan Lou
lou.yanyan@mayo.edu

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *10 September 2021*
Accepted: *18 November 2021*
Published: *10 December 2021*

Citation:
Shi H, Zhou K, Cochuyt J, Hodge D, Qin H, Manochakian R, Zhao Y, Ailawadhi S, Adjei AA and Lou Y (2021) Survival of Black and White Patients With Stage IV Small Cell Lung Cancer. Front. Oncol. 11:773958. doi: 10.3389/fonc.2021.773958

improving the management in non-academic centers could improve the overall survival of stage IV SCLC.

Keywords: stage IV Small cell lung cancer, racial, socioeconomic status, survival, academic program

INTRODUCTION

Blacks bear a disproportionate burden of cancers. Blacks have the lowest survival rate and the highest death rate in comparison to other racial or ethnic groups for most cancer types (1). Socioeconomic status (SES), such as income status, education level, and medical insurance, plays the most critical role in leading to these racial inequalities (1–5). Blacks have a higher rate with stage IV cancer, and the risk of cancer-related death is higher when compared with Whites because more Blacks are uninsured than Whites (6–11). Previous studies reported that the 5-year survival rate of lung cancer is lower in Blacks than those in Whites (1). In addition to socioeconomic status, curative-intent surgery also plays a vital role in survival. Blacks diagnosed with early-stage lung cancer are less likely to perform radical surgery than Whites even after considering the impact of socioeconomic factors (1, 12, 13). While the roles of race and SES disparity have been well-studied in various cancer care settings, their roles and interplay in stage IV SCLC remains primarily unknown. SCLC accounts for 10-15% of all lung cancers, and the prognosis for SCLC patients is poor (14–16). More than 60% of SCLC patients present with stage IV disease at diagnosis (17, 18). Differences in lung cancer incidence still exist among different ethnic groups, with Blacks having a significantly higher lung cancer rate than Whites and Blacks who are diagnosed with more stage IV cancers than Whites (1). In comparison, the incidence of SCLC decreases among all races, and there is no significant difference in stage distributions between Blacks and Whites from 2006 to 2010 (19).

NCDB is a prospectively maintained registry database covering 70% of newly diagnosed cancer cases in the nation (20). It includes 82% of lung cancer cases with an annual follow-up of at least 90% of the patients (21). In this current study, we analyzed the survival differences in stage IV SCLC patients between Blacks and Whites, considering various variables available in NCDB. We also investigated the characteristics of various clinical and treatment-related features among Blacks and Whites and the potential impact on clinical outcomes.

METHODS

SCLC cases diagnosed between 2004 and 2014 were obtained from the National Cancer Database (22). The histology codes were mainly based on the International Agency for Research on Cancer (IARC) classifications: small cell carcinoma [International Classification of Diseases for Oncology Third Edition (ICD-O-3) codes 8002, 8041–8045)]. We identified 214,096 cases with TNM staging data, excluding stage I, II, III, and the stage unknown cases. The final study cohort consisted of 119,611 stage IV SCLC patients, including 110,696 White patients and 8,915 Black patients.

Study subjects were included in the study from the date of diagnosis and were followed until the end of the study period, the date of the last contact, or death, whichever came first. The primary outcome measure in our study was overall survival (OS) (20). Baseline demographic features including sex, age, education, census median income quartiles, insurance, living area, geographic region, distance to treating facility (great circle distance, distance in miles between patient's residence based on ZIP code centroid or city to street address of treating facility); clinical characteristics including the time of diagnosis, tumor size, Charlson-Deyo score, and treatments including radiation therapy, chemotherapy, immunotherapy, palliative care, facility procedure volume, academic/research program between Blacks and Whites were studied. Patients with missing information were excluded from the analysis. Predictors included age (<60, 60-69, 70-79, >=80), insurance at the time of diagnosis (private, government, no insurance, missing), percentage of without high school degree 2007-2012 (<7%, 7-12.9%, 13-20.9%, >=21%), census median income quartile 2007-2012 (<$38,000, $38,000-$47,999, $48,000-$62,999, >$63,000) and Charlson-Deyo comorbidity score (0, 1, 2, >=3). A modified Charlson-Deyo score was calculated from preexisting comorbidities, which up to six conditions (23).

Statistical Analysis

The distribution of demographics, clinical, and treatment features was compared between Blacks *vs*. Whites using Pearson's chi-square test and Wilcoxon rank-sum test, as appropriate. Overall survival was calculated as the time from the initial diagnosis of stage IV SCLC to date of death or the last known alive. Kaplan-Meier analysis was used to compare survival differences between Blacks and Whites. Multivariable Cox regression modeling was used to identify independent features associated with survival in patients with stage IV SCLC and hazard ratios (HR) for mortality were presented (22). Nonproportional variables were used as stratification variables, and the proportional hazards assumptions were performed using Schoenfield residuals in the final analysis. The primary endpoint of the study was the survival difference between Whites and Blacks. The possible association of other features with survival was the secondary analysis. A two-sided P-value of less than 0.05 with a confidence interval limit at 95% was considered as statistical significance.

RESULTS

A total of 119,611 stage IV SCLC patients, including 110,696 White patients and 8,915 Black patients, were included in this study.

Abbreviations: SCLC, small cell lung cancer; NCDB, National Cancer Database; SES, Socioeconomic status; IARC, International Agency for Research on Cancer; OS, overall survival; PCI, prophylactic cranial irradiation; SEER, The Surveillance, Epidemiology, and End Results Program.

The median follow-up is 8.18 months (range 2.37-15.84 months). Patients' demographic, clinical, and treatment features were compared between Blacks and Whites (**Tables 1**–**3**). Blacks were more likely to be male, younger, with lower education and lower annual income than Whites on the logistic regression model. They also had shorter travel distances to their treatment site. The median travel distance between the patient's primary residence and treatment site was 13.8 *vs*. 24.9 miles in Blacks *versus* Whites (P<0.0001). Blacks were found to have larger tumors (mean tumor size 58 *vs*. 54.2 mm, p <0.0001) and were more likely to receive radiation therapy (41.6% *vs*. 38.9%, p<0.0001) and treatment in academic centers (41.4% *vs*. 24.7%, p<0.0001) than Whites. In addition, Blacks were less likely to receive palliative therapy than Whites (20.7% *vs*. 21.7%, p = 0.03). Among stage IV SCLC patients who received the treatments, 7.7% of Blacks and 6.3% of Whites received radiation therapy alone (p<0.0001), 36.2% of Blacks and 38.2% of Whites received chemotherapy alone (p<0.0001), and 33.3% of Blacks and 32% of Whites received both chemotherapy and radiation therapy (p<0.0001, **Supplementary Table 1**). Among the patients who received radiation therapy and chemotherapy, 60.5% of Black and 62.6% White patients received chemotherapy first, followed by radiation therapy (p = 0.0299). In addition, more Blacks were found brain metastasis than Whites (27.7% *vs*. 24.6%, p<0.0001). The incidence of stage IV SCLC increased from 2004 to 2014 in both Blacks and Whites, although the distribution was slightly different. No significant differences were found between Blacks and Whites in insurance status, living area, geographic region, diagnostic confirmation method, Charlson-Deyo score, chemotherapy, immunotherapy, and treatment facility volume.

Although the overall distribution was consistent regarding the patients' and tumor's characteristics, some differences were noticed. Among those with stage IV SCLC, the proportion of female patients was lower than male patients in both Whites (48.8% versus 51.2%) and Blacks (48.3% *versus* 51.7%) (**Figure 1A**). The total number of White stage IV SCLC patients increased by 27.9% in 2014 compared to 2004 (8,393 patients in 2004 and 11,639 patients in 2014). The

TABLE 1 | Demographic characteristics of patients with stage IV SCLC.

	White (N = 110696)	Black (N = 8915)	*P* value
Sex			0.0147
Male	56673 (51.2%)	4606 (51.7%)	
Female	54023 (48.8%)	4309 (48.3%)	
Age			0.0145
<60	26653 (24.1%)	2586 (29.0%)	
60-69	38088 (34.4%)	3061 (34.3%)	
70-79	32960 (29.8%)	2386 (26.8%)	
>=80	12995 (11.7%)	882 (9.9%)	
Percent No High School Degree 2007-2012[a]			0.0122
Missing	2292	140	
>=21%	18633 (17.2%)	3433 (39.1%)	
13-20.9%	32178 (29.7%)	3289 (37.5%)	
7-12.9%	37236 (34.3%)	1561 (17.8%)	
<7%	20357 (18.8%)	492 (5.6%)	
Census Median Income Quartiles 2007-2012[b]			0.0034
Missing	2344	145	
<$38,000	20776 (19.2%)	4458 (50.8%)	
$38,000-$47,999	29978 (27.7%)	1964 (22.4%)	
$48,000-$62,999	30520 (28.2%)	1465 (16.7%)	
$63,000+	27078 (25.0%)	883 (10.1%)	
Patient's Insurance			0.2267
Missing	4618	554	
No insurance	2220 (2.1%)	243 (2.9%)	
Government Insurance	73253 (69.1%)	6201 (74.2%)	
Private insurance	30605 (28.9%)	1917 (22.9%)	
Type of area			0.2763
Missing	4289	215	
Urban	20207 (19.0%)	781 (9.0%)	
Metro	83254 (78.2%)	7814 (89.8%)	
Rural	2946 (2.8%)	105 (1.2%)	
Great Circle Distance			<0.0001
N	108440	8763	
Mean (SD)	24.9 (91.6)	13.8 (53.9)	
Median	9.2	5.0	
Geographic region			0.1088
Missing	339	39	
East Coast	45258 (41.0%)	4225 (47.6%)	
Central	52345 (47.4%)	4223 (47.6%)	
Mountain	4087 (3.7%)	64 (0.7%)	
Pacific	8667 (7.9%)	364 (4.1%)	

[a,b]*Variables refer to the residential region, rather than individual.*

TABLE 2 | Disease characteristics of patients with stage IV SCLC.

	White (N = 110696)	Black (N = 8915)	*P* value
Year of Diagnosis			<0.0001
2004	8393 (7.6%)	645 (7.2%)	
2005	8543 (7.7%)	649 (7.3%)	
2006	8838 (8.0%)	643 (7.2%)	
2007	8947 (8.1%)	708 (7.9%)	
2008	9933 (9.0%)	758 (8.5%)	
2009	10032 (9.1%)	836 (9.4%)	
2010	10816 (9.8%)	914 (10.3%)	
2011	11005 (9.9%)	912 (10.2%)	
2012	11200 (10.1%)	921 (10.3%)	
2013	11350 (10.3%)	971 (10.9%)	
2014	11639 (10.5%)	958 (10.7%)	
Diagnostic confirmation			0.3776
Missing	103	7	
Positive histology	92351 (83.5%)	7239 (81.3%)	
Positive cytology	18242 (16.5%)	1669 (18.7%)	
Tumor size (mm)			<0.0001
N	68419	5580	
Mean (SD)	54.2 (51.6)	58.0 (50.8)	
Median	47.0	50.0	
Charlson-Deyo Score			0.1180
0	61000 (55.1%)	4883 (54.8%)	
1	33422 (30.2%)	2522 (28.3%)	
2	11645 (10.5%)	1035 (11.6%)	
>=3	4629 (4.2%)	475 (5.3%)	

TABLE 3 | Clinical treatment of patients with stage IV SCLC.

	White (N = 110696)	Black (N = 8915)	*P* value
Radiation Therapy			<0.0001
Missing	552	38	
No	67347 (61.1%)	5184 (58.4%)	
Yes	42797 (38.9%)	3693 (41.6%)	
Chemotherapy			0.1874
Missing	1512	146	
No	31134 (28.5%)	2559 (29.2%)	
Yes	78050 (71.5%)	6210 (70.8%)	
Immunotherapy			0.3848
Missing	352	48	
No	110019 (99.7%)	8846 (99.8%)	
Yes	325 (0.3%)	21 (0.2%)	
Palliative Care			0.0331
Missing	502	33	
No	86268 (78.3%)	7040 (79.3%)	
Yes	23926 (21.7%)	1842 (20.7%)	
Facility Procedure Volume			0.1189
Low	8706 (7.9%)	500 (5.6%)	
Medium	49572 (44.8%)	3792 (42.5%)	
High	52418 (47.4%)	4623 (51.9%)	
Academic/Research Program			<0.0001
Missing	339	39	
No	83049 (75.3%)	5201 (58.6%)	
Yes	27308 (24.7%)	3675 (41.4%)	

number of Black stage IV SCLC patients increased by 32.7% in 2014 compared to 2004 (645 in 2004 and 958 in 2014). The percentage of stage IV SCLC among all patients with SCLC had also increased significantly from 2004 to 2014 in both Whites and Blacks, and the increases in Whites were more than those in Blacks (**Figure 1B**). The median overall survival time of stage IV SCLC in Blacks was 6.57 months, higher than 6.21 months in Whites ($p<0.001$). The 6 months, 12 months, 18 months and 24 months' survivals were 51.0%, 23.6%, 11.5%, 6.9% for Whites and 52.4%, 25.7%, 13.2%, 7.9% for Blacks ($p<0.01$) (**Figure 1C**).

We next used the Cox proportional hazards multivariable model to analyze the potential predictors of overall survival of patients with stage IV SCLC. Black was independently associated with a decreased hazard of death in Cox proportional hazards

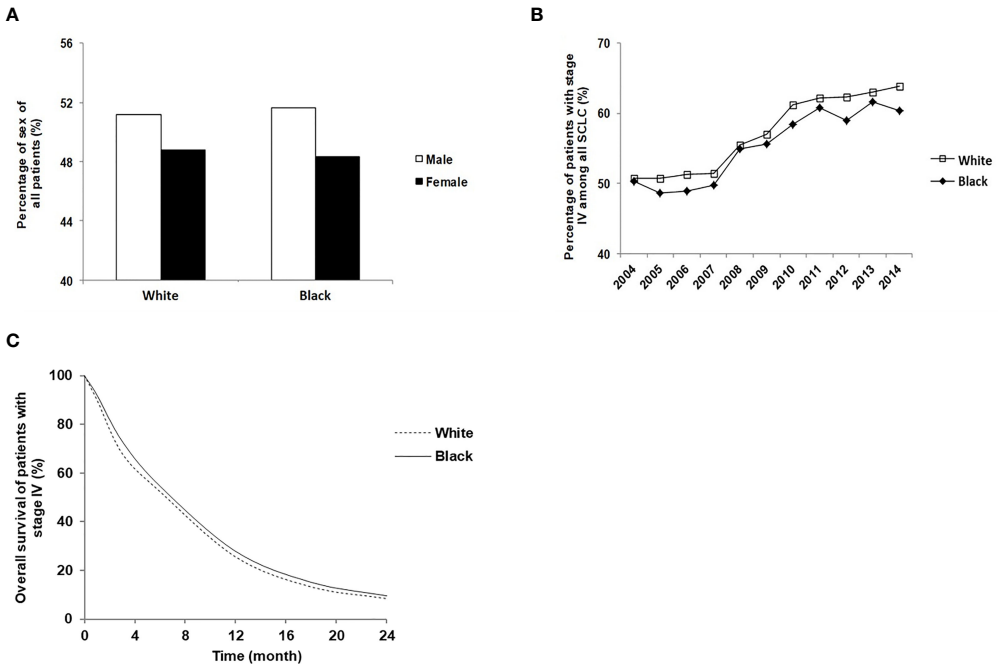


FIGURE 1 | Disease distribution characteristics and survival time of stage IV SCLC patients of Black and White. **(A)** Sex distribution of White and Black stage IV SCLC patients; **(B)** The percentage of White and Black patients who were diagnosed with stage IV SCLC between 2004 and 2014; **(C)** The overall survival of White and Black stage IV SCLC patients.

modeling after controlling for demographic and clinical factors (HR=0.911; 95%CI: 0.884-0.938; P<0.0001). Other favorable factors associated with improved OS included female (HR=0.852; 95% CI: 0.839-0.864, P<0.0001), greater distance to treatment center (HR=0.988; 95%CI: 0.983-0.993, P<0.0001), higher income particularly income>63,000 (HR=0.929; 95% CI: 0.901-0.958, P<0.0001), private insurance (HR=0.888; 95% CI: 0.835-0.945, P<0.0001), radiation therapy (HR=0.774; 95% CI: 0.762-0.787, P<0.0001), chemotherapy (HR=0.406; 95%CI:0.397-0.414, P<0.0001), and receiving treatment in academic/research center (HR=0.978; 95% CI: 0.96-0.997, P=0.02) (**Figure 2**). Treatment with Chemotherapy appeared to be the most favorable predictor of survival in patients with stage IV SCLC. By contrast, factors associated with decreased OS included increase in age (HR=1.136; 95% CI: 1.126-1.147, P<0.0001), high education (HR=1.05; 95% CI: 1.017-1.085, P=0.0029), living in Rural (HR=1.055; 95% CI: 1.005-1.108, P = 0.0315), high Charlson-Deyo score (HR=1.537; 95% CI: 1.481-1.596, P < 0.0001), increase in tumor size (HR=1.007; 95% CI: 1.006-1.009, P<0.0001) and palliative care (HR=1.225; 95% CI: 1.202-1.247, P<0.0001).

DISCUSSION

Cancer screening, advances in surgery and radiation techniques, and developments of novel therapeutic agents have undoubtedly led to improved clinical outcomes in many cancer patients. However, new developments and advances are uneven among different cancer types, and disparities exist in clinical outcomes across multiple cancer types and some attributes to modifiable factors. Stage IV SCLC represented one of the most notorious cancers with only 2-4 months of survival in untreated patients. Chemotherapy using platinum-based doublet remains a cornerstone of first-line treatment until the very recent breakthrough of adding the benefit of immunotherapy in combination with chemotherapy (24, 25). Although modest improvement in survival has been observed in patients with SCLC over time, the differences in clinical outcome, socioeconomic status, clinical features, and treatment patterns between races, particularly Blacks *versus* Whites, remain mostly unknown. Our current study represents the largest real-world analysis of stage IV SCLC patients to investigate the clinical outcomes, the impacts of SES, clinical and treatment factors among Blacks and Whites.

Multiple studies have examined the impact of race and ethnicity on overall lung cancer prognosis (26–28). These studies have demonstrated that non-Black patients have had favorable survival than Blacks over the last several decades. In fact, Blacks have been reported to have the lowest 1-year, and 3-year survival rates in patients with non-small cell lung cancer among all races (29), and the 5-year survival rate for lung cancer is lower in Blacks than in Whites (16% *vs.* 19%, respectively) (1). In a multivariate analysis of SCLC patients treated through Southwest Oncology Group trials, Whites were found associated with a more favorable clinical outcome than Blacks (26). A few factors such as differences in treatments, time of diagnosis, and possible biological variability

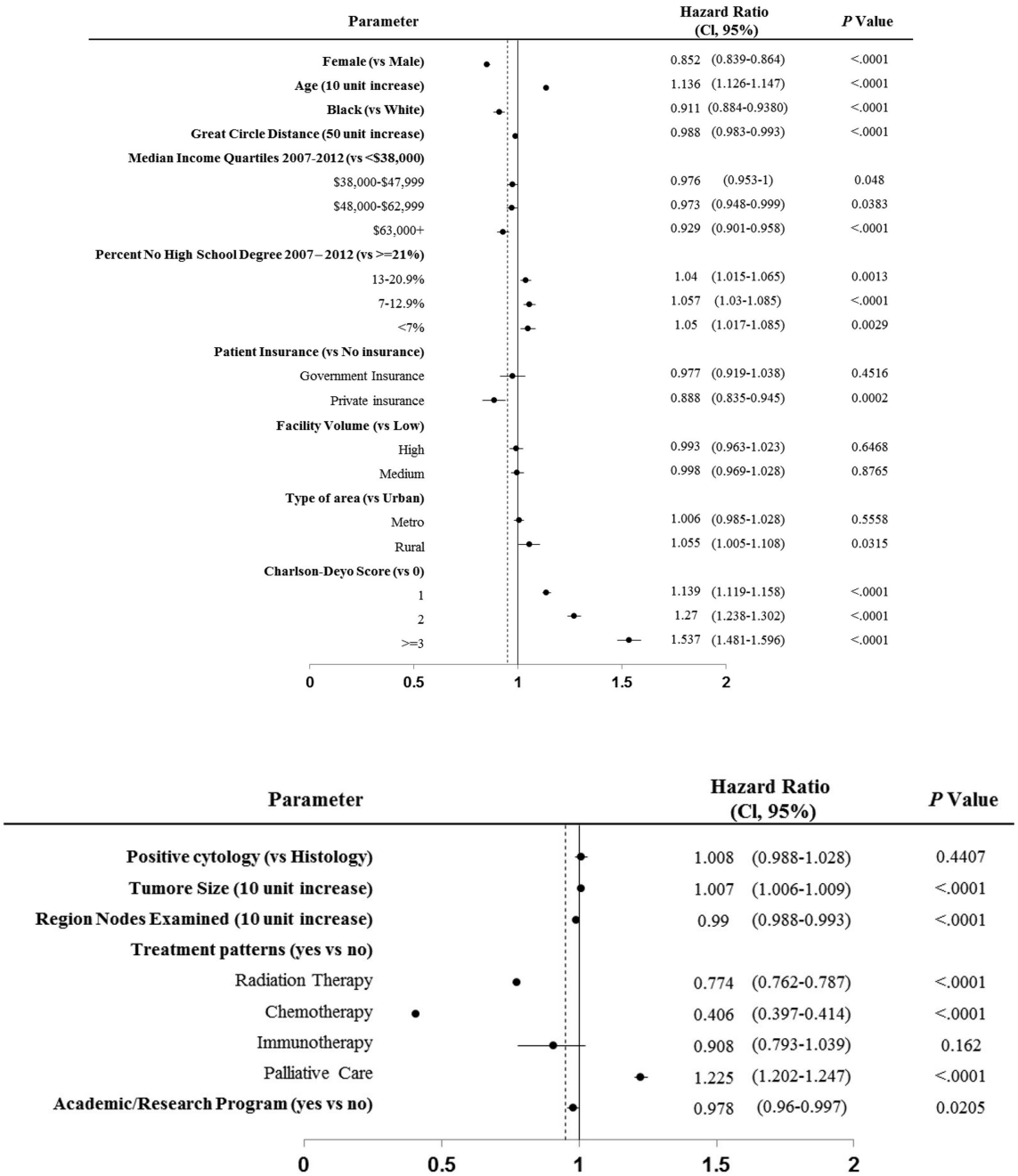


FIGURE 2 | Cox proportional hazards multivariable regression analysis for predictors of overall survival of patients with stage IV SCLC.

were suspected of contributing to the difference observed in Whites *versus* non-Whites. By contrast, using the CALGB lung cancer database, Stock et al. found that the overall survival of Black patients was not significantly different from that of non–Black patients with or without adjustment for histology, treatment, and metastatic site, despite Black patients were more likely to present with worse performance status and lower socioeconomic condition (30). Both studies offered high-quality information built upon the constraints of clinical trial entry criteria, close follow-up, and relatively uniform treatment regimens through the studies. However, the overall numbers of SCLC patients remain relatively small, and the patients who enter the clinical studies may not best represent patients in the real world. Our study specifically focused on stage IV SCLC where the treatment regimens have been relatively uniform to control the potential confounding effects of stages on outcomes. Our study included a total of 119,611 stage IV

SCLC patients. We demonstrated that Blacks with stage IV SCLC had improved overall survival compared to Whites (HR 0.91, CI: 0.884-0.938; P<0.0001). This finding is interesting as our analysis revealed that Blacks were more likely to be associated with unfavorable factors such as lower income levels and larger tumor size. Despite the lower socioeconomic status and larger tumor, Blacks' one-year and two-year survival were qualitatively similar and statistically superior to Whites (25.7% *versus* 23.6% and 7.9% *versus* 6.9% respectively, p<0.01). This might be because more Black patients received radiation therapy (41.6% *vs*. 38.9%, p<0.0001), had a younger age at diagnosis, and received treatments in academic/research programs (41.4% *vs*. 24.7%, p<0.0001) than Whites.

Consolidative thoracic radiotherapy and prophylactic cranial irradiation (PCI) have been recommended in patients with stage IV SCLC based on improved overall survival (15, 31). Particularly, PCI has been recommended as the standard of care during our study period, although a recent study performed in Japan demonstrated inconsistent conclusions. Nevertheless, our data found that radiation therapy led to a 23% reduction in mortality (HR 0.77, CI: 0.762-0.787; P<0.0001) in patients with stage IV, and remained the second most favorable factor. It was not surprising to see that chemotherapy represented the most favorable factor in patients with stage IV SCLC (HR 0.41, CI: 0.397-0.414; P<0.0001). Our data suggested that the improved overall survival observed in Blacks than Whites might be related to more Blacks receiving radiation therapy than Whites. Unfortunately, we are unable to classify further the type of radiation (to the primary site *vs* prophylactic) and location of radiation therapy (thoracic, brain, or other organs) due to limitations of the NCDB dataset. Further studies in this regard are warranted and may be done through claims-based databases. Age was found to be an unfavorable prognosis factor in our study (HR 1.14, CI: 1.126-1.147; P<0.0001), and Blacks were more likely to be diagnosed at an age younger than 60 years (29% *vs*. 24.1%, p=0.0145). This finding is consistent with our previous findings in non-small cell lung cancer (20). It is also found Black patients have superior survival compared to White patients with multiple myeloma, particularly due to diagnosis in the younger population (22, 32).

One of the most striking differences we noticed between Blacks and Whites was that Blacks were more likely to receive their treatments in academic/research programs than Whites. In our study, 41.4% of Black patients with stage IV SCLC received treatment at an academic/research program in contrast to 24.7% in White patients. Our study also demonstrated that treatment in academic/research centers was a favorable factor, therefore likely contributing to the improved survival in Blacks that we observed in our study. This data suggested receiving care at an academic/research center might mitigate the significant SES disadvantages in society and medical care as an actionable approach. This was similar to our previous study in NSCLC where we found the initial therapy at academic centers significantly improved clinical outcomes (20). Similarly, treatment at academic centers was also demonstrated independently associated with improved survival in patients with locally advanced head and neck cancer (22, 33). The underlying factors driving the improved outcomes are likely multifactorial. Academic/research centers more likely provide access to clinical trials, multi-disciplinary expertise, and ancillary services. Unfortunately, we cannot determine each factor's exact impact on the survival of stage IV SCLC patients in our study due to the lack of such information in NCDB (34).

Consistent with previous studies performed in non-small cell lung cancers, our study demonstrated that Blacks were associated with lower socioeconomic status, including lower education and low annual incomes than Whites (20, 35–39). Interestingly, high education was found to not correlate with improved clinical outcomes. Instead, it was associated with worse clinical outcomes, although the difference was small (HR 1.05, CI: 1.017-1.085; P=0.0029). This is different from findings from other cancers, including non-small cell lung cancer, where a high level of education has been a beneficial factor for survival (34, 40, 41). This may be due to the ratio of no high school degree patients referred to the residential region rather than individual. Another reason might be that most Blacks are less educated and have better outcomes. Patients with high annual income, particularly those with > $ 63,000, were associated with favorable outcomes, likely due to more access to treatments. Similarly, females were also found to have better outcomes in our study, consistent with findings from other cancer types (39).

Not surprisingly, high comorbidity status was found closely correlated with poor prognosis. Palliative care was found to be associated with poor prognosis in our study. This may be because patients who received palliative care were more symptomatic and had lower tolerability to therapy. It has been noted previously that palliative care utilization is extremely low even amongst academic cancer centers, suggesting that its use is more associated with an alternative than an adjunct to active anticancer therapeutics (42). During this period of 2004 to 2014, the incidence of SCLC declined (43). However, the percentages of stage IV SCLC were increased in both White and Blacks, indicating potentially later diagnosis of this disease.

Limitations

While our study was the first large-scale data analysis focusing on various factors associated with stage IV SCLC outcomes in Blacks and Whites, several limitations were noticed. This study was performed retrospectively through NCDB database, and therefore selection bias was presented. As shown in our study, Blacks with stage IV SCLC had significant disadvantages in socioeconomic status compared to Whites. Thus, Blacks with stage IV SCLC were more likely to present without complete staging information and might be under-represented in our study. The imbalance in the numbers of Blacks and Whites was also evident. However, the number of Blacks included in this study is still higher than other available databases, such as The Surveillance, Epidemiology, and End Results (SEER) Program. Furthermore, although the NCDB contains relatively comprehensive information on cancer patients in the United States, some detailed information was not available. For example, the details of individual treatment, chemotherapy regimens and the number of cycles, accurate radiation dose and field, and comorbidities were not available, which might impact assessment accuracy. Besides the treatment status, it is also a lack of

the smoking data of patients, which is important for the survival of SCLC patients. In addition, due to the dismal prognosis associated with stage IV SCLC, the magnitude of overall survival difference between Blacks and Whites is relatively small.

CONCLUSIONS

Our study demonstrated that Blacks were associated with significant socioeconomic disadvantages in comparison with Whites. However, despite these unfavorable factors, Blacks had survival outcomes qualitatively similar and statistically superior than those of Whites after co-variable adjustment. Blacks were found associated with younger age at diagnosis, a significantly higher chance of receiving radiation therapy and treatments at the academic/research program. This suggests the importance of radiotherapy and receiving care at an academic center could mitigate the known SES disadvantages in treating small cell lung cancer patients.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: National cancer database.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Mayo Clinic IRB. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

HS and YL drafted the article. HS, KZ, and YL designed the study. JC and DH helped the statistical analysis. HS, KZ, HQ, AA, RM, YZ, AS, and YL give critical revision of the article for important intellectual content. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Institutes of Health [grant number is K12CA090628, YL].

ACKNOWLEDGMENTS

The data used in the study are derived from a de-identified NCDB file. The American College of Surgeons and the Commission on Cancer have not verified and are not responsible for the analytic or statistical methodology used or the conclusions drawn from these data by the investigators.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.773958/full#supplementary-material

Supplementary Table 1 | Chemotherapy/radiation treatment and brain metastasis in patients with stage IV SCLC

REFERENCES

1. DeSantis CE, Miller KD, Sauer AG, Jemal A, Siegel RL. Cancer Statistics for African Americans, 2019. *CA Cancer J Clin* (2019) 69(3):211–33. doi: 10.3322/caac.21555
2. DeSantis C, Jemal A, Ward E. Disparities in Breast Cancer Prognostic Factors by Race, Insurance Status, and Education. *Cancer Causes Control* (2010) 21 (9):1445–50. doi: 10.1007/s10552-010-9572-z
3. Fedewa SA, Etzioni R, Flanders WD, Jemal A, Ward EM. Association of Insurance and Race/Ethnicity With Disease Severity Among Men Diagnosed With Prostate Cancer, National Cancer Database 2004-2006. *Cancer Epidemiol Biomarkers Prev* (2010) 19(10):2437–44. doi: 10.1158/1055-9965.EPI-10-0299
4. Singh GK, Jemal A. Socioeconomic and Racial/Ethnic Disparities in Cancer Mortality, Incidence, and Survival in the United States, 1950-2014: Over Six Decades of Changing Patterns and Widening Inequalities. *J Environ Public Health* (2017) 2017:1–19. doi: 10.1155/2017/2819372
5. Soni A, Sabik LM, Simon K, Sommers BD. Changes in Insurance Coverage Among Cancer Patients Under the Affordable Care Act. *JAMA Oncol* (2018) 4 (1):122–4. doi: 10.1001/jamaoncol.2017.3176
6. Niu XL, Roche LM, Pawlish KS, Henry KA. Cancer Survival Disparities by Health Insurance Status. *Cancer Med-Us* (2013) 2(3):403–11. doi: 10.1002/cam4.84
7. DeRouen MC, Parsons HM, Kent EE, Pollock BH, Keegan THM. Sociodemographic Disparities in Survival for Adolescents and Young Adults With Cancer Differ by Health Insurance Status. *Cancer Causes Control* (2017) 28(8):841–51. doi: 10.1007/s10552-017-0914-y
8. Ward E, Halpern M, Schrag N, Cokkinides V, DeSantis C, Bandi P, et al. Association of Insurance With Cancer Care Utilization and Outcomes. *CA Cancer J Clin* (2008) 58(1):9–31. doi: 10.3322/CA.2007.0011
9. Pan HY, Walker GV, Grant SR, Allen PK, Jiang J, Guadagnolo BA, et al. Insurance Status and Racial Disparities in Cancer-Specific Mortality in the United States: A Population-Based Analysis. *Cancer Epidemiol Biomarkers Prev* (2017) 26(6):869–75. doi: 10.1158/1055-9965.EPI-16-0976
10. Gaffney A, McCormick D, Bor DH, Goldman A, Woolhandler S, Himmelstein DU. The Effects on Hospital Utilization of the 1966 and 2014 Health Insurance Coverage Expansions in the United States. *Ann Intern Med* (2019) 171(3):172–+. doi: 10.7326/M18-2806
11. de Moor JS, Cohen RA, Shapiro JA, Nadel MR, Sabatino SA, Yabroff KR, et al. Colorectal Cancer Screening in the United States: Trends From 2008 to 2015 and Variation by Health Insurance Coverage. *Prev Med* (2018) 112:199–206. doi: 10.1016/j.ypmed.2018.05.001
12. Coughlin SS, Matthews-Juarez P, Juarez PD, Melton CE, King M. Opportunities to Address Lung Cancer Disparities Among African Americans. *Cancer Med-Us* (2014) 3(6):1467–76. doi: 10.1002/cam4.348
13. Check DK, Albers KB, Uppal KM, Suga JM, Adams AS, Habel LA, et al. Examining the Role of Access to Care: Racial/ethnic Differences in Receipt of Resection for Early-Stage non-Small Cell Lung Cancer Among Integrated System Members and non-Members. *Lung Cancer* (2018) 125:51–6. doi: 10.1016/j.lungcan.2018.09.006

14. Govindan R, Page N, Morgensztern D, Read W, Tierney R, Vlahiotis A, et al. Changing Epidemiology of Small-Cell Lung Cancer in the United States Over the Last 30 Years: Analysis of the Surveillance, Epidemiologic, and End Results Database. *J Clin Oncol* (2006) 24(28):4539–44. doi: 10.1200/JCO.2005.04.4859
15. Slotman BJ, van Tinteren H, Praag JO, Knegjens JL, El Sharouni SY, Hatton M, et al. Use of Thoracic Radiotherapy for Extensive Stage Small-Cell Lung Cancer: A Phase 3 Randomised Controlled Trial. *Lancet* (2015) 385(9962):36–42. doi: 10.1016/S0140-6736(14)61085-0
16. Herbst RS, Heymach JV, Lippman SM. Lung Cancer. *N Engl J Med* (2008) 359 (13):1367–80. doi: 10.1056/NEJMra0802714
17. van Meerbeeck JP, Fennell DA, De Ruysscher DKM. Small-Cell Lung Cancer. *Lancet* (2011) 378(9804):1741–55. doi: 10.1016/S0140-6736(11)60165-7
18. Byers LA, Rudin CM. Small Cell Lung Cancer: Where Do We Go From Here? *Cancer-Am Cancer Soc* (2015) 121(5):664–72. doi: 10.1002/cncr.29098
19. Meza R, Meernik C, Jeon J, Cote ML. Lung Cancer Incidence Trends by Gender, Race and Histology in the United States, 1973-2010. *PloS One* (2015) 10(3):e0121323. doi: 10.1371/journal.pone.0121323
20. Lou Y, Dholaria B, Soyano A, Hodge D, Cochuyt J, Manochakian R, et al. Survival Trends Among non-Small-Cell Lung Cancer Patients Over a Decade: Impact of Initial Therapy at Academic Centers. *Cancer Med* (2018) 7 (10):4932–42. doi: 10.1002/cam4.1749
21. Merritt RE, Abdel-Rasoul M, Fitzgerald M, D'Souza DM, Kneuertz PJ. The Academic Facility Type Is Associated With Improved Overall Survival for Early-Stage Lung Cancer. *Ann Thorac Surg* (2021) 111(1):261–8. doi: 10.1016/j.athoracsur.2020.05.051
22. Zhou K, Shi H, Chen R, Cochuyt JJ, Hodge DO, Manochakian R, et al. Association of Race, Socioeconomic Factors, and Treatment Characteristics With Overall Survival in Patients With Limited-Stage Small Cell Lung Cancer. *JAMA Netw Open* (2021) 4(1):e2032276. doi: 10.1001/jamanetworkopen.2020.32276
23. Deyo RA, Cherkin DC, Ciol MA. Adapting a Clinical Comorbidity Index for Use With Icd-9-Cm Administrative Databases. *J Clin Epidemiol* (1992) 45 (6):613–9. doi: 10.1016/0895-4356(92)90133-8
24. Paz-Ares L, Dvorkin M, Chen Y, Reinmuth N, Hotta K, Trukhin D, et al. Durvalumab Plus Platinum-Etoposide Versus Platinum-Etoposide in First-Line Treatment of Extensive-Stage Small-Cell Lung Cancer (CASPIAN): A Randomised, Controlled, Open-Label, Phase 3 Trial. *Lancet* (2019) 394 (10212):1929–39. doi: 10.1016/S0140-6736(19)32222-6
25. Horn L, Mansfield AS, Szczesna A, Havel L, Krzakowski M, Hochmair MJ, et al. First-Line Atezolizumab Plus Chemotherapy in Extensive-Stage Small-Cell Lung Cancer. *N Engl J Med* (2018) 379(23):2220–9. doi: 10.1056/NEJMoa1809064
26. Albain KS, Crowley JJ, LeBlanc M, Livingston RB. Determinants of Improved Outcome in Small-Cell Lung Cancer: An Analysis of the 2,580-Patient Southwest Oncology Group Data Base. *J Clin Oncol* (1990) 8(9):1563–74. doi: 10.1200/JCO.1990.8.9.1563
27. Ries LA, Wingo PA, Miller DS, Howe HL, Weir HK, Rosenberg HM, et al. The Annual Report to the Nation on the Status of Cancer, 1973-1997, With a Special Section on Colorectal Cancer. *Cancer-Am Cancer Soc* (2000) 88 (10):2398–424. doi: 10.1002/(SICI)1097-0142(20000515)88:10<2398::AID-CNCR26>3.0.CO;2-I
28. Wingo PA, Ries LA, Giovino GA, Miller DS, Rosenberg HM, Shopland DR, et al. Annual Report to the Nation on the Status of Cancer, 1973-1996, With a Special Section on Lung Cancer and Tobacco Smoking. *J Natl Cancer Inst* (1999) 91(8):675–90. doi: 10.1093/jnci/91.8.675
29. Tannenbaum SL, Koru-Sengul T, Zhao W, Miao F, Byrne MM. Survival Disparities in Non-Small Cell Lung Cancer by Race, Ethnicity, and Socioeconomic Status. *Cancer J* (2014) 20(4):237–45. doi: 10.1097/PPO.0000000000000058
30. Blackstock AW, Herndon JE, Paskett ED, Miller AA, Lathan C, Niell HB, et al. Similar Outcomes Between African American and non-African American Patients With Extensive-Stage Small-Cell Lung Carcinoma: Report From the Cancer and Leukemia Group B. *J Clin Oncol* (2006) 24(3):407–12. doi: 10.1200/JCO.2005.02.1436
31. Slotman B, Faivre-Finn C, Kramer G, Rankin E, Snee M, Hatton M, et al. Prophylactic Cranial Irradiation in Extensive Small-Cell Lung Cancer. *N Engl J Med* (2007) 357(7):664–72. doi: 10.1056/NEJMoa071780
32. Fillmore NR, Yellapragada SV, Ifeorah C, Mehta A, Cirstea D, White PS, et al. With Equal Access, African American Patients Have Superior Survival Compared to White Patients With Multiple Myeloma: A VA Study. *Blood* (2019) 133(24):2615–8. doi: 10.1182/blood.2019000406
33. David JM, Ho AS, Luu M, Yoshida EJ, Kim S, Mita AC, et al. Treatment at High-Volume Facilities and Academic Centers is Independently Associated With Improved Survival in Patients With Locally Advanced Head and Neck Cancer. *Cancer-Am Cancer Soc* (2017) 123(20):3933–42. doi: 10.1002/cncr.30843
34. Herndon JE, Kornblith AB, Holland JC, Paskett ED. Patient Education Level as a Predictor of Survival in Lung Cancer Clinical Trials. *J Clin Oncol* (2008) 26(25):4116–23. doi: 10.1200/JCO.2008.16.7460
35. Blackstock AW, Herndon JE2nd, Paskett ED, Perry MC, Graziano SL, Muscato JJ, et al. Outcomes Among African-American/non-African-American Patients With Advanced non-Small-Cell Lung Carcinoma: Report From the Cancer and Leukemia Group B. *J Natl Cancer Inst* (2002) 94(4):284–90. doi: 10.1093/jnci/94.4.284
36. Gray PJ, Lin CC, Cooperberg MR, Jemal A, Efstathiou JA. Temporal Trends and the Impact of Race, Insurance, and Socioeconomic Status in the Management of Localized Prostate Cancer. *Eur Urol* (2017) 71(5):729–37. doi: 10.1016/j.eururo.2016.08.047
37. Wong MCS, Lao XQ, Ho KF, Goggins WB, Tse SLA. Incidence and Mortality of Lung Cancer: Global Trends and Association With Socioeconomic Status. *Sci Rep* (2017) 7(1):14300. doi: 10.1038/s41598-017-14513-7
38. Dabbikeh A, Peng Y, Mackillop WJ, Booth CM, Zhang-Salomons J. Temporal Trends in the Association Between Socioeconomic Status and Cancer Survival in Ontario: A Population-Based Retrospective Study. *CMAJ Open* (2017) 5(3): E682–E9. doi: 10.9778/cmajo.20170025
39. Ignatius SH, Ziogas A, Zell JA. Prognostic Factors for Survival in Extensive Stage Small Cell Lung Cancer (ED-SCLC) The Importance of Smoking History, Socioeconomic and Marital Statuses, and Ethnicity. *J Thorac Oncol* (2009) 4(1):37–43. doi: 10.1097/JTO.0b013e31819140fb
40. Cavalli-Bjorkman N, Lambe M, Eaker S, Sandin F, Glimelius B. Differences According to Educational Level in the Management and Survival of Colorectal Cancer in Sweden. *Eur J Cancer* (2011) 47(9):1398–406. doi: 10.1016/j.ejca.2010.12.013
41. Herndon JE, Kornblith AB, Holland JC, Paskett ED. Effect of Socioeconomic Status as Measured by Education Level on Survival in Breast Cancer Clinical Trials. *Psychooncology* (2013) 22(2):315–23. doi: 10.1002/pon.2094
42. Osagiede O, Colibaseanu DT, Spaulding AC, Frank RD, Merchea A, Kelley SR, et al. Palliative Care Use Among Patients With Solid Cancer Tumors: A National Cancer Data Base Study. *J Palliat Care* (2018) 33(3):149–58. doi: 10.1177/0825859718777320
43. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2019. *CA Cancer J Clin* (2019) 69(1):7–34. doi: 10.3322/caac.21551

Conflict of Interest: YL: Advisory board: AstraZeneca, Novocure; Consultant: AstraZeneca; Honorarium: clarion health care; Research Funding Support: Merck, MacroGenics, Tolero Pharmaceuticals, AstraZeneca, Vaccinex, Blueprint Medicines, Harpoon Therapeutics, Sun Pharma Advanced Research, Bristol-Myers Squibb, Kyowa Pharmaceuticals, Tesaro, Bayer HealthCare. RM: Advisory board: AstraZeneca, Guardant Health, Novocure, Takeda; Consulting: AstraZeneca. AS is a consultant with Celgene, Takeda, Janssen, Amgen, AstraZeneca, Glaxo-Smithkline, and has received institutional research support from Pharmacyclics, Bristol Myers Squibb, Janssen, Amgen, Cellectar, Medimmune, and Ascentage.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Oncology

ORIGINAL RESEARCH
published: 04 January 2022
doi: 10.3389/fonc.2021.766939

Construction and Validation of a Lung Cancer Risk Prediction Model for Non-Smokers in China

*Lan-Wei Guo[1], Zhang-Yan Lyu[2], Qing-Cheng Meng[3], Li-Yang Zheng[1], Qiong Chen[1], Yin Liu[1], Hui-Fang Xu[1], Rui-Hua Kang[1], Lu-Yao Zhang[1], Xiao-Qin Cao[1], Shu-Zheng Liu[1], Xi-Bin Sun[1], Jian-Gong Zhang[1] and Shao-Kai Zhang[1]**

[1] Department of Cancer Epidemiology and Prevention, Henan Engineering Research Center of Cancer Prevention and Control, Henan International Joint Laboratory of Cancer Prevention, The Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China, [2] Department of Cancer Epidemiology and Biostatistics, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy of Tianjin, Tianjin's Clinical Research Center for Cancer, Key Laboratory of Molecular Cancer Epidemiology of Tianjin, Key Laboratory of Breast Cancer Prevention and Therapy of the Ministry of Education, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China, [3] Department of Radiology, The Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China

OPEN ACCESS

Edited by:
Yutong He,
Fourth Hospital of Hebei Medical University, China

Reviewed by:
Lei Yang,
Peking University Cancer Hospital, China
Xiaolong Tang,
Shandong University, China

***Correspondence:**
Shao-Kai Zhang
shaokaizhang@126.com

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *30 August 2021*
Accepted: *13 December 2021*
Published: *04 January 2022*

Citation:
Guo L-W, Lyu Z-Y, Meng Q-C, Zheng L-Y, Chen Q, Liu Y, Xu H-F, Kang R-H, Zhang L-Y, Cao X-Q, Liu S-Z, Sun X-B, Zhang J-G and Zhang S-K (2022) Construction and Validation of a Lung Cancer Risk Prediction Model for Non-Smokers in China. Front. Oncol. 11:766939. doi: 10.3389/fonc.2021.766939

Background: About 15% of lung cancers in men and 53% in women are not attributable to smoking worldwide. The aim was to develop and validate a simple and non-invasive model which could assess and stratify lung cancer risk in non-smokers in China.

Methods: A large-sample size, population-based study was conducted under the framework of the Cancer Screening Program in Urban China (CanSPUC). Data on the lung cancer screening in Henan province, China, from October 2013 to October 2019 were used and randomly divided into the training and validation sets. Related risk factors were identified through multivariable Cox regression analysis, followed by establishment of risk prediction nomogram. Discrimination [area under the curve (AUC)] and calibration were further performed to assess the validation of risk prediction nomogram in the training set, and then validated by the validation set.

Results: A total of 214,764 eligible subjects were included, with a mean age of 55.19 years. Subjects were randomly divided into the training (107,382) and validation (107,382) sets. Elder age, being male, a low education level, family history of lung cancer, history of tuberculosis, and without a history of hyperlipidemia were the independent risk factors for lung cancer. Using these six variables, we plotted 1-year, 3-year, and 5-year lung cancer risk prediction nomogram. The AUC was 0.753, 0.752, and 0.755 for the 1-, 3- and 5-year lung cancer risk in the training set, respectively. In the validation set, the model showed a moderate predictive discrimination, with the AUC was 0.668, 0.678, and 0.685 for the 1-, 3- and 5-year lung cancer risk.

Conclusions: We developed and validated a simple and non-invasive lung cancer risk model in non-smokers. This model can be applied to identify and triage patients at high risk for developing lung cancers in non-smokers.

Keywords: lung cancer, risk model, forecasting, validation, non-smokers

HIGHLIGHTS

1. Evidence before this study:
- About 15% of lung cancers in men and 53% in women are not attributable to smoking worldwide.
- Screening people at high risk for lung cancer by low-dose computed tomography (LDCT) has been approved to effective in reducing the burden of this disease.
- Developing lung-cancer risk prediction tools for Chinese non-smokers in large-scale population-based lung screening programs is sparse.

2. Added value of this study:
- Risk factors associated with lung cancer in Chinese non-smokers were identified.
- The model developed has moderate discriminatory accuracy and goodness-of-fit for both men and women, non-passive smokers and passive smokers.

3. Implications of all the available evidence:
- This model can be applied to identify and triage patients at high risk for developing lung cancer in non-smokers.
- The model has potential utility for shared decision making and individualized risk assessment for tailored lung cancer screening in Chinese non-smokers.

INTRODUCTION

Lung cancer is the leading cause of cancer related deaths in both the world and China. The latest data from the International Agency for Research on Cancer (IARC) shows that in 2020, there were about 1.80 million lung cancer deaths worldwide, which China accounts for 39.8% (1). The majority of lung cancer cases in China were found to be clinically advanced, with 64.6% of stage III-IV lung cancers in 2012-2014 (2). The age standardized 5-year survival rate of lung cancer in China increased slightly between 2003 and 2015, but still did not exceed 20.0% (3). The prognosis of lung cancer is closely related to the diagnostic stage, and the 5-year survival rate after surgery is almost 0 for stage IV patients, but >80% for stage I lung cancer patients (4).

The results of the National Lung Screening Trial (NLST), initiated in 2002, suggested that low-dose computed tomography (LDCT) screening could reduce lung cancer mortality by 20% (5). However, this project only screened people at high risk for lung cancer based on age and smoking history (55-74 years, smoked no less than 30 pack-years, and had no more than 15 years of smoking quit time). It is well known that smoking significantly increases the risk of lung cancer. Meta-analysis showed that the risk of lung cancer was 13.1 times higher among smokers than non-smokers in Europe and the United States [Hazard Ratio (HR)=13.1, 95% CI= 9.9-17.3] (6), much higher than the 2.77 times risk in the Chinese population [Odds Ratio (OR)=2.77, 95% CI=2.26-3.40] (7). This suggests that the current international standards for lung cancer screening based on smoking as the main indicator for high-risk populations may not be suitable for the Chinese population, especially for Chinese non-smokers. Therefore, how to effectively predict the risk of lung cancer in non-smokers and then guide the more cost-effective LDCT screening is an effective way to achieve efficient early diagnosis and treatment of lung cancer.

Previous studies have constructed several lung cancer risk prediction models based on different characteristics of populations (8–38), but there is few lung cancer risk prediction models based on non-smokers in mainland of China. To this end, developing lung-cancer risk prediction tools for Chinese non-smokers based on risk factors consistently identified in previous studies becomes a priority (39). However, this is difficult and challenging. Unlike the situation of tobacco-driven lung cancer, there is no established risk factors dominating the development of lung cancer among non-smokers. Numerous risk factors have been suggested and their effects vary greatly by geographical region (40–43). For example, we note that the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) models do not seem to be useful for Asian non-smokers because PLCO only included about 2000 never-smokers of Asian ethnicity, of which 7 cases of lung cancer occurred (44). Indeed, none of the non-smokers in the PLCO (n=65,711) had a six-year risk >0.0151, using the $PLCO_{M2014}$ that is analogous to $PLCO_{M2012}$ and included non-smokers.

The model was developed based on the Cancer Screening Program in Urban China (CanSPUC) (45). With the focus on established risk factors for lung cancer routinely available in general cancer screening settings, we aimed to develop and internally validated a risk prediction model for lung cancer in Chinese non-smokers.

METHODS

Data Source and Subjects

This study was conducted within the framework of CanSPUC, an ongoing, nationwide, population-based cancer screening program in urban China. The purpose of CanSPUC is to screen five most prevalent cancers, including lung cancer, female breast cancer, liver cancer, upper gastrointestinal cancer, and colorectal cancer. The methodology of the CanSPUC has been previously described (45, 46). In brief, after signing a written informed consent, all eligible participants (40-74 years old) were interviewed by trained staffs to collect data on their exposure to risk factors and to evaluate their cancer risk using a defined clinical cancer risk score system. CanSPUC was launched in Henan province of China in October 2013, covering eight cities (Zhengzhou, Zhumadian, Anyang, Luoyang, Nanyang, Jiaozuo, Puyang, and Xinxiang). In this study, we used data from the first six years (from October 2013 to October 2019) in Henan province. Only those non-smokers (except former smokers) were included in this study. Subjects would be excluded if they have been already diagnosed with lung cancer.

Outcome, Variables and Measurements

All new cases of lung cancer in the study were ascertained through local cancer registry databases with a histologically confirmed diagnosis from October 1, 2013 to March 10, 2020

in mainland of China. Newly diagnosed lung cancers were classified by sites according to International Classification of Diseases, 10th version (ICD-10). Lung cancers were identified by ICD-10 of C33-C34. To identify potential risk factors for lung cancer, the following data were collected by self-report:

(1) Demographic characteristics: including age, gender, race, height, weight and level of education. A low education was defined as primary school or below, medium education was defined junior or senior high school, and high-level education was defined as undergraduate or over. Body mass index (BMI) was calculated according to height and weight, and classified as "<18.5 kg/m^2", "18.5-23.9 kg/m^2", "24.0-27.9 kg/m^2"and "≥28.0 kg/m^2".

(2) Dietary habit: a) Dietary intake of the following food in the past two years: vegetables intake (<2.5kg/week, ≥2.5kg/week), fruit intake (<1.25kg/week, ≥1.25kg/week), roughage intake (<0.5kg/week, ≥0.5kg/week). Vegetables referred to green leaf plants and fungi, except for potato, sweet potato, and other starch. Roughage referred to the grains except white flour and rice. Food weight was determined before cooking. b) Taste preferences: heavy-salt diet (yes, no) and heavy-grease diet (yes, no).

(3) Living environment, behavior and habits: a) Cooking oil fume (COF) exposure: exposure is considered as "none or a little", if chimneys, fume extractors, or smoke-less pots was used during cooking; otherwise, it is considered as "a lot". b) Physical activity: activities were categorized as Taijiquan/Qigong/Walking, long distance running/aerobics, ball games (basketball, table tennis, badminton, etc.), fast walking/yangko dance, swimming and other physical exercises (such as mountain climbing, rope skipping, shuttlecock kicking). Subjects who did exercise for at least three days with a total time ≥90 mins per week were categorized as "heavy physical activity"; otherwise, were categorized as "moderate or no physical activity".

(4) Comorbidities: including history of chronic respiratory disease, tuberculosis, chronic bronchitis, emphysema, asthma bronchiectasis and hyperlipidemia. All self-reported comorbidities required a diagnosis from professional medical institutions.

(5) Family history of lung cancer: whether first-degree relatives, second-degree relatives or third-degree relatives had lung cancer or not.

Statistical Analysis

All statistical analyses were performed with the statistical software SAS version 9.4 (SAS Institute, Cary, NC) and R version 4.0.3 (The Free Software Foundation, Boston, MA, USA). The "rms" package was used to draw the nomogram. The "survivalROC" package was used to draw the ROC curves. The "ggplot2" package was used to draw the calibration curves. All tests were two-sided and p-values of 0.05 or less were considered to be statistically significant.

With the help of randomization codes produced by means of the PROC PLAN of the SAS system, the dataset was randomly divided into training set and validation set with a 1:1 assignment ratio. The training set was used to create the risk prediction model, while the validation set was used to validate the performance of the model.

Descriptive statistics, expressed as proportions for categorical variables, were used to compare the characteristics of those with and without the outcome of developing lung cancer. Chi-squared tests for categorical variables were used to determine the univariate association between the baseline factors and lung cancer development. Continuous variables were described by means (standard deviation) or median (interquartile range, IQR).

In this study, the combined model based on all independent prognostic factors selected by the stepwise multivariable Cox regression (P_{entry}=0.15, P_{stay}=0.10) was used to construct a nomogram to assess the 1-, 3-, and 5-year estimates of the lung cancer risk in the training set. The calibration curves were used to evaluate the validity of the nomogram. The Kaplan–Meier curves were plotted for low-, medium-, and high-risk groups using the 33% and 66% quantiles for lung cancer according to the risk prediction model, and differences among the three curves were tested according to the log-rank test. The prediction performance of the 1-, 3- and 5-year estimates of the lung cancer risk was quantified by receiver operating characteristic (ROC) curves and the area under the curve (AUC) in the training set and validation set. The bootstrap sampling approach was used to evaluate the calibration of the present model by comparing the observed and predicted probabilities. Correction for deviation of estimates from observations (overfitting correction) estimates were based on predictions for a subset of the interval. The median absolute error is also used to evaluate the calibration performance.

RESULTS

Characteristics of the Study Population

A total of 214,764 eligible subjects with a mean age of 55.19 years were included into this study, and 70.70% were females. Subjects were randomly divided into the training set (107,382 subjects) and validation set (107,382 subjects) (**Figure 1**). By March 2020, among 214,764 eligible participants, 344 lung cancer cases occurred in the follow-up yielding an incident density of 50.53/100,000 person-years. Compared with participants without lung cancer, lung cancer cases were more likely to have a low education, without passive smoking exposure, have a heavy physical activity and have a family history of lung cancer (all *P* vales <0.05). Additional characteristics are presented in **Supplementary Table 1** and **Table 1**.

Development of the Lung Cancer Risk Assessment Model

Table 2 presents the HRs (95% CI) for each predictor. In the training set, age (≥55 years: 3.68,1.60-8.43; ≥60 years: 5.51, 2.48-12.26; ≥65 years: 7.62, 3.43-16.92; ≥70 years: 9.03, 3.79-21.54), gender (male: 2.07, 1.53-2.79), education (low: 1.87, 1.05-3.33;

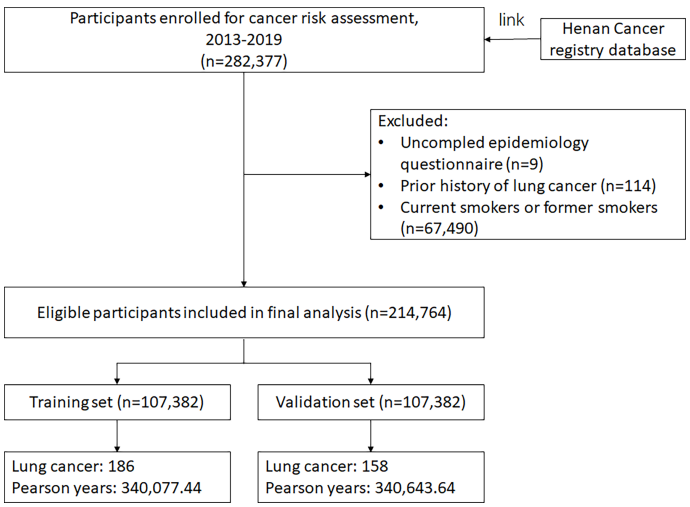


FIGURE 1 | Flow chart of participants included in this analysis.

medium: 1.36, 0.81-2.31), family history of lung cancer (2.00, 1.25-3.20), history of tuberculosis (2.16, 0.87-5.37), and history of hyperlipidemia (0.61, 0.40-0.95) were independent risk factors of lung cancer. Thus, we used these variables to build the model. We plotted 1-year, 3-year, and 5-year lung cancer risk prediction nomogram (**Figure 2A**).

Predictive Performance of the Model

The risk predictions were stratified into low-, medium-, and high-risk groups and visualized by Kaplan–Meier curves, showing statistically significant differences between the groups by a log-rank test (**Figure 2B**, $P<0.001$).

Using this model, the AUC was 0.753, 0.752, and 0.755 for 1-year, 3-year, and 5-year lung cancer risk in the training set, respectively. Stratified analysis by gender showed that the AUC of the model was higher among men (1-year: 0.776, 3-year: 0.780, and 5-year: 0.816) than women (1-year: 0.724, 3-year: 0.707, and 5-year: 0.694). Stratified analysis by age showed that the AUC of the model was higher among younger participants (<60 years) (1-year: 0.740, 3-year: 0.705, and 5-year: 0.664) than elder participants (≥60 years) (1-year: 0.628, 3-year: 0.648, and 5-year: 0.661). When examined by passive smoking status, the model yielded higher AUC for non-passive smokers (1-year: 0.762, 3-year: 0.756, and 5-year: 0.757) than passive smokers (1-year: 0.711, 3-year: 0.726, and 5-year: 0.738) (**Figure 3**). Calibration was satisfactory, with observed risks awfully close to the predicted risks (**Figure 4**).

Validation of the Lung Cancer Risk Model

The model showed a moderate predictive discrimination in the validation set, with the AUC was 0.668, 0.678, and 0.685 for 1-year, 3-year, and 5-year lung cancer risk (**Supplementary Figure 1**) and the satisfactory calibration of relative risk (**Supplementary Figure 2**).

DISCUSSION

In this study, using data from a large perspective lung cancer screening cohort studies, we developed and internally validated a simple risk prediction model for lung cancer in non-smokers, based on six widely available variables, including demographics (age, gender, education), comorbidities (tuberculosis, hyperlipidemia) and family history of lung cancer. Our results showed that the model has good discriminatory accuracy and goodness-of-fit for both men and women, non-passive smokers and passive smokers.

For non-smokers, several risk factors for lung cancer have been identified, including passive smoking (47, 48), previous lung diseases [tuberculosis, chronic bronchitis, emphysema, previous lung diseases (COPD)] (49), indoor radon (50), cooking oil fumes (51) and family history of lung cancer (52). The risk factors for lung cancer identified in our study, such as age, gender, family history of lung cancer, history of tuberculosis, are consistent with the findings. The most dominant risk factors for lung cancer in non-smokers is age, and our study showed that elder age was the main risk factor for lung cancer and the risk was more than 9 times higher in age group of 70-74 years than in the age group of 40-44 years. Besides, being male remains a risk factor for lung cancer in non-smokers in our study, even though more than 50% lung cancers were non-smokers in women in Southeast Asia compared to approximately 2–6% in men in Western series (41, 42, 53). Just like other prediction models, such as Bach model (8), LLP (Liverpool Lung Project) model (10) and $PLCO_{M2012}$ model (54), education levels was included in our model as a protection factor.

Another important finding was that history of hyperlipidemia [increased total cholesterol (TC), or triglycerides (TG), or low-density lipoprotein cholesterol (LDL-C), or decreased high-density lipoprotein cholesterol (HDL-C)] exposure might

TABLE 1 | Baseline characteristics of the study population in the training set.

	Total no. (%)	No lung cancer, n (%)	Lung cancer, n (%)	χ^2	*P*-value
All participants	107382	107196	186		
Person-years, median(IQR)	2.95 (1.71-4.83)	2.95 (1.71-4.83)	1.47 (0.78-2.33)		
Age, mean ± SD, years	55.16 ± 8.78	55.14 ± 8.78	61.53 ± 7.47		
Age (years)				98.24	<0.001
40-44	14028 (13.06)	14021 (99.95)	7 (0.05)		
45-49	20109 (18.73)	20101 (99.96)	8 (0.04)		
50-54	19769 (18.41)	19750 (99.90)	19 (0.10)		
55-59	16083 (14.98)	16054 (99.82)	29 (0.18)		
60-64	17556 (16.35)	17508 (99.73)	48 (0.27)		
65-69	14356 (13.37)	14302 (99.62)	54 (0.38)		
70-74	5481 (5.10)	5460 (99.62)	21 (0.38)		
Gender				16.73	<0.001
Male	31531 (29.36)	31451 (99.75)	80 (0.25)		
Female	75851 (70.64)	75745 (99.86)	106 (0.14)		
Race				0.44	0.506
Han nationality	105549 (98.29)	105365 (99.83)	184 (0.17)		
Others	1833 (1.71)	1831 (99.89)	2 (0.11)		
Education[a]				14.34	0.001
Low	20139 (18.75)	20086 (99.74)	53 (0.26)		
Medium	71634 (66.71)	71517 (99.84)	117 (0.16)		
High	15609 (14.54)	15593 (99.90)	16 (0.10)		
BMI (kg/m^2)				2.82	0.419
<18.5	1381 (1.29)	1377 (99.71)	4 (0.29)		
18.5-23.9	47588 (44.32)	47498 (99.81)	90 (0.19)		
24.0-28.0	46882 (43.66)	46810 (99.85)	72 (0.15)		
≥28.0	11531 (10.74)	11511 (99.83)	20 (0.17)		
Vegetables intake				0.03	0.861
≥2.5kg/week	56467 (52.59)	56368 (99.82)	99 (0.18)		
<2.5kg/week	50915 (47.41)	50828 (99.83)	87 (0.17)		
Fruit intake				0.07	0.785
≥1.25kg/week	63026 (58.69)	62915 (99.82)	111 (0.18)		
<1.25kg/week	44356 (41.31)	44281 (99.83)	75 (0.17)		
Roughage intake				0.47	0.492
≥0.5kg/week	73524 (68.47)	73401 (99.83)	123 (0.17)		
<0.5kg/week	33858 (31.53)	33795 (99.81)	63 (0.19)		
Heavy-slat diet				1.57	0.209
No	88623 (82.53)	88463 (99.82)	160 (0.18)		
Yes	18759 (17.47)	18733 (99.86)	26 (0.14)		
Heavy-grease diet				0.07	0.793
No	89672 (83.51)	89518 (99.83)	154 (0.17)		
Yes	17710 (16.49)	17678 (99.82)	32 (0.18)		
Cooking oil fume exposure				0.03	0.853
None or a little	95144 (88.60)	94980 (99.83)	164 (0.17)		
A lot	12238 (11.40)	12216 (99.82)	22 (0.18)		
Passive smoking				4.65	0.031
No	76618 (71.35)	76472 (99.81)	146 (0.19)		
Yes	30764 (28.65)	30724 (99.87)	40 (0.13)		
Alcohol Drinking				0.46	0.794
Never	96004 (89.40)	95835 (99.82)	169 (0.18)		
Current	9598 (8.94)	9584 (99.85)	14 (0.15)		
Former	1780 (1.66)	1777 (99.83)	3 (0.17)		
Physical activity				3.38	0.066
Moderate or no	56293 (52.42)	56208 (99.85)	85 (0.15)		
Heavy	51089 (47.58)	50988 (99.80)	101 (0.20)		
Family history of lung cancer				4.69	0.030
No	99660 (92.81)	99495 (99.83)	165 (0.17)		
Yes	7722 (7.19)	7701 (99.73)	21 (0.27)		
History of chronic respiratory disease				1.93	0.165
No	93185 (86.78)	93030 (99.83)	155 (0.17)		
Yes	14197 (13.22)	14166 (99.78)	31 (0.22)		
History of tuberculosis				4.58	0.032
No	106230 (98.93)	106049 (99.83)	181 (0.17)		
Yes	1152 (1.07)	1147 (99.57)	5 (0.43)		

(Continued)

TABLE 1 | Continued

	Total no. (%)	No lung cancer, n (%)	Lung cancer, n (%)	χ^2	*P*-value
History of chronic bronchitis				1.55	0.213
No	96485 (89.85)	96323 (99.83)	162 (0.17)		
Yes	10897 (10.15)	10873 (99.78)	24 (0.22)		
History of emphysema				4.51	0.034
No	106543 (99.22)	106361 (99.83)	182 (0.17)		
Yes	839 (0.78)	835 (99.52)	4 (0.48)		
History of asthma bronchiectasis				0.01	0.925
No	104612 (97.42)	104431 (99.83)	181 (0.17)		
Yes	2770 (2.58)	2765 (99.82)	5 (0.18)		
History of hyperlipidemia				0.50	0.481
No	91561 (85.27)	91399 (99.82)	162 (0.18)		
Yes	15821 (14.73)	15797 (99.85)	24 (0.15)		

[a]*Low, primary school or below; Medium, junior or senior high school; High, undergraduate or over.*
IQR, Interquartile range; BMI, body mass index.

decrease the risk of lung cancer, despite a small effect. Since the 1980s, several epidemiological studies have investigated the associations of TC, TG, and HDL-C with lung cancer risk in non-smokers but have shown markedly contrasting results due to differences in the classification of smoking status, lack of prospective cohort study designs, relatively modest sample sizes and other potential bias (55–58). Lyu etc. (58) conducted a prospective cohort study among over 100 thousand Chinese males and found that both low and high TC levels, both low and high TG levels, and low LDL-C levels increased lung cancer risk in non-smokers. Besides, many studies reported an inverse relationship between TC (56, 59), LDL-C (60) and lung cancer incidence, to some extent, consistent with our findings. More epidemiologic, molecular and biochemical studies are needed to test this hypothesis.

In addition to credible predictors, a risk prediction model should also meet performance standards related to discrimination defined as the ability to distinguish lung cancer cases from controls, and calibration defined as the consistency between observed and predicted risk for lung cancer. The rapid increase in the number of lung cancer risk prediction model studies since 2010 reflects the current need for the use of predictive models to guide population splitting. Initially, models focused on the use of traditional epidemiological risk factors such as age, smoking history, personal history of disease and family history of cancer, such as the Bach model (8), Spitz model (9), LLP model (10) and $PLCO_{M2012}$ model

TABLE 2 | Multivariable Cox-regression prediction model of lung cancer risk in training set.

Variables	*β coefficient*	*se*	HR (95% *CI*)	χ^2	*P*
Age (years)					
40-44			1.00		
45-49	-0.24	0.52	0.78 (0.28-2.16)	0.22	0.637
50-54	0.67	0.44	1.96 (0.82-4.66)	2.29	0.130
55-59	1.30	0.42	3.68 (1.60-8.43)	9.47	0.002
60-64	1.71	0.41	5.51 (2.48-12.26)	17.53	<0.001
65-69	2.03	0.41	7.62 (3.43-16.92)	24.92	<0.001
70-74	2.20	0.44	9.03 (3.79-21.54)	24.65	<0.001
Gender					
Male	0.73	0.15	2.07 (1.53-2.79)	22.63	<0.001
Female			1.00		
Education[a]					
Low	0.62	0.30	1.87 (1.05-3.33)	4.45	0.035
Medium	0.31	0.27	1.36 (0.81-2.31)	1.33	0.249
High			1.00		
Family history of lung cancer					
No			1.00		
Yes	0.69	0.24	2.00 (1.25-3.20)	8.24	0.004
History of tuberculosis					
No			1.00		
Yes	0.77	0.46	2.16 (0.87-5.37)	2.75	0.097
History of hyperlipidemia					
No			1.00		
Yes	-0.49	0.22	0.61 (0.40-0.95)	4.80	0.028

[a]*Low, primary school or below; Medium, junior or senior high school; High, undergraduate or over.*

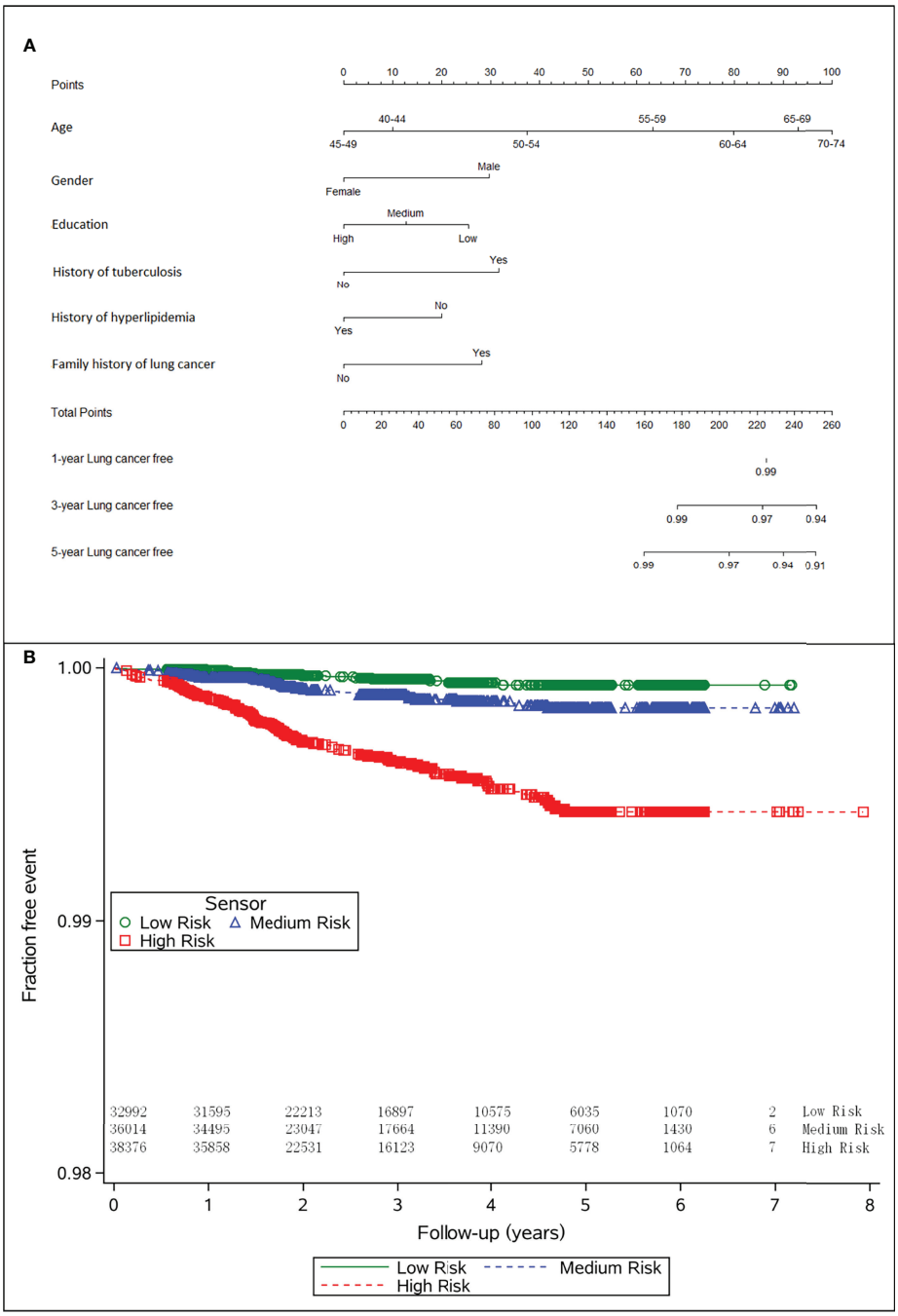


FIGURE 2 | **(A)** Nomogram to calculate the personal 1-, 3- and 5-year risk of lung cancer risk, and **(B)** the lung cancer incidence across different cancer risk categories.

(54). To our knowledge, the present study is one of the few studies to model lung cancer risk prediction among non-smoking men and women in mainland China. It is hard to directly compare the discriminatory performance of risk prediction models as each was developed in different populations with varying baseline risks or lengths of follow-up time. Nevertheless, each of the models' discriminative ability was relatively similar, with C-statistics ranges from 0.72 to 0.86. Our model showed comparable predictive performance compared with previous studies.

Specific strengths and limitations deserve careful attention when interpreting our results. A major strength of our study is the fact that our analyses were based on a large-scale population-based cancer screening program in mainland of China. Furthermore, the variables included in this model could be

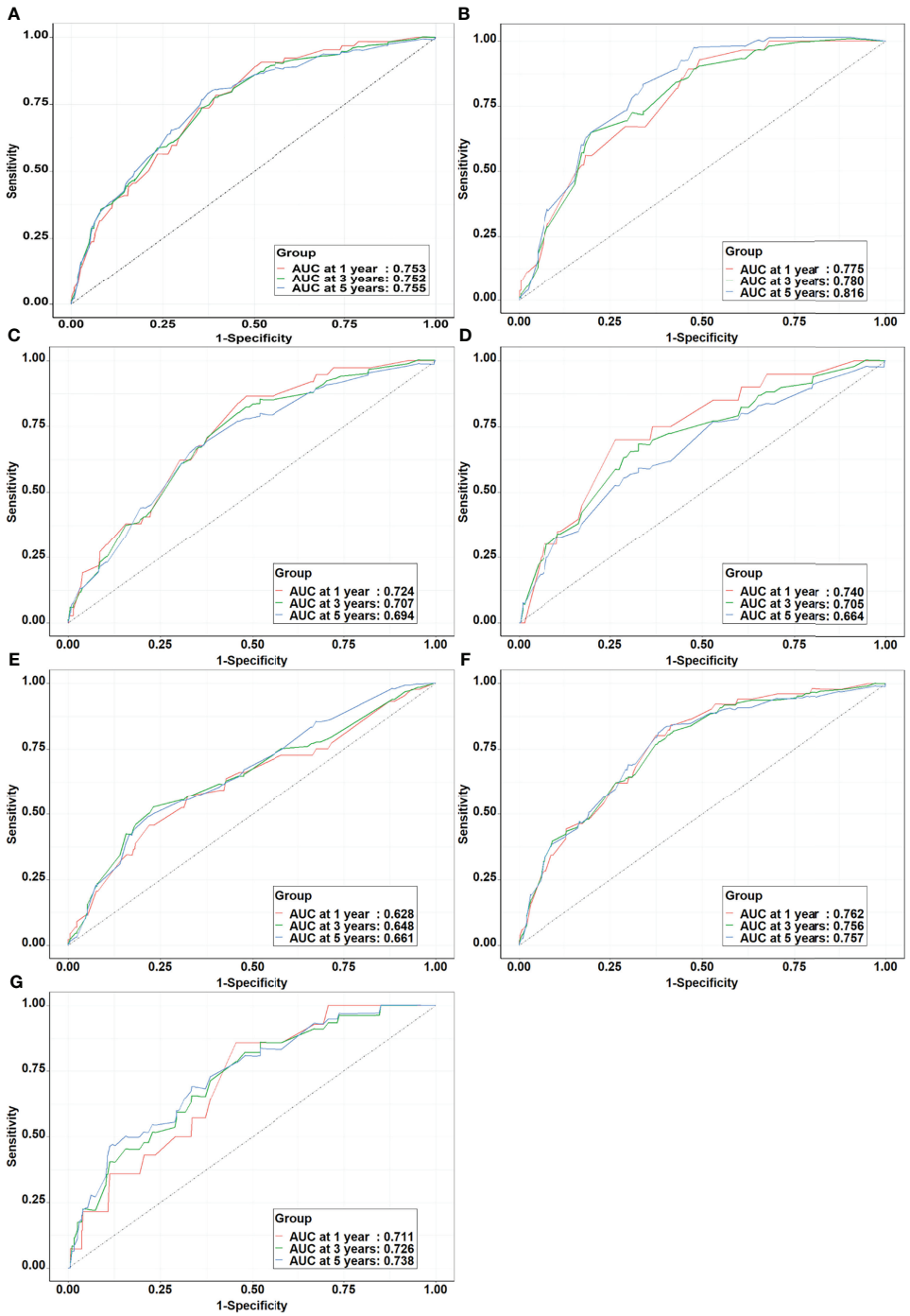


FIGURE 3 | The receiver operating characteristic curves of prediction models in the training set. **(A)** Whole population; **(B)** Male; **(C)** Female; **(D)** Age<60 years; **(E)** Age≥60 years; **(F)** Non-passive smokers; **(G)** Passive smokers.

easily collected and updated without any imaging, sophisticated testing or calculation. Moreover, the model will not only be used as a practical tool to triage high risk patients in non-smokers, but also have implications for public health measures, such as guidelines for the prevention of lung cancer in non-smokers. However, limitations include that the self-report data might subject to social desirability and recall bias. However, given the good data acquisition and quality control, most information is believed to be reliable. Secondly, the performance of our risk prediction model was not validated on an external dataset.

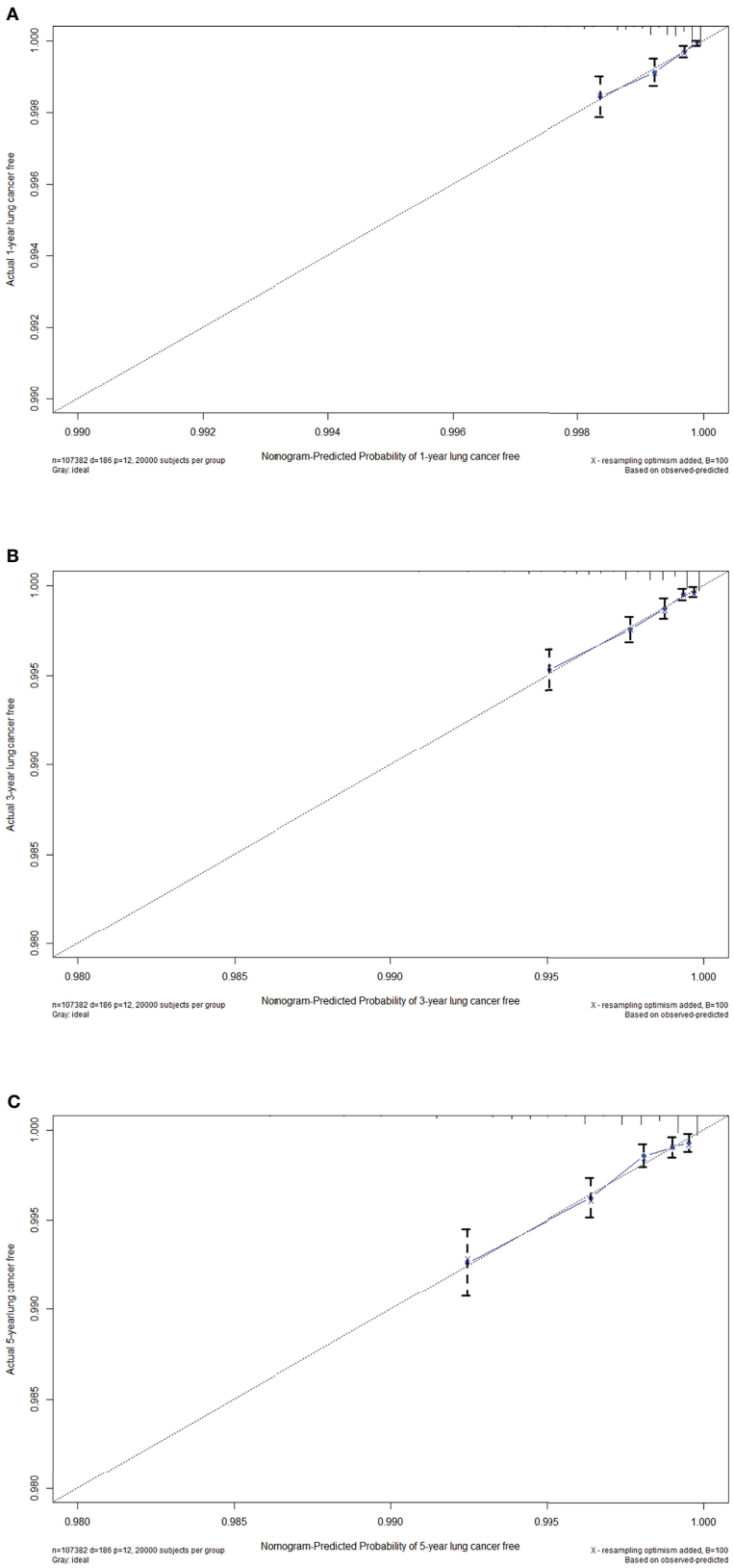


FIGURE 4 | Calibration curves of the nomogram for **(A)** 1-year, **(B)** 3-year and **(C)** 5-year lung cancer free in the training set.

However, the results of the internal validation suggest promisingly that this model will obtain well performance when applied to other populations.

CONCLUSIONS

In summary, we developed and internally validated a simple risk prediction model for lung cancer in non-smokers based on a large-scale lung cancer screening program in China. The model has good discrimination and could be used as a tool for triaging high-risk patients to prevent lung cancer in non-smokers. Further prospective studies are required to validate the model in external populations.

DATA AVAILABILITY STATEMENT

The datasets for this manuscript are not publicly available because all our data are under regulation of both the National Cancer Center of China and Henan Cancer Hospital. Requests to access the datasets should be directed to Shaokai Zhang, shaokaizhang@126.com.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Henan Cancer Hospital (no. 2021-KY-0028-001). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Conception and design: L-WG, Z-YL, J-GZ and S-KZ. Statistical Analyses: L-WG and L-YZ. Data acquisition and data interpretation: L-WG, Q-CM, L-YZ, QC, YL, H-FX, R-HK, L-YZ, X-QC, S-ZL, X-BS and S-KZ. Drafting of the article: L-WG. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Natural Science Foundation of Henan Province (No. 212300410261), the Henan Province Medical Science and Technology Tackling Program (SBGJ202001004), the National Key R&D Program of China (2018YFC1315000/ 2018YFC1315001) and the CAMS Innovation Fund for Medical Science (2019-I2M-2-002).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.766939/full#supplementary-material

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
2. Shi JF, Wang L, Wu N, Li JL, Hui ZG, Liu SM, et al. Clinical Characteristics and Medical Service Utilization of Lung Cancer in Chin-2014: Overall Design and Results From a Multicenter Retrospective Epidemiologic Survey. *Lung Cancer* (2019) 128:91–100. doi: 10.1016/j.lungcan.2018.11.031
3. Zeng H, Chen W, Zheng R, Zhang S, Ji JS, Zou X, et al. Changing Cancer Survival in China During 2003-15: A Pooled Analysis of 17 Population-Based Cancer Registries. *Lancet Glob Health* (2018) 6(5):e555–67. doi: 10.1016/s2214-109x(18)30127-x
4. Detterbeck FC, Boffa DJ, Kim AW, Tanoue LT. The Eighth Edition Lung Cancer Stage Classification. *Chest* (2017) 151(1):193–203. doi: 10.1016/j.chest.2016.10.010
5. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced Lung-Cancer Mortality With Low-Dose Computed Tomographic Screening. *N Engl J Med* (2011) 365(5):395–409. doi: 10.1056/NEJMoa1102873
6. Ordóñez-Mena JM, Schöttker B, Mons U, Jenab M, Freisling H, Bueno-de-Mesquita B, et al. Quantification of the Smoking-Associated Cancer Risk With Rate Advancement Periods: Meta-Analysis of Individual Participant Data From Cohorts of the CHANCES Consortium. *BMC Med* (2016) 14:62. doi: 10.1186/s12916-016-0607-5
7. He J, Li N, Chen WQ, Wu N, Shen HB, Jiang Y, et al. [China Guideline for the Screening and Early Detection of Lung Cancer(2021, Beijing)]. *Zhonghua Zhong Liu Za Zhi* (2021) 43(3):243–68. doi: 10.3760/cma.j.cn112152-20210119-00060
8. Bach PB, Kattan MW, Thornquist MD, Kris MG, Tate RC, Barnett MJ, et al. Variations in Lung Cancer Risk Among Smokers. *J Natl Cancer Inst* (2003) 95 (6):470–8. doi: 10.1093/jnci/95.6.470
9. Spitz MR, Hong WK, Amos CI, Wu X, Schabath MB, Dong Q, et al. A Risk Model for Prediction of Lung Cancer. *J Natl Cancer Inst* (2007) 99(9):715–26. doi: 10.1093/jnci/djk153
10. Cassidy A, Myles JP, van Tongeren M, Page RD, Liloglou T, Duffy SW, et al. The LLP Risk Model: An Individual Risk Prediction Model for Lung Cancer. *Br J Cancer* (2008) 98(2):270–6. doi: 10.1038/sj.bjc.6604158
11. Etzel CJ, Kachroo S, Liu M, D'Amelio A, Dong Q, Cote ML, et al. Development and Validation of a Lung Cancer Risk Prediction Model for African-Americans. *Cancer Prev Res (Phila)* (2008) 1(4):255–65. doi: 10.1158/1940-6207.capr-08-0082
12. Spitz MR, Etzel CJ, Dong Q, Amos CI, Wei Q, Wu X, et al. An Expanded Risk Prediction Model for Lung Cancer. *Cancer Prev Res (Phila)* (2008) 1(4):250–4. doi: 10.1158/1940-6207.capr-08-0060
13. Young RP, Hopkins RJ, Hay BA, Epton MJ, Mills GD, Black PN, et al. Lung Cancer Susceptibility Model Based on Age, Family History and Genetic Variants. *PloS One* (2009) 4(4):e5302. doi: 10.1371/journal.pone.0005302
14. D'Amelio AM Jr, Cassidy A, Asomaning K, Raji OY, Duffy SW, Field JK, et al. Comparison of Discriminatory Power and Accuracy of Three Lung Cancer Risk Models. *Br J Cancer* (2010) 103(3):423–9. doi: 10.1038/sj.bjc.6605759
15. Raji OY, Agbaje OF, Duffy SW, Cassidy A, Field JK. Incorporation of a Genetic Factor Into an Epidemiologic Model for Prediction of Individual Risk of Lung Cancer: The Liverpool Lung Project. *Cancer Prev Res (Phila)* (2010) 3 (5):664–9. doi: 10.1158/1940-6207.capr-09-0141
16. Maisonneuve P, Bagnardi V, Bellomi M, Spaggiari L, Pelosi G, Rampinelli C, et al. Lung Cancer Risk Prediction to Select Smokers for Screening CT–a Model Based on the Italian COSMOS Trial. *Cancer Prev Res (Phila)* (2011) 4 (11):1778–89. doi: 10.1158/1940-6207.capr-11-0026

17. Tammemagi CM, Pinsky PF, Caporaso NE, Kvale PA, Hocking WG, Church TR, et al. Lung Cancer Risk Prediction: Prostate, Lung, Colorectal And Ovarian Cancer Screening Trial Models and Validation. *J Natl Cancer Inst* (2011) 103(13):1058–68. doi: 10.1093/jnci/djr173
18. Tammemagi MC, Lam SC, McWilliams AM, Sin DD. Incremental Value of Pulmonary Function and Sputum DNA Image Cytometry in Lung Cancer Risk Prediction. *Cancer Prev Res (Phila)* (2011) 4(4):552–61. doi: 10.1158/1940-6207.capr-10-0183
19. Hoggart C, Brennan P, Tjonneland A, Vogel U, Overvad K, Ostergaard JN, et al. A Risk Model for Lung Cancer Incidence. *Cancer Prev Res (Phila)* (2012) 5(6):834–46. doi: 10.1158/1940-6207.capr-11-0237
20. Li H, Yang L, Zhao X, Wang J, Qian J, Chen H, et al. Prediction of Lung Cancer Risk in a Chinese Population Using a Multifactorial Genetic Model. *BMC Med Genet* (2012) 13:118. doi: 10.1186/1471-2350-13-118
21. Raji OY, Duffy SW, Agbaje OF, Baker SG, Christiani DC, Cassidy A, et al. Predictive Accuracy of the Liverpool Lung Project Risk Model for Stratifying Patients for Computed Tomography Screening for Lung Cancer: A Case-Control and Cohort Validation Study. *Ann Intern Med* (2012) 157(4):242–50. doi: 10.7326/0003-4819-157-4-201208210-00004
22. Park S, Nam BH, Yang HR, Lee JA, Lim H, Han JT, et al. Individualized Risk Prediction Model for Lung Cancer in Korean Men. *PloS One* (2013) 8(2): e54823. doi: 10.1371/journal.pone.0054823
23. Spitz MR, Amos CI, Land S, Wu X, Dong Q, Wenzlaff AS, et al. Role of Selected Genetic Variants in Lung Cancer Risk in African Americans. *J Thorac Oncol* (2013) 8(4):391–7. doi: 10.1097/JTO.0b013e318283da29
24. Tammemagi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, et al. Selection Criteria for Lung-Cancer Screening.[Erratum Appears in N Engl J Med. 2013 Jul 25;369(4):394]. *New Engl J Med* (2013) 368(8):728–36. doi: 10.1056/NEJMoa1211776
25. Veronesi G, Maisonneuve P, Rampinelli C, Bertolotti R, Petrella F, Spaggiari L, et al. Computed Tomography Screening for Lung Cancer: Results of Ten Years of Annual Screening and Validation of Cosmos Prediction Model. *Lung Cancer* (2013) 82(3):426–30. doi: 10.1016/j.lungcan.2013.08.026
26. El-Zein RA, Lopez MS, D'Amelio AM Jr, Liu M, Munden RF, Christiani D, et al. The Cytokinesis-Blocked Micronucleus Assay as a Strong Predictor of Lung Cancer: Extension of a Lung Cancer Risk Prediction Model. *Cancer Epidemiol Biomarkers Prev* (2014) 23(11):2462–70. doi: 10.1158/1055-9965.epi-14-0462
27. Li K, Huesing A, Sookthai D, Bergmann M, Boeing H, Becker N, et al. Selecting High-Risk Individuals for Lung Cancer Screening: A Prospective Evaluation of Existing Risk Models and Eligibility Criteria in the German EPIC Cohort. *Cancer Prev Res* (2015) 8(9):777–85. doi: 10.1158/1940-6207.CAPR-14-0424
28. Marcus MW, Chen Y, Raji OY, Duffy SW, Field JK. LLPi: Liverpool Lung Project Risk Prediction Model for Lung Cancer Incidence. *Cancer Prev Res (Phila)* (2015) 8(6):570–5. doi: 10.1158/1940-6207.capr-14-0438
29. Wang X, Ma K, Cui J, Chen X, Jin L, Li W. An Individual Risk Prediction Model for Lung Cancer Based on a Study in a Chinese Population. *Tumori J January/February* (2015) 101(1):16–23. doi: 10.5301/tj.5000205
30. Marcus MW, Raji OY, Duffy SW, Young RP, Hopkins RJ, Field JK. Incorporating Epistasis Interaction of Genetic Susceptibility Single Nucleotide Polymorphisms in a Lung Cancer Risk Prediction Model. *Int J Oncol* (2016) 49(1):361–70. doi: 10.3892/ijo.2016.3499
31. Wang X, Ma K, Chi L, Cui J, Jin L, Hu J-F, et al. Combining Telomerase Reverse Transcriptase Genetic Variant Rs2736100 With Epidemiologic Factors in the Prediction of Lung Cancer Susceptibility. *J Of Cancer* (2016) 7(7):846–53. doi: 10.7150/jca.13437
32. Wu X, Wen CP, Ye Y, Tsai M, Wen C, Roth JA, et al. Personalized Risk Assessment in Never, Light, and Heavy Smokers in a Prospective Cohort in Taiwan. *Sci Rep* (2016) 6:36482. doi: 10.1038/srep36482
33. Muller DC, Johansson M, Brennan P. Lung Cancer Risk Prediction Model Incorporating Lung Function: Development and Validation in the UK Biobank Prospective Cohort Study. *J Clin Oncol* (2017) 35(8):861–9. doi: 10.1200/jco.2016.69.2467
34. Weber M, Yap S, Goldsbury D, Manners D, Tammemagi M, Marshall H, et al. Identifying High Risk Individuals for Targeted Lung Cancer Screening: Independent Validation of the PLCOm2012 Risk Prediction Tool. *Int J Cancer* (2017) 141(2):242–53. doi: 10.1002/ijc.30673
35. Charvat H, Sasazuki S, Shimazu T, Budhathoki S, Inoue M, Iwasaki M, et al. Development of a Risk Prediction Model for Lung Cancer: The Japan Public Health Center-Based Prospective Study. *Cancer Sci* (2018) 109(3):854–62. doi: 10.1111/cas.13509
36. Katki HA, Kovalchik SA, Petito LC, Cheung LC, Jacobs E, Jemal A, et al. Implications of Nine Risk Prediction Models for Selecting Ever-Smokers for Computed Tomography Lung Cancer Screening. *Ann Intern Med* (2018) 169 (1):10–9. doi: 10.7326/m17-2701
37. Markaki M, Tsamardinos I, Langhammer A, Lagani V, Hveem K, Røe OD. A Validated Clinical Risk Prediction Model for Lung Cancer in Smokers of All Ages and Exposure Types: A HUNT Study. *EBioMedicine* (2018) 31:36–46. doi: 10.1016/j.ebiom.2018.03.027
38. Lyu Z, Li N, Chen S, Wang G, Tan F, Feng X, et al. Risk Prediction Model for Lung Cancer Incorporating Metabolic Markers: Development and Internal Validation in a Chinese Population. *Cancer Med* (2020) 9(11):3983–94. doi: 10.1002/cam4.3025
39. Lam S. Lung Cancer Screening in Never-Smokers. *J Thorac Oncol* (2019) 14 (3):336–7. doi: 10.1016/j.jtho.2018.12.019
40. Toh CK, Gao F, Lim WT, Leong SS, Fong KW, Yap SP, et al. Never-Smokers With Lung Cancer: Epidemiologic Evidence of a Distinct Disease Entity. *J Clin Oncol* (2006) 24(15):2245–51. doi: 10.1200/jco.2005.04.8033
41. Sun S, Schiller JH, Gazdar AF. Lung Cancer in Never Smokers–a Different Disease. *Nat Rev Cancer* (2007) 7(10):778–90. doi: 10.1038/nrc2190
42. Couraud S, Zalcman G, Milleron B, Morin F, Souquet PJ. Lung Cancer in Never Smokers–a Review. *Eur J Cancer* (2012) 48(9):1299–311. doi: 10.1016/j.ejca.2012.03.007
43. Sisti J, Boffetta P. What Proportion of Lung Cancer in Never-Smokers can be Attributed to Known Risk Factors? *Int J Cancer* (2012) 131(2):265–75. doi: 10.1002/ijc.27477
44. Tammemägi MC, Church TR, Hocking WG, Silvestri GA, Kvale PA, Riley TL, et al. Evaluation of the Lung Cancer Risks at Which to Screen Ever- and Never-Smokers: Screening Rules Applied to the PLCO and NLST Cohorts. *PloS Med* (2014) 11(12):e1001764. doi: 10.1371/journal.pmed.1001764
45. Guo LW, Chen Q, Shen YC, Meng QC, Zheng LY, Wu Y, et al. Evaluation of a Low-Dose Computed Tomography Lung Cancer Screening Program in Henan, China. *JAMA Netw Open* (2020) 3(11):e2019039. doi: 10.1001/jamanetworkopen.2020.19039
46. Guo L, Zhang S, Liu S, Zheng L, Chen Q, Cao X, et al. Determinants of Participation and Detection Rate of Upper Gastrointestinal Cancer From Population-Based Screening Program in China. *Cancer Med* (2019) 8 (16):7098–107. doi: 10.1002/cam4.2578
47. Hackshaw AK, Law MR, Wald NJ. The Accumulated Evidence on Lung Cancer and Environmental Tobacco Smoke. *Bmj* (1997) 315(7114):980–8. doi: 10.1136/bmj.315.7114.980
48. Stayner L, Bena J, Sasco AJ, Smith R, Steenland K, Kreuzer M, et al. Lung Cancer Risk and Workplace Exposure to Environmental Tobacco Smoke. *Am J Public Health* (2007) 97(3):545–51. doi: 10.2105/ajph.2004.061275
49. Brenner DR, McLaughlin JR, Hung RJ. Previous Lung Diseases and Lung Cancer Risk: A Systematic Review and Meta-Analysis. *PloS One* (2011) 6(3): e17479. doi: 10.1371/journal.pone.0017479
50. Darby S, Hill D, Auvinen A, Barros-Dios JM, Baysson H, Bochicchio F, et al. Radon in Homes and Risk of Lung Cancer: Collaborative Analysis of Individual Data From 13 European Case-Control Studies. *Bmj* (2005) 330 (7485):223. doi: 10.1136/bmj.38308.477650.63
51. Zhao Y, Wang S, Aunan K, Seip HM, Hao J. Air Pollution and Lung Cancer Risks in China–a Meta-Analysis. *Sci Total Environ* (2006) 366(2-3):500–13. doi: 10.1016/j.scitotenv.2005.10.010
52. Lissowska J, Foretova L, Dabek J, Zaridze D, Szeszenia-Dabrowska N, Rudnai P, et al. Family History and Lung Cancer Risk: International Multicentre Case-Control Study in Eastern and Central Europe and Meta-Analyses. *Cancer Causes Control* (2010) 21(7):1091–104. doi: 10.1007/s10552-010-9537-2
53. Scagliotti GV, Longo M, Novello S. Nonsmall Cell Lung Cancer in Never Smokers. *Curr Opin Oncol* (2009) 21(2):99–104. doi: 10.1097/CCO.0b013e328321049e
54. Tammemägi MC, Katki HA, Hocking WG, Church TR, Caporaso N, Kvale PA, et al. Selection Criteria for Lung-Cancer Screening. *N Engl J Med* (2013) 368(8):728–36. doi: 10.1056/NEJMoa1211776
55. Ulmer H, Borena W, Rapp K, Klenk J, Strasak A, Diem G, et al. Serum Triglyceride Concentrations and Cancer Risk in a Large Cohort Study in Austria. *Br J Cancer* (2009) 101(7):1202–6. doi: 10.1038/sj.bjc.6605264

56. Kitahara CM, Berrington de González A, Freedman ND, Huxley R, Mok Y, Jee SH, et al. Total Cholesterol and Cancer Risk in a Large Prospective Study in Korea. *J Clin Oncol* (2011) 29(12):1592–8. doi: 10.1200/jco.2010.31.5200
57. Chandler PD, Song Y, Lin J, Zhang S, Sesso HD, Mora S, et al. Lipid Biomarkers and Long-Term Risk of Cancer in the Women's Health Study. *Am J Clin Nutr* (2016) 103(6):1397–407. doi: 10.3945/ajcn.115.124321
58. Lyu Z, Li N, Wang G, Feng X, Chen S, Su K, et al. Independent and Joint Associations of Blood Lipids and Lipoproteins With Lung Cancer Risk in Chinese Males: A Prospective Cohort Study. *Int J Cancer* (2019) 144 (12):2972–84. doi: 10.1002/ijc.32051
59. Ahn J, Lim U, Weinstein SJ, Schatzkin A, Hayes RB, Virtamo J, et al. Prediagnostic Total and High-Density Lipoprotein Cholesterol and Risk of Cancer. *Cancer Epidemiol Biomarkers Prev* (2009) 18(11):2814–21. doi: 10.1158/1055-9965.Epi-08-1248
60. Benn M, Tybjærg-Hansen A, Stender S, Frikke-Schmidt R, Nordestgaard BG. Low-Density Lipoprotein Cholesterol and the Risk of Cancer: A Mendelian Randomization Study. *J Natl Cancer Inst* (2011) 103(6):508–19. doi: 10.1093/jnci/djr008

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

CORRECTION
published: 03 March 2022
doi: 10.3389/fonc.2022.871848

Corrigendum: Construction and Validation of a Lung Cancer Risk Prediction Model for Non-Smokers in China

*Lan-Wei Guo[1], Zhang-Yan Lyu[2], Qing-Cheng Meng[3], Li-Yang Zheng[1], Qiong Chen[1], Yin Liu[1], Hui-Fang Xu[1], Rui-Hua Kang[1], Lu-Yao Zhang[1], Xiao-Qin Cao[1], Shu-Zheng Liu[1], Xi-Bin Sun[1], Jian-Gong Zhang[1] and Shao-Kai Zhang[1]**

[1] Department of Cancer Epidemiology and Prevention, Henan Engineering Research Center of Cancer Prevention and Control, Henan International Joint Laboratory of Cancer Prevention, The Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China, [2] Department of Cancer Epidemiology and Biostatistics, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy of Tianjin, Tianjin's Clinical Research Center for Cancer, Key Laboratory of Molecular Cancer Epidemiology of Tianjin, Key Laboratory of Breast Cancer Prevention and Therapy of the Ministry of Education, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China, [3] Department of Radiology, The Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China

Keywords: lung cancer, risk model, forecasting, validation, non-smokers

A Corrigendum on:

Construction and Validation of a Lung Cancer Risk Prediction Model for Non-Smokers in China
by Guo L-W, Lyu Z-Y, Meng Q-C, Zheng L-Y, Chen Q, Liu Y, Xu H-F, Kang R-H, Zhang L-Y, Cao X-Q, Liu S-Z, Sun X-B, Zhang J-G and Zhang S-K (2022). Front. Oncol. 11:766939. doi: 10.3389/fonc.2021.766939

In the originally published article, affiliations 1 and 3 were presented incorrectly.

Affiliation 1 was presented as "Department of Cancer Epidemiology and Prevention, Henan Engineering Research Center of Cancer Prevention and Control, Henan International Joint Laboratory of Cancer Prevention, Henan Cancer Hospital, The Affiliated Cancer Hospital of Zhengzhou University, Zhengzhou, China"; it should be "Department of Cancer Epidemiology and Prevention, Henan Engineering Research Center of Cancer Prevention and Control, Henan International Joint Laboratory of Cancer Prevention, The Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China".

Affiliation 3 was presented as "Department of Radiology, Henan Cancer Hospital, The Affiliated Cancer Hospital of Zhengzhou University, Zhengzhou, China"; it should be "Department of Radiology, The Affiliated Cancer Hospital of Zhengzhou University, Henan Cancer Hospital, Zhengzhou, China".

The authors apologize for these errors and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

OPEN ACCESS

Approved by:
Frontiers Editorial Office, Frontiers Media SA, Switzerland

***Correspondence:**
Shao-Kai Zhang
shaokaizhang@126.com

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *08 February 2022*
Accepted: *09 February 2022*
Published: *03 March 2022*

Citation:
Guo L-W, Lyu Z-Y, Meng Q-C, Zheng L-Y, Chen Q, Liu Y, Xu H-F, Kang R-H, Zhang L-Y, Cao X-Q, Liu S-Z, Sun X-B, Zhang J-G and Zhang S-K (2022) Corrigendum: Construction and Validation of a Lung Cancer Risk Prediction Model for Non-Smokers in China. Front. Oncol. 12:871848. doi: 10.3389/fonc.2022.871848

frontiers in Oncology

ORIGINAL RESEARCH
published: 10 January 2022
doi: 10.3389/fonc.2021.795528

Participation and Yield of a Lung Cancer Screening Program in Hebei, China

*Di Liang, Jin Shi, Daojuan Li, Siqi Wu, Jing Jin and Yutong He**

Cancer Institute in Hebei Province, The Fourth Hospital of Hebei Medical University, Shijiazhuang, China

Objective: Lung cancer screening has been widely conducted in Western countries. However, population-based lung cancer screening programs in Hebei in China are sparse. Our study aimed to assess the participation rate and detection rate of positive nodules and lung cancer in Hebei province.

Method: In total, 228 891 eligible participants aged 40–74 years were enrolled in the Cancer Screening Program in Hebei from 2013 to 2019. A total of 54 846 participants were evaluated as the lung cancer high-risk population by a risk score system which basically followed the Harvard Risk Index and was adjusted for the characteristics of the Chinese population. Then this high-risk population was recommended for low-dose computed tomography (LDCT) screening. And all participants attended annual passive follow-up, and the active follow-up interval was based on radiologist's suggestion. All participants were followed-up until December 31, 2020. The overall, group-specific participation rates were calculated, and its associated factors were analyzed by a multivariable logistic regression model. Participation rates and detection of positive nodules and lung cancer were reported.

Results: The overall participation rate was 52.69%, where 28 899 participants undertook LDCT screening as recommended. The multivariable logistic regression model demonstrated that a high level of education, having disease history, and occupational exposure were found to be associated with the participation in LDCT screening. The median follow-up time was 3.56 person-years. Overall, the positive identification of lung nodules and suspected lung cancer were 12.73% and 1.46% through LDCT screening. After the native and passive follow-up, 257 lung cancer cases were diagnosed by lung cancer screening, and the detection rate of lung cancer was 0.89% in the screening group. And its incidence density was 298.72 per 100,000. Positive lung nodule rate and detection rate were increased with age.

Conclusion: Our study identified personal and epidemiological factors that could affect the participation rate. Our findings could provide the guideline for precise prevention and control of lung cancer in the future.

Keywords: lung cancer, screening, Hebei province, participation rate, detection rate

OPEN ACCESS

Edited by:
Lizza E. L. Hendriks, Maastricht University Medical Centre, Netherlands

Reviewed by:
Shaokai Zhang, Henan Provincial Cancer Hospital, China
Jing Jiang, First Affiliated Hospital of Jilin University, China

***Correspondence:**
Yutong He
hytong69@yahoo.com

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *15 October 2021*
Accepted: *09 December 2021*
Published: *10 January 2022*

Citation:
Liang D, Shi J, Li D, Wu S, Jing J and He Y (2022) Participation and Yield of a Lung Cancer Screening Program in Hebei, China. Front. Oncol. 11:795528. doi: 10.3389/fonc.2021.795528

INTRODUCTION

Lung cancer is the second most diagnosed cancer, and it is also the leading cause of cancer death in the world. According to GLOBOCAN 2020, there were approximately 2 206 771 newly diagnosed lung cancer cases and 1 796 144 cancer deaths in 2020, accounting for 11.4% and 18.0% of all new cases from cancer, respectively (1). As reported by the Chinese National Cancer Center (CNCC), with a 36.05/100,000 age-standardized incidence rate and a 28.06/100,000 age-standardized mortality rate, lung cancer was the most common cancer and the leading cause of cancer death in 2016 in China. It also showed an increasing trend in China (2). While the five-year survival rate of lung cancer was only 19.7% (3).

A series of randomized controlled trials, cohort studies, and case-control studies have demonstrated that low-dose computed tomography (LDCT) screening in a high-risk population could reduce mortality due to lung cancer (4–7). By now, lung cancer screening programs have been organized by many countries, such as the national lung cancer screening trial (NLST), National Cancer Institute Prostate, Lung, Colorectal & Ovarian Cancer Screening Trial (PLCO), and others (4, 8, 9). These trails were mainly carried out in Western countries. However, the effectiveness evaluation of lung cancer screening programs in China, in which the lifestyle is different from Western countries, is still rare.

The population-based Cancer Screening Program in Urban China (CanSPUC) was conducted in 2012. It included five type common kinds of cancer: lung cancer, female breast cancer, liver cancer, colorectal cancer, and upper digestive tract cancer (esophagus cancer and gastric cancer). Participants were invited to take a cancer risk assessment using an established clinical cancer risk score system, and those who were evaluated to be at high risk for specific types of cancer were recommended to take the appropriate screening intervention by the study design. Individuals who were found to be at high risk of lung cancer were recommended to undergo LDCT at tertiary-level hospitals.

Combined with follow-up, we aimed to assess the participation rate, screening effectiveness, and results of lung cancer screening in a high-risk population in Hebei province, China. It could provide reliable and effective data support for lung cancer prevention and control.

MATERIALS AND METHODS

Study Population

The study was conducted in Shijiazhuang and Tangshan City which are located in Hebei province (North China), and screenings took place in six tertiary-level hospitals (the Fourth Hospital of Hebei Medical University, the First Hospital of Hebei Medical University, the first Hospital of Shijiazhuang, Hebei Cheat Hospital, Tangshan People's Hospital, and Kailuan Hospital). The participants who met the following conditions became the screening objects: (1) the residents of the program's city; (2) residents' age is 40-74 years old. The program used a cluster sampling method to select the screening participants. And selecting the screening participants was based on the community. The staff of the community mobilized eligible residents of the area under their jurisdiction to participate in the program. Eligible residents took part in face-to-face interviews in the selected communities. After obtaining signed informed consent, all the eligible participants were interviewed by trained staff to complete an epidemiological questionnaire and to assess their cancer risk using an established risk score system. In this study, to maximize the use of limited health resources and increase the detection rate of lung cancer, participants who were put into the high-risk groups of lung cancer were recommended free LDCT examinations in those tertiary hospitals. The present study was approved by the Ethics Board of the Fourth Hospital of Hebei Medical University. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline. A flow diagram showing the recruitment of the study population is shown in **Figure 1**.

Risk Assessment

The rationale of the cancer risk score system was based on the Harvard Risk Index (10). According to the Chinese characteristics, the risk score system included risk factors, relative risks, and exposure rates of risk factors that were adjusted. Each risk factor was allocated a score by the expert panel based on the magnitude of its association with lung cancer. The cumulative risk scores were calculated and were then divided by the average risk score in the general population to get the final individual relative risks (11). People who smoke more than one cigarette a day for more than 6 months were defined as smokers. Second-hand smoking exposure was identified in participants living with a smoker on a regular basis in the workplace or at home. The database was established by professional trained community doctors with double-entry and high-quality control to ensure consistency. The questionnaires completed every day required a random sample of 2% for re-examination, and the compliance rate of each item after the re-examination could not be less than 90%.

LDCT Scanning

All participants undertaking the LDCT screening used the 64-section CT machine. The parameters were set as follows: (1) Scan parameters: 120 kVp and ≤30 mAs; scanning thickness: 5 mm and scanning spacing: 5 mm; the reconstructed layer thickness was 1.0-1.25 mm continuous (layer interval is 0); (2) the scanning range was from the lung tip to the costophrenic angle (including all lungs); (3) nodule measurement: Using an electronic measuring ruler to measure the maximum of the nodule length and wide diameter; (4) positive nodule: The mean diameter of solid or partly nodular nodules ≥ 5 mm, or non-solid nodules ≥ 8 mm in average diameter, or endobronchial nodules; and (5) suspicious lung cancer: A suspicious lung cancer case was identified when cases were diagnosed as suspected lung cancer or malignant lesions by senior thoracic radiologists.

Follow-Up of Participants

All participants were followed-up by active and passive methods until December 31, 2020. An annual passive and regular active

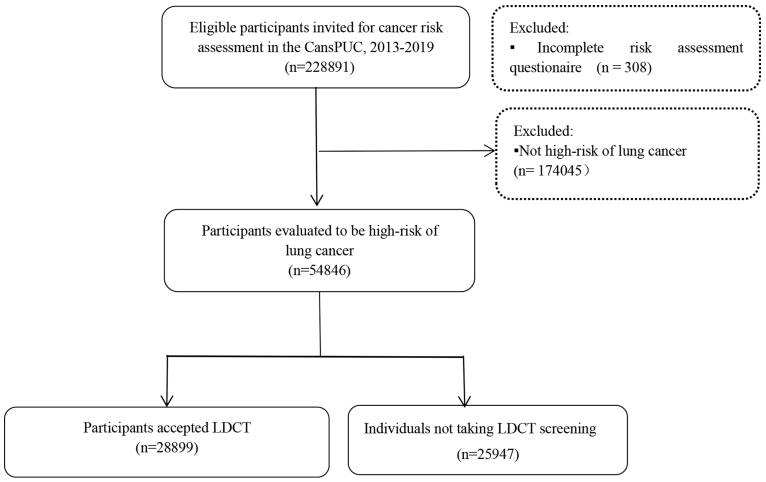


FIGURE 1 | The flow chart of lung cancer screening in Hebei province, 2013-2019.

follow-up mechanism for the entire cohort population was established and carried out in our program based on the cancer registration system. Through telephone, home visits, and retrieval of medical record information from medical institutions, positive cases were actively followed-up to obtain the final diagnosis and outcome. For people with positive results, regular active follow-up was conducted by radiologist's suggestion after the LDCT screening.

For passive follow-up, all participants who completed the questionnaire survey were matched by a personal identification number with the local cancer registration database and the all-cause mortality database in 2013-2020. The information of cancer incidence, subsite, topography, and morphology were obtained from these databases. Newly diagnosed cases of lung cancer were classified by sites according to International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (codes C33 and C34).

Statistical Analysis

The overall and group-specific participation rates by different characteristics were calculated and compared by χ^2 test. Categorical variables were presented as numbers and percentages. The relationship of variables with participation rate of lung cancer screening were quantified by a multivariable logistics regression model with odds ratios (ORs) and their 95% confidence intervals (CIs). All statistical analyses were performed using R, version 3.4. Statistical significance was established at $P \leq 0.05$ on two-sided probabilities.

RESULTS

Characteristics of the Study Population

In the lung cancer screening program, 228 891 participants were recruited and had completed a risk assessment questionnaire in 2013-2019. There were 54 846 high-risk participants for lung cancer accounting for 23.96% of the total population. More women took part in the screening program, while the high-risk rate in women (43.73%) was less than that in men (56.27%). The majority of participants were between 50 and 64 years old. Most participants had junior school education level or below. In the high-risk population group, half (54.86%) had first degree relatives who had history of lung cancer, and three-quarters were smokers (**Table 1**).

Participation Rate for LDCT Screening

In the 54 846 participants in the lung cancer high-risk population, 28 899 undertook LDCT screening. The total participation rate was 52.69%. The screening program in Shijiazhuang (64.97%) had a higher participation rate than that in Tangshan (48.20%). Although there was a higher high-risk rate in men, the participation rate in men (42.44%) was less than that in women (65.88%). Participants aged 45-49 had the higher participation rate (56.57%), and the participation rates decreased along with the increasing age. It was found that participants with higher educational level, who worked as technical staff, had occupational exposure, never smoked, had second-hand smoke exposure, a history of lung diseases, and family history of lung cancer had relatively higher participant rates (**Table 1**).

In multivariable analysis, we found that participants who had occupational exposure had 45% higher odds of undertaking screening than other participants (OR: 1.45; 95%CI: 1.39-1.51). Smokers and former smokers were less willing to accept the screening, in which the ORs were 0.87 (95%CI: 0.81-0.92) and 0.83 (95%CI: 0.74-0.93), respectively. After adjusting for year of recruitment, study areas, married condition, Body Mass Index (BMI), drinking consumption, heating methods, and cooking fuels, we found that age, sex, educational level, occupation, occupational exposure, smoke condition, second-hand smoking exposure, history of lung diseases, and family history of lung cancer were associated with participation rate (**Table 2**).

TABLE 1 | Characteristics of the study population and participation rates between different groups.

Variables		Numbers of questionnaire	High-risk participants (%)	LDCT screening participants (%)	Participation rates (%)	P value
Total		228891	54846	28899	52.69	
Area	Shijiazhuang	76508	14692 (26.79)	9545 (33.03)	64.97	<0.001
	Tangshan	152383	40154 (73.21)	19354 (66.97)	48.20	
Years	2013-2014	26171	4951 (9.03)	1899 (6.57)	38.36	<0.001
	2014-2015	33616	8317 (15.16)	2759 (9.55)	33.17	
	2015-2016	24124	5938 (10.83)	2773 (9.60)	46.70	
	2016-2017	46942	10026 (18.28)	5875 (20.33)	58.60	
	2017-2018	34942	8421 (15.35)	5204 (18.01)	61.80	
	2018-2019	39937	11788 (21.49)	7108 (24.60)	60.30	
	2019-2020	23159	5405 (9.85)	3281 (11.35)	60.70	
Sex	Male	109946	30863 (56.27)	13099 (45.33)	42.44	<0.001
	Female	118945	23983 (43.73)	15800 (54.67)	65.88	
Age	40-	30630	6142 (11.20)	3321 (11.49)	54.07	<0.001
	45-	36604	8793 (16.03)	4974 (17.21)	56.57	
	50-	40565	10737 (19.58)	5787 (20.02)	53.90	
	55-	38939	10256 (18.70)	5323 (18.42)	51.90	
	60-	42795	10901 (19.88)	5537 (19.16)	50.79	
	65-	28800	6538 (11.92)	3180 (11.00)	48.64	
	70-	10558	1479 (2.70)	777 (2.69)	52.54	
BMI*	<18.5	2090	620 (1.13)	279 (0.97)	45.00	0.002
	18.5-	91013	20662 (37.67)	10797 (37.36)	52.26	
	24-	109451	26204 (47.78)	13878 (48.02)	52.96	
	28-	26337	7360 (13.42)	3945 (13.65)	53.60	
Educational level	Junior school and less	140559	30411 (55.45)	13938 (48.23)	45.83	<0.001
	Senior high school	58757	15313 (27.92)	8913 (30.84)	58.21	
	College and above	29575	9122 (16.63)	6048 (20.93)	66.30	
Job	Technician/employee	39267	10927 (19.92)	6711 (23.22)	61.42	<0.001
	Farmer	46041	9539 (17.39)	4539 (15.71)	47.58	
	Worker	107990	27394 (49.95)	13761 (47.62)	50.23	
	Others	35593	6986 (12.74)	3888 (13.45)	55.65	
Occupational exposure	No	180570	27229 (49.65)	11889 (41.14)	43.66	<0.001
	Yes	48321	27617 (50.35)	17010 (58.86)	61.59	
Fuels for heating	clean	186540	41473 (75.62)	21258 (73.56)	51.26	<0.001
	coal	30374	11011 (20.08)	6291 (21.77)	57.13	
	Other	11977	2362 (4.31)	1350 (4.67)	57.15	
Fuels for cooking	Natural/liquefied gas	200943	42933 (78.28)	21436 (74.18)	49.93	<0.001
	Coal	17127	9311 (16.98)	6245 (21.61)	67.07	
	Other	10821	2602 (4.74)	1218 (4.21)	46.81	
Smoking	Never	168602	13262 (24.18)	8583 (29.70)	64.72	<0.001
	Smoke	52564	39853 (72.66)	19450 (67.30)	48.80	
	Ever smoke	7725	1731 (3.16)	866 (3.00)	50.03	
Second-hand smoking exposure	No	154473	14633 (26.68)	5457 (18.88)	37.29	<0.001
	Yes	74418	40213 (73.32)	23442 (81.12)	58.29	
Drinking	Never	176367	25480 (46.46)	13717 (47.47)	53.83	<0.001
	Current	46718	27227 (49.64)	14189 (49.10)	52.11	
	Former	5806	2139 (3.90)	993 (3.44)	46.42	
History of respiratory disease	No	186420	19796 (36.09)	6094 (21.09)	30.78	<0.001
	Yes	42471	35050 (63.91)	22805 (78.91)	65.06	
Family history of cancer	No	168441	18778 (34.24)	5431 (18.79)	28.92	<0.001
	Yes	60448	36067 (65.76)	23467 (81.21)	65.07	
Family history of lung cancer	No	190626	24760 (45.14)	8452 (29.25)	33.46	<0.001
	Yes	38265	30086 (54.86)	20447 (70.75)	67.96	

**BMI, Body mass index.*

Positive Rates in Study

In the screening program, 3 679 positive nodules and 421 suspected lung cancer cases were detected, yielding rates of 12.73% and 1.46%, respectively. Comparing the results in different genders, the positive nodules rate in men (1757, 13.41%) was higher than that in women (1922, 12.16%). With increasing age, the positive rates gradually increased. The highest positive nodule rate was reached at 70-74 years old in both genders, which was 21.79% in men and 18.35% in women. In the positive nodules rates in ages 40-44 and 65-69, the rates in men were higher than the respective rates in women at the same age range. Along with an increasing age, the suspected lung cancer

TABLE 2 | Factors associated with participation rate in lung cancer screening.

Variables		Model 1*			Model 2#		
		OR	95%CI	*P* value	OR	95%CI	*P* value
Age							
	40-	Reference			Reference		
	45-	1.24	1.16-1.33	<0.001	1.18	1.1-1.27	<0.001
	50-	1.23	1.15-1.32	<0.001	1.16	1.08-1.25	<0.001
	55-	1.31	1.22-1.41	<0.001	1.27	1.18-1.36	<0.001
	60-	1.38	1.28-1.48	<0.001	1.28	1.19-1.37	<0.001
	65-	1.34	1.24-1.45	<0.001	1.16	1.07-1.26	<0.001
	70-	1.45	1.27-1.64	<0.001	1.02	0.9-1.17	0.707
Sex							
	Male	Reference			Reference		
	Female	1.30	1.23-1.37	<0.001	1.40	1.33-1.48	<0.001
Educational level							
	Junior school and less	Reference			Reference		
	Senior high school	1.37	1.31-1.43	<0.001	1.24	1.18-1.29	<0.001
	College and above	1.67	1.57-1.78	<0.001	1.56	1.46-1.66	<0.001
Job							
	Technician/employee	Reference			Reference		
	Farmer	0.83	0.78-0.9	<0.001	0.70	0.65-0.76	<0.001
	Worker	0.84	0.79-0.89	<0.001	0.88	0.83-0.93	<0.001
	Others	0.95	0.88-1.01	0.119	0.88	0.82-0.94	<0.001
Occupational exposure							
	No	Reference			Reference		
	Yes	1.31	1.26-1.37	<0.001	1.45	1.39-1.51	<0.001
Smoking							
	Never	Reference			Reference		
	Smoke	0.77	0.73-0.82	<0.001	0.87	0.81-0.92	<0.001
	Ever smoke	0.77	0.68-0.86	<0.001	0.83	0.74-0.93	0.002
Second-hand smoking exposure							
	No	Reference			Reference		
	Yes	1.26	1.2-1.32	<0.001	1.14	1.08-1.19	<0.001
History of respiratory disease							
	No	Reference			Reference		
	Yes	1.81	1.72-1.89	<0.001	1.74	1.66-1.83	<0.001
Family history of lung cancer							
	No	Reference			Reference		
	Yes	1.50	1.41-1.60	<0.001	1.62	1.52-1.73	<0.001

**Adjusted for married condition, BMI, fuels for heating, fuels for cooking, drinking, and family history of any cancer.*
#Adjusted for areas, year of recruitment, married condition, BMI, fuels for heating, fuels for cooking, drinking, and family history of any cancer.

rates had an increasing trend. At 70-74 years old in both men and women, the rates reached the top which were 5.38% and 3.10%, respectively (**Figures 2** and **3**).

Detection Rate of Positive Pulmonary Nodules

The characteristics of the nodules are shown in **Table 3**. The mean diameter of the nodule demonstrated the significant difference in benign nodule and lung cancer groups in which the median sizes were 6.00 mm and 12.25 mm, respectively. The majority of nodules were solid (83.30% in the benign nodule group and 36.63% in the lung cancer group). Non-solid and part-solid nodules accounted for 5.94% and 10.76% in the benign nodule group and 27.91% and 35.47% in the lung cancer group, respectively. A larger number of cancers were observed in the left upper (23.30%) and right upper lobes (38.07%) than in the other lobe (**Table 1**). In the lung cancer group, the proportion of nodules with stretched pleura and spiculation was higher than those in the benign nodule group.

Follow-Up Results

From 2013 to 2020, the median follow-up time was 3.56 years and the total follow-up time was 828 252.5 person-year. By follow-up, 257 lung cancer cases were screened in the screening group, in which the detection rate in the screening group was 0.89% and incidence density was 298.72/100,000. In the screening group, the participants with positive results (positive nodules and suspicious lung cancer) had the higher detection rate of lung cancer than participants with negative results (4.73% vs. 0.31%). In the high risk of lung cancer population, the detection rate of the screening group (0.89%) was significantly higher than those in the non-screening group (0.44%). **Figure 4** shows that the detection rates from lung cancer increased with age and those were higher in men than in women. In the screening and non-screening groups, the most common subsite of lung cancer was the upper lobe. And adenocarcinoma was the main histologic type, followed by squamous cell carcinoma and small-cell carcinoma (**Table S1**).

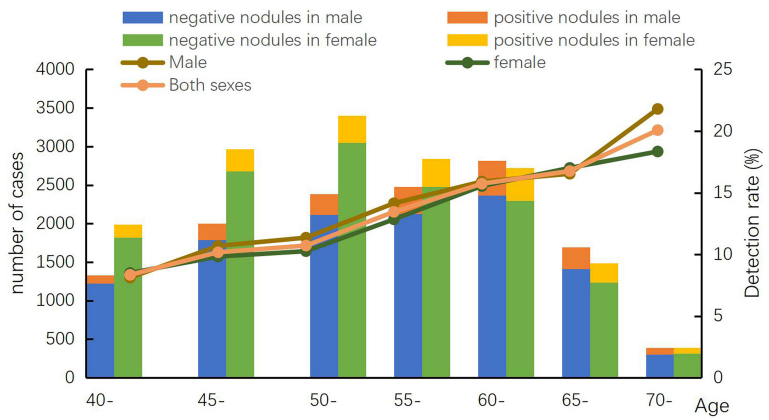


FIGURE 2 | Age-specified positive nodules detection rate in Hebei province, 2013-2019.

DISCUSSION

This study reported the 228 891 participants undertaking LDCT screening among a large-scale population-based screening program. This is the first study in Hebei province in China that combined epidemiological investigation, risk assessment stratification, and LDCT for participants. Although great efforts have been made by previous studies to develop effective screening, the majority of studies aimed to optimization risk scores and few were truly implemented in large-scale lung cancer screening, especially in Hebei province. The overall participation rate was 52.69% in LDCT screening among the lung cancer high-risk population. The detection rate of lung cancer in the screening group was 0.89%. And we found that the population of nodules with a relatively large mean diameter (6.00 mm vs. 12.25 mm in the benign nodule group vs. lung cancer group), non-solid, spiculation, non-calcification, and stretched pleura would more likely to develop into lung cancer. This study could provide a reliable, reasonable, and precise management strategy for lung cancer prevention and control in Hebei.

The overall participation rate was 52.69% in Hebei province. The participation rate of lung cancer screening in the high-risk population varies in different programs. It might relate to the local management and personal factors. Smoking is one of the most important factors for lung cancer and smokers were more likely to develop lung cancer (5, 12). While we found that the participation rates of lung cancer screening in smokers and former smokers were lower than that in non-smokers, in which the adjusted OR was 0.87 (95%CI: 0.81-0.92) and 0.83 (95%CI: 0.74-0.93), respectively. In smokers, the aversion to encountering adverse screening results might prevent test uptake (13–16). We also found that the population with higher educational level, who were technicians or employees, had second-hand smoking exposure, history of respiratory disease and family history of lung cancer had a higher participation rate. Previous studies demonstrate that the level of education was significantly positively correlated with the level of compliance with screening (17, 18). Our study was consistent with that of Henan province where people with undergraduate degrees or more had higher compliance (OR = 1.34, 95%CI: 1.24-1.44). It

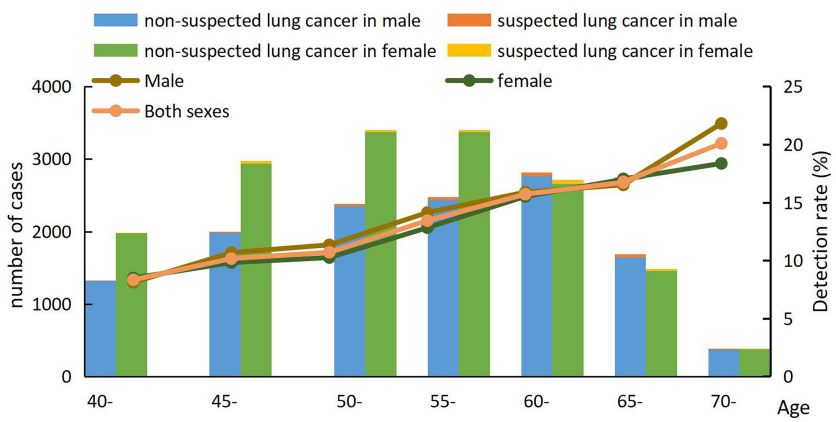


FIGURE 3 | Age-specified suspected lung cancer detection rate in Hebei province, 2013-2019.

TABLE 3 | Distribution of nodule characteristics in lung cancer screening in Hebei province, 2013-2019.

	Benign nodule		Lung cancer		Total		*P* value
	N	%	N	%	N	%	
Nodule size (mm)							
Median	6.00		12.25		6.00		<0.001
Interquartile range	(5.00,7.50)		(8.00,17.50)		(5.00,7.50)		
Nodule type							
Solid	2803	83.30	63	36.63	2866	81.03	<0.001
Non-solid	200	5.94	48	27.91	248	7.01	
Part-solid	362	10.76	61	35.47	423	11.96	
Unknown	170	5.05	8	4.65	178	5.03	
Nodule location							
Right upper lobe	984	28.23	67	38.07	1051	28.70	<0.001
Right middle lobe	497	14.26	12	6.82	509	13.90	
Right lower lobe	816	23.41	32	18.18	848	23.16	
Left upper lobe	469	13.45	41	23.30	510	13.93	
Left lower lobe	720	20.65	24	13.64	744	20.32	
Others	20		0		20		
Unknown	119		4		123		
Nodule's edge							<0.001
Spiculation	491	14.29	90	52.33	581	16.10	
Smooth	2946	85.71	82	47.67	3028	83.90	
unknown	98		8		106		
	3437		172		3609		
Calcification							0.008
No	3234	94.81	164	99.39	3398	95.02	
Yes	177	5.19	1	0.61	178	4.98	
Unknown	124		15		139		
Stretched pleura							<0.001
No	3213	94.86	125	75.30	3338	93.95	
Yes	174	5.14	41	24.70	215	6.05	
Unknown	148		14		162		

might be that the participants with a higher level of education, disease history, and occupational exposure have better understanding, self-health awareness, and pay more attention to self-care. In NLST and the European Dutch-Belgian Randomized Lung Cancer Screening Trial (NELSON), the compliance rate of screening reached 90% (4, 19). And some studies showed that the rates of participation were more than 50% (9, 20). While the overall participation rate was 34.86% in LDCT screening in three provinces (Zhejiang province, Anhui province, and Liaoning province) in China (21). And in the same program in Henan province in China, the overall participation rate was 40.16% which was lower than that in our studies. Different regional compliance was not at the same level, and other factors included the publicity and mobilization of the communities and hospitals involved in the program, the organization and mobilization process, and the health

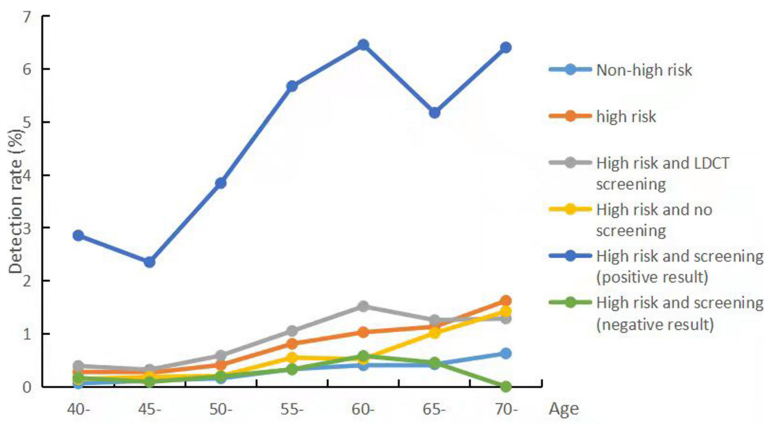


FIGURE 4 | Detection rates of lung cancer in different groups, 2013-2019.

awareness of residents. Another survey conducted among family physicians in South Carolina in 2015 showed that most people had a knowledge gap and there were limited referrals of patients eligible for LDCT screening (16, 22). We conducted multiple training sessions for community physicians to educate them on the necessity and importance of lung cancer screening. As community physicians could influence screening uptake, issues related to penetration and educational outreach around LDCT screening to physicians should be examined. These studies confirmed that community physicians can help improve the compliance to a screening program, especially for people with low educational level and high age. Strengthening health education in the community system and improving the awareness rate of residents' cancer knowledge will have a positive influence on the compliance of lung cancer screening.

In our study, the positive nodule rate was 12.73% in 2013-2019. After active and passive follow-up, the lung cancer detection rate was 0.89% in the screening group. The study of lung cancer screening in 2013-2017 in China showed that the positive rate of nodules detected by LDCT screening in high-risk groups of lung cancer was 11.36% (23). The detection rate of lung positive nodules reported in various provinces in China showed that Zhejiang province was the highest at 21.61%, followed by Beijing with 10.99%, Chongqing City, Yunnan province, Hunan province, and Henan province with 12.91%, 6.90%, 5.92%, and 5.87%, respectively (24–29). The one reason for the different levels in positive nodules rates is the different skill level of diagnosis of cancer in the early stage. During the implementation of the program, our province conducted multiple clinical diagnosis training sessions and unified the diagnostic standards to ensure the homogeneity of the data. Some of our findings with respect to the initial low-dose CT screening are not fully consistent with previous studies. The prevalence of lung cancer (0.89%) was at the middle of the reported range in some prior large studies [NLST, Early Lung Cancer Action Project (ELCAP) (30), International Early Lung Cancer Action Program (I-ELCAP) (31), NELSON (32), Rural China Screening Programme (RuraCSP) (33), Sone (34)[, which ranged from 0.4% to 2.7%. But it was close to the rate of 1.0% in the NELSON trial and 1.1% in NLST. This relatively low rate may be due to some combination of the following factors: participants in the program were healthier than the general population, and were younger in our study than in other studies. For example, our study included participants aged 40-74 and the NLST criteria included 55-74-year-old and heavy smoker participants. The other reason is that the definitions of a high-risk population were different. Following the NLST age entry criteria, the detection rate of our study in ages 55-74 was 1.28%. If the population only includes smokers, the detection rate will increase. It means that the risk assessment system of our study could concentrate on the high-risk lung cancer population and it could increase the screening effects.

Lung nodules can be effectively detected by LDCT. But discrimination between benign and malignant nodules, and which type of nodule had the greater probability of developing lung cancer are the medical concern (7). Among the positive nodules, 4.85% were malignant in our study, and this corresponded with other studies. In the Pan-Canadian Early Detection of Lung Cancer Study (PanCan) and British Columbia Cancer Agency (BCCA), the rates of cancer in nodules in the two datasets were 5.5% and 3.7% (35). We confirmed that the right upper lobes were the most common sub-site in lung cancer; they accounted for 38.07% of all diagnosed lung cancer cases. Among the screen-detected lung cancers, about three-quarters were adenocarcinomas. And the screening methods for small cell lung cancer and squamous cell carcinoma need to be improved. Lung adenocarcinomas are more likely to be located at the periphery of the lung. And the cancer in the lung periphery had a greater probability of being measured than central lung cancer (36). Lung cancer is most likely to occur in the upper lobe. It is a known phenomenon in non–small cell lung cancer cases and can be explained as the maximum airflow when breathing begins, mainly towards the upper right lobe bronchus. So, tobacco smoke and its carcinogenic toxins accumulates the most in the right upper lobe (37–39). Through our study, we confirmed that nodules with the following characteristics should be paid more attention to in future clinical treatment and diagnosis: larger nodule size, location of the nodule in the upper lobe, non-solid and part-solid nodule type, spiculation, non-calcification, and stretched pleura nodules (35). These nodules were more likely to develop into lung cancer.

This study has strength and limitations. The strengths were as follows: this study was population-based, and it involved a large-scale sample size. Detailed epidemiological questionnaire information was collected in a standardized manner by trained study staff to ensure the quality of the data. A sound annual passive and active follow-up mechanism for the entire cohort population was established and carried out in our program based on the cancer registration system. We obtained information regarding each participant's cancer incidence in the study. This study has the limitation that some variables, such as smoking status and other variables, were self-reported and it might lead to misclassification. Another limitation is that follow-up work for patients diagnosed with lung cancer is still under way, therefore clinical disease information was not fully obtained. And the study population was a pre-selected high-risk population ascertained by the risk assessment system which might not represent the general population of Hebei province, and selection bias cannot be ruled out.

In summary, in this large-scale lung cancer screening in Hebei, we found that some variables, which were age, sex, educational level, job, smoker, secondhand smoking exposure, history of respiratory, and family history of lung cancer contributed to the participation rate. And the detection rate in the screening group was higher than that in other groups. Our finding may provide data support for lung cancer prevention and it is useful for optimizing screening strategies.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Board of the Fourth Hospital of Hebei Medical University (No. 2012KY102). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DiL and YH wrote the main text and conducted data analysis. YH designed the study. JS, DaL, SW, and JJ collected the data. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We gratefully thank the cooperation partners of all cancer screening hospitals and communities in Hebei province. We are grateful to the participants for taking part in this study. The authors assume full responsibility for the analyses and interpretations of the data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.795528/full#supplementary-material

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
2. He J, Wei WQ. *2019 China Cancer Registry Annual Report*. Beijing, China: People's Medical Publishing House (2021).
3. Zeng H, Chen W, Zheng R, Zhang S, Ji JS, Zou X, et al. Changing Cancer Survival in China During 2003-15: A Pooled Analysis of 17 Population-Based Cancer Registries. *Lancet Global Health* (2018) 6(5):e555–67. doi: 10.1016/S2214-109X(18)30127-X
4. National Lung Screening Trial Research T, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced Lung-Cancer Mortality With Low-Dose Computed Tomographic Screening. *N Engl J Med* (2011) 365(5):395–409. doi: 10.1056/NEJMoa1102873
5. Moyer VA. USPSTF. Screening for Lung Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Ann Intern Med* (2014) 160(5):330–8. doi: 10.7326/m13-2771
6. Yousafkhan U, Carlijn VDA, De Jong PA, Heuvelmans M, Scholten E, Walter J, et al. Risk Stratification Based on Screening History: The NELSON Lung Cancer Screening Study. *Thorax* (2017) 72(9):819–24. doi: 10.1136/thoraxjnl-2016-209892
7. Jett J. Screening for Lung Cancer: Who Should be Screened? *Arch Pathol Lab Med* (2012) 136(12):1511–4. doi: 10.5858/arpa.2012-0259-RA
8. Tammemagi CM, Pinsky PF, Caporaso NE, Kvale PA, Hocking WG, Church TR, et al. Lung Cancer Risk Prediction: Prostate, Lung, Colorectal And Ovarian Cancer Screening Trial Models and Validation. *J Natl Cancer Inst* (2011) 103(13):1058–68. doi: 10.1093/jnci/djr173
9. Quaife SL, Ruparel M, Dickson JL, Beeken RJ, McEwen A, Baldwin DR, et al. Lung Screen Uptake Trial (LSUT): Randomized Controlled Clinical Trial Testing Targeted Invitation Materials. *Am J Respir Crit Care Med* (2020) 201(8):965–75. doi: 10.1164/rccm.201905-0946OC
10. Colditz GA, Atwood KA, Emmons K, Monson RR, Willett WC, Trichopoulos D, et al. Harvard Report on Cancer Prevention Volume 4: Harvard Cancer Risk Index. Risk Index Working Group, Harvard Center for Cancer Prevention. *Cancer Causes Control* (2000) 11(6):477–88. doi: 10.1023/a:1008984432272
11. Chen H, Li N, Ren J, Feng X, Lyu Z, Wei L, et al. Participation and Yield of a Population-Based Colorectal Cancer Screening Programme in China. *Gut* (2019) 68(8):1450–7. doi: 10.1136/gutjnl-2018-317124
12. Jaklitsch MT, Jacobson FL, Austin JHM, Field JK, Jett JR, Keshavjee S, et al. The American Association for Thoracic Surgery Guidelines for Lung Cancer Screening Using Low-Dose Computed Tomography Scans for Lung Cancer Survivors and Other High-Risk Groups. *J Thorac Cardiov Sur* (2012) 144(1):33–8. doi: 10.1016/j.jtcvs.2012.05.060
13. Carter-Harris L, DuyKhanh PC, Hanna N, Rawl SM. Lung Cancer Screening: What do Long-Term Smokers Know and Believe? *Health Expect* (2017) 20(1):59–68. doi: 10.1111/hex.12433
14. Raju S, Khawaja A, Han X, Wang X, Mazzone PJ. Lung Cancer Screening: Characteristics of Nonparticipants and Potential Screening Barriers. *Clin Lung Cancer* (2020) 21(5):e329–36. doi: 10.1016/j.cllc.2019.11.016
15. Sung JJ, Choi SY, Chan FK, Ching JY, Lau JT, Griffiths S. Obstacles to Colorectal Cancer Screening in Chinese: A Study Based on the Health Belief Model. *Am J Gastroenterol* (2008) 103(4):974–81. doi: 10.1111/j.1572-0241.2007.01649.x
16. Wong MCS. Health Behavioral Models to Find Reasons for Low Rates of Lung Cancer Screening by Low-Dose Computed Tomography. *JAMA Oncol* (2018) 4(3):425. doi: 10.1001/jamaoncol.2017.0493
17. Dong P, Shi JF, Qiu WQ, Liu CC, Wang K, Huang HY, et al. Analysis on the Health Literacy of the Cancer Prevention and Treatment and Its Related Factors Among Urban Residents in China From 2015 to 2017. *Chin J Prev Med* (2020) 54(1):76–83. doi: 10.3760/cma.j.issn.0253-9624.2020.01.015
18. Dong P, Qiu WQ, Shi JF, Mao AY, Huang HY, Sun ZX, et al. Cancer Screening Service Utilization Amd Willingness-to-Pay of Urban Populations in China: A Cross-Sectional Survey From Potential Service Demander's Perspective. *Zhonghua Liu Xing Bing Xue Za Zhi* (2018) 39(2):165–72. doi: 10.3760/cma.j.issn.0254-6450.2018.02.006
19. de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Oudkerk M. Reduced Lung-Cancer Mortality With Volume CT Screening in a Randomized Trial. *N Engl J Med* (2020) 382(6):503–13. doi: 10.1056/NEJMoa1911793
20. Kinsinger LS, Anderson C, Kim J, Larson M, Chan SH, King HA, et al. Implementation of Lung Cancer Screening in the Veterans Health Administration. *JAMA Intern Med* (2017) 177(3):399–406. doi: 10.1001/jamainternmed.2016.9022
21. Wen Y, Yu L, Du LB, Wei DH, Liu YY, Yang ZY, et al. Analysis of Low−Dose Computed Tomography Compliance and Related Factors Among High−Risk Population of Lung Cancer in Three Provinces Participating in the Cancer Screening Program in Urban China. *Chin J Prev Med* (2021) 55(5):633–9. doi: 10.3760/cma.j.cn112150-20201015-01286
22. Ersek JL EJ, McDonnell KK. Knowledge of Attitudes Toward and Use of Low-Dose Computed Tomography for Lung Cancer Screening Among Family Physicians. *Cancer* (2016) 122(15):2324–31. doi: 10.1002/cncr.29944
23. Chen WQ, Cao MM. Strengthening Cancer Early Diagnosis and Treatment, Implementing the Strategy of Healthy China. *China Cancer* (2019) 28(9):643. doi: 10.11735/j.issn.1004-0242.2019.09.A001
24. Du J, He M, Qiu H. Results of Lung Cancer Screening Among Urban Residents in Chongqing, 2012~2017. *China Cancer* (2018) 27(5):328–32. doi: 10.11735/j.issn.1004-0242.2018.05.A002
25. Guo LW, Liu SZ, Zhang SK, Yang FN, Wu Y, Zheng LY, et al. Analysis of the Efficacy of Lung Cancer Screening in Urban Areas of Henan Province by Low-Dose Computed Tomography From 2013 to 2017. *Chin J Oncol* (2020) 2:155–9. doi: 10.3760/cma.j.issn.0253-3766.2020.02.013
26. Yang L, Zhang X, Liu S, Li HC. J F Ji. Lung Cancer Screening in Urban Beijing From 2014 to 2019. *Chin J Prev Med* (2021) 3:339–45. doi: 10.3760/cma.j.cn112150-20200817-01126

27. Xiao HF, Yan SP, Xu KK, Zou YH, Shi ZH, Zhu SL, et al. Analysis of Cancer Screening Program in Changsha Urban Area From 2012 to 2018. *China Cancer* (2019) 11:807–15. doi: 10.11735/j.issn.1004-0242.2019.11.A001
28. Zhang Q, Hhuang YC, Shen LD, Zhao YH, Jiang GX, Zhou H, et al. Analysis of Cancer Risk Assessment and Screening Results Among Urban Residents in Kunming City. *China Cancer* (2018) 9:641–6. doi: 10.11735/j.issn.1004-0242.2018.09.A001
29. Zhang LH, Du LB, Sun XH, Gao YM, Lei LV, Wang XH, et al. An Analysis on the Result of Early Detection and Treatment of Cancer in Zhejiang Urban Population. *Zhejiang Prev Med* (2018) 27(12):1189–93. doi: CNKI:SUN:ZYFX.0.2015-12-001
30. Henschke DIM CI, Yankelevitz DF. Early Lung Cancer Action Project: Overall Design and Findings From Baseline Screening. *Lancet* (1999) 354(9173):99–105. doi: 10.1016/S0140-6736(99)06093-6
31. International Early Lung Cancer Action Program I, Henschke CI, Yankelevitz DF, Libby DM, Pasmantier MW, Smith JP, et al. Survival of Patients With Stage I Lung Cancer Detected on CT Screening. *N Engl J Med* (2006) 355 (17):1763–71. doi: 10.1056/NEJMoa060476
32. Rob J, van Klaveren MO, Prokop M. Management of Lung Nodules Detected by Volume CT Scanning. *N Engl J Med* (2009) 361(23):2221–9. doi: 10.1056/NEJMoa0906085
33. Zhou QH, Fan YG, Wu N, Huang YC, Wang Y, Li L, et al. Demonstration Program of Population-Based Lung Cancer Screening in China: Rationale and Study Design. *Thorac Cancer* (2014) 5(3):197–203. doi: 10.1111/1759-7714.12078
34. Sone S, Takashima S, Li F, Yang Z, Honda T. Y Maruyama,Maruyama YMass Screening for Lung Cancer With Mobile Spiral Computed Tomography Scanner. *Lancet* (1998) 351(9111):1242–5. doi: 10.1016/S0140-6736(97)08229-9
35. McWilliams A, Tammemagi MC, Mayo JR, Roberts H, Liu G, Soghrati K, et al. Probability of Cancer in Pulmonary Nodules Detected on First Screening CT. *N Engl J Med* (2013) 369(10):910–9. doi: 10.1056/NEJMoa1214726
36. Horeweg N, van der Aalst CM, Thunnissen E, Nackaerts K, Weenink C, Groen HJ, et al. Characteristics of Lung Cancers Detected by Computer Tomography Screening in the Randomized NELSON Trial. *Am J Respir Crit Care Med* (2013) 187(8):848–54. doi: 10.1164/rccm.201209-1651OC
37. Subramaniam RP, Asgharian B, Freijer JI, Miller FJ, Anjilvel S. Analysis of Lobar Differences in Particle Deposition in the Human Lung. *Inhal Toxicol* (2003) 15(1):1–21. doi: 10.1080/08958370304451
38. Lince L, Lulu DJ. Carcinoma of the Lung. A Comparative Series of 687 Cases. *Arch Surg* (1971) 102(2):103–7. doi: 10.1001/archsurg.1971.01350020013004
39. Parkash O. Lung Cancer. A Statistical Study Based on Autopsy Data From 1928 to 1972. *Respiration* (1977) 34(5):295–304. doi: 10.1159/000193839

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Oncology

ORIGINAL RESEARCH
published: 13 January 2022
doi: 10.3389/fonc.2021.785777

Clinical Features and Surgical Treatment of Synchronous Multiple Primary Lung Adenocarcinomas With Different EGFR Mutations

Rirong Qu, Fan Ye, Dehao Tu, Yixin Cai and Xiangning Fu *

Department of Thoracic Surgery, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

OPEN ACCESS

Edited by:
Yingsong Lin,
Aichi Medical University, Japan

Reviewed by:
Francesco Pepe,
University of Naples Federico II, Italy
Yoshihiro Miyata,
Hiroshima University, Japan

***Correspondence:**
Xiangning Fu
fuxn2006@aliyun.com

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *29 September 2021*
Accepted: *22 December 2021*
Published: *13 January 2022*

Citation:
Qu R, Ye F, Tu D, Cai Y and Fu X (2022) Clinical Features and Surgical Treatment of Synchronous Multiple Primary Lung Adenocarcinomas With Different EGFR Mutations. Front. Oncol. 11:785777. doi: 10.3389/fonc.2021.785777

Background: With the popularity of lung cancer screening and advances in imaging technology, more and more synchronous multiple primary lung adenocarcinomas (SMPLA) are being diagnosed clinically, however, the clinical characteristics and prognosis of SMPLA with different EGFR mutations remains unclear. We aimed to explore clinical features and surgical outcomes of these patients to aid in the diagnosis and treatment of SMPLA.

Methods: Medical records of patients with different EGFR mutations who have been diagnosed as SMPLA and underwent surgical resection from March 2015 to December 2019 were retrospectively analyzed. Clinical characteristics, surgical outcomes, recurrence-free survival (RFS) and overall survival (OS) were investigated.

Results: A total of 70 patients (68.6% female and 77.1% non-somkers) were included. Total of 161 lesions in all patients, 84.4% were ground-glass opacity (GGO) lesions. EGFR mutations were detected in 108 lesions, most of which were L858R (35.4%) and 19Del (20.5%). The mutation rate of mixed GGO is significantly higher than that of pure GGO and solid nodules (SN); the mutation rate of invasive adenocarcinoma is significantly higher than that of other histology subtypes; the mutation rate of lesions >20 mm was significantly higher than that of ≤20 mm. However, there is no significant difference in the mutation rate of specific driver gene between different radiological features, pathological characteristics and sizes. After a median follow-up time of 29 months, the 3-year OS and RFS were 94.4% and 86.0%, respectively.

Conclusions: A high discordance of EGFR mutations were identified between tumors in patients with SMPLA. Synchronous multiple lung adenocarcinomas with predominantly

multiple GGO should be considered as SMPLA, and surgery may be aggressively performed for these patients due to a good prognosis.

Keywords: synchronous multiple primary lung adenocarcinomas, surgical treatment, epidermal growth factor receptor, lung cancer, clinical features

INTRODUCTION

Synchronous multiple primary lung cancer (SMPLC) defined as two or more primary tumors simultaneously identified in ipsilateral or contralateral lung, is a special type of lung cancer. According to previous studies, its incidence varies from 0.2% to 20% (1), of which 40.3%-91.3% (2–4) are multiple primary lung adenocarcinomas. In recent years, the detection rate of SMPLC has shown a steady increase with the popularity of lung cancer screening and advances in imaging technology, especially the widespread use of HRCT and PET-CT (5, 6). Although Surgical resection has become the mainstay of treatment for SMPLC, its 3-year survival rate roughly ranges from 40% to 92% (7). The wide variation in the efficacy of surgical resection is due not only to differences in the timing of treatment, the specific surgical procedures, and demographic characteristics of patients, but more importantly, to the lack of standard criteria for differential diagnosis from intrapulmonary metastasis.

Martini and Melamed's criteria and ACCP guidelines are most commonly used to distinguish multiple primary lung cancers from intrapulmonary metastases. Because there was no genetic approach to consider in 1975, the Martini-Melamed criteria (8) relied heavily on the clinicopathological features, which can make diagnosis extremely difficult when the tumors are of the similar pathologic type. The ACCP guidelines (9), however, took into account the differences in tumor driver mutation genes, which led to a greater improvement in the diagnosis of multiple primary lung cancers. In addition, the widespread use of next-generation gene sequencing (NGS) in recent years has made the diagnosis of multiple primary lung cancers more accurate (10, 11). More and more SMPLA with different EGFR mutations are being diagnosed clinically, however, the clinical characteristics and prognosis of such patients with surgical treatment remains unknown. Moreover, the reports of such patients are mostly case reports (12–14), and there are few studies with larger samples.

Therefore, in the present study, we focused on the clinical characteristics, surgical outcomes, recurrence-free survival (RFS) and overall survival (OS) of SMPLA with different EGFR mutations to aid in the diagnosis and treatment of these patients.

MATERIALS AND METHODS

Patients

This study retrospectively analyzed the clinical data of patients with SMPLA who underwent simultaneous surgical resection in the Department of Thoracic Surgery at Wuhan Tongji Hospital from March 2015 to December 2019. The criteria for diagnosis of SMPLA in this study are based on the Martini-Melamed criteria (8) and incorporate elements of the new international multidisciplinary lung adenocarcinoma classification (15): (1) major histologic subtypes of tumors are significantly different; (2) major histologic subtypes are similar, but all tumors have lepidic growth component to a certain proportion, or immunohistological features or genetic profiles of tumors are different. The inclusion criteria were as follows: (1) number of lesions ≥2; (2) all lesions of the patient were tested for EGFR and the mutations were different; (3) postoperative pathology of the patient's lesions were all lung adenocarcinoma; (4) the patient did not have adjuvant therapy before surgery; (5) cardiopulmonary function was acceptable and patients could tolerate surgery; (6) no previous history of tumors; (7) no distant metastases on preoperative examinations. The exclusion criteria were as follows: (1) incomplete patient data information; (2) the postoperative pathology of the lesion is not lung adenocarcinoma. Flowchart of participant selection was shown in **Figure 1**. This study was approved by the institutional review board of Tongji Medical College of Huazhong University of Science and Technology and and consent was given by all patients before their clinical records were used.

Surgical Approach

All patients underwent combined intravenous and inhalation general anesthesia with double-lumen endotracheal intubation to maintain single-lung ventilation. The surgery was performed using a 3cm small single-port approach: a 3cm incision was made between the 5th ribs in the mid-axillary line of the patient's surgery side to place a thoracoscope, an elbow laparoscopic suction device, electrocoagulation hooks, and a bipartite clamp was placed to hold the lung lobe if necessary. In bilateral surgery, one side of the surgery is completed and the contralateral surgery is performed in the same way. Systemic lymph node dissection was performed if the dominant lesion was diagnosed as invasive tumor by intraoperative rapid frozen pathology. At the end of the operation, pleural drainage tubes were placed in pleural cavity. The patient received a chest radiograph on the second day after surgery, and pleural drainage tubes could be removed if there was no active bleeding or air leakage. Our specific surgical strategy was: (1) for lesions that were all in the same lobe, we performed direct lobectomy; (2) for lesions that are on the same side but not in the same lobe, we performed anatomical lobectomy or segmentectomy for the dominant lesion ≥2 cm, and sublobar resection for the remaining lesions as much as possible; (3) In patients with multiple lesions in both lungs, we give preference to the side with less lung tissue removed to start the procedure, but if the left side is the dominant lesion and lobectomy is required, the surgery should start on the right side. (4) For all peripheral lesions, intraoperative rapid pathology should be performed as

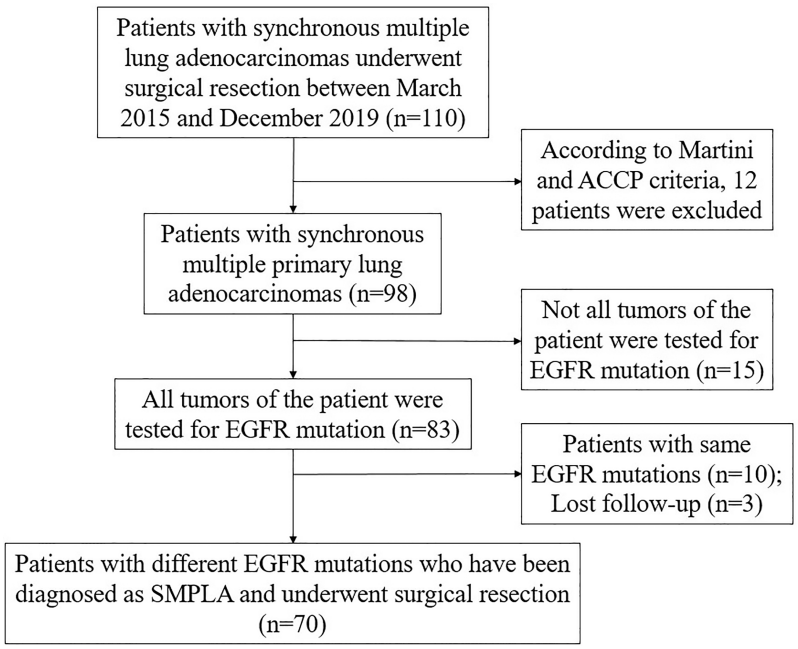


FIGURE 1 | Flowchart of participant selection.

much as possible, and the extent of resection should be determined based on the rapid pathology results and imaging of the lesion, but sublobar resection should be performed as much as possible. (5) For GGO lesions, try to adopt sublobar resection as much as possible. In conclusion, we should take into account the characteristics of the tumor, the patient's physical condition, and the decisions of the physician to try to develop the best individualized treatment plan for each patient. Specific surgical procedures and strategies for selecting the extent of surgical resection are described in our previous study (16, 17).

Tissue Samples and EGFR Mutation Analysis

Genomic DNA was extracted from Formalin-fixed paraffin-embedded (FFPE) samples using QIAamp DNA Tissue Kit (Qiagen, Germany). EGFR mutation was detected using commercially available kits from YZY Medical (Wuhan, China) based on amplification refractory mutation system real-time polymerase chain reaction technology. Twenty-nine kinds of EGFR mutation in exon 18-21 were detected in all lesions of these patients.

Follow-Up

Follow-up was performed by outpatient or telephone follow-up. The follow-up time was calculated from the day after surgery and was followed up until November 2020. In the first year after surgery, chest CT, tumor markers and abdominal ultrasound were reviewed every 3 months; in the second year after surgery, the above indicators were reviewed every 6 months; the above indicators were reviewed annually.

Statistical Analysis

Measured data were expressed as mean ± standard deviation (SD) and differences between groups were analysed by t-tests. Counted data were expressed as number or percent, and differences were analysed using X^2 or Fisher's exact tests. The above data was analyzed using Statistical Product and Service Solutions (version 23; SPSS Inc., Chicago, IL, USA). OS was defined as the time from surgery until death from any cause or last follow-up. Recurrence-free survival (RFS) was defined as the time from surgery until recurrence, death from any cause, or last follow-up. The Kaplan–Meier method was used to analyze OS and RFS by GraphPad Prism software version 7.0. $P<0.05$ was considered statistically significant.

RESULTS

Clinical Characteristics of Patients and Tumors

Clinical characteristics of patients and tumors are shown in **Tables 1**, **2** respectively. Based on the inclusion criteria, 70 patients were ultimately included in this study, of whom 48 were female (68.6%). The mean age of patients was 58.6 ± 8.40 years (range, 41-74 years). A total of 16 patients had a history of smoking or were current smokers, and the rest had no history of smoking. Eleven patients had a family history of cancer, mainly lung cancer (7 patients). Twenty-six patients had preoperative co-morbidities, mainly hypertension and diabetes mellitus. Ten patients had a preoperative mild increase in serum CEA. All

TABLE 1 | Clinical characteristics of patients.

Variables	Number (%)	Mean value
Age (years)		58.6±8.40
≥60	29 (41.4)	
<60	41 (58.6)	
Sex		
Male	22 (31.4)	
Female	48 (68.6)	
Smoking status		
Current and former	16 (22.9)	
Never	54 (77.1)	
Family history of tumor		
Yes	11 (15.7)	
No	59 (84.3)	
Comorbidity		
Yes	26 (37.1)	
No	44 (62.9)	
Preoperative CEA level		3.42±1.33
≥5.0 ng/ml	10 (14.3)	
<5.0 ng/ml	60 (85.7)	
Ejection fraction		63.23±3.57
≥60	58 (82.9)	
55-59	12 (17.1)	
FEV1 (L)		2.74±0.53
p-FEV1%		97.34±15.91
≥100	18 (25.7)	
80-100	39 (55.7)	
≤80	13 (18.6)	
Distribution of tumors		
Unilateral	58 (82.9)	
Bilateral	12 (17.1)	
Number of tumors		
2	54 (77.1)	
3	11 (15.7)	
≥4	5 (7.2)	
Highest pT		
T1	57 (81.4)	
T2	13 (18.6)	
Highest pN		
N0	63 (90.0)	
N1-2	7 (10.0)	
Adjuvant therapy		
Yes	8 (11.4)	
No	62 (88.6)	

patients had a good preoperative cardiopulmonary assessment and could tolerate the procedure. Of these patients, 12 had lesions in both lungs, 11 had three lesions, and 5 had no less than four lesions. Sixty-seven patients had GGO lesions, and 3 patients had only solid nodules. The lesions of 47 patients were located in different lobes. A total of 161 lesions with an average diameter of 20.67 ± 11.7 mm; 84.4% were GGO lesions (pGGO, 42.2%; mGGO, 42.2%) and 80.1% of the lesions were ≤20 mm in diameter. Of all lesions, 71.4% were located in the right lung, with the most lesions in RLL (39.8%), the least in RML (11.2%), and the second least in LLL (11.8%). The postoperative pathology of all lesions was dominated by invasive adenocarcinoma (59.6%), followed by *in situ* adenocarcinoma and microinvasive adenocarcinoma. EGFR mutations were present in 67.1% of the lesions, with L858R (35.4%) and 19Del (20.5%) mutations predominating. Among 70 patients, the highest pathological T stage was mainly pT1 (81.4%) and only seven patient had lymph node metastasis.

Surgical Procedure and Perioperative Results

Twelve and 58 patients respectively underwent bilateral and unilateral surgical resection. The postoperative complications included 3 cases of pulmonary infection, 2 cases of atrial fibrillation, and persistent air leakage for more than 3 days was observed in 4 cases. After treatment, they were all discharged smoothly. No severe perioperative complications or deaths occurred. The average operation time was 205.88 ± 61.94 minutes, the average intraoperative blood loss was 273.83 ± 238.60 ml, the mean postoperative daily drainage of chest tube was 163.52 ± 29.46 ml, the mean postoperative chest tube duration was 6.76 ± 3.43 days, and the average postoperative hospital stay 8.43 ± 3.56 days. Details of surgical procedure are described in **Table 3**.

Detail of EGFR Mutation in 161 Tumors of 70 Patients

EGFR detection of all lesions revealed that 108 lesions had mutations, mainly L858R and 19DEL, and their mutation rates were 35.4% and 20.5%, respectively. Among the different radiologic features, the mutation rate of mGGO was significantly higher than that of pGGO and SN ($P<0.001$); among the different histology features, the mutation rate of invasive adenocarcinoma was significantly higher than that of other histology subtypes ($P<0.001$); among the different size, the mutation rate for lesions >20 mm was significantly higher than that of lesions ≤20 mm ($P<0.001$). However, the mutation rate of specific types were not significantly different among radiologic features, pathology types, or sizes ($P>0.05$). The results of EGFR mutation are presented in **Table 4**.

Postoperative Treatment and Follow-Up of Patients

Eight patients received adjuvant therapy after surgery, of which seven patients received targeted therapy due to the presence of lymph node metastases, and one patient received chemotherapy because the lesions invaded the pleura and were larger than 4 cm. As of November 30, 2020, the average follow-up time was 30.6 ± 13.5 months. Except for three patients with recurrent metastases, one of whom died due to extensive postoperative pleural metastases, the rest of the patients did not develop new lesions or metastases, and all of them are alive. The 3-year OS and RFS in all patients were 94.4% and 86.0%, respectively (**Figure 2**).

DISCUSSION

Lung cancer has been a leading cause of cancer-related death worldwide for decades, with adenocarcinoma representing the most prevalent subtype (18). In recent years, with the popularity of low-dose spiral CT screening, more and more SMPLC are being diagnosed. At the same time, research on the diagnosis and

TABLE 2 | Clinical data of tumors.

Variables	Number (%)	Mean value
Total number of tumors	**161**	
Tumor characteristics (mm)		
pGGO	**68 (42.2)**	**12.52±8.06**
mGGO	**68 (42.2)**	**19.27±8.21**
SN	**25 (15.6)**	**31.29±11.88**
Size of tumors (mm)		**20.67±11.7**
≤10mm	**70 (43.5)**	
<10mm, ≤20mm	**59 (36.6)**	
>20mm	**32 (19.9)**	
Tumor type pattern per patient		
Multiple pGGO	**11 (15.7)**	
Multiple mGGO	**9 (12.9)**	
pGGO+mGGO	**30 (42.9)**	
SN+GGO	**17 (24.3)**	
Multiple SN	**3 (4.2)**	
Location of tumors		
RUL	**64 (39.8)**	
RML	**18 (11.2)**	
RLL	**33 (20.4)**	
LUL	**27 (16.8)**	
LLL	**19 (11.8)**	
Location of lobe		
Same lobe	**23 (32.9)**	
Different lobe	**47 (67.1)**	
Histology in all tumors		
AIS	**36 (22.4)**	
MIA	**29 (18.0)**	
IAC	**96 (59.6)**	
EGFR in all tumors		
WT	**53 (32.9)**	
Mutation	**108 (67.1)**	
L858R	**57 (35.4)**	
19Del	**33 (20.5)**	
Double mutations[a]	**5 (3.1)**	
other	**13 (8.1)**	

SN, solid nodule; pGGO, pure ground-glass opacity; mGGO, mixed ground-glass opacity; RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe; LUL, left upper lobe; LLL, Left lower lobe; AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma; IAC, invasive adenocarcinoma.
[a]*There are two types of mutations in a tumor.*

treatment of multiple primary lung cancers has increased significantly. It is difficult to distinguish multiple primary lung cancers from intrapulmonary metastases, but the treatment options and prognosis of multiple primary lung cancers are completely different compared with intrapulmonary metastases, so the evaluation of diagnosis, treatment, and prognosis of multiple primary lung cancers is particularly important. Unlike previous studies, in this study, we retrospectively examined the clinical characteristics, surgical treatment, and long-term prognosis of all patients with postoperative lesions diagnosed as multiple primary lung adenocarcinoma by EGFR testing. We found a high heterogeneity of EGFR driver genes between tumors in patients with multiple primary lung adenocarcinoma, suggesting the importance of EGFR testing in the diagnosis of such patients.

In the present study, we classified lesions by their radiology, pathology type, and diameter size of the lesions, and then compared EGFR mutations in lesions in different categories. We found that the mutation rate of mixed GGO is significantly higher than that of pure GGO and solid nodules (SN); the mutation rate of invasive adenocarcinoma is significantly higher than that of other histology subtypes (AIS as well as MIA). These results are in general agreement with the findings of Liu et al. (19). The result also concurred with the hypothesis for the progression of lung adenocarcinoma that EGFR-mutated AAH follows a linear progression schema, whereby AAH progresses to AIS and is followed by MIA (20, 21). Sun and colleagues (22) found that diameter size of GGO lesion correlated with EGFR mutation rates, with lesions ≥20 mm in diameter being more likely to be mutated than lesions <20 mm in diameter, which is similar to the results of our study.

For multiple primary lung adenocarcinoma, we should routinely test for EGFR mutations in all lesions. Although multiple lung cancers with predominantly multiple GGO lesions or containing GGO lesions should be considered more often as multiple primary lung cancers, one study found the presence of similar somatic mutations by exon sequencing in multiple GGO lesions in two patients with multiple lung adenocarcinomas, including two pure GGO lesions in one patient (23). The result suggests that intrapulmonary metastases can occur in patients with multiple GGO lesions. In the current study, all but three patients had solid lesions, and the rest contained at least one GGO lesion, yet

TABLE 3 | Surgical procedure and perioperative results of patients.

Variables	Number
Surgical procedure	
Unilateral	**58**
Single lobectomy	**19**
Lobectomy-wedge resection	**13**
Wedge resection-wedge resection	**8**
Segmentectomy-wedge resection	**5**
Lobectomy-lobectomy	**5**
Lobectomy-segmentectomy	**4**
Segmentectomy-segmentectomy	**2**
Single segmentectomy	**2**
Bilateral	**12**
Segmentectomy-wedge resection	**5**
Lobectomy-wedge resection	**4**
Segmentectomy-segmentectomy	**1**
Lobectomy-segmentectomy	**1**
Lobectomy+segmentectomy-wedge resection	**1**
Perioperative results	
Operation time (min)	**205.88±61.94**
Intraoperative blood loss (ml)	**273.83±238.60**
Postoperative chest tube duration (day)	**6.76 ± 3.80**
Daily drainage of chest tube (ml)	**163.58 ± 28.93**
Postoperative hospital stay (day)	**8.43 ± 3.56**

postoperative pathology showed lymph node metastases in seven patients, all of whom should be considered to have multiple intrapulmonary metastases according to previous Martini-Melamed criterion (8). Ye and colleagues (24) reported a case of a patient with multiple primary lung adenocarcinoma in whom two tumors, one with EGFR mutation and one with KRAS mutation, were identified by genetic testing, and the lesion with the KRAS mutation was resected and followed by gefitinib-targeted therapy, after which the remaining lesion disappeared. Therefore, we believe that EGFR mutation can be a good supplement to histological, imaging and morphological evidence of tumor, so as to better distinguish multiple primary lesions from metastatic lesions and provide patients with a more accurate staging. In this study, analysis of EGFR testing results for all lesions revealed the presence of EGFR mutations in 108 lesions (67.1%), including 35.4% and 20.5% for L858R and 19DEL, respectively. This may be related to the fact that the patients in this study were non-smokers (77.1%) and the majority of female patients (68.6%).

Surgery remains the most effective treatment option for multiple primary lung cancers, but the specific surgical method is still controversial. Several studies (7, 25–27) have shown that for multiple primary lung adenocarcinoma, lobectomy should be performed as far as possible for the primary lesion, while sublobar resection (segmentectomy or wedge resection) can be performed flexibly for the secondary lesion, especially for patients with multiple bilateral lung lesions, which ensures adequate distance between the tumor margins and maximizes preservation of more lung function. Nakata and colleagues reported (25) that 26 patients with SMPLA, only 5 patients underwent lobectomy alone, and the 3-year OS and DFS were 92.9% and 77.9%, respectively. Ishikawa and colleagues also found (7) that 93 patients with SMPLA, sublobar resection was used during surgery in 58% of patients, and the 3-year OS and RFS were 93.6% and 87% respectively. In the current study, since most of lesions were distributed in different lobes, we tried to adopt a combined sublobar resection approach during surgery, and the OS and RFS at 3 years reached 94.4% and 86%, respectively, which was comparable to the results of the above study.

There are several limitations to our study. First, it is a retrospective study and selection bias cannot be avoided. Second, it is a single-center study with a small sample size, which needs to be further confirmed by multicenter study with larger sample size. However, our study is currently the largest cohort of SMPLA with different EGFR mutations. Third, we did not test other tumor's driver genes, such as KRAS, ALK, ROS1, and BRAF. A whole genome sequencing would be more accurate to identify the source of multiple tumors. Finally, the follow-up time is not long enough to appropriately assess long-term survival. In future studies, we will provide longer-term follow-up data.

In summary, a high discordance of EGFR mutations were identified between tumors in patients with SMPLA, so the detection of EGFR mutation may be used routinely to prevents unnecessary adjuvant treatment for patients with histologically similar synchronous primary lung cancers. Synchronous multiple

TABLE 4 | Distribution of EGFR mutations in 161 tumors of 70 patients.

Variables	Total	WT	EGFR+	P-value	L858R	19Del	Other[a]	P-value
Different radiology								
pGGO	**68**	**34 (50.0)**	**34 (50.0)**		**15 (44.1)**	**14 (41.2)**	**5 (14.7)**	
mGGO	**68**	**13 (19.1)**	**55 (80.9)**		**33 (60.0)**	**14 (25.5)**	**8 (14.5)**	
SN	**25**	**6 (24.0)**	**19 (76.0)**	**0.000**	**9 (47.4)**	**5 (26.3)**	**5 (26.3)**	**0.386**
Different histology								
AIS	**36**	**22 (61.1)**	**14 (38.9)**		**7 (50.0)**	**5 (35.7)**	**2 (14.3)**	
MIA	**29**	**13 (44.8)**	**16 (55.2)**		**8 (50)**	**5 (31.3)**	**3 (18.7)**	
IAC	**96**	**18 (18.7)**	**78 (81.3)**	**0.000**	**42 (53.8)**	**23 (29.5)**	**13 (16.7)**	**0.959**
Different size								
≤10mm	**70**	**34 (48.6)**	**36 (51.4)**		**17 (47.2)**	**12 (33.3)**	**7 (19.5)**	
<10mm, ≤20mm	**59**	**16 (27.1)**	**43 (72.9)**		**24 (55.8)**	**14 (32.6)**	**5 (11.6)**	
>20mm	**32**	**3 (9.4)**	**29 (90.6)**	**0.000**	**16 (55.2)**	**7 (24.1)**	**6 (20.7)**	**0.749**

SN, solid nodule; pGGO, pure ground-glass opacity; mGGO, mixed ground-glass opacity; AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma; IAC, invasive adenocarcinoma; [a]Refers to other rare mutations including L861Q, G719X, 20Ins and T790M.

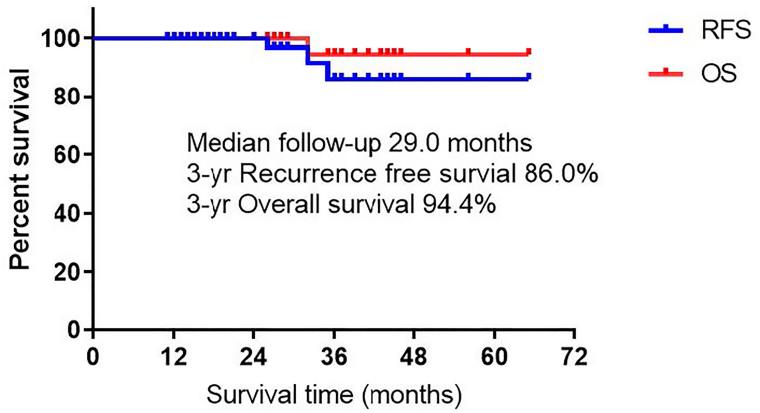


FIGURE 2 | Recurrence free survival and overall survival of all patients.

lung adenocarcinomas with predominantly multiple GGO should be considered as SMPLA, and surgery may be aggressively performed for these patients due to a good prognosis.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the institutional review board of Tongji Medical College of Huazhong University of Science and Technology. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RQ, YC, and XF contributed to the design of the study and the performing of the procedure. RQ, FY, and DT acquired and analyzed the data. RQ drafted the manuscript. RQ, FY, DT, and XF revised and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Tongji Hospital Clinical Research Flagship Program (No. 2019CR107).

REFERENCES

1. Murphy SJ, Aubry MC, Harris FR, Halling GC, Johnson SH, Terra S, et al. Identification of Independent Primary Tumors and Intrapulmonary Metastases Using DNA Rearrangements in Nonsmall-Cell Lung Cancer. *J Clin Oncol* (2014) 32:4050–8. doi: 10.1200/JCO.2014.56.7644
2. Chang YL, Wu CT, Lee YC. Surgical Treatment of Synchronous Multiple Primary Lung Cancers: Experience of 92 Patients. *J Thorac Cardiovasc Surg* (2007) 134:630–7. doi: 10.1016/j.jtcvs.2007.06.001
3. De Leyn P, Moons J, Vansteenkiste J, Verbeken E, Van Raemdonck D, Nafteux P, et al. Survival After Resection of Synchronous Bilateral Lung Cancer. *Eur J Cardiothorac Surg* (2008) 34:1215–22. doi: 10.1016/j.ejcts.2008.07.069
4. Finley DJ, Yoshizawa A, Travis W, Zhou Q, Seshan VE, Bains MS, et al. Predictors of Outcomes After Surgical Resection of Synchronous Primary Lung Cancers. *J Thorac Oncol* (2010) 5:197–205. doi: 10.1097/JTO.0b013e3181c814c5
5. Bak SH, Lee HY, Kim JH, Um SW, Kwon OJ, Han J, et al. Quantitative CT Scanning Analysis of Pure Ground-Glass Opacity Nodules Predicts Further CT Scanning Change. *Chest* (2016) 149:180–91. doi: 10.1378/chest.15-0034
6. National Lung Screening Trial Research Team, Church TR, Black WC, Aberle DR, Berg CD, Clingan KL, et al. Results of Initial Low-Dose Computed Tomographic Screening for Lung Cancer. *N Engl J Med* (2013) 368:1980–91. doi: 10.1056/NEJMoa1209120
7. Ishikawa Y, Nakayama H, Ito H, Yokose T, Tsuboi M, Nishii T, et al. Surgical Treatment for Synchronous Primary Lung Adenocarcinomas. *Ann Thorac Surg* (2014) 98:1983–8. doi: 10.1016/j.athoracsur.2014.07.006
8. Martini N, Melamed MR. Multiple Primary Lung Cancers. *J Thorac Cardiovasc Surg* (1975) 70:606–12. doi: 10.1016/S0022-5223(19)40289-4
9. Kozower BD, Larner JM, Detterbeck FC, Jones DR. Special Treatment Issues in Non-Small Cell Lung Cancer: Diagnosis and Management of Lung Cancer, 3rd Ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* (2013) 143:e369S–99S. doi: 10.1378/chest.12-2362
10. Liu J, Mao G, Li Y, Tao L, Wang W, Peng X, et al. Targeted Deep Sequencing Helps Distinguish Independent Primary Tumors From Intrapulmonary Metastasis for Lung Cancer Diagnosis. *J Cancer Res Clin Oncol* (2020) 146 (9):2359–67. doi: 10.1007/s00432-020-03227-5
11. Zheng R, Shen Q, Mardekian S, Solomides C, Wang ZX, Evans NR 3rd. Molecular Profiling of Key Driver Genes Improves Staging Accuracy in Multifocal Non-Small Cell Lung Cancer. *J Thorac Cardiovasc Surg* (2020) 160:e71–9. doi: 10.1016/j.jtcvs.2019.11.126
12. Haratake N, Takenoyama M, Edagawa M, Shimamatsu S, Toyozawa R, Nosaki K, et al. A Case of Different EGFR Mutations in Surgically Resected

Synchronous Triple Lung Cancer. *J Thorac Dis* (2018) 10:E255–9. doi: 10.21037/jtd.2018.03.105
13. Sakai H, Saji H, Kimura H, Tsuda M, Wakiyama Y, Miyazawa T, et al. Different EGFR Gene Mutations in Two Patients With Synchronous Multiple Lung Cancers: A Case Report. *Thorac Cancer* (2018) 9:189–92. doi: 10.1111/1759-7714.12554
14. Yang Y, Xie X, Jiang G, Liu H. Different Driver Gene Mutations in Patients With Synchronous Multiple Primary Lung Cancers: A Case Report. *J Cardiothorac Surg* (2020) 15:196. doi: 10.1186/s13019-020-01178-z
15. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Classification of Lung Adenocarcinoma. *J Thorac Oncol* (2011) 6:244–85. doi: 10.1097/JTO.0b013e318206a221
16. Qu R, Hao Z, Zhang Y, Bie L, Fu X, Zhang N. Single-Center Experience of Simultaneous Bilateral Uni-Portal Video-Assisted Thoracoscopic Surgery for Multiple Ground-Glass Opacities. *J Cardiothorac Surg* (2020) 15:69. doi: 10.1186/s13019-020-01107-0
17. Qu R, Tu D, Ping W, Cai Y, Zhang N, Fu X. Surgical Outcomes of One-Stage Resection for Synchronous Multiple Primary Lung Adenocarcinomas With No Less Than Three Lesions. *J Cardiothorac Surg* (2021) 16:265. doi: 10.1186/s13019-021-01647-z
18. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71:209–49. doi: 10.3322/caac.21660
19. Liu M, He WX, Song N, Yang Y, Zhang P, Jiang GN. Discrepancy of Epidermal Growth Factor Receptor Mutation in Lung Adenocarcinoma Presenting as Multiple Ground-Glass Opacities. *Eur J Cardiothorac Surg* (2016) 50:909–13. doi: 10.1093/ejcts/ezw113
20. Kobayashi Y, Mitsudomi T, Sakao Y, Yatabe Y. Genetic Features of Pulmonary Adenocarcinoma Presenting With Ground-Glass Nodules: The Differences Between Nodules With and Without Growth. *Ann Oncol* (2015) 26:156–61. doi: 10.1093/annonc/mdu505
21. Yatabe Y, Borczuk AC, Powell CA. Do All Lung Adenocarcinomas Follow a Stepwise Progression? *Lung Cancer* (2011) 74:7–11. doi: 10.1016/j.lungcan.2011.05.021
22. Sun F, Xi J, Zhan C, Yang X, Wang L, Shi Y, et al. Ground Glass Opacities: Imaging, Pathology, and Gene Mutations. *J Thorac Cardiovasc Surg* (2018) 156:808–13. doi: 10.1016/j.jtcvs.2018.02.110
23. Li R, Li X, Xue R, Yang F, Wang S, Li Y, et al. Early Metastasis Detected in Patients With Multifocal Pulmonary Ground-Glass Opacities (GGOs). *Thorax* (2018) 73:290–2. doi: 10.1136/thoraxjnl-2017-210169
24. Ye C, Wang J, Li W, Chai Y. Novel Strategy for Synchronous Multiple Primary Lung Cancer Displaying Unique Molecular Profiles. *Ann Thorac Surg* (2016) 101:e45–7. doi: 10.1016/j.athoracsur.2015.06.042
25. Nakata M, Sawada S, Yamashita M, Saeki H, Kurita A, Takashima S, et al. Surgical Treatments for Multiple Primary Adenocarcinoma of the Lung. *Ann Thorac Surg* (2004) 78:1194–9. doi: 10.1016/j.athoracsur.2004.03.102
26. Zhang Y, Wang Y, Lv C, Shu X, Wang J, Yang Q. Clinical Analysis of 56 Cases of Simultaneous Bilateral Video-Assisted Thoracoscopic Surgery for Bilateral Synchronous Multiple Primary Lung Adenocarcinoma. *J Thorac Dis* (2018) 10:6452–7. doi: 10.21037/jtd.2018.11.10
27. Chen TF, Xie CY, Rao BY, Shan SC, Zhang X, Zeng B, et al. Surgical Treatment to Multiple Primary Lung Cancer Patients: A Systematic Review and Meta-Analysis. *BMC Surg* (2019) 19:185. doi: 10.1186/s12893-019-0643-0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Oncology

ORIGINAL RESEARCH
published: 02 February 2022
doi: 10.3389/fonc.2022.772770

Three-Dimensional Convolutional Neural Network-Based Prediction of Epidermal Growth Factor Receptor Expression Status in Patients With Non-Small Cell Lung Cancer

*Xuemei Huang[1], Yingli Sun[1], Mingyu Tan[1], Weiling Ma[1], Pan Gao[1], Lin Qi[1], Jinjuan Lu[1], Yuling Yang[1], Kun Wang[1], Wufei Chen[1], Liang Jin[1], Kaiming Kuang[2], Shaofeng Duan[3] and Ming Li[1]**

[1] Department of Radiology, Huadong Hospital Affiliated With Fudan University, Shanghai, China, [2] Dianei Technology, Shanghai, China, [3] Precision Health Institution, GE Healthcare, Shanghai, China

OPEN ACCESS

Edited by:
Xue Qin Yu,
The University of Sydney, Australia

Reviewed by:
Stephen Yip,
Janssen Pharmaceuticals, Inc., United States
Xuan Wu,
Peking University Shenzhen Hospital, China

***Correspondence:**
Ming Li
minli77@163.com

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *08 September 2021*
Accepted: *10 January 2022*
Published: *02 February 2022*

Citation:
Huang X, Sun Y, Tan M, Ma W, Gao P, Qi L, Lu J, Yang Y, Wang K, Chen W, Jin L, Kuang K, Duan S and Li M (2022) Three-Dimensional Convolutional Neural Network-Based Prediction of Epidermal Growth Factor Receptor Expression Status in Patients With Non-Small Cell Lung Cancer. Front. Oncol. 12:772770. doi: 10.3389/fonc.2022.772770

Objectives: EGFR testing is a mandatory step before targeted therapy for non-small cell lung cancer patients. Combining some quantifiable features to establish a predictive model of EGFR expression status, break the limitations of tissue biopsy.

Materials and Methods: We retrospectively analyzed 1074 patients of non-small cell lung cancer with complete reports of *EGFR* gene testing. Then manually segmented VOI, captured the clinicopathological features, analyzed traditional radiology features, and extracted radiomic, and deep learning features. The cases were randomly divided into training and test set. We carried out feature screening; then applied the light GBM algorithm, Resnet-101 algorithm, logistic regression to develop sole models, and fused models to predict EGFR mutation conditions. The efficiency of models was evaluated by ROC and PRC curves.

Results: We successfully established $Model_{clinical}$, $Model_{radiomic}$, $Model_{CNN}$ (based on clinical-radiology, radiomic and deep learning features respectively), $Model_{radiomic+clinical}$ (combining clinical-radiology and radiomic features), and $Model_{CNN+radiomic+clinical}$ (combining clinical-radiology, radiomic, and deep learning features). Among the prediction models, $Model_{CNN+radiomic+clinical}$ showed the highest performance, followed by $Model_{CNN}$, and then $Model_{radiomic+clinical}$. All three models were able to accurately predict EGFR mutation with AUC values of 0.751, 0.738, and 0.684, respectively. There was no significant difference in the AUC values between $Model_{CNN+radiomic+clinical}$ and $Model_{CNN}$. Further analysis showed that $Model_{CNN+radiomic+clinical}$ effectively improved the efficacy of $Model_{radiomic+clinical}$ and showed better efficacy than $Model_{CNN}$. The inclusion of clinical-radiology features did not effectively improve the efficacy of $Model_{radiomic}$.

Conclusions: Either deep learning or radiomic signature-based models can provide a fairly accurate non-invasive prediction of EGFR expression status. The model combined both features effectively enhanced the performance of radiomic models and provided

marginal enhancement to deep learning models. Collectively, fusion models offer a novel and more reliable way of providing the efficacy of currently developed prediction models, and have far-reaching potential for the optimization of noninvasive EGFR mutation status prediction methods.

Keywords: NSCLC, EGFR, tomography, radiogenomics, deep learning, machine learning

INTRODUCTION

Lung cancer is the leading cause of cancer-related deaths, with incidence and mortality rates of approximately 11.4% and 18%, respectively, and is the second-highest incidence rate in the world (1). Non-small cell lung cancer is the main pathological form and accounts for approximately 80-90% of all lung cancers (2). Targeted therapy has become one of the first-line standard treatments for non-small cell lung cancer patients; because this form of treatment can effectively improve their prognosis, prolong the PFS and OS, compared with traditional means of treatment, like chemotherapy (3–6). In patients with non-small cell lung cancer, EGFR is responsible for approximately 10-20% of all and is the most predominant driver mutations target for targeted therapy (7). As a consequence, EGFR-TKI therapy plays a pivotal role in the targeted therapy of patients with non-small cell lung cancer.

Prior to EGFR-TKI treatment, it is essential to perform EGFR genetic testing to clarify the presence of EGFR mutations. There are several methods that can be used to detect EGFR mutations, including tissue biopsy, liquid biopsy, and radiogenomics.

Histopathological biopsy has been the gold standard in terms of high sensitivity and specificity in clinical disease and genetic diagnosis. However, it still has the following restrictions: 1. High sample size threshold, requiring at least 20% of tumor cells in the sample to be detectable (8). 2. As the tumor genotype itself possesses heterogeneity (9–11), while part of the samples are taken from puncture biopsies, so there is a risk of sampling bias, which means that the gene mutation status detection result may not correspond to the authentic condition and is not representative of the whole gene expression profile of the cancer spot. 3. Because of heterogeneity of neoplastic cell genetic status, disease progression or drug resistance commonly occurs in terminal period of the disease, so that re-biopsy is necessitated to evaluate the disease and clarify if a drug resistant mutation such as T790M (12) has evolved to instruct subsequent treatment, yet the biopsy is an invasive operation with complications including pneumothorax and bleeding, and often not feasible due to the patient's physiological issues in terminal course of the disease, thus blocking the personalized health maintenance strategy. 4. More expensive, with higher standards of material storage and instrumentation, which is not conducive to applying and promoting in certain impoverished and remote areas.

Liquid biopsy refers to the extraction of tumor gene-carrying agents from body fluids, such as Circulating tumor-derived DNA, cell-free DNA, etc., for detecting the relevant genetic alterations, and it has the merits of real-time detection and minor invasion, however, due to the existence of tumor spatial heterogeneity, it may not be capable of accurate localization or representing the true mutation level in the whole tumor; besides, in the early stage, there are often no circulating tumor cells in body fluids, and their concentration is susceptible to influence, resulting in an insufficient sample size. Presently, cell-free DNA is the only liquid biopsy marker recommended for insufficient volume of pathology biopsies or to monitor the presence of EGFR T790M mutations with disease progression or drug resistance (13, 14). Moreover, a recent study (15) indicates that the sensitivity and specificity of this technique are poor and that the practical use of this method remains controversial.

Hence enabling a holistic and comprehensive analysis of the lesion by surmounting the obstacle posed by genetic heterogeneity is now a much desirable claim.

Regarding the aforementioned downsides of tissue biopsy pathology and liquid biopsy, researchers have exploited the advancing artificial intelligence to provide a technology with promising clinical applicability - radiogenomics (16–19). It is a group of imaging biomarkers that can offset the constraints of tissue biopsies and liquid biopsies by effectively and non-invasively projecting the mutational status of genes such as EGFR and ALK *via* artificial intelligence methodology, enabling high-throughput molecular biological information, as tumor heterogeneity and genotype, which is not visible to the naked eye, and converting them into digital signals (deep learning features or radiomics features), quantifying and characterizing them to facilitate disease diagnosis as well as monitoring and guiding targeted therapy decision-making. Several researchers have reported that radiogenomics represents a promising application for EGFR gene detection. Both deep learning models (20, 21) and radiomic models (22–24) have been shown to be more precise in predicting the mutational status of EGFR. However, most studies have applied deep learning and radiomic features in an independent manner; far fewer studies have attempted to combine these two features. A previous study reported the successful creation of an EGFR mutation prediction model based on the fusion of these two

Abbreviations: NSCLC, Non-Small Cell Lung Cancer; *EGFR*, Epidermal Growth Factor Receptor; TKI, Tyrosine Kinase Inhibitors; PFS, Progression-Free Survival; OS, Overall Survival; CT, Computed Tomography; MM, Millimeter; WHO, World Health Organization; PCR, Polymerase Chain Reaction; DICOM, Digital Imaging Communications in Medicine; VOI, Volume Region of Interest; ICC, Intraclass Correlation Coefficient; Light GBM, Light Gradient Boosting Machine; HU, Hounsfield Unit; Res-Net, Residual Network; ROC, Receiver Operating Characteristic Curve; PRC, Precision Recall Curve; AUC, Area Under the Curve; GLCM, Gray-Level Co-occurrence Matrix; *ALK*, Anaplastic Lymphoma Kinase; *KRAS*, Kirsten Rat Sarcoma Viral Oncogene Homolog.

features (25). However, this study only included patients with solid lung adenocarcinoma. Furthermore, some of the images used were thick; this may have led to the loss of valuable features. Moreover, the EGFR mutation sites described in this previous study only contained exons 19 and 21. This is a concern because the ground-glass component within a cancer site maybe can provide more heterogeneous information than the solid component (26).

In the present study, we aimed to investigate and validate whether a prediction model incorporating deep learning features and radiomic features can improve the performance of the current mainstream models for the non-invasive prediction of EGFR mutations. To expand the application of radiomic features and deep learning features for non-invasive gene detection, we recruited a large number of patients with ground glass non-small cell lung cancers and used thin-layer images to avoid or minimize the loss of effective features.

MATERIALS AND METHODS

Figure 1 shows a schematic for how the models were constructed.

Population and Clinicopathological Data

Before initiating the research, we derived the AUC value of the radiogenomic model from that of several previous studies, which was about 0.70-0.95, and made a sample size estimation based on this data, which resulted in a predicted maximum number of 104 people needed. Later, after reminded by deep learning experts, and given the demand for large data samples for deep learning, it was decided to extend the sample on the pre-estimated sample size. We ultimately retrospectively recruited patients with pathologically confirmed primary non-small cell lung cancer between 4th June 2019 and 21st January 2021 at the Huadong Hospital, Fudan University, Shanghai, China. All patients were screened according to strict inclusion and exclusion criteria; this process led to the inclusion of 1074 eligible patients. The inclusion criteria were as follows: (1) detailed EGFR gene test reports were available, (2) the interval between chest CT examination and surgery was within 1 month, and (3) pathological samples were obtained from surgically resected specimens. The exclusion criteria were as follows: (1) image layer thickness greater than 1.5 mm, (2) images with severe motion artifacts or conditions such as pleural effusion or obstructive pneumonia that may affect detailed observation, (3) preoperative history of tumors or a history of lung surgery, and (4) an inability to convert image format or extract features for unknown reasons. For each patient, we collated a complete range of clinicopathological data, including age, gender, smoking history, invasive degree, and EGFR mutation status. The basic principle of the training/test split is to maintain a general fraction of positive samples in each subset. We used the train_ test_ split function in Scitkit-learn 0.24.2 to perform a random selection of training/test data while maintaining roughly the same proportion of positives/negatives in both subsets, and to guarantee reproducibility, we kept the seed of the random number generator fixed at 42, which is a prevalent alternative among deep learning researchers. All cases were randomly divided into a training set (770 cases) and a test set (304 cases).

CT Instrument and Parameters

All patients were scanned with a GE Discovery CT750HD or LightSpeed VCT or Somatom Sensation 16 CT system, operating with the following parameters: tube voltage: 120 kV; tube current: 200 mA; reconstruction algorithm: STND/medium sharp; and layer thickness: 1.00/1.25/1.5 mm. Three apparatus distribution for Discovery: VCT: Somatom (training set-340:184:246; test set-135:83:86) The scan phase was set to the

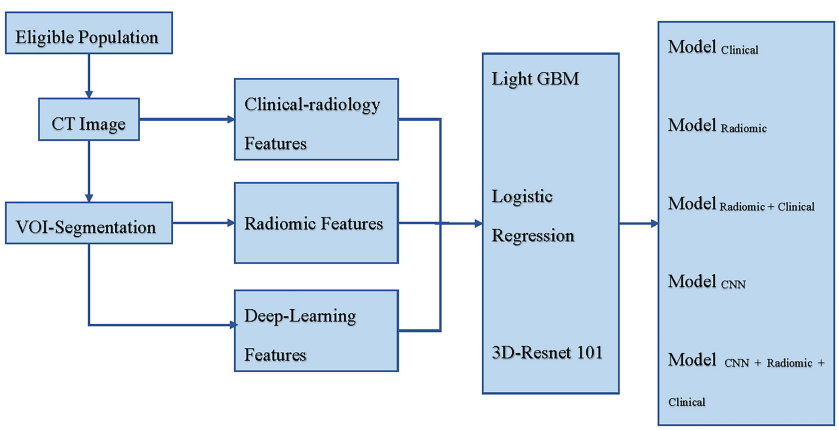


FIGURE 1 | Schematic for the models' construction. CT, Computed Tomography; VOI, Volume Region of Interest; Light GBM, Light Gradient Boosting Machine; Res-Net, Residual Network; Modelclinical incorporated clinical-radiology features, Modelradiomic incorporated radiomic features, Modelradiomic+clinical combined clinical-radiology and radiomic features, ModelCNN incorporated deep learning features, and ModelCNN+radiomic+clinical combined clinical-radiology, radiomic, and deep learning features.

deep inspiratory phase and the patient was scanned in the supine position. Images were acquired in the DICOM format. Further details of the parameters used for CT are shown in **Supplementary Table 1**.

Histopathology and the Diagnosis of EGFR Status

The histopathological type of non-small cell lung cancer was identified by our diagnostic pathologists for secondary diagnosis using the 2011 international and multidisciplinary classification guidelines proposed by the International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society (27) and the World Health Organization (WHO) 2015 guidelines for lung cancer classification (28). The mutation status of EGFR exons 18, 19, 20 and 21 (which are associated with drug targets) was detected using a real-time fluorescent PCR-based amplification refractory mutation system and a human EGFR gene mutation real-time reverse transcription-polymerase chain reaction diagnostic kit (AmoyDx, Xiamen, China).

VOI Segmentation and Radiology Features

First, the pixels in the raw DICOM images were uniformly transformed to a layer thickness of 1 mm. Then, the VOI of the cancer was manually segmented by a junior diagnostician (Reader 1) using the open-source software 3D-slicer (https://www.slicer.org/) ensuring that large blood vessels and fibrous connective tissue was avoided during contouring. A secondary manual correction was performed by a senior physician (Reader 2). Another senior diagnostician (Reader 3) analyzed and recorded the CT radiology features of the tumor while remaining blinded to the EGFR mutation status and pathological subtypes. Reader 3 recorded a range of data, including location, cancer density, border, vacuole sign, air bronchogram sign, spiculation sign, lobulation sign, halo sign, vascular alteration, pleural indentation, and umbilicated indentation. In case of disagreement, a second evaluation was performed by another senior diagnostician (Reader 4); the results were recorded after discussion and agreement. All images were observed with a window position of -500 HU and a window width of 1500 HU. In the following features description, for the sake of brevity, we merge the radiology features with the clinical features, and use the description of the clinical features uniformly.

Analysis of Radiomic Features

The outlined VOIs were placed into Pyradiomics (29) (version 3.0 software) to extract radiomic features. Pyradiomics is an open-source python package for extracting radiomic features from medical imaging.

Reproducibility Analysis

To assess the reproducibility and stability of the radiomic features, 60 patients were randomly selected by the diagnostician (Reader 1) for secondary manual segmentation of the tumor VOI after one month. The radiomic features were extracted and subjected to ICC analysis; features with an ICC index≥0.95 were selected for subsequent model construction.

Clinical and Radiomic Models

To further identify redundant features and improve the performance of the radiomic model, we re-screened the initial radiomic features by considering mutual information between each feature and the mutation status of the EGFR gene. The mutual information between two random variables is a non-negative value that measures the dependence between the two variables (30). This function relies on a non-parametric approach based on entropy estimation from K-nearest neighbor distances and can be used for the univariate selection of features. Ultimately, we filtered the top 10% of features with the highest mutual information in the training set to develop the model. Then, we retained the same 10% of features in the test set to evaluate model performance. Based on the screened radiomic features and clinical features, we established $Model_{radiomic}$ and a fusion model ($Model_{radiomic+clinical}$) using the Light GBM algorithm (31). To avoid overfitting, during model construction, we adjusted several hyperparameters, including learning rate, data down-sampling ratio, feature down-sampling ratio, and L1/L2 regularization strength. The learning rate was tuned before the steady convergence of the training and validation losses of the model was observed. Intensity of overfitting prevention enhances when we decrease the data down-sampling rate, feature down-sampling rate, or augment the L1/L2 regularization strength.

Deep Learning Model

Both the original CT images and the mask of the VOI were resampled to a space-occupying 1 mm × 1 mm × 1 mm. Next, we counted the spanning distribution of the cancer in three dimensions, and selected 64 mm × 64 mm × 64 mm as the input size for deep learning to ensure that the cropped input size could cover the extent of all lung nodules. The HU values of this patch were processed using the clip of the lung window [(-1000, 400)] and subjected to the minimum-maximum normalization process. Next, the resultant data were imported into the Ampyx 3D ResNet101 network to facilitate the creation of $Model_{CNN}$, a model that featured only deep learning features. 3D ResNet101 (32) is a well-characterized and broad applicable neural network in the field of deep learning, and remains considered as a strong comparative baseline in computational vision research. Compared to its successor, its network is relatively simplistic, which further alleviates overfitting and thus enables a more robust model ultimately. The model was optimized with AdamW (33) with a maximum learning rate of 0.001. We also used a cosine annealing schedule (34) to gradually reduce the model to 10^{-6} within 500 epochs. To further suppress overfitting and enhance the robustness of the model, we performed data augmentation using random rotation, random flip, and mix-up (35) with an α of 0.2. Since the objective of this study was not to innovate new neural network structures, the hyperparameters of this ResNet101 model were adjusted following the configuration given in the Torch Vision Python package.

Fusion of Clinical-Radiomic-Deep Learning Features Model

Since the deep learning features and clinical/radiomic features are totally different in terms of both data distribution and expressed meaning, and the number of filtered clinical/radiomic features is larger than that of deep learning features, the weight of clinical/radiomic features tends to be greater if the features are simply combined, and the model performance is poor. Therefore, we finally opted to model the prediction probability of $Model_{CNN}$ and that of $Model_{radiomic+clinical}$, and constructed a metamodel $Model_{CNN+radiomic+clinical}$ using logistic regression. Essentially, we perform 5-fold cross-validation on the $Model_{CNN}$ and the $Model_{radiomic+clinical}$ respectively in the training set, and build a logistic regression $Model_{CNN+radiomic+clinical}$ by weighting the probabilities calculated from the two models.

Model Evaluation

Next, the ROC curve, AUC value, and PRC curve were used to evaluate the predictive performance of each model. To verify whether the fusion model performs better than the sole model and whether the improvement in model performance is statistically significant, the De-long test is applied to compare the performance variation of each model.

Statistical Analysis

This research was carried out with Python (version 3.8.10). Modeling of radiomics features, clinical features, and the concatenation of both was done using Light GBM (version3.2.1). CNN experiments were conducted using PyTorch (version1.8.1). The logistic regression model fusing clinical, radiomic, and deep learning features were provided by Scitkit-Learn (version0.24.2.). DeLong tests were done in MedCalc (version20.0009). The sample size was calculated in PASS 15 (Power: 0.90; Alpha: 0.05; AUC1:0.7; Two-Sided).Univariate analysis and multivariate logistic analysis using SPSS (version23.0). The normality distribution of the continuous variables was verified with the Kolmogorov-Smirnov test($P<0.001$). Continuous variables were analyzed using Mann-Whitney U test. Categorical variables were analyzed using chi-square tests or Fisher's exact test. p-values less than 0.05 were considered statistically significant.

RESULTS

A total of 1074 eligible non-small cell lung cancer cases were enrolled in this study, including 527 wild-type EGFR cases and 547 EGFR mutant cases; there were 443 males and 631 females. Analysis of between-group discrepancy showed that there was no significant difference in the clinical-radiology characteristics when compared between the training and test sets, as detailed in **Supplementary Table 2**. The distribution of the clinical-radiology characteristics of EGFR mutant-type and wild-type cases within the training set is shown in **Table 1**. Screening of the training set revealed that six items (gender, age, invasive degree, cancer density, vacuole sign, and smoking history) were all independent predictors for EGFR mutation. Detailed statistics of clinical characteristics are shown in **Table 2**. In contrast, location, border, air bronchogram sign, spiculation sign, lobulation sign, halo sign, vascular alteration, pleural indentation, and umbilicated indentation, could not specifically identify EGFR mutation. For each case, 1218 radiomic features were extracted from the VOI; ICC analysis yielded a mean correlation coefficient of 0.96 ± 0.07. Subsequently, 243 radiomic features with coefficients <0.95 were excluded, and the top 10% of the radiomic features with the highest mutual information were identified, and used to build the model. Finally, six clinical features and 108 radiomic features were used to build the predictive models. The top 20 radiomic features selected are shown in **Supplementary Table 3**.

Next, we successfully built five prediction models: $Model_{CNN+radiomic+clinical}$, $Model_{CNN}$, $Model_{radiomic+clinical}$, $Model_{radiomic}$, and $Model_{clinical}$. The performance of each model was verified in the test set, as shown in **Figure 2**. In the test set, the most effective prediction model, as based on the ROC curve, was $Model_{CNN+radiomic+clinical}$ with an AUC of 0.751; this was followed by $Model_{CNN}$, $Model_{radiomic+clinical}$, and finally $Model_{clinical}$. Our analysis showed that deep learning models and radiomic models both can predict EGFR mutations with the best levels of accuracy. $Model_{CNN+radiomic+clinical}$, which featured both deep learning and radiomic features, showed more effective improvement than the mainstream radiomic models ($Model_{radiomic+clinical}$ and $Model_{radiomic}$), with p-values of 0.0067 and 0.0063, respectively. Although the Delong Test revealed that the difference in efficacy between the two models was not statistically significant, detailed analysis of the ROC and PRC curves showed that the fusion model ($Model_{CNN+radiomic+clinical}$) was slightly more effective than the deep learning model ($Model_{CNN}$). The Delong Test also showed that the difference in efficacy between $Model_{radiomic+clinical}$ and $Model_{radiomic}$ was also not statistically significant, and that the addition of clinical information did not enhance the efficacy of $Model_{radiomic}$ ($p = 0.876$).

DISCUSSION

In this study, we developed a fusion model for predicting EGFR mutation levels in 1074 patients with non-small cell lung cancer by analyzing the clinical, radiology, radiomic, and deep learning features. The value of the combined model ($Model_{CNN+radiomic+clinical}$) was more efficient than models based on radiomic or deep learning features alone, particularly those based on radiomic features. The general objectives of this study were to investigate the feasibility of improving the efficacy of prevalent models to date (predictive models based on radiomic or deep learning features alone) and to provide a new approach for constructing models for non-invasive detection of EGFR mutations, a and there may be a promise for future extensions to develop models for predicting other genotypes or other tasks.

Tumor heterogeneity (36–38) is the leading driver of drug resistance and disease progression in the post-EGFR-TKI treatment course, and the underlying factor that liquid biopsy and puncture pathology may not reflect the overall truly mutated

TABLE 1 | The distribution of clinical-radiology features for EGFR mutant and wild type cases in the training set.

Characteristics	EGFR wild	EGFR mutation	*p-value*
Gender			***0.004****
Male	176 (46.6)	142 (36.2)	
Female	202 (53.4)	250 (63.8)	
Age	57.5 (18.0)	60.0 (16.0)	***0.004****
Invasive Degree			***<0.001****
Non-invasive	64 (16.9)	27 (6.9)	
Micro-invasive	158 (41.8)	131 (33.4)	
Invasive	156 (41.3)	234 (59.7)	
Location			***0.248***
RUL	137 (36.2)	145 (37.0)	
RML	23 (6.1)	39 (9.9)	
RLL	64 (16.9)	59 (15.1)	
LUL	98 (25.9)	103 (26.3)	
LLL	56 (14.8)	46 (11.7)	
Cancer density			***<0.001****
Pure GGO	72 (19.0)	29 (7.4)	
Mixed GGO	221 (58.5)	313 (79.8)	
Solid	85 (22.5)	50 (12.8)	
Border			***0.121***
Well-define	271 (71.7)	254 (64.8)	
Less-define	61 (16.1)	79 (20.2)	
Ill-define	46 (12.2)	59 (15.1)	
Vacuolation			***<0.001****
Present	112 (29.6)	202 (51.5)	
Absent	266 (70.4)	190 (48.5)	
Air Bronchogram			***<0.001****
Present	120 (31.7)	194 (49.5)	
Absent	258 (68.3)	198 (50.5)	
Spiculation			***0.025****
Short	77 (20.4)	99 (25.3)	
Deep	26 (6.9)	32 (8.2)	
Mixed	60 (15.9)	81 (20.7)	
Absent	215 (56.9)	180 (45.9)	
Lobulation			***0.133***
Shallow	125 (33.1)	114 (29.1)	
Deep	4 (1.1)	11 (2.8)	
Mixed	248 (65.6)	263 (67.1)	
Absent	1 (0.3)	4 (1.0)	
Halo			***0.006****
Present	46 (12.2)	76 (19.4)	
Absent	332 (87.8)	316 (80.6)	
Vascular- Alteration			***0.675***
Present	191 (50.5)	204 (52.0)	
Absent	187 (49.5)	188 (48.0)	
Pleural- Indentation			***<0.001****
Present	133 (35.2)	187 (47.7)	
Absent	245 (64.8)	205 (52.3)	
Umbilicated- Indentation			***0.001****
Present	29 (7.7)	59 (15.1)	
Absent	349 (92.3)	333 (84.9)	
Smoke History			***0.020****
Yes	172 (45.5)	146 (37.2)	
No	206 (54.5)	246 (62.8)	

*RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe; LUL, left upper lobe; LLL, left lower lobe; GGO, ground glass opacity; Categorical variables (e.g. gender) are expressed by a number (percentage), continuous variables (e.g. age) are expressed by the Median (interquartile range). *p<0.05 (significant), P-values taken with three decimal places equal to 0.000 are expressed as <0.001.*

The bolded values in the left column refer to the clinical-radiological features included in the statistical analysis of this study, and the bolded values in the right column refer to the P values, with P less than 0.05 as the criterion to evaluate whether they are statistically significant and whether they are included in the subsequent statistical sub-analysis. The data are bolded for the purpose of making them more prominent and clear only.

status of the lesion in the process of disease genetic identification, therapeutic efficacy monitoring and follow-up.

However, Radiogenomics can effectively discern the heterogeneous patterns within tumors through artificial intelligence and mathematical statistics, bridging the limitations of pathological biopsies and liquid biopsies and assisting clinicians in conducting more precise clinical decisions. For remote and impoverished area and countries,

TABLE 2 | | Statistical analysis outcome of clinical-radiology characteristics.

Selected Features	Univariate Analysis		Multivariate Analysis	
	Z or $\chi 2$	*P*	Regression coefficient	*P*
Gender	8.481	*0.004*	-0.649	***<0.001***
Age	-2.826	*0.004*	0.015	***0.037***
Invasive Degree	32.923	*<0.001*	-1.158	***<0.001***
Cancer density	42.991	*<0.001*	1.510	***<0.001***
Vacuolation	38.221	*<0.001*	0.571	***0.001***
Air-Bronchogram	25.088	*<0.001*	0.251	*0.165*
Spiculation	9.348	*0.025*	0.313	*0.253*
Halo	7.520	*0.006*	-0.506	*0.051*
Pleural- Indentation	12.418	*<0.001*	0.145	*0.493*
Umbilicated- Indentation	10.352	*0.001*	0.481	*0.093*
Smoke History	5.413	*0.020*	-0.335	***0.038***

Univariate Analysis: Continuous variables were analyzed using Mann-Whitney U test. Categorical variables were analyzed using chi-square tests or Fisher's exact test. Features with bolded numbers of the P-value column are independent predictors.

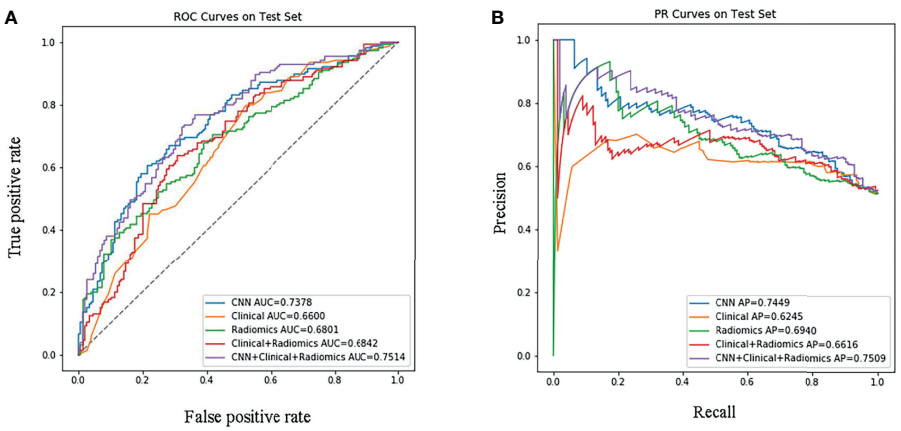


FIGURE 2 | Performance evaluation of the models in the test set. **(A)** Receiver Operating Characteristic curve; **(B)** Precision-Recall curve. 'CNN+Clinical+Radiomic' refers to ModelCNN+radiomic+clinical, 'Clinical+Radiomic' refers to Modelradiomic+clinical, 'Radiomic' refers to Modelradiomic, 'Clinical' refers to Modelclinical, and 'CNN' refers to ModelCNN.

this is an inexpensive, low-cost and efficient genetic diagnostic weapon if the radiogenomic model can be brought to clinical practice successfully by future.

The results of this study confirmed the reliability of radiomic and deep learning models for the non-invasive prediction of EGFR mutation status in lung adenocarcinoma with a high degree of accuracy. In lung adenocarcinoma patients, two previous studies (39, 40)combined both radiomic and clinical features to successfully build a radiomic-clinical model that could efficiently identify EGFR mutant phenotypes from wild types with good AUCs of 0.779 and 0.823. However, two other studies (41, 42) also successfully built a combined radiomic-clinical prediction model but also found that a deep learning feature-based model could also predict EGFR gene mutation status in patients with lung adenocarcinoma in a more accurate manner, achieving AUCs of 0.810 and 0.758. These previous findings are consistent with the results of our current study. However, our present differs from these previous studies in that they predominantly applied radiomic and deep learning features separately to build radiomic-clinical models or deep learning models. In this study, we innovatively developed a fusion prediction model to diagnose EGFR mutations in patients with non-small cell lung cancer by fusing the most widely accepted clinicopathological, radiology, and radiomic features with deep learning features. A previous study published findings for a fusion model that were similar to our present results; the efficacy of this previous fusion model was also more efficient than the radiomic model (AUC: 0.831 *vs* 0.758) (25). Comparing to this study, which enrolled only solid lung adenocarcinoma cases, had incomplete coverage of the mutant site, and used thick layers of images, our study also included a significant number of ground glass type non-small cell lung cancer cases and new radiology features. All of the images used in the present study had a layer thickness of <1.5 mm, thus making our models more realistic to the actual clinical scenario, thus providing more applicable data that could support the wider use of these models clinically.

Our current findings confirm the concept of fusing multiple features to build prediction models to enhance the efficacy of

individual models. We prove that this strategy is feasible and may be applied to the prediction of other genetic targets in the future, and even to other fields, including the identification of benign and malignant nodes, prediction of the degree of infiltration, as well as the prognosis of survival analysis.

Both Clinicopathology features that gender and smoking history, degree of invasion, and morphology features like cancer density and the vacuole sign, were independent predictors for the EGFR mutant phenotype. The present study reconfirmed that EGFR mutant phenotype is more prevalent in women and non-smoking patients (43–45). In addition, the tumor invasion degree and density are highly associated with the EGFR mutation status. The higher the degree of tumor infiltration and density, the more likely the mutation of EGFR will occur. A greater degree of invasion indicates more heterogeneous cells, faster gene duplication and an increased mutation frequency. This is in line with prior research (46) study 1 where the mutation frequency of EGFR was observed to be much larger and more distinct in IAC, than in MIA, AIS, and AAH. Compared to pure ground glass nodules, mixed ground glass nodules and solid nodules with greater density had significantly better EGFR mutation rates, which also is aligned with previous studies (46, 47) posting that the solid component is remarkably sensitive for diagnosing invasiveness and has a superior EGFR mutation profile. Both vacuole sign and age were correlated with EGFR mutation condition, yet unfortunately, this discovery was not in accordance with the results of earlier studies (48–50), probably because our research center specializes in geriatrics, so the population enrolled is mostly elderly, so there might be a sample error, while the studies correlating vacuole sign and EGFR are fewer, both of which have to be further verified by subsequent research.

Two previous studies incorporated two EGFR-related predictors, gender and smoking history, into the construction of a fused clinical-radiomic model; however, the efficacy of the final separate radiomic model was not improved (51, 52). We also found that several radiology features were not significantly correlated with EGFR mutations, including air bronchogram sign, spiculation sign, and lobulation sign. The involvement of relevant features in model construction did not effectively augment the efficacy of the radiomic model. These highly subjective and time-consuming features should be considered carefully in future studies; deletion of these features may help to streamline the development procedures of radiogenomic predictive models.

Some limitations need to be considered. First, this was a retrospective study. Firstly, EGFR frequently merges with tumor suppressor genes mutations (53), like TP53 (incidence >5%), but in the clinical setting, tumor suppressor genes testing is not routinely conducted, thus the genetic data in this study only contains EGFR synapses, and there is no investigation yet to elucidate whether the effect of the remaining co-alteration mutations upon the radiogenomic model, so more information should be collected on combined mutations for rigorous prospective trials in the future. Second, EGFR mutation prevalence is varying across ethnics, such that it is generally of a higher rate for the Asian population than that of the American and European ones (54), hence the model may be more generalizable to Asia; also, there are large regional diversity in lifestyle practices, which may sometimes change the structuring composition of the model, such as clinical features smoking history. This is why in coming future, multi-center, multi-ethnic studies are expected to validate the robustness and generalization power of radiogenomics models. And lastly: in this study, a time-consuming manual segmentation pattern was implemented; the future semi-automatic or fully automatic segmentation mode with deep learning algorithms should be applied to streamline the whole process.

CONCLUSION

Both radiomic models and deep learning models can predict EGFR gene mutation status relatively efficiently and non-invasively. By integrating radiomics and deep learning features, it is possible to build prediction models that can significantly upgrade the performance of the basic radiomic models and help to improve the performance of deep learning models. Models featuring deep learning techniques have the potential for broader application in the non-invasive diagnosis of lung cancer genes mutation.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

This study involved human participants and was reviewed and approved by the Ethics Committee of the Huadong Hospital Affiliated with Fudan University (number:2019K134). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

ML and YS conceived the concept of this study. YY and KW collected the data. PG, MT, WM, LQ, JL, WC, LJ, KK, and SD analyzed the data. XH and YS drafted and revised the manuscript. All authors reviewed the manuscript, and ML made corrections to the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Nature Science Foundation of China (Reference: 61976238; ML), and the

Shanghai "Rising Stars of Medical Talent" and Youth Development Program "Outstanding Youth Medical Talents" (Reference: SHWSRS(2021)_99), and Scientific Research Program of Shanghai Science and Technology Commission (Reference: 20Y11902900).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.772770/full#supplementary-material

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
2. Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman J, Chirieac LR, et al. Non–Small Cell Lung Cancer, Version 5.2017, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* (2017) 15(4):504–35. doi: 10.6004/jnccn.2017.0050
3. Ramalingam SS, Yang JC-H, Lee CK, Kurata T, Kim D-W, John T, et al. Osimertinib as First-Line Treatment of EGFR Mutation–Positive Advanced Non–Small-Cell Lung Cancer. *J Clin Oncol* (2018) 36(9):841–9. doi: 10.1200/jco.2017.74.7576
4. Yoshida K, Yamada Y. Erlotinib Alone or With Bevacizumab as First-Line Therapy in Patients With Advanced non-Squamous non-Small-Cell Lung Cancer Harboring EGFR Mutations (JO25567): An Open-Label, Randomized, Multicenter, Phase II Study. *Transl Lung Cancer Res* (2015) 4 (3):217–9. doi: 10.3978/j.issn.2218-6751.2015.03.04
5. Rosell R, Carcereny E, Gervais R, Vergnenegre A, Massuti B, Felip E, et al. Erlotinib Versus Standard Chemotherapy as First-Line Treatment for European Patients With Advanced EGFR Mutation-Positive Non-Small-Cell Lung Cancer (EURTAC): A Multicentre, Open-Label, Randomised Phase 3 Trial. *Lancet Oncol* (2012) 13(3):239–46. doi: 10.1016/s1470-2045(11)70393-x
6. Maemondo M, Inoue A, Kobayashi K, Sugawara S, Oizumi S, Isobe H, et al. Gefitinib or Chemotherapy for Non-Small-Cell Lung Cancer With Mutated EGFR. *N Engl J Med* (2010) 362(25):2380–8. doi: 10.1056/NEJMoa0909530
7. Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, Wistuba II, et al. Clinical and Biological Features Associated With Epidermal Growth Factor Receptor Gene Mutations in Lung Cancers. *J Natl Cancer Inst* (2005) 97 (5):339–46. doi: 10.1093/jnci/dji055
8. Kalemkerian GP, Narula N, Kennedy EB, Biermann WA, Donington J, Leighl NB, et al. Molecular Testing Guideline for the Selection of Patients With Lung Cancer for Treatment With Targeted Tyrosine Kinase Inhibitors: American Society of Clinical Oncology Endorsement of the College of American Pathologists/International Association for the Study of Lung Cancer/Association for Molecular Pathology Clinical Practice Guideline Update. *J Clin Oncol* (2018) 36(9):911–9. doi: 10.1200/jco.2017.76.7293
9. Skoulidis F, Heymach JV. Co-Occurring Genomic Alterations in Non-Small-Cell Lung Cancer Biology and Therapy. *Nat Rev Cancer* (2019) 19(9):495–509. doi: 10.1038/s41568-019-0179-8
10. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, et al. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N Engl J Med* (2012) 366(10):883–92. doi: 10.1056/NEJMoa1113205
11. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the Evolution of Non-Small-Cell Lung Cancer. *N Engl J Med* (2017) 376(22):2109–21. doi: 10.1056/NEJMoa1616288
12. 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. doi: 10.1016/j.critrevonc.2020.103194
13. Maheswaran S, Sequist LV, Nagrath S, Ulkus L, Brannigan B, Collura CV, et al. Detection of Mutations in EGFR in Circulating Lung-Cancer Cells. *N Engl J Med* (2008) 359(4):366–77. doi: 10.1056/NEJMoa0800668
14. Siravegna G, Marsoni S, Siena S, Bardelli A. Integrating Liquid Biopsies Into the Management of Cancer. *Nat Rev Clin Oncol* (2017) 14(9):531–48. doi: 10.1038/nrclinonc.2017.14
15. Marquette CH, Boutros J, Benzaquen J, Ferreira M, Pastre J, Pison C, et al. Circulating Tumour Cells as a Potential Biomarker for Lung Cancer Screening: A Prospective Cohort Study. *Lancet Respir Med* (2020) 8(7):709–16. doi: 10.1016/s2213-2600(20)30081-3
16. Zhou M, Leung A, Echegaray S, Gentles A, Shrager JB, Jensen KC, et al. Non-Small Cell Lung Cancer Radiogenomics Map Identifies Relationships Between Molecular and Imaging Phenotypes With Prognostic Implications. *Radiology* (2018) 286(1):307–15. doi: 10.1148/radiol.2017161845
17. Napel S, Mu W, Jardim-Perassi BV, Aerts H, Gillies RJ. Quantitative Imaging of Cancer in the Postgenomic Era: Radio(Geno)Mics, Deep Learning, and Habitats. *Cancer* (2018) 124(24):4633–49. doi: 10.1002/cncr.31630
18. Iwatate Y, Hoshino I, Yokota H, Ishige F, Itami M, Mori Y, et al. Radiogenomics for Predicting P53 Status, PD-L1 Expression, and Prognosis With Machine Learning in Pancreatic Cancer. *Br J Cancer* (2020) 123 (8):1253–61. doi: 10.1038/s41416-020-0997-1
19. Rios Velazquez E, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfield O, et al. Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer. *Cancer Res* (2017) 77(14):3922–30. doi: 10.1158/0008-5472.Can-17-0122
20. Yin G, Wang Z, Song Y, Li X, Chen Y, Zhu L, et al. Prediction of EGFR Mutation Status Based on (18)F-FDG PET/CT Imaging Using Deep Learning-Based Model in Lung Adenocarcinoma. *Front Oncol* (2021) 11:709137. doi: 10.3389/fonc.2021.709137
21. Qin R, Wang Z, Qiao K, Hai J, Jiang L, Chen J, et al. Multi-Type Interdependent Feature Analysis Based on Hybrid Neural Networks for Computer-Aided Diagnosis of Epidermal Growth Factor Receptor Mutations. *IEEE Access* (2020) 8:38517–27. doi: 10.1109/access.2020.2971281
22. Rossi G, Barabino E, Fedeli A, Ficarra G, Coco S, Russo A, et al. Radiomic Detection of EGFR Mutations in NSCLC. *Cancer Res* (2021) 81(3):724–31. doi: 10.1158/0008-5472.CAN-20-0999
23. Zhang L, Chen B, Liu X, Song J, Fang M, Hu C, et al. Quantitative Biomarkers for Prediction of Epidermal Growth Factor Receptor Mutation in non-Small Cell Lung Cancer. *Transl Oncol* (2018) 11(1):94–101. doi: 10.1016/j.tranon.2017.10.012
24. Lu X, Li M, Zhang H, Hua S, Meng F, Yang H, et al. A Novel Radiomic Nomogram for Predicting Epidermal Growth Factor Receptor Mutation in Peripheral Lung Adenocarcinoma. *Phys Med Biol* (2020) 65(5):055012. doi: 10.1088/1361-6560/ab6f98
25. Li XY, Xiong JF, Jia TY, Shen TL, Hou RP, Zhao J, et al. Detection of Epithelial Growth Factor Receptor (EGFR) Mutations on CT Images of Patients With Lung Adenocarcinoma Using Radiomics and/or Multi-Level Residual Convolutionary Neural Networks. *J Thorac Dis* (2018) 10(12):6624–35. doi: 10.21037/jtd.2018.11.03
26. Wu G, Woodruff HC, Shen J, Refaee T, Sanduleanu S, Ibrahim A, et al. Diagnosis of Invasive Lung Adenocarcinoma Based on Chest CT Radiomic Features of Part-Solid Pulmonary Nodules: A Multicenter Study. *Radiology* (2020) 297(2):451–8. doi: 10.1148/radiol.2020192431
27. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger K, Yatabe Y, et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society: International Multidisciplinary Classification of Lung Adenocarcinoma: Executive Summary. *Proc Am Thorac Soc* (2011) 8(5):381–5. doi: 10.1513/pats.201107-042ST
28. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol* (2015) 10(9):1243–60. doi: 10.1097/JTO.0000000000000630
29. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic

Phenotype. *Cancer Res* (2017) 77(21):e104–e7. doi: 10.1158/0008-5472.CAN-17-0339
30. Kraskov A, Stgbauer H, Grassberger PJPRE. Estimating Mutual Information. *Phys Rev E* (2004) 69(6 Pt 2):066138. doi: 10.1103/PhysRevE.69.066138
31. Bentéjac C, Csörgő A, Martínez-Muñoz G. A Comparative Analysis of Gradient Boosting Algorithms. *Artif Intell Rev* (2020) 54(3):1937–67. doi: 10.1007/s10462-020-09896-5
32. Kaiming He XZ, Ren S, Sun J. Deep Residual Learning for Image Recognition. *Proc Comput Vis Pattern Recognit (CVPR)* (2018) 19:770–8. doi: 10.1109/CVPR.2016.90
33. Loshchilov I, Hutter F. Decoupled Weight Decay Regularization. *Proc ICLR* (2019), 1–15.
34. Loshchilov I, Hutter F. Sgdr: Stochastic Gradient Descent With Warm Restarts. *ProcICLR* (2017), 1–16.
35. Zhang H, Cisse M, Dauphin YN, Lopez-Paz D. Mixup: Beyond Empirical Risk Minimization. *Proc ICLR* (2018), 1760–72. Vancouver, BC, Canada.
36. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* (2017) 168(4):613–28. doi: 10.1016/j.cell.2017.01.018
37. Lim ZF, Ma PC. Emerging Insights of Tumor Heterogeneity and Drug Resistance Mechanisms in Lung Cancer Targeted Therapy. *J Hematol Oncol* (2019) 12(1):134. doi: 10.1186/s13045-019-0818-2
38. Shibue T, Weinberg RA. EMT. Cscs, and Drug Resistance: The Mechanistic Link and Clinical Implications. *Nat Rev Clin Oncol* (2017) 14(10):611–29. doi: 10.1038/nrclinonc.2017.44
39. Yang X, Dong X, Wang J, Li W, Gu Z, Gao D, et al. Computed Tomography-Based Radiomics Signature: A Potential Indicator of Epidermal Growth Factor Receptor Mutation in Pulmonary Adenocarcinoma Appearing as a Subsolid Nodule. *Oncologist* (2019) 24(11):e1156–e64. doi: 10.1634/theoncologist.2018-0706
40. Zhang B, Qi S, Pan X, Li C, Yao Y, Qian W, et al. Deep CNN Model Using CT Radiomics Feature Mapping Recognizes EGFR Gene Mutation Status of Lung Adenocarcinoma. *Front Oncol* (2020) 10:598721. doi: 10.3389/fonc.2020.598721
41. Wang S, Shi J, Ye Z, Dong D, Yu D, Zhou M, et al. Predicting EGFR Mutation Status in Lung Adenocarcinoma on Computed Tomography Image Using Deep Learning. *Eur Respir J* (2019) 53(3). doi: 10.1183/13993003.00986-2018
42. Zhao W, Yang J, Ni B, Bi D, Sun Y, Xu M, et al. Toward Automatic Prediction of EGFR Mutation Status in Pulmonary Adenocarcinoma With 3D Deep Learning. *Cancer Med* (2019) 8(7):3532–43. doi: 10.1002/cam4.2233
43. Rizzo S, Petrella F, Buscarino V, De Maria F, Raimondi S, Barberis M, et al. CT Radiogenomic Characterization of EGFR, K-RAS, and ALK Mutations in non-Small Cell Lung Cancer. *Eur Radiol* (2016) 26(1):32–42. doi: 10.1007/s00330-015-3814-0
44. Bell DW, Lynch TJ, Haserlat SM, Harris PL, Okimoto RA, Brannigan BW, et al. Epidermal Growth Factor Receptor Mutations and Gene Amplification in Non-Small-Cell Lung Cancer: Molecular Analysis of the IDEAL/INTACT Gefitinib Trials. *J Clin Oncol* (2005) 23(31):8081–92. doi: 10.1200/jco.2005.02.7078
45. Hong SJ, Kim TJ, Choi YW, Park JS, Chung JH, Lee KW. Radiogenomic Correlation in Lung Adenocarcinoma With Epidermal Growth Factor Receptor Mutations: Imaging Features and Histological Subtypes. *Eur Radiol* (2016) 26(10):3660–8. doi: 10.1007/s00330-015-4196-z
46. Li Y, Li X, Li H, Zhao Y, Liu Z, Sun K, et al. Genomic Characterisation of Pulmonary Subsolid Nodules: Mutational Landscape and Radiological Features. *Eur Respir J* (2020) 55(2):1901409. doi: 10.1183/13993003.01409-2019
47. Aherne EA, Plodkowski AJ, Montecalvo J, Hayan S, Zheng J, Capanu M, et al. What CT Characteristics of Lepidic Predominant Pattern Lung Adenocarcinomas Correlate With Invasiveness on Pathology? *Lung Cancer (Amsterdam Netherlands)* (2018) 118:83–9. doi: 10.1016/j.lungcan.2018.01.013
48. Nie Y, Liu H, Tan X, Wang H, Li F, Li C, et al. Correlation Between High-Resolution Computed Tomography Lung Nodule Characteristics and EGFR Mutation in Lung Adenocarcinomas. *OncoTargets Ther* (2019) 12:519–26. doi: 10.2147/ott.S184217
49. Lu L, Sun SH, Yang H, Linning E, Guo P, Schwartz LH, et al. E LRadiomics Prediction of EGFR Status in Lung Cancer-Our Experience in Using Multiple Feature Extractors and the Cancer Imaging Archive Data. *Tomography (Ann Arbor Mich)* (2020) 6(2):223–30. doi: 10.18383/j.tom.2020.00017
50. Tu W, Sun G, Fan L, Wang Y, Xia Y, Guan Y, et al. Radiomics Signature: A Potential and Incremental Predictor for EGFR Mutation Status in NSCLC Patients, Comparison With CT Morphology. *Lung Cancer (Amsterdam Netherlands)* (2019) 132:28–35. doi: 10.1016/j.lungcan.2019.03.025
51. Xiong JF, Jia TY, Li XY, Wen Y, Xu ZY, Cai XW, et al. Identifying Epidermal Growth Factor Receptor Mutation Status in Patients With Lung Adenocarcinoma by Three-Dimensional Convolutional Neural Networks. *Br J Radiol* (2018) 91(1092):20180334. doi: 10.1259/bjr.20180334
52. Zhang J, Zhao X, Zhao Y, Zhang J, Zhang Z, Wang J, et al. Value of Pre-Therapy (18)F-FDG PET/CT Radiomics in Predicting EGFR Mutation Status in Patients With Non-Small Cell Lung Cancer. *Eur J Nucl Med Mol Imaging* (2020) 47(5):1137–46. doi: 10.1007/s00259-019-04592-1
53. Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, et al. Somatic Mutations Affect Key Pathways in Lung Adenocarcinoma. *Nature* (2008) 455(7216):1069–75. doi: 10.1038/nature07423
54. Dearden S, Stevens J, Wu YL, Blowers D. Mutation Incidence and Coincidence in Non Small-Cell Lung Cancer: Meta-Analyses by Ethnicity and Histology (Mutmap). *Ann Oncol: Off J Eur Soc Med Oncol* (2013) 24(9):2371–6. doi: 10.1093/annonc/mdt205

Conflict of Interest: Author KK was employed by Dianei Technology. Author SD was employed by GE Healthcare.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Oncology

ORIGINAL RESEARCH
published: 03 February 2022
doi: 10.3389/fonc.2022.779395

Clinical Features and Surgical Treatment of Primary Pulmonary Lymphoma: A Retrospective Study

*Hui Shen and Yaodong Zhou**

Department of Thoracic Surgery, Fudan University Shanghai Cancer Center, Shanghai, China

Background: Primary pulmonary lymphoma (PPL) is a rare clonal lymphoproliferative lung disease. The present study analyzes the clinical features, imaging data, pathologic characteristics, treatment, and prognosis of PPL patients, with the aim to discuss the appropriate diagnosis and therapy of PPL patients in thoracic surgery.

Methods: We performed a retrospective analysis on 36 patients with PPL confirmed by postoperative pathology between 2006 and 2020. We divided the patients into low-stage (IE) and high-stage (IIE) groups using modified Ann Arbor staging. The clinical manifestations, imaging findings, treatment modalities, and outcomes were evaluated.

Results: The female to male ratio was 1.57:1 and the median age was 55 (31–69) years old. The majority of the patients had stage IE disease (75%; 27 of 36) and 9 patients had stage IIE disease. Patients with advancing stage were more likely to have respiratory symptoms. The imaging findings presented solid nodule or mass, pneumonia-like consolidative pattern, ground-glass opacity, and mixed pattern. There were 31 cases of mucosa-associated lymphoid tissue lymphoma (MALT), 2 diffuse large B-cell lymphoma (DLBCL), 2 nodular sclerosing Hodgkin's lymphoma, and 1 marginal zone B-cell lymphoma. Two patients were diagnosed with PPL and non-small cell lung cancer (NSCLC) synchronously (one AIS and MIS and one lung adenocarcinoma). All the patients received surgery. Nine patients received adjuvant therapy after surgery (five radiotherapy, two chemotherapy, and two chemoradiotherapy). Thirty-four patients had a median follow-up time of 31 months (follow-up range: 7–152 months). Of the 34 patients, 1 patient died of liver metastases and 1 patient died of intestinal metastases.

Conclusions: Our retrospective analysis suggested that most PPLs were indolent and had favorable prognosis, but the discrimination of PPL with other lung diseases was difficult. Preoperative biopsy and intraoperative frozen section examination might help in the surgical choice. Limited lung resection was enough for peripherally localized PPL.

Keywords: biopsy, pathology, lung resection, primary pulmonary lymphoma, prognosis

OPEN ACCESS

Edited by:
Xue Qin Yu,
The University of Sydney, Australia

Reviewed by:
Michael Ried,
University Hospital Regensburg, Germany
Jelena Stojsic,
University of Belgrade, Serbia

***Correspondence:**
Yaodong Zhou
yaodongzhou@shca.org.cn

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *18 September 2021*
Accepted: *13 January 2022*
Published: *03 February 2022*

Citation:
Shen H and Zhou Y (2022) Clinical Features and Surgical Treatment of Primary Pulmonary Lymphoma: A Retrospective Study. Front. Oncol. 12:779395. doi: 10.3389/fonc.2022.779395

BACKGROUND

Lung cancer is the most frequent cancer and the leading cause of cancer death in the world (1). Different from non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC), primary pulmonary lymphoma (PPL) is a rare clonal lymphoproliferative lung disease with no detectable extrapulmonary involvement at primary diagnosis or the subsequent 3 months, accounting for about 3.6% of extranodal lymphoma and 0.3% of primary pulmonary malignancies (2, 3). PPL can be classified into Hodgkin's lymphoma (HL) and non-Hodgkin's lymphoma (NHL) (4). Mucosa-associated lymphoid tissue lymphoma (MALT), B-cell lymphoma, diffuse large B-cell lymphoma (DLBCL), and T/NK-cell lymphoma are the main types of NHL (5).

It is reported that MALT is the most common subtype of PPL, accounting for up to 90% of all primary cases (6, 7). Because of its rarity, the clinical presentation and prognosis of PPL are little known. Patients with PPL often have non-specific symptoms such as cough, fever, chest pain, and dyspnea (8, 9). Most PPLs are indolent in nature and have good long-term survival (10). With the lack of a large cohort study, the Ann Arbor staging system which is developed for Hodgkin's lymphoma is extensively used for clinical staging of PPL (11). The radiographic features of PPL vary from masses, nodules, and consolidation to ground-glass opacity (GGO) (12). As a result, PPL is easily misdiagnosed as other lung diseases, like lung cancer, pneumonia, or tuberculosis. In consideration of the indolent characteristic of PPL, many experts suggest that observation and regular follow-up are the first choice for PPL patients. Besides observation, radiotherapy and chemotherapy can also be used. It is conflicting whether PPL should make surgical resection or not. To improve our understanding of PPL, we conducted this retrospective study. In this study, we retrospectively analyzed 36 patients with primary pulmonary lymphoma treated at the Department of Thoracic Surgery of Shanghai Cancer Center to investigate the clinical features, surgical treatment, and prognosis of PPL.

MATERIALS AND METHODS

Patient Selection

From 2006 to 2020, we retrospectively included 36 patients with primary pulmonary lymphoma from the Department of Thoracic Surgery of Shanghai Cancer Center. The inclusion criteria were as follows: 1) patients initially diagnosed at the Department of Thoracic Surgery, 2) clear diagnosis based on the postoperative pathology, 3) pulmonary (unilaterally or bilaterally) with or without hilar or mediastinal involvement, 4) no evidence of extrathoracic and extranodal tissue invasion at primary diagnosis or the subsequent 3 months, and 5) patients with surgical resection or biopsy.

Data Collection

Medical records concerning the demographic and clinical characteristics of all patients were obtained. All patients underwent a contrast-enhanced computed tomography (CT) scan before surgery. Two radiologists were required to review the same CT data. If they disagreed with each other, a third radiologist would make a decision. In consideration of the high costs of PET-CT, not all patients received PET-CT. Instead, patients would receive radionuclide bone scan and ultrasound of the neck and abdomen. Some patients underwent pathologic evaluation by CT-guided percutaneous lung biopsy or bronchoscopy before surgery. The postoperative pathology was also reviewed by two pathologists. If pathologists could not get a clear diagnosis by routine paraffin pathology, gene rearrangement analysis was required. All patients received surgical resection or biopsy from the Department of Thoracic Surgery. Some patients received chemotherapy, radiotherapy, or chemoradiotherapy after surgery according to the histological type and tumor stages obtained by surgery. The research was approved by the Research Ethics Committee of Shanghai Cancer Center of Fudan University. All patients consented to participate in this research.

Diagnostic and Therapeutic Approach

The preoperative diagnosis of MALT depends on the medical history, physical examination, and radiologic imaging like PET-CT, contrast-enhanced CT scan, and CT-guided percutaneous lung biopsy or bronchoscopic biopsy. If the preoperative transbronchial or CT-guided biopsy does not show malignancy, wedge resection would be performed for the suspicious nodules at first. Moreover, the intraoperative frozen section could help decide whether a more extensive procedure should be performed. The postoperative pathology was evaluated by H&E (hematoxylin–eosin staining) and immunohistochemical staining in all the patients. The interstitial infiltrate of small lymphocytes forming a mass-like lesion in the bronchiolar mucosa was the distinct morphology with the unique immunohistochemical staining of CD20(+), CD5(−), and CD10(−). When immunohistochemical staining was not sufficient for diagnosis, Ig gene rearrangement was required for further verification. In general, stage IE or IIE pulmonary MALT lymphoma with unilateral pulmonary involvement was treated only by surgical resection if the tumor was resectable. When PPL involved bilateral pulmonary or a complete resection could not be performed, adjuvant therapy was suggested. Usually, segmentectomy and lobectomy were often combined with lymph node dissection if the mediastinal lymph nodes were completely resectable.

Statistical Analysis

Continuous data were analyzed using Student's *t*-test or Mann–Whitney *U* test, while categorical data were analyzed using chi-square test. The survival curves of overall survival (OS) and relapse-free survival (RFS) were calculated by Kaplan–Meier survival analysis and log-rank test. Univariate and multivariate Cox regression analyses were performed to estimate the potential prognostic factors. Statistical analysis was performed using SPSS 23. *P*-value less than 0.05 was considered to be statistically significant.

RESULTS

General Characteristics

In this study, we retrospectively reviewed patients in our department with histologically proven PPL diagnosed from January 2006 to November 2020. Finally, 36 patients who met the inclusion criteria were included in this study. According to the modified Ann Arbor staging system (13), the 36 patients were classified into two groups: low-stage (stage IE, n = 27, 75%) and high-stage (stage IIE, n = 9, 25%). In the high-stage patients, there were five cases of stage II 2E, one cases of stage II 1E, and three cases of stage II 2EW. The baseline characteristics of the 36 patients with PPL are summarized in **Table 1**. Of the 36 patients, there were 14 men and 22 women with a median age of 55 (31–69) years. Twenty-two patients were accidentally diagnosed by routine radiographic examination with no clinical symptoms, but 14 other patients had respiratory symptoms like cough (n = 7, 50%), fever (n = 6, 42.9%), chest distress (n = 4, 28.6%), chest pain (n = 1, 7.1%), and bloody sputum (n = 1, 7.1%). Twenty-one patients received antibiotics for 1 to 4 weeks before hospitalization. Only seven patients had a smoking history. As shown by the results in **Table 1**, there was no significant difference between tumor stage and clinical characteristics.

Lung Lesion Features in Radiographic Images and Preoperative Biopsy

Fourteen patients received PET-CT before surgery. Of the 14 patients, 9 patients had data of SUVmax absorption (1 patient with no absorption, 1 patient with 17.3 SUVmax, the other 7 patients with less than 6 SUVmax). All the 36 patients had contrast-enhanced CT scan before surgery. The main features could be divided into four groups: I) solid nodule or mass, II) pneumonia-like consolidative pattern, III) ground-glass opacity, and IV) mixed pattern (**Figure 1**). Most of the CT images of PPL featured a solid nodule or mass. As shown in **Table 1**, there was no significant difference between tumor stage and radiologic findings. There were seven patients with GGO featured in the CT images, which should be distinguished from early-stage NSCLC.

Thirteen patients had preoperative pathology by CT-guided percutaneous lung biopsy (n = 10) or bronchoscopy (n = 3). Of the 10 patients with percutaneous lung biopsy, 1 patient got a cytological pathology of NSCLC, but the paraffin pathology of the surgical specimen was DLBCL. Of the three patients with bronchoscopic biopsy, one patient got a cytological pathology of chronic mucositis and inflammatory cells in the stroma. Compared with bronchoscopic biopsy, percutaneous lung biopsy was much more accurate, but still could not be fully trusted. Paraffin pathology of the surgical specimens is still the golden standard. Radiographic images like PET-CT or CT and preoperative pathology could help assist with clinical diagnosis.

Surgical Options, Pathology Characteristics, and Prognosis

All the 36 patients received surgery besides surgical biopsy. Regarding surgical type, 16 patients received muscle-sparing lung resection (2 wedge resection, 1 segmentectomy, 13 lobectomy). Nineteen patients received video-assisted thoracoscopic surgery (11 wedge resection, 3 segmentectomy, 5 lobectomy). One patient only received thoracoscopic biopsy. Moreover, 22 patients received mediastinal lymphadenectomy, and 4 patients received only biopsy of mediastinal lymph node.

TABLE 1 | Clinical characteristics of the patients with primary pulmonary lymphoma.

Characteristic	Total (n = 36)	Stage I (n = 27)	Stage II (n = 9)	P-value
Age (range)	55 (31–69)	54 (31–69)	61 (44–66)	0.087
≤60 years	25 (69.4)	21 (77.8)	4 (44.4)	0.096
Male sex	14 (38.9)	10 (37.0)	4 (44.4)	0.712
Never smoker	29 (80.6)	22 (81.5)	7 (77.8)	1.000
Comorbidities				
Hypertension	5 (13.9)	4 (14.8)	1 (11.1)	1.000
Diabetes mellitus	2 (5.6)	2 (7.4)	0 (0)	1.000
Respiratory symptoms				
Cough	7 (19.4)	3 (11.1)	4 (44.4)	0.050
Fever	6 (16.7)	4 (14.8)	2 (22.2)	0.627
Chest distress	4 (11.1)	2 (7.4)	2 (22.2)	0.255
Chest pain	1 (2.8)	0 (0)	1 (11.1)	0.206
Bloody sputum	1 (2.8)	1 (3.7)	0 (0)	1.000
Radiologic findings				0.169
Solid nodule or mass	19 (52.8)	14 (38.9)	5 (13.9)	
Pneumonia-like consolidative pattern	8 (22.2)	4 (11.1)	4 (11.1)	
Ground-glass opacity	7 (19.4)	7 (19.4)	0 (0)	
Mixed pattern	2 (5.6)	2 (5.6)	0 (0)	
IPI score				0.250
0–1	35 (97.2)	27 (100.0)	8 (88.9)	
2–3	1 (2.8)	0 (0)	1 (11.1)	
LDH				0.443
Normal	34 (94.4)	26 (96.3)	8 (88.9)	
Elevated	2 (5.6)	1 (3.7)	1 (11.1)	

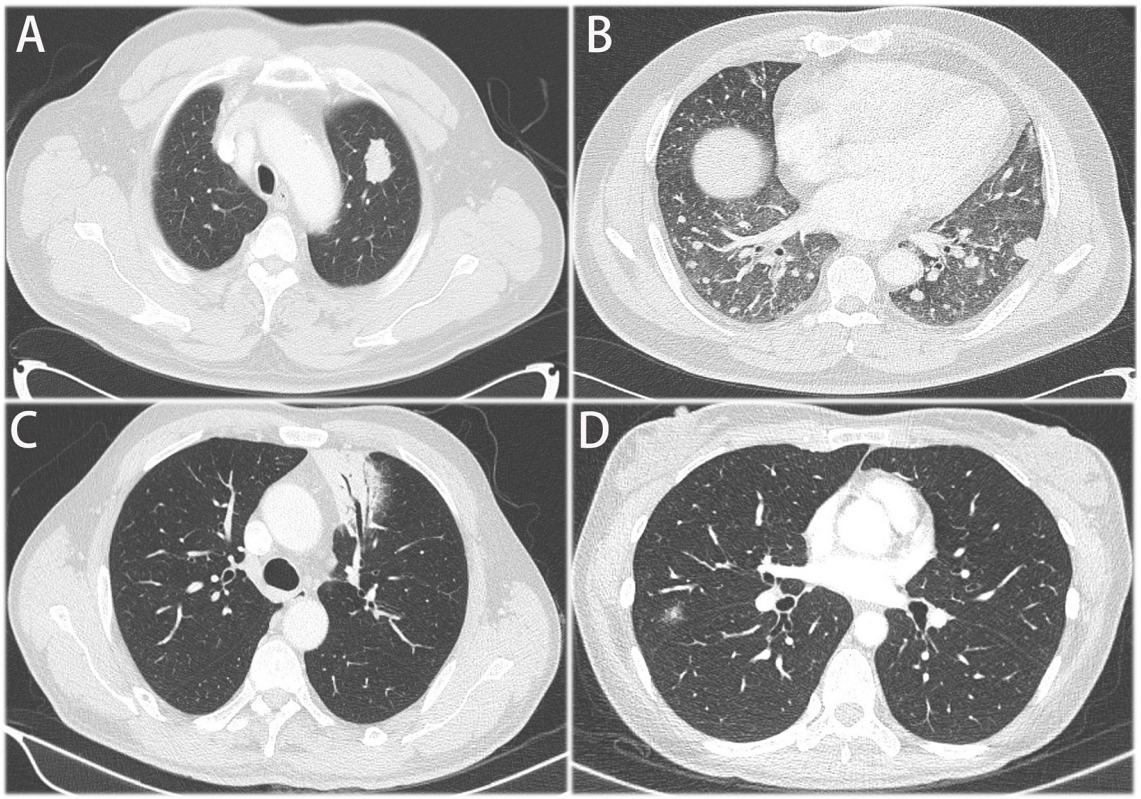


FIGURE 1 | Representative CT images of primary pulmonary. **(A)** solid nodule or mass, **(B)** multiple nodules or masses **(C)** pneumonia-like consolidative pattern, **(D)** ground glass opacity.

Postoperative pathology verified that 8 patients had mediastinal lymph node or hilar lymph node metastasis. The postoperative histopathology showed that there were 31 patients with MALT, 2 with DLBCL, 2 with nodular sclerosing Hodgkin's lymphoma and 1 with marginal zone B-cell lymphoma. **Figure 2** shows H&E and immunohistochemical staining of pulmonary DLBCL (**Figure 2A**) and pulmonary MALT (**Figure 2B**). Two patients with DLBCL received chemotherapy (rituximab plus CHOP) after surgery. Two patients with nodular sclerosing Hodgkin's lymphoma received chemoradiotherapy (ABVD plus radiotherapy) after surgery. Five patients (one marginal zone B-cell lymphoma and four MALT) with residual tumor or bilateral pulmonary involvement received local radiotherapy.

Two patients with MALT were lost to follow-up. The remaining 34 patients had a median follow-up time of 31 months (follow-up range: 7–152 months). During the follow-up period, two patients died and five patients developed progressive disease. One patient with marginal zone B-cell lymphoma (stage II 2E) died of liver metastases, and one patient with MALT (stage II 1E) died of intestinal metastases. The log-rank analysis identified that both tumor stage and surgery type were not associated with OS and RFS (**Figure 3**). There was also no significant predictor found in the univariate and multivariate Cox regression analysis (**Supplementary Tables 1**, **2**).

Interestingly, two patients were diagnosed with PPL and NSCLC synchronously (**Figure 4**). One patient had a 10- and 17-mm GGO, confirmed to be lung adenocarcinoma and MALT. Both nodules were resected in one surgery. The lung adenocarcinoma had the size of 1 * 0.8 * 0.3 cm. It had high–moderate differentiation and the lepidic subtype predominated with no invasion of pleura, lymph, and nerve. Its stage was IA (T1aN0M0). The other one patient had a 7-mm GGO featuring AIS, a 35-mm consolidation featuring MALT in the right lung, and a 10-mm GGO featuring MIS in the left lung. The patient received VATS wedge resection of the right lung for the first time and received another VATS wedge resection of the left lung after 5 months. As both patients had early-stage lung cancer, regular follow-up after surgery was suggested for them.

Moreover, we divided patients into central tumor and peripheral tumor groups according to tumor location (**Table 2**). Thirty-six patients were classified into 10 central tumors and 26 peripheral tumors. Patients with central tumors had larger tumors than patients with peripheral tumors (P = 0.003). Of the 10 patients with central tumors, 60% patients (6 of 10) received lobectomy and 3 patients received palliative resection. Of the 26 patients with peripheral tumors, 61.5% patients (16 of 26) received limited resection, including VATS wedge resection (n = 10), muscle-sparing wedge resection (n = 2), VATS segmentectomy (n = 3), and muscle-sparing segmentectomy (n = 1). Patients with

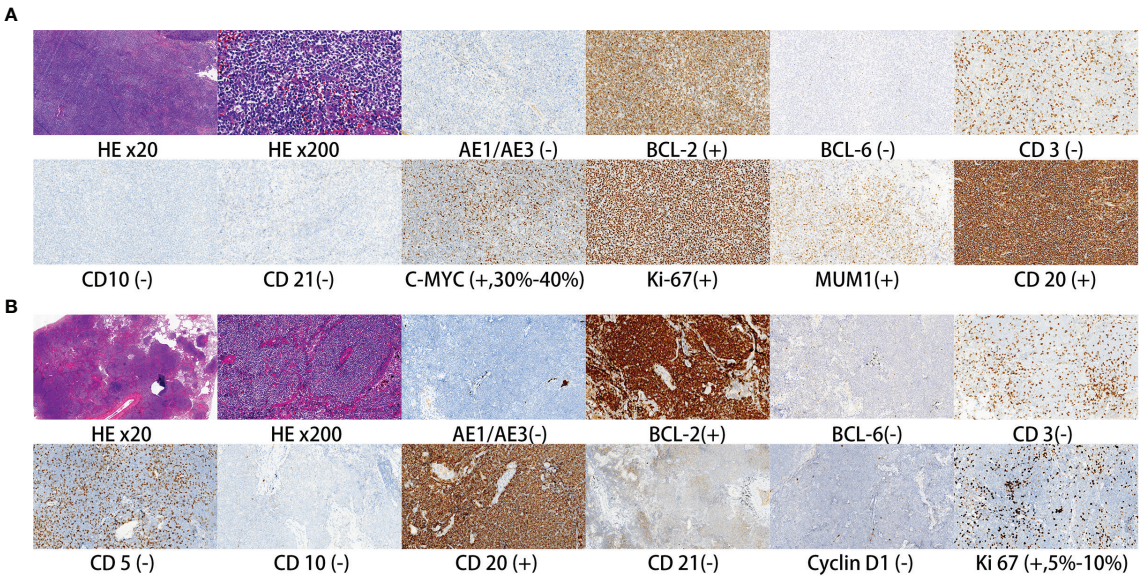


FIGURE 2 | Representative H&E and immunohistochemical staining of pulmonary DLBCL **(A)** and pulmonary MALT **(B)**.

peripheral tumors were more prone to receive limited resection, compared with patients with central tumors (P = 0.003). For patients with central tumors, 60% of the patients received mediastinal lymph node dissection. For patients with peripheral tumors, 38.5% of the patients received mediastinal lymph node dissection and 19.2% of the patients received lymph node biopsy. Moreover, only two patients had postoperative complications in the group of peripheral tumors. There was no significant difference

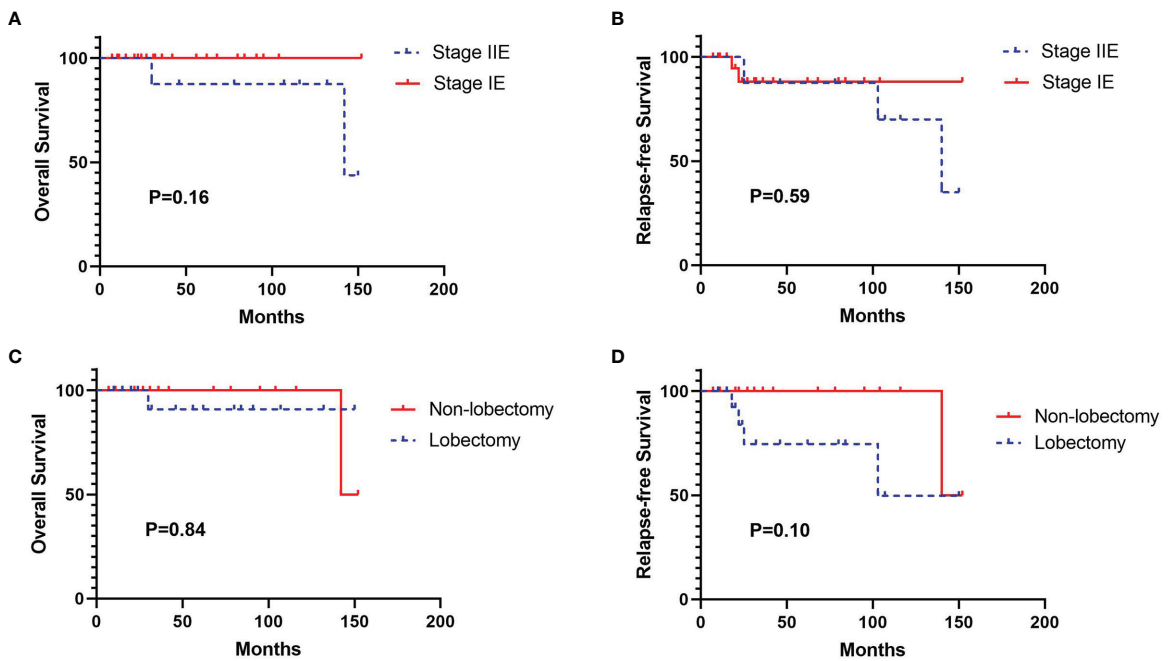


FIGURE 3 | Kaplan-Meier survival curves for overall survival (OS) and relapse-free survival (RFS) according to tumor stage and surgery type. **(A)** OS curve according to tumor stage, **(B)** RFS curve according to tumor stage, **(C)** OS curve according to surgery type, **(D)** RFS curve according to surgery type.

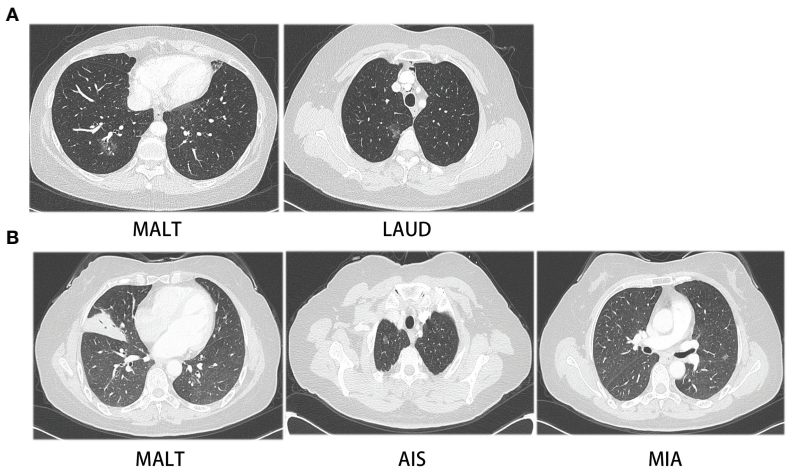


FIGURE 4 | CT images of two patients with synchronous PPL and NSCLC. **(A)** CT images of patient with MALT and LAUD, **(B)** CT images of patient with MALT, AIS and MIA.

between these two groups. One patient had arrhythmia and chylothorax after VATS segmentectomy, and another patient had air leak after muscle-sparing lobectomy.

DISCUSSION

PPL is a rare disease of clonal lymphoid proliferation affecting one or both lungs (14). It has a large difference from lymphoma which is a malignant tumor of the immune system. Most PPLs are indolent with 5-year survival rates of over 85% and median survival of over 10 years in previous studies (13). Regular follow-up is thought to be the proper therapy for most PPL patients. A study from the Sylvester Comprehensive Cancer Center showed that patients with clinically asymptomatic disease followed by observation had no progression during a median follow-up of 40.5 months (15). In this present study, we retrospectively reviewed 36 PPL patients who were first diagnosed at our thoracic department from 2006 to 2020. Most PPLs were MALT which were indolent and associated with favorable outcomes in previous studies. Five PPLs were marginal zone B-cell lymphoma, DLBCL, and HL. Different from previous studies, all the 36 PPL patients received surgery-based modality. We assessed the clinical manifestations, lung lesion features of preoperative examinations, therapy modality, pathology characteristics, and prognosis.

Previous studies showed that gender ratio was controversial. In some literature, males were slightly higher in number than females; however, some Chinese studies got the opposite results (16, 17). In our study, the female to male ratio was 1.57:1. The clinical manifestations of PPL are non-specific. A total of 61.1% of the

TABLE 2 | Clinical characteristics and treatment options in patients with central tumors and peripheral tumors.

Variable	Central tumor (*n* = 10)	Peripheral tumor (*n* = 26)	*P*-value
Age	52 (31–64)	56.5 (38–69)	0.200
Male sex	3 (30.0)	11 (42.3)	0.706
Radiologic findings			
Longest diameter, cm	5.5 (3–10)	2.75 (0.8–13)	0.003
Classification			1.000
Single nodular or consolidation	7 (70.0)	18 (69.2)	
Multiple nodular or consolidation	3 (30.0)	8 (30.8)	
Treatment characteristics			0.003
Lobectomy	6 (60.0)	9 (34.6)	
Limited resection	1 (10.0)	16 (61.5)	
Palliative resection	3 (30.0)	0 (0)	
Biopsy	0 (0)	1 (3.8)	
Mediastinal lymph node dissection	6 (60.0)	10 (38.5)	0.355
Lymph node biopsy	0 (0)	5 (19.2)	
No lymph node dissection	4 (40.0)	11 (42.3)	
Complications	0 (0)	2 (7.7)	1.000

patients had no clinical symptoms, and 14 other patients had respiratory symptoms like cough, fever, chest distress, chest pain, and bloody sputum. A total of 77.8% (7/9) of patients and 25.9% (7/27) of patients had respiratory symptoms in stage IIE and stage IE, separately. Patients in stage IIE were more prone to have respiratory symptoms than stage IE patients ($P = 0.014$). Twenty-one patients received antibiotics for 1 to 4 weeks before hospitalization. Radiographic images showed no significant change after antibiotic treatment. Consistent with previous studies, our study showed that patients with PPL had favorable outcomes. During the follow-up period, 86% of the patients were in stable condition. Only two patients died of lymphoma progression.

Preoperative examinations like PET-CT, contrast-enhanced CT scan, CT-guided percutaneous lung biopsy, or bronchoscopic biopsy offered helpful information. However, it is still hard to clearly discriminate PPL from lung cancer and pneumonia. PET-CT results suggested that patients with indolent PPL like MALT often had slightly increased SUVmax absorption (from 0 to 6), and the absorption value had no relationship with tumor size. One patient with nodular sclerosing Hodgkin's lymphoma had a relatively high SUVmax absorption (17.3). Most preoperative biopsy offered the same results with histopathology obtained from surgery. However, due to the small sample size and tumor heterogeneity, preoperative biopsy could not 100% represent the real pathology.

In our study, CT-guided percutaneous lung biopsy and bronchoscopic biopsy got 90% (9/10) and 66.7% (2/3) accuracy rate, separately.

Regarding the CT images, PPL could be divided into four groups. Consistent with previous studies, single or multiple nodules and pneumonia-like with air space consolidation were the main patterns of radiographic features (18, 19). As the clinical features of PPL were poorly defined, most of the patients were initially misdiagnosed as pneumonia, pulmonary tuberculosis, anaerobic bacteria or fungal infections, and lung cancer (9). For the single or multiple solid nodules or masses, PPL should be discriminated from primary lung cancer and metastatic lung cancer. For the pneumonia-like consolidative pattern, PPL should be discriminated from pneumonia, especially in these COVID-19 pandemic days. For the ground-glass opacity, PPL should be discriminated from early-stage lung cancer.

It is conflicting whether PPL should undergo surgical resection or not. Some experts have suggested that surgery brought no survival benefits to PPL patients and it might cause periprocedural complications (20). The aim of surgery for PPL patients should be to preserve lung function. The others consider that surgical resection could suppress PPL progression and prolong complete response, especially for young patients (15, 21). As this study only included patients with lung resection or just surgical biopsy, we could not compare the results of surgical treatment with those of non-surgical treatment. In the future, we would expand the inclusion criteria and include patients with only non-surgical treatment. However, we could get some evidence from the comparison of non-radical resection and radical resection. In the present study, four patients received palliative resection or just surgical biopsy. Among them, one died of tumor progression and one was lost to follow-up. The other 32 patients received radical resection. Among them, 1 was lost to follow-up, 1 patient died of tumor progression, and 3 patients developed progressive disease. Patients with non-radical resection seemed to have a worse survival than patients with radical resection. However, the result should be further verified in a large population. For the peripherally located tumors, restricted lung resection removes tumors and preserves most of the lung function. Moreover, for a large medical center, restricted lung resection like VATS wedge resection or VATS segmentectomy is proficient in manipulating. Periprocedural complications or death could be reduced in low proportion. In our department, only one patient had arrhythmia and chylothorax, and one patient had air leak. No patients died in the perioperative period. In addition, restricted lung resection is also an important method to make a clear diagnosis, because preoperative examinations like PET-CT, contrast-enhanced CT scan, CT-guided percutaneous lung biopsy, or bronchoscopic biopsy cannot get a precise diagnosis. The misdiagnosis could delay the best time of surgery for patients with malignant cancer. Particularly, the surgical option for GGO featuring PPL is hard to make. For those patients, the therapy modality could follow the guidelines of GGO (22). High-risk patients with GGO could be followed up for 3 or 6 months. If the tumor size has no change or it increases, patients are suggested to have a resection. For patients with tumor size less than 6 mm or pure GGO less than 3 cm, routine follow-up may be an appropriate choice. When the preoperative examination cannot discriminate PPL from lung cancer, surgical resection is the appropriate treatment modality.

It is not clear whether lymph node dissection should be performed for PPL. In our department, 60% of the patients received mediastinal lymph node dissection for central tumors and 57.7% of the patients received mediastinal lymph node dissection or lymph node biopsy for peripheral tumors. Of the eight patients with mediastinal lymph node or hilar lymph node metastasis, only three patients had mediastinal or hilar lymph node enlargement in CT images. Lymph node features in the CT images had a poor match with pathology characteristics. Although lymph node dissection can help evaluate tumor stage, there was no significant difference in overall survival rate between patients with stage I or II disease according to previous studies (23). Lymph node dissection may not make a difference in PPL treatment. But for the patient with uncertain diagnosis, intraoperative frozen section examination is indispensable. We suggest that intraoperative frozen section is an effective method to guide resection strategy (24). Only with the precise diagnosis of PPL, lymph node dissection is not mandatory.

Two patients were diagnosed with PPL and NSCLC synchronously. Simultaneous operations were performed in our department. The tumor sizes of synchronous NSCLC were not larger than 1 cm. Preoperative biopsy was hard to perform. For patients with PPL, multiple pulmonary nodules are common. Some of the nodules are PPL of the same origin, but some may be lung cancer of second origin. We recommend to perform restricted lung resection for suspicious nodules of lung cancer.

There were some limitations in our study. Firstly, the study only included 36 patients with PPL confirmed by surgical pathology.

The small sample size from one single-center may restrict the spread of the clinical experience. Secondly, the study was a retrospective research. The shortage of retrospective research was inevitable. Thirdly, the surgical modality of the patients did not follow the same strategy. Some patients received mediastinal lymphadenectomy, but some patients only received lymph node biopsy. Lastly, some patients did not have regular follow-up in our hospital. Hence, their information of subsequent treatment may not be credible.

CONCLUSION

Our retrospective analysis suggested that most PPLs were indolent. As PPL has no specific clinical and radiological manifestations, the definite diagnosis depends on pathology. Intraoperative frozen section weighs over preoperative biopsy owing to its high accuracy. Wedge resection could be performed for the suspicious nodules at first. The intraoperative frozen section helps to decide whether a more extensive procedure should be performed. For the peripherally located tumors, restricted lung resection like VATS wedge resection or VATS segmentectomy is recommended. For small hospitals with no proficient pathologists, preoperative biopsy should be carefully evaluated. When PPL is hard to discriminate from lung cancer, the multidisciplinary team could make a better therapeutic regimen.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

YZ contributed to the study concept and design and the critical revision of the manuscript. HS performed the data collection and analysis and drafted the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Natural Science Foundationof China (81902310).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.779395/full#supplementary-material

REFERENCES

1. Bade BC, Dela Cruz CS. Lung Cancer 2020: Epidemiology, Etiology, and Prevention. *Clinics Chest Med* (2020) 41(1):1–24. doi: 10.1016/j.ccm.2019.10.001
2. Cadranel J, Wislez M, Antoine M. Primary Pulmonary Lymphoma. *Eur Respir J* (2002) 20(3):750–62. doi: 10.1183/09031936.02.00404102
3. Poletti V, Ravaglia C, Tomassetti S, Gurioli C, Casoni G, Asioli S, et al. Lymphoproliferative Lung Disorders: Clinicopathological Aspects. *Eur Respir Rev* (2013) 22(130):427–36. doi: 10.1183/09059180.00004313
4. Wannesson L, Cavalli F, Zucca E. Primary Pulmonary Lymphoma: Current Status. *Clin Lymphoma Myeloma* (2005) 6(3):220–7. doi: 10.3816/CLM.2005.n.049
5. Sirajuddin A, Raparia K, Lewis VA, Franks TJ, Dhand S, Galvin JR, et al. Primary Pulmonary Lymphoid Lesions: Radiologic and Pathologic Findings. *Radiographics* (2016) 36(1):53–70. doi: 10.1148/rg.2016140339
6. Li G, Hansmann ML, Zwingers T, Lennert K. Primary Lymphomas of the Lung: Morphological, Immunohistochemical and Clinical Features. *Histopathology* (1990) 16(6):519–31. doi: 10.1111/j.1365-2559.1990.tb01157.x
7. Ahmed S, Siddiqui AK, Rai KR. Low-Grade B-Cell Bronchial Associated Lymphoid Tissue (BALT) Lymphoma. *Cancer Invest* (2002) 20(7-8):1059–68. doi: 10.1081/cnv-120005924
8. Parissis H. Forty Years Literature Review of Primary Lung Lymphoma. *J Cardiothorac Surg* (2011) 6:23. doi: 10.1186/1749-8090-6-23
9. Yao D, Zhang L, Wu PL, Gu XL, Chen YF, Wang LX, et al. Clinical and Misdiagnosed Analysis of Primary Pulmonary Lymphoma: A Retrospective Study. *BMC Cancer* (2018) 18(1):281. doi: 10.1186/s12885-018-4184-1
10. Piña-Oviedo S, Weissferdt A, Kalhor N, Moran CA. Primary Pulmonary Lymphomas. *Adv Anatomic Pathol* (2015) 22(6):355–75. doi: 10.1097/pap.0000000000000090
11. Hasenclever D, Diehl V. A Prognostic Score for Advanced Hodgkin's Disease. International Prognostic Factors Project on Advanced Hodgkin's Disease. *N Engl J Med* (1998) 339(21):1506–14. doi: 10.1056/nejm199811193392104
12. Cardenas-Garcia J, Talwar A, Shah R, Fein A. Update in Primary Pulmonary Lymphomas. *Curr Opin Pulm Med* (2015) 21(4):333–7. doi: 10.1097/mcp.0000000000000180
13. Ferraro P, Trastek VF, Adlakha H, Deschamps C, Allen MS, Pairolero PC. Primary Non-Hodgkin's Lymphoma of the Lung. *Ann Thorac Surg* (2000) 69 (4):993–7. doi: 10.1016/s0003-4975(99)01535-0
14. Freeman C, Berg JW, Cutler SJ. Occurrence and Prognosis of Extranodal Lymphomas. *Cancer* (1972) 29(1):252–60. doi: 10.1002/1097-0142(197201)29:1<252::aid-cncr2820290138>3.0.co;2-#
15. Stefanovic A, Morgensztern D, Fong T, Lossos IS. Pulmonary Marginal Zone Lymphoma: A Single Centre Experience and Review of the SEER Database. *Leuk Lymphoma* (2008) 49(7):1311–20. doi: 10.1080/10428190802064933
16. Cordier JF, Chailleux E, Lauque D, Reynaud-Gaubert M, Dietemann-Molard A, Dalphin JC, et al. Primary Pulmonary Lymphomas. A Clinical Study of 70 Cases in Nonimmunocompromised Patients. *Chest* (1993) 103(1):201–8. doi: 10.1378/chest.103.1.201
17. Wang L, Ye G, Liu Z, Shi L, Zhan C, Gu J, et al. Clinical Characteristics, Diagnosis, Treatment, and Prognostic Factors of Pulmonary Mucosa-Associated Lymphoid Tissue-Derived Lymphoma. *Cancer Med* (2019) 8 (18):7660–8. doi: 10.1002/cam4.2683
18. King LJ, Padley SP, Wotherspoon AC, Nicholson AG. Pulmonary MALT Lymphoma: Imaging Findings in 24 Cases. *Eur Radiol* (2000) 10(12):1932–8. doi: 10.1007/s003300000491
19. Wislez M, Cadranel J, Antoine M, Milleron B, Bazot M, Mayaud C, et al. Lymphoma of Pulmonary Mucosa-Associated Lymphoid Tissue: CT Scan Findings and Pathological Correlations. *Eur Respir J* (1999) 14(2):423–9. doi: 10.1034/j.1399-3003.1999.14b30.x
20. Oh SY, Kim WS, Kim JS, Kim SJ, Kwon HC, Lee DH, et al. Pulmonary Marginal Zone B-Cell Lymphoma of MALT Type–What Is a Prognostic Factor and Which Is the Optimal Treatment, Operation, or Chemotherapy?: Consortium for Improving Survival of Lymphoma (CISL) Study. *Ann Hematol* (2010) 89(6):563–8. doi: 10.1007/s00277-009-0875-7
21. Sammassimo S, Pruneri G, Andreola G, Montoro J, Steffanoni S, Nowakowski GS, et al. A Retrospective International Study on Primary Extranodal Marginal Zone Lymphoma of the Lung (BALT Lymphoma) on Behalf of International Extranodal Lymphoma Study Group (IELSG). *Hematol Oncol* (2016) 34(4):177–83. doi: 10.1002/hon.2243

22. Zhang Y, Fu F, Chen H. Management of Ground-Glass Opacities in the Lung Cancer Spectrum. *Ann Thorac Surg* (2020) 110(6):1796–804. doi: 10.1016/j.athoracsur.2020.04.094
23. Lee H, Yang B, Nam B, Jeong BH, Shin S, Zo JI, et al. Treatment Outcomes in Patients With Extranodal Marginal Zone B-Cell Lymphoma of the Lung. *J Thorac Cardiovasc Surg* (2017) 154(1):342–9. doi: 10.1016/j.jtcvs.2017.03.043
24. Liu S, Wang R, Zhang Y, Li Y, Cheng C, Pan Y, et al. Precise Diagnosis of Intraoperative Frozen Section Is an Effective Method to Guide Resection Strategy for Peripheral Small-Sized Lung Adenocarcinoma. *J Clin Oncol* (2016) 34(4):307–13. doi: 10.1200/jco.2015.63.4907

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 23 February 2022
doi: 10.3389/fonc.2022.797823

Application Value of PET/CT and MRI in the Diagnosis and Treatment of Patients With Synchronous Multiple Pulmonary Ground-Glass Nodules

*Shaonan Xie[1], Shaoteng Li[2], Huiyan Deng[3], Yaqing Han[1], Guangjie Liu[1] and Qingyi Liu[1]**

[1] Department of Thoracic Surgery, The Fourth Hospital of Hebei Medical University, Shijiazhuang, China, [2] Department of Diagnostic Radiology, The People's Hospital of Xingtai, Xingtai, China, [3] Department of Pathology, The Fourth Hospital of Hebei Medical University, Shijiazhuang, China

OPEN ACCESS

Edited by:
Yutong He,
Fourth Hospital of Hebei Medical University, China

Reviewed by:
Jian-Wei Wang,
Chinese Academy of Medical Sciences and Peking Union Medical College, China
Lei Shi,
University of Chinese Academy of Sciences, China

***Correspondence:**
Qingyi Liu
zhongmeijian-lqy@163.com

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *19 October 2021*
Accepted: *21 January 2022*
Published: *23 February 2022*

Citation:
Xie S, Li S, Deng H, Han Y, Liu G and Liu Q (2022) Application Value of PET/CT and MRI in the Diagnosis and Treatment of Patients With Synchronous Multiple Pulmonary Ground-Glass Nodules. Front. Oncol. 12:797823. doi: 10.3389/fonc.2022.797823

Background: Synchronous multiple ground-glass nodules (SMGGNs) in synchronous multiple lung cancers are associated with specific imaging findings. It is difficult to distinguish whether multiple nodules are primary tumors or metastatic lesions in the lungs. The need for PET/CT and contrast-enhanced brain MRI for these patients remains unclear. This study investigated the necessity of these two imaging examinations for SMGGN patients by means of retrospective analysis.

Methods: SMGGN patients who were diagnosed and treated in our hospital from October 2017 to May 2020 and underwent whole-body PET/CT(Cranial excepted) and/or contrast-enhanced brain MRI+DWI were enrolled in this study. We analyzed the imaging and clinical characteristics of these patients to evaluate SMGGN patients' need to undergo whole-body PET/CT and brain MRI examination.

Results: A total of 87 SMGGN patients were enrolled. 51 patients underwent whole-body PET/CT examinations and did not show signs of primary tumors in other organs, metastatic foci in other organs, or metastasis to surrounding lymph nodes. 87 patients underwent whole-brain MRI, which did not reveal brain metastases but did detect an old cerebral infarction in 23 patients and a new cerebral infarction in one patient. 87 patients underwent surgical treatment in which 219 nodules were removed. All nodules were diagnosed as adenocarcinoma or atypical adenomatous hyperplasia. No lymph node metastasis was noted.

Conclusion: For SMGGN patients, PET/CT and enhanced cranial MRI are unnecessary for SMGGNs patients, but from the perspective of perioperative patient safety, preoperative MRI+DWI examination is recommended for SMGGNs patients.

Keywords: synchronous multiple ground-glass nodules (SMGGNs), positron-emission tomography and computed tomography (PET/CT), magnetic resonance imaging (MRI), lung cancer, adenocarcinoma

INTRODUCTION

The incidence of synchronous multiple lung cancers (SMLCs) accounts for approximately 0.2% of all lung cancers, but the incidence of SMLCs has tended to gradually increase worldwide (1). The reason for this situation may be due to the popularization of low-dose computed tomography (LDCT) and high-resolution CT (HRCT), as well as its promotion and application in early lung cancer screening. Particularly, LDCT and HRCT can be used to find ground-glass nodules (GGNs) in the lungs that cannot be found on traditional chest X-ray (2). However, lung cancer patients with multiple lung lesions have long been difficult to classify due to the inability to distinguish between independent primary tumors and lung cancer with intrapulmonary metastasis, in addition, several patterns of radiological expression are associated with SMLCs (3). To provide better clarity, the 8th edition of the tumor–node–metastasis (TNM) classification for lung cancer developed by the International Association for the Study of Lung Cancer (IASLC) shows that lung cancers that manifest as multiple foci in imaging studies are classified into four categories: secondary primary lung cancer, isolated tumor nodules (intrapulmonary metastasis), multiple GGNs, and pneumonic-type lung adenocarcinoma (4, 5).

The National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines in Oncology (NCCN guidelines, the 2nd edition, 2020) recommend performing positron-emission tomography/computed tomography (PET/CT) and contrast-enhanced brain magnetic resonance imaging(MRI) for patients with multiple-nodule lung cancer (6). However, the guidelines do not specify anything about the type of multifocal lung cancer. The NCCN guidelines suggest using PET/CT as a preoperative evaluation for patients with multifocal lung cancer to assess whether there is mediastinal lymph node metastasis or distant metastasis (6). Due to the deficiencies of PET/CT in brain imaging, contrast-enhanced brain MRI is used to assess whether patients have neurological metastases (5).

Although the incidence of synchronous multiple GGNs (SMGGNs) has not been quantified, it is becoming more common as one of the imaging manifestations of SMLCs. It is mostly considered to be multiple early primary lung adenocarcinomas or precancerous lesions. Whether patients with SMGGNs can benefit from preoperative PET/CT and brain MRI needs further study. Therefore, to understand the effectiveness and necessity of these modalities in these patients, we conducted a retrospective analysis and evaluation of patients with SMGGNs who underwent routine PET/CT(Cranial excepted) and contrast-enhanced brain MRI+dispersion weighted images sequences(DWI) before surgery to determine whether there was mediastinal lymph node metastasis or organ metastasis, including determining whether SMGGNs were intrapulmonary metastasis.

METHODS

This retrospective study was reviewed and approved by the Ethics Committee of the Fourth Hospital of Hebei Medical University. Since the data of this study are retrospective and anonymous, no informed consent is required.

Enrollment of Patients

SMGGN patients who were diagnosed and treated in the Department of Thoracic Surgery in our hospital from October 2017 to May 2020 and underwent whole-body PET/CT(Cranial excepted) and contrast-enhanced brain MRI+DWI examinations were enrolled in this study. SMGGNs include multiple pure GGNs (pGGNs) and multiple mixed GGNs (mGGNs). pGGNs are seen as focal ground-glass shadows on the lung window on CT, and the nodules must not contain solid components that can block the structure of blood vessels or bronchi (7). mGGNs are GGNs that show up as shadows and contain solid components that block the structure of blood vessels or bronchi. In this study, the patients were selected based on their thin-slice CT images and the above definitions. The CT diagnosis of SMGGNs was defined as two or more GGNs shown on the images in which the maximum diameter of the solid component of mGGNs is not greater than 5 mm. The patients were diagnosed with precancerous lesions or early lung cancer if they had a relevant medical history. Two radiologists and two thoracic surgeons evaluated and compared the CT images longitudinally and reached an agreement on patient enrollment.

Image Acquisition and Analysis

All scanning was performed on a Gemini GXL 16-slice PET/CT system (Philips) with ^{18}F-fluorodeoxyglucose (FDG) (radiochemical identity/purity > 95%) provided by Andico. The patient had fasted for more than 6 hours. The patient's height and weight and level of fasting blood glucose (<6.1 mmol/L) were measured. In the resting state, 222-492.1 MBq (6-13.3 mCi) ^{18}F-FDG was injected *via* the dorsal vein of the hand. PET/CT was performed 50-60 minutes after injection, during which the patient was resting in a dark room. The patient was in the supine position with both hands on the head. Multislice spiral CT scan was performed first. The scan range was from the neck to the upper segment of the femur. The scan conditions were as follows: voltage 120 kV, current 160 mA, slice thickness 5 mm, interslice gap 5 mm, matrix 512 × 512, helical pitch 0.813, and single rotation time of the tube 0.5 s. The patient was asked to breathe calmly to ensure the scanning images to synchronize with the PET images. Then, the PET scan was performed in the 3D acquisition mode with an acquisition speed of 2.5 min/frame for a total of 8-10 frames. PET images of PET/CT were three-dimensionally reconstructed using the 3D line of response reconstruction algorithm. At the same time, CT data were used for attenuation compensation of the PET images. Both the slice thickness and interslice gap were 5 mm. During a breath-hold, a thin-slice CT scan was carried out on all GGNs in the lungs for reconstruction, with a slice thickness of 0.8 mm. When the pulmonary nodules were suspected of being malignant on PET or CT images, delayed PET scan of the chest was performed 120 minutes after injection of ^{18}F-FDG.

Contrast-enhanced brain MRI was performed by a 1.5- or 3.0-T MRI scanner (GE), with the parameters as follows: axial FSE T2WI/FLAIR (repetition time (TR), 9000 ms; echo time

(TE), 96 ms; number of excitations (NEX), 1; slice thickness, 5 mm), intravenous contrast agent Gd-DTPA (TR, 1700 ms; TE, 2.32 ms; NEX, 1; slice thickness, 5 mm) for axial and sagittal FSE T1WI scanning; DWI sequence (TR, 4100 ms; TE, 64 ms; NEX, 1; slice thickness, 5 mm); and MRA (TR, 22 ms; TE, 3.67 ms; NEX, 1; slice thickness, 5 mm).

Treatment and Pathological Staging

Pathological diagnosis and staging were done in patients undergoing surgical treatment. The postoperative pathological diagnosis was performed according to the standards developed by IASLC/American Thoracic Society/European Respiratory Society classification. Pathological staging was based on the 8th edition of the IASLC lung cancer staging system. Molecular pathological analysis on surgical specimens was performed to investigate the mutation status of the epidermal growth factor receptor (*EGFR*) gene.

RESULTS

Patient Selection

A total of 109 patients who were diagnosed and treated in the Department of Thoracic Surgery in our hospital from October 2017 to May 2020. 87 patients underwent surgical treatment. All of them underwent enhanced MRI+DWI of the head, and 51 of them underwent whole-body PET/CT(Cranial excepted). All 51 whole-body PET/CT examinations showed no signs of primary tumors in other organs, metastatic foci of other organs, or metastasis to surrounding lymph nodes (no abnormal high FDG uptake was found). All patients underwent contrast-enhanced brain MRI+DWI scan, and it did not reveal brain metastases, though it did detect an old cerebral infarction in 23 patients and new cerebral infarction in one patient by DWI sequences.

22 patients did not undergo surgical treatment. 10 patients' dominant nodules did not meet surgical indications, 3 patients could not tolerate surgery because of underlying disease, and 9 patients refused surgical treatment. All of them underwent whole-body PET/CT(Cranial excepted) too. And All of the whole-body PET/CT examinations showed no signs of primary tumors in other organs, metastatic foci of other organs, or metastasis to surrounding lymph nodes (no abnormal high FDG uptake was found).No patients with SMGGNs abandoned surgery because they were considered for intrapulmonary metastasis (**Figure 1**).

Characteristics and Imaging Data of the Patient

The median age of the 87 SMGGN patients at onset was 58 years (43-75 years). There were 26 men (29.9%) and 61 women (70.1%). A total of 351 GGNs were observed in the whole group of patients (nodules less than 5 mm were not included). The number of GGNs in each patient was two in 51 patients, 3-5 in 19 patients, 6-10 in 6 patients, and more than 10 in 11 patients. The diameter of the largest nodule on CT was ≤ 10.0 mm in 13 patients, > 10.0 mm and ≤ 20 mm in 42 patients, > 20.0 mm and ≤ 30 mm in 29 patients, and > 30 mm in 3 patients (**Table 1**).

Surgical and Pathological Outcomes After Surgery

63 patients underwent surgery directly after completing preoperative examination. Of the 23 patients diagnosed with chronic cerebral infarction,14 patients underwent surgery after short-term (7 days) antiplatelet aggregation therapy. 9 patients underwent surgery after cranial MRA and carotid ultrasound to screen blood vessels and control for risk factors. The patient with a new cerebral infarction was treated for cerebral infarction and underwent surgery 6 months later. No cerebrovascular accident occurred in all patients during perioperative period.

A total of 219 GGNs were resected in 87 patients. Postoperative pathology showed invasive adenocarcinoma (AC) in 75 GGNs, minimally invasive adenocarcinoma (MIA) in 47 GGNs, adenocarcinomas *in situ* (AIS) in 47 GGNs, and atypical adenomatous hyperplasia (AAH) in 50 GGNs. There were no cases of pleural invasion or vascular tumor thrombus, and no metastasis was found in the sampled lymph nodes. One GGN was removed in 19 patients, multiple GGNs were removed from the same lobe in five patients, GGNs in different lobes on the same side were removed in 36 patients, and bilateral GGNs were removed in 27 patients at the same time or in stages. The number of the resected GGNs was one in 19 patients, two in 36 patients, three in 17 patients, four in seven patients, five in four patients, six in one patient, seven in two patients, and nine in one patient (**Table 2**).

In patients with more than three (including five) GGNs removed, the largest three GGNs were tested for *EGFR* gene mutation. Among the 68 patients with multiple GGNs removed, 57 patients had one of various *EGFR* mutations (83.8%, 57/68) in their lung GGNs (**Figure 2**).

DISCUSSION

We conducted this retrospective analysis to evaluate the necessity of preoperative whole-body PET/CT and contrast-enhanced brain MRI in patients with SMGGNs. Although false-positive uptake of ^{18}F-FDG is often seen in daily PET/CT studies (8), none of our 51 patients undergoing PET/CT had other primary or secondary lesions in distant organs or metastasis to mediastinal lymph nodes. In all patients undergoing surgery, no metastatic lymph nodes were found in postoperative pathological examination. The preoperative screening section of the NCCN Guidelines for non-small cell lung cancer in the United States recommends that patients with SMLC undergo a whole-body PET/CT and a contrast-enhanced MRI examination of the head before surgery (9). For patients with multiple pGGNs, whether these two examinations are required before surgery is inconclusive, and no scientific consensus has been reached.

On CT images, lung nodules can be divided into solid nodules (SNs), part-solid nodules (PSNs) or mGGNs, and nonsolid nodules or pGGNs. Both pGGNs/NSNs and mGGNs/PSNs are called subsolid nodules (SSNs) (10–12). We added an additional high-resolution CT scan at breath-hold for patients who had GGNs identified on the thin-slice CT in the subsequent PET/CT scan, so the diagnosis of pGGN is reliable.

Lung GGNs that persist in CT scans is considered an imaging manifestation of early lung adenocarcinoma or precancerous

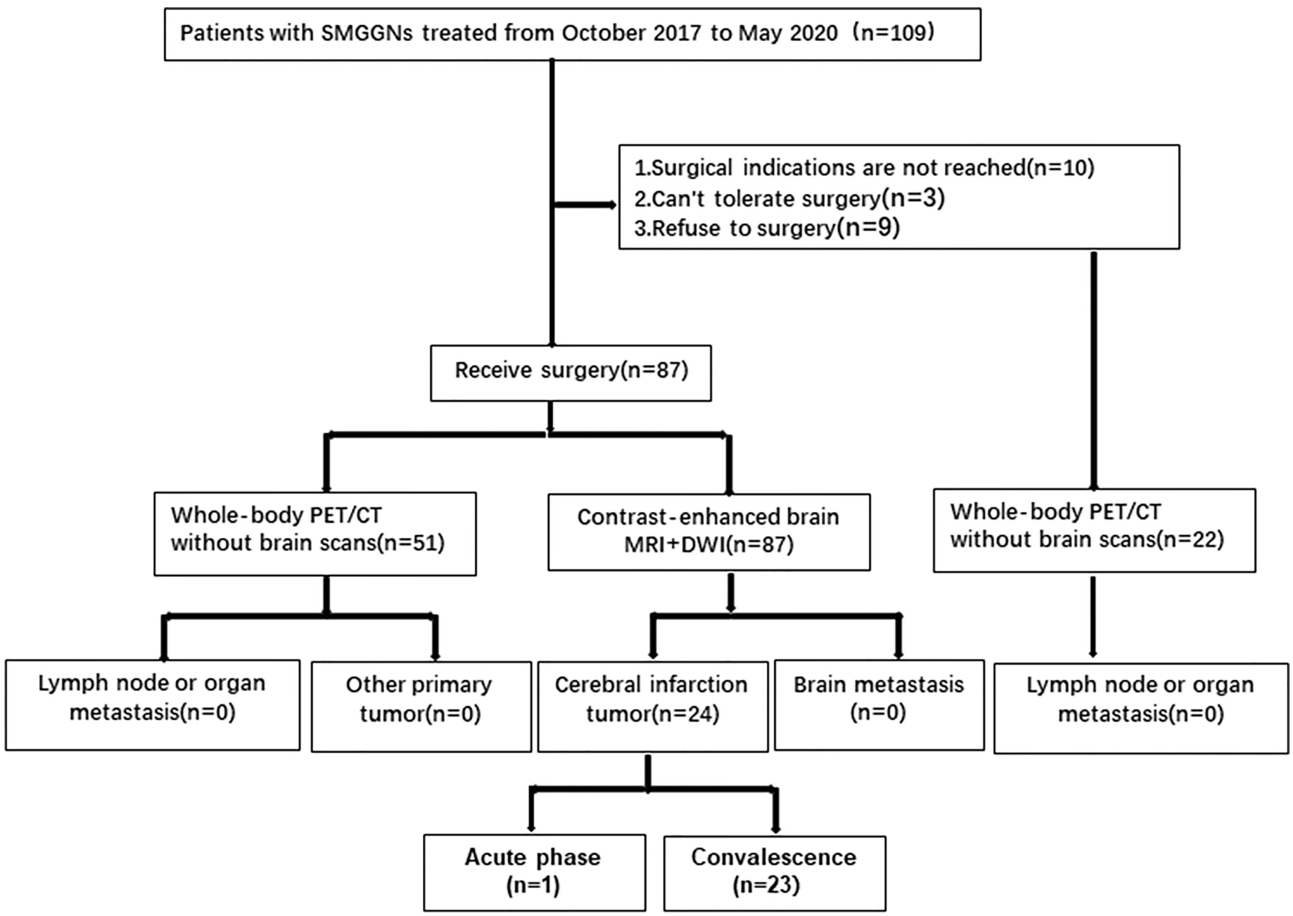


FIGURE 1 | Flow chat of SMGGN patient enrollment and examination.

lesions (13). According to the new lung adenocarcinoma staging system (4), GGNs can show a growth pattern of attachment to the alveolar wall under the microscope. This indicates a lower invasiveness. This growth pattern has little effect on alveolar ventilation. This feature of GGNs can be observed on CT images (14). The 219 nodules we resected were pathologically diagnosed as invasive adenocarcinoma, minimally invasive adenocarcinoma, adenocarcinoma *in situ*, or lung atypical adenomatous hyperplasia after surgery. Lung adenocarcinoma is the most common pathological type of lung cancer and is associated with mutations in a variety of oncogenes. The most commonly mutated genes include *EGFR*, *ALK*, *BRAF*, and *KRAS*. We performed *EGFR* mutation detection on 187 of these nodules, and 109 *EGFR* mutations were detected, for a mutation rate of 58.3%.

The cause of GGNs still needs to be investigated. GGNs are somewhat different from typical lung cancer. There is no obvious relationship between the occurrence of GGNs and smoking (a carcinogen). Most GGN patients do not smoke. The occurrence and development of GGNs are relatively slow, mostly in the peripheral part of the lung, and multiple primary lesions may be present. Household air pollution may be related to the incidence of GGNs. Household air pollution includes exhaust gas from burning solid fuels for heating and cooking and oil fumes that come from cooking (15). In our study, females (61/87) and nonsmoking patients (69/87) patients accounted for most of the enrolled patients.

Like other SMLCs, SMGGNs cause confusion for clinicians. It is hard to tell whether an SMGGNs are a lung metastasis of the same primary cancer or are multiple primary lung cancers (MPLCs) of different origins. Moreover, this confusion has not been resolved by advancements in pathology, because the SMGGNs usually have the same histological type, even if they have different growth patterns (16). According to the traditional definition, MPLCs with the same histological results must be evaluated according to the following criteria: 1. The histological origin is carcinoma *in situ*. 2. No lymph nodes in the conventional lymph node metastasis pathways are involved. 3. There is no extrathoracic metastasis (17, 18). In our study, we found no signs of metastasis on any PET/CT images, including lymph node metastasis and organ metastasis. These findings confirm that the multiple pGGNs were MPLCs with multiple primary lesions. In the Fleischner Society and IASLC statements, SMGGNs are considered to be the early stages of MPLC.

TABLE 1 | Clinical characteristics of 87 patients.

Variable/characteristic	Result/No. of patients (%)
Median age, years	58 (range, 43-75)
Sex	
Male	26 (29.9)
Female	61 (70.1)
Smoking history	
Yes	18 (20.7)
No	69 (79.3)
Initial symptoms	
Asymptomatic	69 (79.3)
Cough	14 (16.1)
Other symptoms	4 (4.6)
No. of GGNs per patient	477
2	51 (58.6)
3-5	19 (21.8)
6-10	6 (6.9)
>10	11 (12.6)
Tumor size of dominant lesion on CT	
≤10 mm	13 (14.9)
>10 mm, ≤20 mm	42 (48.3)
>20 mm, ≤30 mm	29 (33.3)
>30 mm	3 (3.4)
Average SUV_{max} of dominant lesion on PET/CT	1.1

TABLE 2 | Pathological and molecular-biological characteristics of 87 patients.

Variable/characteristic	Result/No. of patients (%)
Surgical procedure	87 (100)
pGGNs resected	219
Operation	
Ipsilateral	60 (69.0)
Bilateral	27 (31.0)
Histological type (219 pGGNs)	
AAH	50 (22.8)
AIS	47 (21.5)
MIA	47 (21.5)
AC	75 (34.2)
Pathological lymph nodal metastasis	
No (N_0)	87 (100)
Yes (N_{1-2})	0 (0)
Molecular pathology (194 pGGNs)	
***EGFR* wild-type**	78 (42.7)
***EGFR* mutated**	109 (58.3)
21L858R	54 (27.8)
21L861Q	3 (1.5)
19deletion	44 (22.7)
18G719X	2 (1.0)
20T790M	2 (1.0)
20S768I	1 (0.5)
21L858R+2L62R	1 (0.5)
21L858R+20T790M	1 (0.5)
19deletion+21L858R	1 (0.5)

In addition to identifying and classifying SMGGNs based on histopathological findings, researchers and clinicians have also used molecular biology methods. The molecular biology test results of tissue specimens strongly support the view that they are all independent primary tumors (19, 20). In our study, among the 68 patients with multiple nodules removed, 57 patients had lung nodules with *EGFR* mutation status (83.8%, 57/68), confirming the above results. Base on this, we believe that we believe that for a molecular biology point of view, all SMGGNs are independent primary tumors, it is invalid to use PET/CT to determine whether the SMGGNs are metastatic nodules or to look for other primary tumor. However, Li et al. (21) published an article about the occurrence of intrapulmonary metastases in the form of multiple GGNs. They performed whole-exome sequencing on each of the removed nodules in two patients with multiple GGNs in the lungs. They found in each of these two patients that two GGNs shared multiple rare nonsynonymous and synonymous mutations, which strongly suggested that they were intrapulmonary metastases. In contrast, the remaining GGNs showed different clonal origins. The reason for the early metastasis of GGNs may be the dissemination of tumor cells in the alveolar cavity. Although this new metastasis model of lung cancer has been well accepted, whether there are metastases in multiple GGNs in the lungs, especially multiple GGNs in the bilateral lungs, still needs to be studied and verified.

Lesions (including single lesions and multiple lesions) that appear as pure GGNs on CT images and tumors with growth patterns of attachment to the wall show indolent biological behavior and are associated with a relatively good prognosis (22, 23). Based on the multiple origins of multiple GGNs in the lungs confirmed by our research and the absence of lymph node metastasis, we have reason to believe that for patients with multiple GGNs in the lungs, limited lung resection with close follow-up should be recommended over lobectomy to observe the remaining nodules after surgery. This is the best diagnostic and treatment strategy for patients with multiple GGNs in the lung. Next-generation gene sequencing technology can be used for whole-genome sequencing, whole-exome sequencing, or target gene sequencing on surgically resected specimens to analyze whether multiple GGNs have the same origin (24). This is necessary for the overall management of the disease in patients with multiple GGNs.

It has been report (25, 26) that the sensitivity of crania enhanced MRI in screening for brain metastasis of lung cancer is not inferior to or even higher than that of whole-body PET/CT, so in our study cranial enhanced MRI was used to screen for intracranial metastasis. Although the preoperative MRI of all 87 patients did not show intracranial metastases, but combined with DWI sequence, it did reveal an old cerebral infarction in 23 patients and a new cerebral infarction in one patient. The new onset of cerebral infarction indicates the patient is in a period of hemodynamic instability, so cerebrovascular accident is more likely to recur during the perioperative period after video-assisted thoracoscopic lung resection (27). Patients with old cerebral infarction and abnormal cerebrovascular stenosis are more likely to have cerebrovascular accidents when undergoing lung (especially upper lobe) resection than the healthy population (28). Compared with PET/CT, head MRI+DWI has significant advantages in the diagnosis of cerebral infarction (29). Therefore, we believe that for patients with multiple GGNs, preoperative contrast-enhanced brain MRI is necessary.

This study has some limitations. First, the sample size of this study was relatively small. The high cost of PET/CT examinations made it hard to enroll many cases. Second, the study is limited by its retrospective design, which may have caused selection bias.

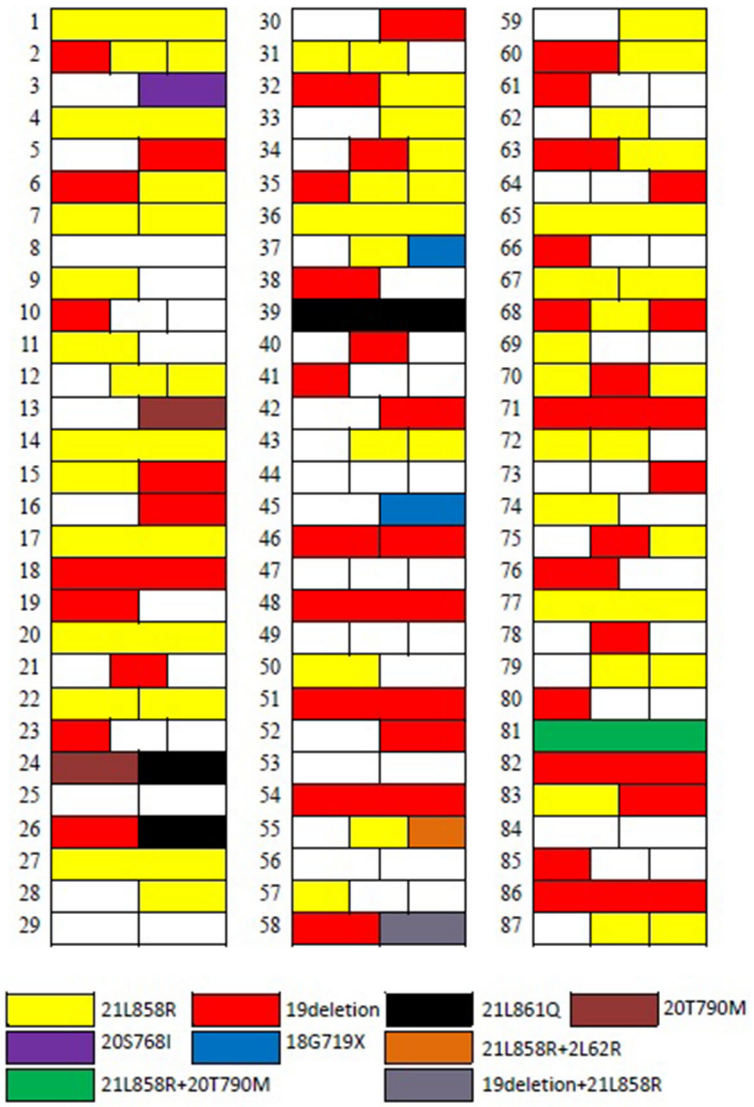


FIGURE 2 | The distribution of EGFR mutations in 87 patients.

Prospective studies should evaluate the practicality of preoperative PET/CT and brain MRI. Finally, we only included patients with multiple pGGNs and mGGNs with solid components ≤ 5 mm and excluded patients with mGGNs with solid components > 5 mm. Since mGGNs with a solid component > 5 mm are more invasive, PET/CT and brain MRI in SMGGNs of patients with a solid component > 5 mm before surgery require further study.

CONCLUSION

Whole-body PET/CT and contrast-enhanced brain MRI didn't provide additional information to determine whether patients with SMGGNs had intrapulmonary metastases and the presence of lymph node and organ metastases., but the DWI sequence can help us find some surgical risk factors before surgery, screen some patients who are not suitable for surgery, and intervene in advance for those patients who still have the chance of surgery, so as to reduce the perioperative surgical risk of patients. Based on the results of our study, we believe that PET/CT and enhanced cranial MRI are unnecessary for SMGGNs patients, but from the perspective of perioperative patient safety, preoperative MRI+DWI examination is recommended for SMGGNs patients. Of course, this conclusion is based on our retrospective study, and we expect RCT studies to prove this conclusion.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

SX: Conceptualization, Data curation, Writing - original draft. SL: Data curation. HD: Formal analysis, Writing - review & Editing. YH: Data curation; Formal analysis. GL: Investigation; Methodology. QL: Conceptualization, Funding acquisition, Writing - review & Editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Medical science research project of Hebei province (Grant No. 20201076), The programme of the government funding Clinical Excellence of Hebei province (2019 Grant No. 139).

ACKNOWLEDGMENTS

We thank American Journal Experts for language editorial assistance, Yueping Liu M.D. for pathological image assistance, Hui Feng M.D. for MRI image assistance, Shuai Liu for the PET/CT image assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.797823/full#supplementary-material

REFERENCES

1. Jung EJ, Lee JH, Jeon K, Koh WJ, Suh GY, Chung MP, et al. Treatment Outcomes for Patients With Synchronous Multiple Primary Non-Small Cell Lung Cancer. *Lung Cancer* (2011) 73(2):237–42. doi: 10.1016/j.lungcan.2010.11.008
2. Henschke CI, Yip R, Shaham D, Zulueta JJ, Aguayo SM, Reeves AP, et al. The Regimen of Computed Tomography Screening for Lung Cancer: Lessons Learned Over 25 Years From the International Early Lung Cancer Action Program. *J Thorac Imaging* (2020) 36:6–23. doi: 10.1097/RTI.0000000000000538
3. Yu YC, Huang CS, Huang BS. Separate or Intrapulmonary Metastasis. *J Thorac Dis* (2018) 10(Suppl 26):S3128–30. doi: 10.21037/jtd.2018.08.74
4. Detterbeck FC, Nicholson AG, Franklin WA, Marom EM, Travis WD, Girard N, et al. The IASLC Lung Cancer Staging Project: Summary of Proposals for Revisions of the Classification of Lung Cancers With Multiple Pulmonary Sites of Involvement in the Forthcoming Eighth Edition of the TNM Classification. *J Thorac Oncol* (2016) 11(5):639–50. doi: 10.1016/j.jtho.2016.01.024
5. Detterbeck FC, Boffa DJ, Kim AW, Tanoue LT. The Eighth Edition Lung Cancer Stage Classification. *Chest* (2017) 151(1):193–203. doi: 10.1016/j.chest.2016.10.010
6. Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman JR, Bharat A, et al. Non-Small Cell Lung Cancer, Version 2.2020, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* (2020) 15(4):504–35. doi: 10.6004/jnccn.2021.0013
7. Callister ME, Baldwin DR, Akram AR, Barnard S, Cane P, Draffan J, et al. British Thoracic Society Guidelines for the Investigation and Management of Pulmonary Nodules. *Thorax* (2015) 70(Suppl 2):ii1–54. doi: 10.1136/thoraxjnl-2015-207168
8. Bustos García de Castro A, Ferreirós Domínguez J, Delgado Bolton R, Fernández Pérez C, Cabeza Martínez B, García García-Esquinas M, et al. PET/CT in Presurgical Lymph Node Staging in Non-Small Cell Lung Cancer: The Importance of False-Negative and False-Positive Findings. *Radiologia* (2017) 59(2):147–58. doi: 10.1016/j.rx.2016.12.001
9. Kalemkerian GP, Loo BW, Akerley W, Attia A, Bassetti M, Boumber Y, et al. NCCN Guidelines Insights: Small Cell Lung Cancer, Version 2.2018. *J Natl Compr Canc Netw* (2018) 16(10):1171–82. doi: 10.6004/jnccn.2018.0079
10. Kim TJ, Kim CH, Lee HY, Chung MJ, Shin SH, Lee KJ, et al. Management of Incidental Pulmonary Nodules: Current Strategies and Future Perspectives. *Expert Rev Respir Med* (2020) 14(2):173–94. doi: 10.1080/17476348.2020.1697853
11. Raad RA, Suh J, Harari S, Naidich DP, Shiau M, Ko JP. Nodule Characterization: Subsolid Nodules. *Radiol Clin North Am* (2014) 52(1):47–67. doi: 10.1016/j.rcl.2013.08.011
12. Truong MT, Ko JP, Rossi SE, Rossi I, Viswanathan C, Bruzzi JF, et al. Update in the Evaluation of the Solitary Pulmonary Nodule. *Radiographics* (2014) 34(6):1658–79. doi: 10.1148/rg.346130092
13. Kim HY, Shim YM, Lee KS, Han J, Yi CA, Kim YK. Persistent Pulmonary Nodular Ground-Glass Opacity at Thin-Section CT: Histopathologic Comparisons. *Radiology* (2007) 245(1):267–75. doi: 10.1148/radiol.2451061682
14. Rossi G, Nappi O. What's New in the WHO Classification of Tumors of Lung and Pleura. *Pathologica* (2018) 110(1):3–4.
15. Hosgood HD3rd, Song M, Hsiung CA, Yin Z, Shu XO, Wang Z, et al. Interactions Between Household Air Pollution and GWAS-Identified Lung Cancer Susceptibility Markers in the Female Lung Cancer Consortium in Asia (FLCCA). *Hum Genet* (2015) 134(3):333–41. doi: 10.1007/s00439-014-1528-z
16. Homer RJ. Pathologists' Staging of Multiple Foci of Lung Cancer: Poor Concordance in Absence of Dramatic Histologic or Molecular Differences. *Am J Clin Pathol* (2015) 143(5):701–6. doi: 10.1309/AJCPNBWF55VGKOIW
17. Shinozaki-Ushiku A, Kohsaka S, Kage H, Oda K, Miyagawa K, Nakajima J, et al. Genomic Profiling of Multiple Primary Cancers Including Synchronous Lung Adenocarcinoma and Bilateral Malignant Mesotheliomas: Identification of a Novel BAP1 Germline Variant. *Pathol Int* (2020) 70:775–80. doi: 10.1111/pin.12977
18. Kozower BD, Larner JM, Detterbeck FC, Jones DR. Special Treatment Issues in non-Small Cell Lung Cancer: Diagnosis and Management of Lung Cancer, 3rd Ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* (2013) 143(5 Suppl):e369S–99S. doi: 10.1378/chest.12-2362
19. Liu M, He WX, Song N, Yang Y, Zhang P, Jiang GN. Discrepancy of Epidermal Growth Factor Receptor Mutation in Lung Adenocarcinoma Presenting as Multiple Ground-Glass Opacities. *Eur J Cardiothorac Surg* (2016) 50(5):909–13. doi: 10.1093/ejcts/ezw113
20. Chung JH, Choe G, Jheon S, Sung SW, Kim TJ, Lee KW, et al. Epidermal Growth Factor Receptor Mutation and Pathologic-Radiologic Correlation Between Multiple Lung Nodules With Ground-Glass Opacity Differentiates Multicentric Origin From Intrapulmonary Spread. *J Thorac Oncol* (2009) 4(12):1490–5. doi: 10.1097/JTO.0b013e3181bc9731
21. Li R, Li X, Xue R, Yang F, Wang S, Li Y, et al. Early Metastasis Detected in Patients With Multifocal Pulmonary Ground-Glass Opacities (GGOs). *Thorax* (2018) 73(3):290–2. doi: 10.1136/thoraxjnl-2017-210169
22. Lee HW, Jin KN, Lee JK, Kim DK, Chung HS, Heo EY, et al. Long-Term Follow-Up of Ground-Glass Nodules After 5 Years of Stability. *J Thorac Oncol* (2019) 14(8):1370–7. doi: 10.1016/j.jtho.2019.05.005
23. Lee JH, Park CM, Lee SM, Kim H, McAdams HP, Goo JM. Persistent Pulmonary Subsolid Nodules With Solid Portions of 5 Mm or Smaller:

Their Natural Course and Predictors of Interval Growth. *Eur Radiol* (2016) 26 (6):1529–37. doi: 10.1007/s00330-015-4017-4
24. Saab J, Zia H, Mathew S, Kluk M, Narula N, Fernandes H. Utility of Genomic Analysis in Differentiating Synchronous and Metachronous Lung Adenocarcinomas From Primary Adenocarcinomas With Intrapulmonary Metastasis. *Transl Oncol* (2017) 10(3):442–9. doi: 10.1016/j.tranon.2017.02.009
25. Lee HY, Lee KS, Kim BT, Cho YS, Lee EJ, Yi CA, et al. Diagnostic Efficacy of PET/CT Plus Brain MR Imaging for Detection of Extrathoracic Metastases in Patients With Lung Adenocarcinoma. *J Korean Med Sci* (2009) 24(6):1132–8. doi: 10.3346/jkms.2009.24.6.1132
26. Ohno Y, Koyama H, Nogami M, Takenaka D, Yoshikawa T, Yoshimura M, et al. Whole-Body MR Imaging vs. FDG-PET: Comparison of Accuracy of M-Stage Diagnosis for Lung Cancer Patients. *J Magn Reson Imaging* (2007) 26 (3):498–509. doi: 10.1002/jmri.21031
27. Wang Z, Zhang J, Cheng Z, Li X, Wang Z, Liu C, et al. Factors Affecting Major Morbidity After Video-Assisted Thoracic Surgery for Lung Cancer. *J Surg Res* (2014) 192(2):628–34. doi: 10.1016/j.jss.2014.07.051
28. Matsumoto K, Sato S, Okumura M, Niwa H, Hida Y, Kaga K, et al. Left Upper Lobectomy Is a Risk Factor for Cerebral Infarction After Pulmonary Resection: A Multicentre, Retrospective, Case-Control Study in Japan. *Surg Today* (2020) 50:1383–92. doi: 10.1007/s00595-020-02032-4
29. Zhang MJ, Zhang X, Xu YX. Analysis on Value of CT and MRI Clinical Application in Diagnosis of Middle-Aged Patients With Multiple Cerebral Infarction. *Int J Clin Exp Med* (2015) 8(10):17123–7.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Oncology

ORIGINAL RESEARCH
published: 02 March 2022
doi: 10.3389/fonc.2022.853801

A Classifier for Improving Early Lung Cancer Diagnosis Incorporating Artificial Intelligence and Liquid Biopsy

Maosong Ye[1†], Lin Tong[1,2†], Xiaoxuan Zheng[3,4†], Hui Wang[5,6†], Haining Zhou[7†], Xiaoli Zhu[8†], Chengzhi Zhou[9†], Peige Zhao[10†], Yan Wang[11†], Qi Wang[12†], Li Bai[13†], Zhigang Cai[14†], Feng-Ming (Spring) Kong[15†], Yuehong Wang[16†], Yafei Li[17†], Mingxiang Feng[18†], Xin Ye[19,20†], Dawei Yang[1], Zilong Liu[1], Quncheng Zhang[6], Ziqi Wang[6], Shuhua Han[8], Lihong Sun[11], Ningning Zhao[11], Zubin Yu[21], Juncheng Zhang[19,20], Xiaoju Zhang[6], Ruth L. Katz[22], Jiayuan Sun[3,4] and Chunxue Bai[1*]*

[1] Department of Pulmonary and Critical Care Medicine, Zhongshan Hospital, Fudan University, Shanghai, China, [2] Shanghai Respiratory Research Institute, Shanghai, China, [3] Department of Respiratory Endoscopy, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China, [4] Department of Respiratory and Critical Care Medicine, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China, [5] Xinxiang Medical University, Xinxiang, China, [6] Department of Respiratory and Critical Care Medicine, Henan Provincial People's Hospital, People's Hospital of Zhengzhou University, Zhengzhou, China, [7] Department of Thoracic Surgery, Respiratory Center of Suining Central Hospital, Suining, China, [8] Department of Pulmonary and Critical Care Medicine, Zhongda Hospital, Southeast University, Nanjing, China, [9] State Key Laboratory of Respiratory Disease, National Clinical Research Center of Respiratory Disease, Guangzhou Institute of Respiratory Health, First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, [10] Department of Respiratory and Critical Care Medicine, Affiliated Hospital of Qingdao University, Qingdao, China, [11] Department of Respiratory and Critical Care Medicine, Liaocheng People's Hospital, Liaocheng, China, [12] Department of Respiratory Medicine, The Second Affiliated Hospital of Dalian Medical University, Dalian, China, [13] Department of Respiratory Disease, Xinqiao Hospital, Army Medical University, Chongqing, China, [14] The First Department of Pulmonary and Critical Care Medicine, The Second Hospital of Hebei Medical University, Shijiazhuang, China, [15] Clinical Oncology Center, The University of Hong Kong-Shenzhen Hospital, Shenzhen, China, [16] Department of Respiratory Medicine, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, China, [17] Department of Epidemiology, College of Preventive Medicine, Army Medical University, Chongqing, China, [18] Division of Thoracic Surgery, Zhongshan Hospital, Fudan University, Shanghai, China, [19] Joint Research Center of Liquid Biopsy in Guangdong, Hong Kong, and Macao, Zhuhai, China, [20] Zhuhai Sanmed Biotech Ltd., Zhuhai, China, [21] Department of Thoracic Surgery, Xinqiao Hospital, Army Medical University, Chongqing, China, [22] Chaim Sheba Hospital, Tel Aviv University, Ramat Gan, Israel*

Lung cancer is the leading cause of cancer-related deaths worldwide and in China. Screening for lung cancer by low dose computed tomography (LDCT) can reduce mortality but has resulted in a dramatic rise in the incidence of indeterminate pulmonary nodules, which presents a major diagnostic challenge for clinicians regarding their underlying pathology and can lead to overdiagnosis. To address the significant gap in evaluating pulmonary nodules, we conducted a prospective study to develop a prediction model for individuals at intermediate to high risk of developing lung cancer. Univariate and multivariate logistic analyses were applied to the training cohort (n = 560) to develop an early lung cancer prediction model. The results indicated that a model integrating clinical characteristics (age and smoking history), radiological characteristics of pulmonary nodules (nodule diameter, nodule count, upper lobe location, malignant sign at the nodule edge, subsolid status), artificial intelligence analysis of LDCT data, and liquid

OPEN ACCESS

Edited by:
Yutong He,
Fourth Hospital of Hebei Medical University, China

Reviewed by:
Baishen Chen,
Sun Yat-sen Memorial Hospital, China
Jianhua Gao,
Chinese People's Armed Police General Hospital, China
Wookjin Choi,
Thomas Jefferson University, United States

***Correspondence:**
Chunxue Bai
cxbai@fudan.edu.cn
Jiayuan Sun
xkyyjysun@163.com

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *13 January 2022*
Accepted: *07 February 2022*
Published: *02 March 2022*

Citation:
Ye M, Tong L, Zheng X, Wang H, Zhou H, Zhu X, Zhou C, Zhao P, Wang Y, Wang Q, Bai L, Cai Z, Kong F-M(S), Wang Y, Li Y, Feng M, Ye X, Yang D, Liu Z, Zhang Q, Wang Z, Han S, Sun L, Zhao N, Yu Z, Zhang J, Zhang X, Katz RL, Sun J and Bai C (2022) A Classifier for Improving Early Lung Cancer Diagnosis Incorporating Artificial Intelligence and Liquid Biopsy. Front. Oncol. 12:853801. doi: 10.3389/fonc.2022.853801

biopsy achieved the best diagnostic performance in the training cohort (sensitivity 89.53%, specificity 81.31%, area under the curve [AUC] = 0.880). In the independent validation cohort (n = 168), this model had an AUC of 0.895, which was greater than that of the Mayo Clinic Model (AUC = 0.772) and Veterans' Affairs Model (AUC = 0.740). These results were significantly better for predicting the presence of cancer than radiological features and artificial intelligence risk scores alone. Applying this classifier prospectively may lead to improved early lung cancer diagnosis and early treatment for patients with malignant nodules while sparing patients with benign entities from unnecessary and potentially harmful surgery.

Clinical Trial Registration Number: ChiCTR1900026233, URL: http://www.chictr.org.cn/showproj.aspx?proj=43370.

Keywords: lung cancer, artificial intelligence, liquid biopsy, prediction model, early diagnosis

INTRODUCTION

Approximately 22% of the newly diagnosed cancer cases worldwide and 27% of cancer-related deaths occur in China (1). In 2018, the 5-year survival rate for lung cancer in China was 19.7% (2). Based on the results of the National Lung Screening Trial (NLST) (3, 4), low-dose computed tomography (LDCT) is the recommended test for lung cancer screening, but the high false-positive rate has diminished the benefits of the test; indeed, in a previous study, only 3.6% of the participants who had pulmonary nodules were confirmed to have lung cancer (3). Therefore, clinicians use diagnostic decision tools to stratify the malignancy risk of patients with positive LDCT results (5). The Mayo Clinic Model has been extensively validated worldwide and includes factors such as age, smoking history, extra-thoracic cancer history, spiculation, nodule diameter, and upper lobe location (6). However, because of the variation in ethnicity and environment, some risk factors might have different impacts on the Chinese population. For example, the diagnostic significance of the malignant risk factor "upper lobe location" is weakened owing to the high prevalence of tuberculosis (7).

New technologies have resulted in the emergence of several tools for early cancer diagnosis. Artificial intelligence (AI) approaches combined with deep learning technology have been adopted for image analysis in clinical settings. The use of AI can help clinicians reduce the risk of human errors caused by classifying a large number of medical images (8), which may lead to improved diagnostic efficacy of LDCT for lung cancer (9). Several studies have demonstrated that the application of deep learning technology may improve the performance of lung cancer diagnosis by the precise recognition of specific malignant features from LDCT images (10, 11). In general, AI can analyze the whole pulmonary nodule, looking for features characteristic of invasion, as opposed to histopathological evaluation of a small biopsy taken from an intermediate- or high-risk pulmonary nodule, which may not be representative (8, 11, 12). In addition, testing for early lung cancer *via* liquid biopsy using novel, sensitive, and specific biomarkers to examine cancer-related proteins or abnormal DNA (13, 14). Liquid biopsy for early lung cancer detection has been extensively investigated with various biomarkers and platforms. Indeed, previous studies (15–17) demonstrated that a fluorescent *in situ* hybridization (FISH) liquid biopsy approach to detect cells with cytogenetic abnormalities may be used to rule out lung cancer in individuals with intermediate pulmonary nodules (18, 19).

Guidelines for the early diagnosis of lung cancer in China recommend that prediction models be established based on data retrieved from Chinese populations (20), based on a broad range of preliminary information and evidence (21, 22). We hypothesized that the integration of clinical and radiological characteristics, together with AI interpretation of LDCT images and liquid biopsy testing for cells with cytogenetic abnormalities *via* a 4-color FISH array, might improve the ability to diagnose early lung cancer in individuals with intermediate and high-risk pulmonary nodules on LDCT. To this end, we conducted a prospective multicenter study in China to establish an effective early lung cancer prediction model to improve the diagnosis of pulmonary nodules with an intermediate and high risk of lung cancer detected by LDCT.

MATERIAL AND METHODS

Study Population

The study was approved by the Institutional Review Board of Zhongshan Hospital of Fudan University. A total of 1,663 individuals were recruited to the study from consecutive outpatients of 12 tertiary hospitals across mainland China. Pulmonary nodules detected by LDCT were identified as intermediate and high-risk for lung cancer by physicians in the usual care routine. Intermediate risk was defined as individuals requiring follow up to rule out malignancy, while high-risk was defined as individuals with a clinical suspicion of lung cancer. The flow chart in **Figure 1** describes the criteria for patient recruitment in this study. Written informed consent was obtained from all participants.

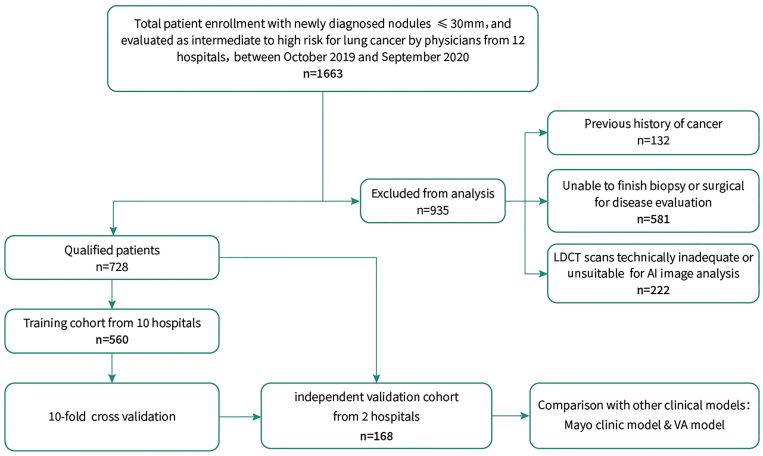


FIGURE 1 | Schematic Diagram of the Study Design.

Eligible patients recruited from ten hospitals between September 2019 and September 2020 were enrolled in the training set to establish an early lung cancer prediction model. Subsequently, an independent validation set composed of participants evaluated between March 2020 and October 2020 from the remaining two hospitals was used to test the diagnostic performance of the comprehensive lung cancer risk prediction model. The final selection of the individuals comprising the training set (n = 560) and independent validation set (n = 168) was based on the exclusion criteria shown in **Figure 1**.

Data Collection

All participants completed a demographic survey to obtain clinical information. LDCT images in the 6 months prior to enrollment of individuals were obtained for AI analysis. Following AI of LDCT scans and liquid biopsy, patients with intermediate and high-risk pulmonary nodules who met the inclusion criteria were subjected to fiberoptic bronchoscopy, fine needle biopsy, and/or surgical resection of their nodules for pathological examination. The World Health Organization classification for lung tumors was used to classify lung masses, and staging was based on the 8th edition of the TNM Classification for Lung Cancer of the International Cancer Control and the American Joint Committee on Cancer staging system.

AI Analysis Tool Development

An automated diagnostic platform comprising a deep-learning-based AI algorithm with a three-stage end-to-end deep conventional neural network (DCNNs) was developed to analyze the LDCT images of the patients. First, a 3D U-net-based DCNN was used for the patch segmentation of lung nodules to identify suspicious nodules. The LDCT images with labels were cropped in a sliding window style and feed into a 3-layer 3D U-Net segmentation model for training. Then the predicted segmentation patches were combined to generate final segmentation results. Next, the 3D patches of the suspicious nodules were forwarded to a false positive reduction network (FPRN) to discriminate the true clinically positive nodules from the false positive nodules. Then, the patches that were labeled positive were forwarded to a CNN-based classifier to determine whether the nodule was malignant or benign. This 3D U-net segmentation network was initially trained with the publicly available The Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) dataset and then further trained on a dataset of about approximately 20,000 samples from hospitals in the U.S. and China with histopathological results. Through further evaluation by experienced radiologists, the patches identified by the U-net in the first stage were segmented by manually marking the true clinically positive nodules and false positive nodules. The FPRN and malignant/benign (M/B) classifier were then trained at the patch level according to the true malignancy status confirmed by pathology results (**Figure 2**). All networks were trained with Python 3.6 and Tensorflow 1.10 on a NVIDIA DGX station. The LDCT data of the 728 participants were saved in DICOM format and uploaded to the AI lung nodule analysis platform for analysis. After the images were analyzed, the AI model provided a risk score for developing lung cancer (ranging from 0 to 100%) and a diagnosis statement for each participant.

Liquid Biopsy

To detect genetically circulating abnormal cells, we used a peripheral blood 4-color FISH assay developed to generate data for this study (23). This multiplex interphase FISH assay is composed of four DNA probes that are universally deleted in non-small cell lung cancer (NSCLC) and have been implicated in the pathogenesis of NSCLC (14, 23). This assay has previously shown a high degree of accuracy in detecting cells containing chromosomal abnormalities at *10q22.3* and *3p22.1* and in the internal control genes *CEP 10* and *3q29* (14) in several studies involving the detection of early lung cancer (24). Abnormal cells that were discovered by the 4-color FISH assay were identified as intact cells with a nucleus larger than a lymphocyte nucleus and polysomy of at least two probes per nucleus. The FISH assay was

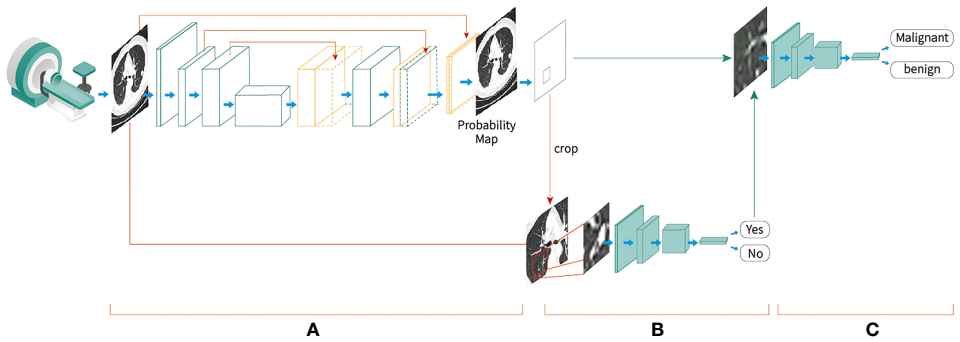


FIGURE 2 | End-to-end deep convolutional neural network-based Artificial Intelligence low-dose computed tomography analysis toll development procedures, **(A)** A three-dimensional (3D) U-net-based convolutional neural network was used for the segmentation of lung nodules to identify suspicious nodules; **(B)** the 3D patches of the suspicious nodules were cropped and forwarded to a false-positive reduction network to discriminate the true clinically positive nodules from the false-positive nodules; **(C)** the patches that were labeled as positive were forwarded to a convolutional neural network-based classifier to determine whether the nodule was malignant or benign.

performed according to the manufacturer's instructions as previously described (**Figure 3**) (25).

Statistical Analysis

Descriptive analyses of the variables are expressed as means, ranges, or numbers, expressed as percentages (%). Statistical analysis was performed using Python version 3.8.5 (Python Software Foundation, USA) and MedCalc version 19.0.4 (MedCalc Software Ltd., Ostend, Belgium). All tests were 2-sided, and statistical significance was set at $p < 0.05$.

Receiver operating curves (ROCs) were used to determine the individual performance of AI and liquid biopsy using the 4-color FISH assay. Univariate logistic regression analyses were used to determine the individual factors associated with early lung cancer in the training cohort. Variables with $p < 0.05$ in the univariate analysis were included in a multivariate logistic regression analysis to examine the independent predictive factors for inclusion in the early lung cancer diagnostic models with different sets of predictors. Cohen's kappa (κ) statistic was used to measure the reliability of the individual predictors. The mean sensitivity, specificity, and area under the curve (AUC) from the 10-fold cross validation were used to determine the diagnostic power of multiple early lung cancer prediction models. Sensitivity and specificity were used to evaluate the ability of the best-performing model to classify malignancy in an independent validation cohort. AUCs were also applied to

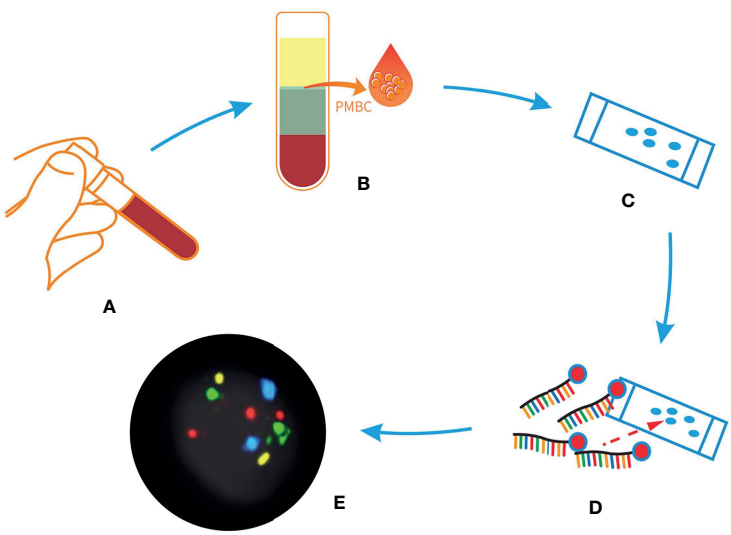


FIGURE 3 | Sample process procedures of liquid biopsy *via* 4-color fluorescent *in situ* hybridization (FISH) assay. **(A)** Peripheral blood from patients with indeterminate or high-risk nodules. **(B)** The peripheral blood mononuclear cells layer was isolated after configuration. **(C)** The peripheral blood mononuclear cells were applied to a glass slide. **(D)** Hybridization with 4-color FISH probes. **(E)** The result of the assay, scanned with a Duet microscope system.

display the classification performance of the individual validation set in Model 4, the Mayo Clinic Model, and the Veteran Affairs (VA) model.

RESULTS

Patient Characteristics

Table 1 describes the clinical characteristics of the training and independent validation cohorts according to whether the underlying pathology was benign or malignant.

Diagnostic Performance of the AI Risk Score and Liquid Biopsy

We evaluated the diagnostic ability of the AI risk score and liquid biopsy results to discriminate between benign and malignant nodules. According to the Youden index, the AI risk score had the best performance when the threshold value was set to >71%. This threshold was associated with a sensitivity of 73.77% (95% confidence interval [CI]: 69.81–77.47%) and a specificity of 65.15% (95% CI: 58.07–71.77%) in the overall cohort.

Similarly, when the cutoff value for the number of abnormal cells was set to ≥3, the sensitivity and specificity were 78.11% (95% CI: 74.35–81.56%) and 73.23% (95% CI: 66.49–79.26%), respectively. Based on the ROC curves of both tools, the AUC was 0.740 (95% CI: 0.698–0.782) for the AI risk score and 0.765 (95% CI: 0.727–0.803) for liquid biopsy in the overall cohort (**Figure 4**). Weak internal validity between the AI risk score and liquid biopsy data ($\kappa = 0.16$, 95% CI: 0.072–0.247) was observed, indicating the good complementary value of the two tools in early lung cancer diagnosis.

Relationship Between Individual Predictors and Lung Cancer

Next, individual radiological and clinical predictive factors were evaluated in a univariate logistic regression analysis using data from 560 patients in the training cohort. It was demonstrated that nodule diameter ($p < 0.001$), nodule count ($p < 0.001$), subsolid status ($p < 0.001$), upper lobe location ($p = 0.005$), and malignant features, namely, lobulation, spiculation, vacuole sign, pleural indentation, and vessel convergence sign or other radiological malignant signs at the nodule edge ($p < 0.001$),

TABLE 1 | Clinical characteristics of the study participants.

Variables	Benign Nodule		Malignant Nodule	
	Training Cohort n = 135	Validation Cohort n = 63	Training Cohort n = 425	Validation Cohort n = 105
Age, y, mean, range	55 (18–81)	57 (30–82)	60 (25–82)	57 (25–81)
Sex, no. of participants (%)				
Male	76 (56%)	37 (59%)	204 (48%)	46 (44%)
Female	59 (44%)	26 (41%)	221 (52%)	59 (56%)
Smoking history, no. of participants (%)				
Current or past smoker^	25 (19%)	17 (27%)	251 (59%)	73 (70%)
Nonsmoker	110 (81%)	46 (73%)	174 (41%)	32 (30%)
Family history, no. of participants (%)				
Yes	8 (6%)	24 (38%)	42 (10%)	39 (37%)
No	127 (94%)	39 (62%)	383 (90%)	66 (63%)
Diameter of the nodule, millimeter, mean, range	12 (1–29)	9 (2–23)	17 (1–30)	14 (4–30)
Nodule count, no. of participants (%)				
Single	80 (59%)	19 (30%)	335 (79%)	33 (31%)
Multiple	55 (41%)	44 (70%)	90 (21%)	72 (69%)
Type of nodule, no. of participants (%)				
Solid	97 (72%)	27 (43%)	207 (49%)	20 (19%)
Subsolid	38 (28%)	36 (57%)	218 (51%)	85 (81%)
Nodule location, no. of participants (%)				
Upper lobe	61 (45%)	24 (38%)	251 (59%)	68 (65%)
Non-upper lobe	74 (55%)	39 (62%)	174 (41%)	37 (35%)
Nodule edge, no. of participants (%)				
Entirely smooth	85 (63%)	35 (56%)	144 (34%)	16 (15%)
Malignant signs*	50 (37%)	28 (44%)	281 (66%)	89 (85%)
Malignant subtypes				
Adenocarcinoma			361 (85%)	97 (92%)
Squamous cell carcinoma			23 (5%)	3 (3%)
Others			41 (10%)	5 (5%)
Cancer stage				
IA1			103 (24%)	45 (43%)
IA2			176 (42%)	45 (43%)
IA3			146 (34%)	15 (14%)

^*Current and past smokers were identified as 20 pack-years and a quit time of <15 years, respectively.*

**Signs of malignancy indicate nodules with one or more of the following: lobulation, spiculation, vacuole sign, pleural indentation, vessel convergence sign, or other radiological signs of malignancy.*

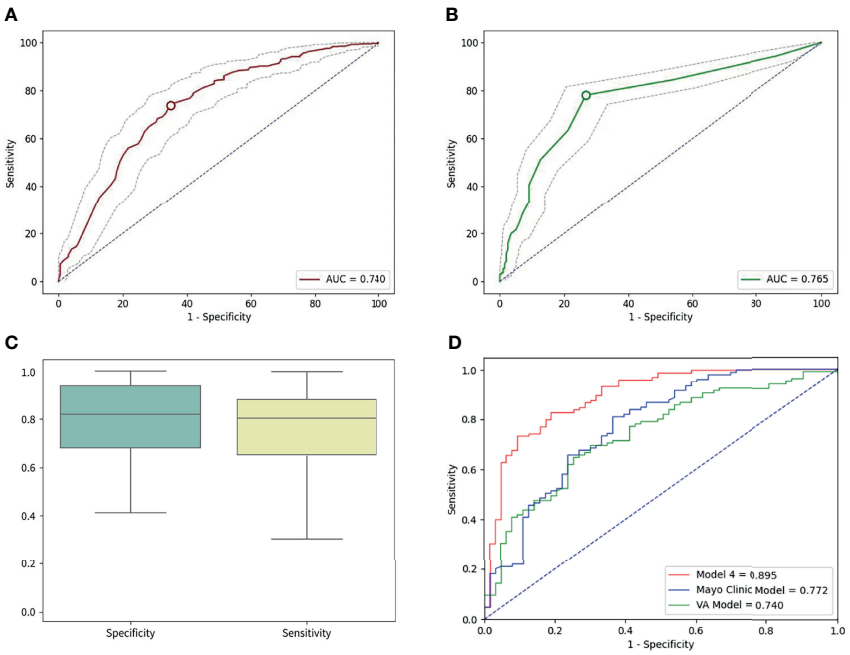


FIGURE 4 | **(A)** The area under the curve (AUC) of AI was 0.740 in the overall cohort. **(B)** The AUC of liquid biopsy was 0.765 on the overall cohort. **(C)** The sensitivity was 82.8%, and the specificity was 80.95 in the independent validation cohort for the best performing model (model 4). **(D)** In the validation cohort, the areas under the curve were 0.895, 0.772, and 0.740 for model 4, the Mayo Clinic Model, and the VA model, respectively.

were independent radiological predictors of malignancy. Age ($p < 0.001$), current smokers with 20 pack-years, or past smokers with quit time <15 years ($p < 0.001$) were clinical characteristics that correlated with lung cancer. Both the risk score predicted by AI LDCT image analysis ($p < 0.001$) and quantitation of abnormal cells identified by liquid biopsy ($p < 0.001$) were strongly associated with malignancy (**Table 2**).

Multivariate Logistic Regression Analysis to Build Early Lung Cancer Prediction Models

Before building the early lung cancer prediction models, we applied correlation analyses to test the internal validation of the individual early lung cancer risk predictors. The correlation heat maps showed that the correlations between age, smoking, AI risk factors, liquid biopsy results, and radiological predictors that were significantly associated with malignancy in the univariate analysis were very weak (**Figure 5**), revealing that there was no multicollinearity association between each predictor.

Using multivariate logistic regression analysis based on the malignancy predictors identified using the univariate statistical method, we first built four models, each with a different set of predictors (**Table 3**). Next, we calculated the diagnostic powers of the four models using 10-fold cross validation. The lowest diagnostic performance was found in model 1, which comprised only radiological characteristics (diameter, nodule count,

TABLE 2 | Univariate analyses of predictors of malignancy.

Variable	Odds Ratio(95% CI)	P-value
Age*	1.041 (1.024–1.059)	<0.001
Sex	0.717 (0.485–1.058)	0.094
Current or past smoking*	6.347 (3.946–10.210)	<0.001
Family history	1.741 (0.796–3.806)	0.165
Nodule diameter*	1.106 (1.073–1.140)	<0.001
Nodule count*	0.786 (0.703–0.879)	<0.001
Subsolid status*	2.713 (1.780–4.133)	<0.001
Upper lobe*	1.750 (1.85–2.585)	0.005
Malignant signs at the nodule edge*	3.247 (2.159–4.882)	<0.001
AI risk score*	36.891 (15.745–86.441)	<0.001
Liquid biopsy result*	1.379 (1.260–1.511)	<0.001

**Indicates significantly associated with lung cancer.*

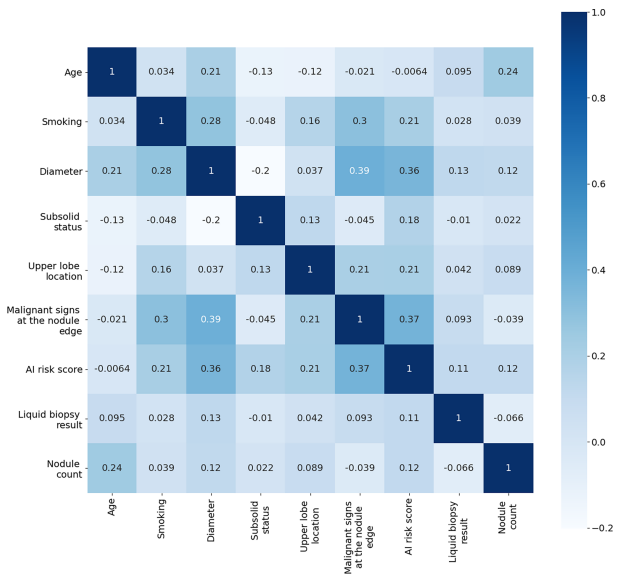


FIGURE 5 | Correlation Heat Map of Individual Predictors in the Training Cohort.

subsolid status, upper lobe location, and malignant signs at the nodule edge), with sensitivity, specificity, and AUC of 89.01% (95% CI: 82–96.03%), 62.52% (95% CI: 50.33–74.70%), and 0.769 (95% CI: 0.719–0.820), respectively. In model 2, when predictors were also consistent with radiological characteristics, with the addition of the AI risk score, there was a slight increase in the AUC to 0.791 (95% CI: 0.737–0.845), with a sensitivity of 89.18% (95% CI: 81.30–97.09%) and a specificity of 65.96% (95% CI: 53.13–78.80%). For model 3, we attempted to integrate clinical characteristics (age and smoking), radiological characteristics, and the quantitation of abnormal cells identified by the 4-color FISH test to determine the power of the risk prediction model without AI. The AUCs of model 3 achieved 0.872 (95% CI: 0.846–0.900), with 86.29% (95% CI: 77.32–95.25%) sensitivity and 83.25% (95% CI: 76.70–89.80%) specificity. The best diagnostic performance appeared to be model 4, which combined clinical and radiological characteristics, the AI risk score, and liquid biopsy results, with 89.53% (95% CI: 81.79–97.26%) sensitivity, 81.31% (95% CI: 76.43–86.18%) specificity, and an AUC of 0.880 (95% CI: 0.852–0.910), respectively (**Table 3**).

Performance of the Best Model in Independent Validation Cohort & Comparison With Other Clinical Models

Based on the perimeters that we developed from the training cohort, we tested the power of the best early lung cancer prediction model that combined clinical characteristics (age and smoking), radiological characteristics (diameter, nodule count, subsolid status, upper lobe location, and malignant signs at the nodule edge), AI risk score, and liquid biopsy results of the 4-color FISH assay in the independent validation cohort (n = 168) (**Table 1**). This model reached 82.86% (95% CI: 74.27–89.51%) sensitivity and 80.95% (95% CI: 69.09–89.75%) specificity for classifying malignant and benign nodules. ROC calculations on model 4, the Mayo Clinic Model, and the VA model were utilized. The AUCs of model 4 were 0.895 (95% CI: 0.844–0.946) in the same cohort compared to 0.772 for the Mayo Clinic Model (95% CI: 0.696–0.848) and 0.740 (95% CI: 0.663–0.817) for the VA model (**Figure 4**).

DISCUSSION

In this prospective Chinese cohort study, clinical and radiological characteristics, together with the AI risk score of LDCT image analysis and quantitation of abnormal cells detected *via* a 4 color FISH-based liquid biopsy assay, were used to build an early lung cancer prediction model to diagnose malignant pulmonary nodules in individuals evaluated as having an intermediate and high risk of lung cancer from outpatient clinics at 12 tertiary hospitals across China with newly diagnosed pulmonary nodules. Our study was a diagnostic study and not a screening study as the study population did not comprise a typical screening population with the set criteria according to the NLST. Instead, we focused on detecting lung cancer in individuals with intermediate and high-risk pulmonary nodules as confirmed by pathological examination following subsequent surgical resection. The training set was comprised of data from 560 patients and was used to establish the model. Subsequently, the efficacy of the model was tested in a validation study using data from a different set of 168 participants. We only included patients with pulmonary nodules ≤30 mm, which means that individuals with malignant pulmonary nodules were all diagnosed with stage IA (T1N0M0) lung cancer according to the TNM classification.

TABLE 3 | Ten-fold cross validation result of classifiers with different predictors.

	Predictors	Sensitivity (mean, 95% CI)	Specificity (mean, 95% CI)	AUC (mean, 95% CI)
Model 1	Diameter + nodule count + subsolid status + upper lobe location + malignant signs at the nodule edge	89.01% (82.00–96.03%)	62.52% (50.33–74.70%)	0.769 (0.719–0.820)
Model 2	Diameter + nodule count + subsolid status + upper lobe location + malignant signs at the nodule edge + AI risk score	89.18% (81.30–97.09%)	65.96% (53.13–78.80%)	0.791 (0.737–0.845)
Model 3	Age + smoking + diameter + nodule count + subsolid status + upper lobe location + malignant signs at the nodule edge + liquid biopsy result	86.29% (77.32–95.25%)	83.25% (76.70–89.80%)	0.872 (0.846–0.900)
Model 4	Age + smoking + diameter + nodule count + subsolid status + upper lobe location + malignant signs at the nodule edge + AI risk score + liquid biopsy result	89.53% (81.79–97.26%)	81.31% (76.43–86.18%)	0.880 (0.852–0.910)

CI, confidence interval; AUC, area under the curve.

To the best of our knowledge, this may be one of the first studies to integrate AI for LDCT image analysis and liquid biopsy to build a prediction model to diagnose malignant pulmonary nodules in individuals with intermediate and high risks of lung cancer in a prospective cohort. We observed an improvement in the AUC in the ability to diagnose early lung cancer when combining the AI risk score with radiological characteristics. However, when using only this information, the sensitivity of the first two models was over 80% in the two cohorts, but the specificity rates were only between 62.52% and 65.96%. As indicated by the AUCs, model 3, which included clinical characteristics, radiological characteristics, and the liquid biopsy result, performed better than models 1 and 2, which only considered information provided by LDCT with and without the assistance of AI. The highest diagnostic value was attained in a model that combined clinical and radiological characteristics, AI analysis of LDCT data, and liquid biopsy results with over 80% sensitivity and specificity. Compared to models 1 and 2, the enhancement in specificity in models 3 and 4, which combined multiple predictors, namely, liquid biopsy data and clinical data, has the potential to reduce harmful side effects such as pneumothorax and bleeding, which may be caused by invasive biopsy, suggesting that the liquid biopsy result and LDCT may complement one another. These findings provide evidence that using a classifier with a broad range of validated predictors may improve the diagnostic accuracy for early lung cancer.

The use of AI in cancer diagnosis is gaining acceptance and has been investigated for its ability to assist physicians in early lung cancer detection. AI can assist clinicians in expediting the interpretation of different pathological diagnoses and reducing the mental fatigue caused by classifying a large number of medical images (26). With the increasing incidence of lung cancer in rural China and the lack of skilled physicians (27), AI may be an excellent tool for clinicians to use as a supplement to the interpretation of LDCT images. To date, the performance metrics of AI in diagnosing lung cancer have not been verified in either retrospective data, such as the NLST dataset (28–30), or relatively small datasets (31). This prospective study evaluated the diagnostic power of AI in a large cohort of 728 patients with validated lung cancer histopathology.

We chose the 4-color FISH assay for this study as we had previously demonstrated that this assay was superior to serum protein biomarkers such as carcinoembryonic antigen, neuron-specific enolase, and cytokeratin 19 fragment (32). Furthermore, certain assays for circulating tumor cells, circulating tumor DNA, and exosomes have been measured in research studies (33, 34); however, most of these assay technologies are insensitive to early-stage lung cancer and are not commercially available for detecting early lung cancer (35–37). The FISH-based liquid biopsy assay was approved for commercial use by the China National Medical Products Administration. The performance of the test was verified in a 10-year study conducted in the USA with an accuracy rate of 94.2% in 207 participants (107 patients with lung cancer, 26 patients with benign nodules, and 80 control participants) who were at high risk of developing lung cancer (25). Additionally, in a study conducted in China, the same assay yielded sensitivities of 66.7 and 73.0% for 339 participants with pure ground-glass nodules and mixed ground-glass nodules who were diagnosed with early NSCLC (32). The results of these studies indicate that the FISH assay is a reliable tool for early lung cancer diagnosis.

According to the American College of Chest Physicians guidelines, upper lobe location is a risk factor for lung cancer, as indicated by the Mayo Clinical Model, with an odds ratio (OR) of 2.2 (38). The OR of upper lobe location in our study was 1.750 ($p = 0.005$). This finding may indicate that, in the Chinese population, the presence of pulmonary nodules located in the upper lobe is associated with a higher risk of malignancy than those discovered in other lobes, even when considering the high prevalence of pulmonary nodules in the upper lobe secondary to tuberculosis. In addition, the AUC of our best performance model was 0.895 in the independent validation cohort, which was superior to that of the Mayo Clinic Model (0.772) and the VA model (0.740). These results demonstrate that it is necessary to develop an early lung cancer classifier based on data retrieved from a Chinese population.

Our study has some limitations. First, because the participants traveled from various locations in the country prior to visiting our outpatient clinics to seek help in evaluating their nodule status, we were unable to calculate the disease prevalence in the general population. Patients in China are more likely to visit tertiary hospitals in big cities after they have discovered pulmonary nodules by LDCT in their hometowns. Since electronic health records are not shared between hospitals, we cannot track back how many people went for lung cancer screening before those with an intermediate and high risk of lung cancer went to the 12 outpatient clinics in the main cities of China. Second, our study cohort was small compared to national-scale data sets, such as those derived from the NLST and the Dutch–Belgian Randomized Lung Cancer Screening Trial (NELSON), and therefore might not be representative of the early lung cancer characteristics of the entire Chinese population; however, this is a diagnostic study and not a screening study in the general population, we have included individuals with positive LDCT results and evaluated as intermediate and high-risk for lung cancer by physicians in the usual care routine.

In the future, we hope to apply this methodology in a prospective study with a larger sample size to continue to validate and refine our classifier to improve early lung cancer diagnosis. Given the high number of pulmonary nodules discovered by LDCT scans, many patients with nodules might need to wait for a long period for physicians to interpret CT images to evaluate the significance of these lung nodules. If nodules are suspicious for malignancy, these patients may require surgical excision, biopsy, or stereotaxic radiation; however, if benign, these patients should undergo serial CT scans. The use of a multivariate lung cancer prediction model as proposed herein can help relieve the patients' anxiety by reducing the follow-up time to a definitive diagnosis if the risk score is high or delaying the follow-up time to less frequent LDCT scans if the classifier returns a low-risk score. This will

help to streamline clinical decision making by physicians for a large number of patients. We believe that a noninvasive tool such as this classifier will be a good complementary tool for physicians in the assessment of early lung cancer.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Zhongshan Hospital, Fudan University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Conception and design: JS and CB Development of methodology: MY, LT, XZhe, HW, HZ, XZhu, CZ, PZ, YaW, QW, LB, ZC FK, YuW, MF, and XY. Provision of study materials: DY, ZL, QZ, ZW, SH, LS, NZ, ZY, JZ, XZha, and JS. Statistical analysis: YL, Revised the manuscript: RK, JS, and CB. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This study was supported by the Program for the Guangdong Introducing Innovative and Entrepreneurial Teams (2019ZT08Y297) and the Shanghai Engineering & Technology Research Center of the Internet of Things for Respiratory Medicine (20DZ2254400).

ACKNOWLEDGMENTS

We acknowledge Xianjun Fan for assistance with the liquid biopsy technology. We also wish to thank the following: Chuoji Huang for excellent editing; Xing Lu for support with artificial intelligence; Xiaozheng Yang for contributing to statistical analysis and scientific writing; Yanling Zhou, Yanci Chen, and Meng Huang for their outstanding laboratory contributions; the Sanmed image analysis team for 4 color-FISH cell image analysis; the Sanmed clinical sample team for handling the blood sample processing.

REFERENCES

1. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer Statistics in China, 2015. *CA: Cancer J Clin* (2016) 66:115–32. doi: 10.3322/caac.21338
2. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, et al. Articles Global Surveillance of Trends in Cancer Survival 2000–14 (CONCORD-3): Analysis of Individual Records for 37 513 025 Patients Diagnosed With One of 18 Cancers From 322 Population-Based Registries in 71 Countries. *Lancet* (2018) 14(17):1023–75. doi: 10.1016/S0140-6736(17)33326-3
3. The National Lung Screening Trial Research Team. Reduced Lung-Cancer Mortality With Low-Dose Computed Tomographic Screening. *N Eng J Med* (2011) 365:395–409. doi: 10.1056/nejmoa1102873
4. Smith RA, Andrews KS, Brooks D, Fedewa SA, Manassaram-Baptiste D, Saslow D, et al. Cancer Screening in the United States, 2018: A Review of Current American Cancer Society Guidelines and Current Issues in Cancer Screening. *CA: Cancer J Clin* (2018) 68:297–316. doi: 10.3322/caac.21446
5. Tanner NT, Aggarwal J, Gould MK, Kearney P, Diette G, Vachani A, et al. Management of Pulmonary Nodules by Community Pulmonologists a Multicenter Observational Study. *Chest* (2015) 148:1405–14. doi: 10.1378/chest.15-0630
6. Swensen SJ, Silverstein MD, Ilstrup DM, Schleck CD, Edell ES. The Probability of Malignancy in Solitary Pulmonary Nodules. *Arch Int Med* (1997) 157(8):849–55. doi: 10.1001/archinte.1997.00440290031002
7. Bai C, Choi CM, Chu CM, Anantham D, Ho JC, Khan AZ, et al. Evaluation of Pulmonary Nodules. *Chest* (2016) 150:877–93. doi: 10.1016/j.chest.2016.02.650
8. Ahuja AS. The Impact of Artificial Intelligence in Medicine on the Future Role of the Physician. *PeerJ* (2019) 7:e7702. doi: 10.7717/peerj.7702
9. Goryński K, Safian I, Grądzki W, Marszałł MP, Krysiński J, Goryński S, et al. Artificial Neural Networks Approach to Early Lung Cancer Detection. *Cent Eur J Med* (2014) 9:632–41. doi: 10.2478/s11536-013-0327-6
10. Espinoza JL. Artificial Intelligence Tools for Refining Lung Cancer Screening. *Am J Cancer Res* (2020) 9:3860. doi: 10.3390/jcm9123860
11. Yu KH, Lee TLM, Yen MH, Kou SC, Rosen B, Chiang JH, et al. Reproducible Machine Learning Methods for Lung Cancer Detection Using Computed Tomography Images: Algorithm Development and Validation. *J Med Internet Res* (2020) 22:e16709. doi: 10.2196/16709
12. Varghese C, Rajagopalan S, Karwoski RA, Bartholmai BJ, Maldonado F, Boland JM, et al. Computed Tomography-Based Score Indicative of Lung Cancer Aggression (SILA) Predicts the Degree of Histologic Tissue Invasion and Patient Survival in Lung Adenocarcinoma Spectrum. *J Thorac Oncol* (2019) 14:1419–29. doi: 10.1016/j.jtho.2019.04.022
13. Kaiser J. 'Liquid Biopsy' for Cancer Promises Early Detection. *Science* (2018) 359:259. doi: 10.1126/science.359.6373.259
14. Katz RL, Zaidi TM, Pujara D, Shanbhag ND. Identification of Circulating Tumor Cells Using 4-Color Fluorescence *in Situ* Hybridization: Validation of a Noninvasive Aid for Ruling Out Lung Cancer in Patients With Low-Dose Computed Tomography-Detected Lung Nodules. *Cancer Cytopathol* (2020) 128:553–62. doi: 10.1002/cncy.22278
15. Mu Z, Benali-Furet N, Uzan G, Ye Z, Austin L, Wang C, et al. Abstract P2-02-14: Detection and Characterization of CTCs Isolated by ScreenCell®-Filtration in Metastatic Breast Cancer. *Cancer Res* (2016) 76(4 Suppl):Abstract P2-02-14. doi: 10.1158/1538-7445.sabcs15-p2-02-14
16. Perakis S, Speicher MR. Emerging Concepts in Liquid Biopsies. *BMC Med* (2017) 15:75. doi: 10.1186/s12916-017-0840-6
17. Zheng H, Wu X, Yin J, Wang S, Li Z, You CX. Clinical Applications of Liquid Biopsies for Early Lung Cancer Detection. *Am J Cancer Res* (2019) 9:2567–79.
18. Marquette CH, Boutros J, Benzaquen J, Ferreira M, Pastre J, Pison C, et al. Circulating Tumour Cells as a Potential Biomarker for Lung Cancer Screening: A Prospective Cohort Study. *Lancet Respir Med* (2020) 8:709–16. doi: 10.1016/S2213-2600(20)30081-3
19. Asghar S, Waqar W, Umar M, Manzoor S. Tumor Educated Platelets, a Promising Source for Early Detection of Hepatocellular Carcinoma: Liquid Biopsy an Alternative Approach to Tissue Biopsy. *Clin Res Hepatol Gastroenterol* (2020) 44:836–44. doi: 10.1016/j.clinre.2020.03.023
20. Humphrey LL, Deffebach M, Pappas M, Baumann C, Artis K, Mitchell JP, et al. Screening for Lung Cancer With Low-Dose Computed Tomography: A

Systematic Review to Update the U.S. Preventive Services Task Force Recommendation. *Ann Int Med* (2013) 159:411–20. doi: 10.7326/0003-4819-159-6-201309170-00690
21. Tranvåg EJ, Norheim OF, Ottersen T. Clinical Decision Making in Cancer Care: A Review of Current and Future Roles of Patient Age. *BMC Cancer* (2018) 18(1):546. doi: 10.1186/s12885-018-4456-9
22. Glatzer M, Panje CM, Sirén C, Cihoric N, Putora PM. Decision Making Criteria in Oncology. *Oncol (Switzerland)* (2020) 98:370–8. doi: 10.1159/000492272
23. Katz RL, He W, Khanna A, Fernandez R, Zaidi TM, Krebs M, et al. Genetically Abnormal Circulating Cells in Lung Cancer Patients: An Antigen-Independent Fluorescence *in Situ* Hybridization-Based Case-Control Study. *Clin Cancer Res* (2010) 16:3976–87. doi: 10.1158/1078-0432.CCR-09-3358
24. Ye X, Yang XZ, Carbone R, Barshack I, Katz RL. Diagnosis of non-Small Cell Lung Cancer *via* Liquid Biopsy Highlighting a Fluorescence-in-Situ-Hybridization Circulating Tumor Cell Approach. *Pathology-from classics to innovations Intechopen* (2021) 129. doi: 10.5772/intechopen.97631
25. Katz RL, Zaidi TM, Pujara D, Shanbhag ND, Truong D, Patil S, et al. Identification of Circulating Tumor Cells Using 4-Color Fluorescence *in Situ* Hybridization: Validation of a Noninvasive Aid for Ruling Out Lung Cancer in Patients With Low-Dose Computed Tomography–Detected Lung Nodules. *Cancer Cytopathol* (2020) 128:553–62. doi: 10.1002/cncy.22278
26. Shen D, Wu G, Suk H IL. Deep Learning in Medical Image Analysis. *Ann Rev BioMed Eng* (2017) 19:221–48. doi: 10.1146/annurev-bioeng-071516-044442
27. Cao M, Chen W. Epidemiology of Lung Cancer in China. *Thorac Cancer* (2019) 10:3–7. doi: 10.1111/1759-7714.12916
28. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-To-End Lung Cancer Screening With Three-Dimensional Deep Learning on Low-Dose Chest Computed Tomography. *Nat Med* (2019) 25:954–61. doi: 10.1038/s41591-019-0447-x
29. Riquelme D, Akhloufi MA. Deep Learning for Lung Cancer Nodules Detection and Classification in CT Scans. *AI* (2020) 1:28–67. doi: 10.3390/ai1010003
30. Yoo H, Kim KH, Singh R, Digumarthy SR, Kalra MK. Validation of a Deep Learning Algorithm for the Detection of Malignant Pulmonary Nodules in Chest Radiographs. *JAMA Netw Open* (2020) 3:e2017135. doi: 10.1001/jamanetworkopen.2020.17135
31. Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A Comparison of Deep Learning Performance Against Health-Care Professionals in Detecting Diseases From Medical Imaging: A Systematic Review and Meta-Analysis. *Lancet Digit Health* (2019) 1:E271–97. doi: 10.1016/S2589-7500(19)30123-2
32. Feng M, Xin Y, Chen B, Zhang J, Lin M, Zhou H, et al. Detection of Circulating Genetically Abnormal Cells Using 4-Color Fluorescence *In Situ* Hybridization for the Early Detection of Lung Cancer. *J Cancer Res Clin Oncol* (2021) 147:2397–405. doi: 10.1007/s00432-021-03517-6
33. Palmirotta R, Lovero D, Cafforio P, Felici C, Mannavola F, Pellè E, et al. Liquid Biopsy of Cancer: A Multimodal Diagnostic Tool in Clinical Oncology. *Ther Adv Med Oncol* (2018) 10:1–24. doi: 10.1177/1758835918794630
34. Rijavec E, Coco S, Genova C, Rossi G, Longo L, Grossi F. Liquid Biopsy in non-Small Cell Lung Cancer: Highlights and Challenges. *Cancers* (2020) 12:17. doi: 10.3390/cancers12010017
35. Cowling T, Loshak H. An Overview of Liquid Biopsy for Screening and Early Detection of Cancer. In: *CADTH Issues in Emerging Health Technologies*. Canadian Agency for Drugs and Technologies in Health (2016). p. 179.
36. Moding EJ, Liu Y, Nabet BY, Chabon JJ, Chaudhuri AA, Hui AB, et al. Circulating Tumor DNA Dynamics Predict Benefit From Consolidation Immunotherapy in Locally Advanced non-Small-Cell Lung Cancer. *Nat Cancer* (2020) 1:176–83. doi: 10.1038/s43018-019-0011-0
37. Hou JM, Krebs MG, Lancashire L, Sloane R, Backen A, Swain RK, et al. Clinical Significance and Molecular Characteristics of Circulating Tumor Cells and Circulating Tumor Microemboli in Patients With Small-Cell Lung Cancer. *J Clin Oncol* (2012) 30:525–32. doi: 10.1200/JCO.2010.33.3716
38. Gould MK, Donington J, Lynch WR, Mazzone PJ, Midthun DE, Naidich DP, et al. Evaluation of Individuals With Pulmonary Nodules: When is it Lung Cancer? Diagnosis and Management of Lung Cancer, 3rd Ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* (2013) 143:93–120. doi: 10.1378/chest.12-2351

Conflict of Interest: Authors XY and JZ are employees of Zhuhai Sanmed Biotech Ltd. RK is a consultant of Sanmed Biotech Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 28 April 2022
doi: 10.3389/fonc.2022.800811

A Multi-Classification Model for Predicting the Invasiveness of Lung Adenocarcinoma Presenting as Pure Ground-Glass Nodules

Fan Song[1†], Lan Song[2†], Tongtong Xing[1], Youdan Feng[1], Xiao Song[3], Peng Zhang[1], Tianyi Zhang[1], Zhenchen Zhu[2,4], Wei Song[2] and Guanglei Zhang[1]*

[1] Beijing Advanced Innovation Center for Biomedical Engineering, School of Biological Science and Medical Engineering, Beihang University, Beijing, China, [2] Department of Radiology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [3] School of Medical Imaging, Shanxi Medical University, Taiyuan, China, [4] 4 + 4 MD Program, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

OPEN ACCESS

Edited by:
Lizza E.L. Hendriks, Maastricht University Medical Centre, Netherlands

Reviewed by:
Chen Chen, Central South University, China
Damiano Caruso, Sapienza University of Rome, Italy

****Correspondence:***
Guanglei Zhang
guangleizhang@buaa.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *24 October 2021*
Accepted: *04 April 2022*
Published: *28 April 2022*

Citation:
Song F, Song L, Xing T, Feng Y, Song X, Zhang P, Zhang T, Zhu Z, Song W and Zhang G (2022) A Multi-Classification Model for Predicting the Invasiveness of Lung Adenocarcinoma Presenting as Pure Ground-Glass Nodules. Front. Oncol. 12:800811. doi: 10.3389/fonc.2022.800811

Objectives: To establish a multi-classification model for precisely predicting the invasiveness (pre-invasive adenocarcinoma, PIA; minimally invasive adenocarcinoma, MIA; invasive adenocarcinoma, IAC) of lung adenocarcinoma manifesting as pure ground-glass nodules (pGGNs).

Methods: By the inclusion and exclusion criteria, this retrospective study enrolled 346 patients (female, 297, and male, 49; age, 55.79 ± 10.53 (24-83)) presenting as pGGNs from 1292 consecutive patients with pathologically confirmed lung adenocarcinoma. A total of 27 clinical were collected and 1409 radiomics features were extracted by *PyRadiomics* package on *python*. After feature selection with L2,1-norm minimization, logistic regression (LR), extra w(ET) and gradient boosting decision tree (GBDT) were used to construct the three-classification model. Then, an ensemble model of the three algorithms based on model ensemble strategy was established to further improve the classification performance.

Results: After feature selection, a hybrid of 166 features consisting of 1 clinical (short-axis diameter, ranked 27th) and 165 radiomics (4 shape, 71 intensity and 90 texture) features were selected. The three most important features are wavelet-HLL_firstorder_Minimum, wavelet-HLL_ngtdm_Busyness and square_firstorder_Kurtosis. The hybrid-ensemble model based on hybrid clinical-radiomics features and the ensemble strategy showed more accurate predictive performance than other models (hybrid-LR, hybrid-ET, hybrid-GBDT, clinical-ensemble and radiomics-ensemble). On the training set and test set, the model can obtain the accuracy values of 0.918 ± 0.022 and 0.841, and its F1-scores respectively were 0.917 ± 0.024 and 0.824.

Conclusion: The multi-classification of invasive pGGNs can be precisely predicted by our proposed hybrid-ensemble model to assist patients in the early diagnosis of lung adenocarcinoma and prognosis.

Keywords: adenocarcinoma of lung, pure ground-glass nodule, computer-assisted diagnosis, neoplasm invasiveness, early diagnosis, prognosis

INTRODUCTION

At present, with the widespread clinical application of computed tomography (CT) and the popularity of early lung cancer screening, more and more ground-glass nodules (GGNs) are detected. GGN is a nodule showing hazy increased density on thin-slice CT, with preservation of bronchial and vascular margins (1, 2). According to whether there are solid components in the lesion, GGN can be further divided into pure GGN (pGGN) and part-solid GGN. The appearance of a persistent invasive pGGN may suggest a high risk of early malignant tumor, so distinguishing the invasiveness of pGGNs is critical. A pathological classification was established in 2011 with respect to the degree of invasion: atypical adenomatous hyperplasia (AAH), adenocarcinoma *in situ* (AIS), minimally invasive adenocarcinoma (MIA) and invasive adenocarcinoma (IAC) (3).

In general, the tumor doubling time of pre-invasive adenocarcinoma (PIA, namely AAH/AIS) can reach more than two years, and through partial resection, the 5-year survival rate of patients can reach 100% (4–7). For MIA, sublobectomy or lobectomy is commonly used, and the 5-year survival rate is close to 100%. For IAC, unless the lesion diameter is less than 2 cm or the ground-glass component is greater than 75%, the 5-year survival rate is only 60%-80% even if lobectomy and lymph node dissection are performed. Therefore, the preoperative differentiation of PIA, MIA and IAC appearing as pGGNs is very important for clinical decision making.

At present, the invasiveness of pGGNs is usually diagnosed clinically based on conventional qualitative and quantitative CT parameters that can be recognized by radiologists with naked eyes, such as the average CT value, lesion size, lobulation and spiculation et al. (8–11). However, the recognition of these features largely depends on the experience of radiologists, which is subjective and time-consuming. Radiomics, as an emerging technology, transforms medical images into quantitative data and then extracts many quantitative features that can be used to accurately and quickly evaluate tumor characteristics (12). It has the advantages of strong explanation and more stable performance on a large number of small-scale medical data sets. At present, it is still widely studied in the field of clinical computer-aided detection (CAD). The domain of investigation in radiomics consists of large-scale radiological image analysis and association with biological or clinical endpoints such as differential diagnosis, survival time prediction, disease metastasis prediction and so on (13–15). Many studies have confirmed that radiomics had high clinical application value in the invasiveness classification of lung adenocarcinoma manifesting as GGNs (2, 16–19). Our previous research also established an efficient clinical-radiomics model to classify the invasiveness of pGGNs (20). However, current studies mainly predicted the invasiveness of lung adenocarcinoma as invasive or non-invasive, and multi-classification studies with more clinical application value were rarely conducted to distinguish the degree of invasion in more detail.

Therefore, this study aims to use quantitative imaging and clinical semantic features to establish a multi-classification radiomics model that can accurately predict different invasion grades (PIA, MIA, IAC) of pGGNs, and assist patients in the early diagnosis of lung cancer and prognosis. We used a large number of clinical features provided by radiologists and radiomics features extracted from CT images. The model ensemble strategy can integrate results obtained from multiple classifiers, and has been proven to obviously improve classification and generalization performance in various research fields (21, 22). So in this work, we introduced this strategy to integrate the classification results of three algorithms, and finally constructed a multi-classification model to effectively distinguish the degree of invasion for pGGNs. The framework of our proposed model is shown in **Figure 1**.

METHODS AND MATERIALS

Patients

Our study was approved by the institutional review board (No. S-K1061), and informed consent was waived. This retrospective study reviewed the CT images of lung adenocarcinoma patients confirmed by the surgical pathology of Peking Union Medical College Hospital from November 2016 to August 2020. The inclusion criteria were as follows: (1) CT examination within one month before surgery; (2) isolated nodules with pure GGN Section (maximum long-axis diameter < 3 cm); (3) Tumor lesions in the clinical stage of T1N0M0. The exclusion criteria were as follows: (1) Radiotherapy or chemotherapy before CT examination; (2) pGGNs with very small size (maximum long-axis diameter < 3 mm). The demographic and clinical data (such as gender, age, smoking history, etc.) of patients were also recorded.

Image Acquisition

Non-contrast enhanced chest CT scans were carried out using multidetector CT scanners from Siemens (Somatom Definition Flash or Somatom Force), General Electric (Discovery CT750 HD), Philips (IQon CT) or Toshiba (Aquilion 64). Breath-hold training was carried out before each examination. The following scanning parameters were used: slice thickness/slice increment 1 mm, 0.625 mm or 0.5 mm; rotation time 0.5 or 0.6 second; pitch 0.984 or 1.2; matrix 512*512; field of view (FOV): 350 mm; standard algorithm reconstruction; tube voltage 120 kVp, tube current adjusted automatically.

Volumes of Interest (VOIs) Segmentation

The anonymized thin-slice CT images (≤1 mm, DICOM format) was delineated and segmented on lung window (window width, 1200 HU; window level, -500 HU) using ITK-SNAP (www.itk-snap.org). Two radiologists (with 15 and 4 years of experience in chest CT image interpretation) manually segmented the nodules slice by slice, both of them were blinded to the clinical data of each subject. Finally, segmentation results were output as three-dimensional VOI files (NRRD format) for subsequent feature extraction.

Radiomics Feature Extraction

A total of 1409 radiomics features were extracted from the three-dimensional VOI of each tumor by *PyRadiomics* package

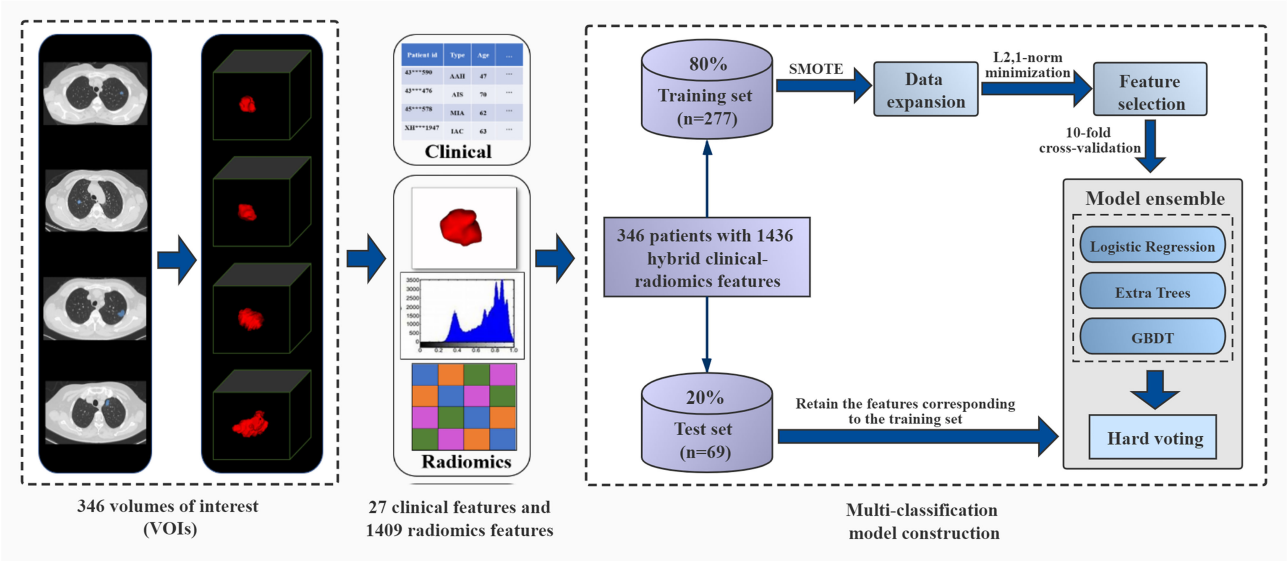


FIGURE 1 | The framework of ensemble multi-classification model based on hard voting. It includes volumes of interest (VOIs) segmentation, clinical feature collection and radiomics feature extraction, division of training set and test set, data expansion on the training set, feature selection, parameter training of three models, model ensemble with the hard voting and model performance testing.

(version 2.1.2) (23) on *python* (version 3.7.1). We extracted three categories consisting of 1409 radiomics features (**Figure 2**): (I) Tumor shape features (n = 14). They were used to quantify the degree of regularity of tumor volume shape, and all 14 features were only from the original image. (II) Tumor intensity features (n = 270). They included 18 original image features and 252 filtered image features to describe the overall density information of each tumor volume. Each original image feature was recalculated through 14 filters, so 252 filtered image features were obtained (18 * 14 = 252). (III) Tumor texture features (n = 1125).

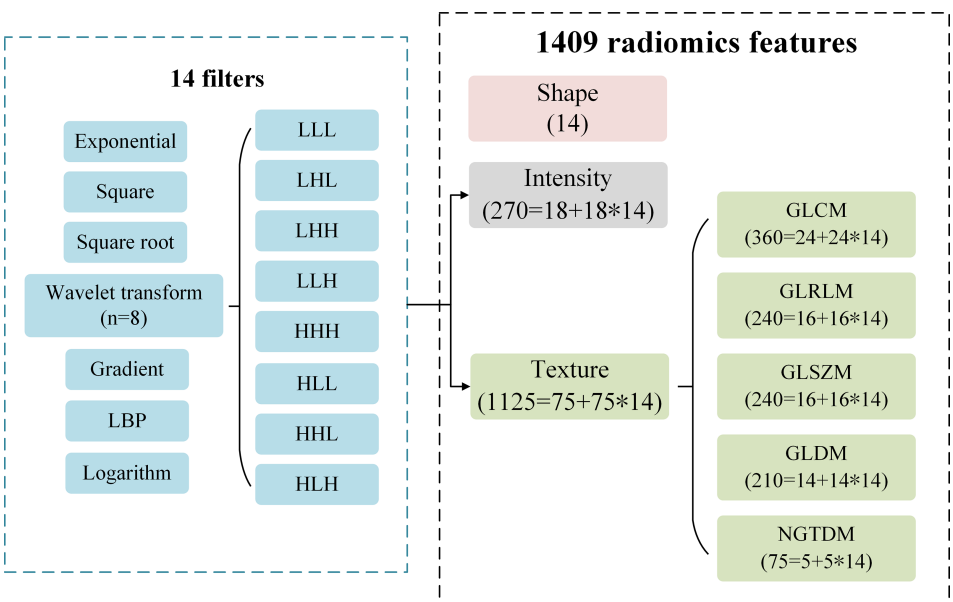


FIGURE 2 | The type description of 1409 radiomics features. A total of 1409 features consisting of intensity, shape and texture features are extracted from the original images and filtered images. A total of 14 filters are used to calculate the original intensity and texture features, respectively.

They were used to describe the heterogeneity within the tumor volume by gray level co-occurrence matrix (GLCM, n = 336), gray level run length matrix (GLRLM, n = 224), gray level size zone matrix (GLSZM, n = 224), gray level dependence matrix (GLDM, n = 196) and neighbourhood gray-tone difference matrix (NGTDM, n = 70). Among them, there were 75 features from the original image (GLCM = 24, GLRLM = 16, GLSZM = 16, GLDM = 14, NGTDM = 5). Similar to the intensity features, original texture features were also calculated through 14 filters, and a total of 1050 filtered features were obtained (75 * 14 = 1050).

Data Division and Expansion

In this work, a total of 346 pGGNs were randomly assigned to the training set (n = 277) and test set (n = 69) at a ratio of 8:2. Due to the existing problem of data imbalance (PIAs: MIAs: IACs = 88: 71: 118) on the training set, the synthetic minority oversampling technique (SMOTE) was used to expand and balance the number of samples (24). It is a commonly used data augmentation technology to deal with unbalanced data, by calculating the Euclidean distance between samples and then inserting new samples to original dataset automatically. On the training set, the 277 cases of three categories was expanded to 606 cases (PIAs: MIAs: IACs = 202: 202: 202), in which the ratio of the three categories was 1:1:1. The cases on the test set (PIAs: MIAs: IACs = 21: 18: 30) must maintain independence and no data expansion.

Feature Selection

After collecting 27 clinical features and extracting 1409 radiomics features, a total of 1436 hybrid clinical-radiomics features were obtained. Since a large number of redundant features could reduce the classification effect and cause the model to be highly complex, this study used the L2,1-norm minimization (25) for feature selection. The total 1436 features were first sorted from high to low according to their importance (weight coefficients) to the classification label (26), and then the top features were selected to participate in the classification. The number of selected features was determined according to the classification results of 10-fold cross-validation (27) on the training set.

Construction of Multi-Classification Models

In this study, we first respectively used logistic regression (LR), extra trees (ET) and gradient boosting decision tree (GBDT) algorithms to construct the three-classification model for predicting the invasiveness of pGGNs based on the selected hybrid clinical-radiomics features. Furthermore, in order to improve the classification performance, we adopted the model ensemble strategy of hard voting (22) to integrate the prediction results of the three algorithms. In addition, we also used independent clinical features and independent radiomics features to respectively construct ensemble models of the three algorithms as the comparisons. These algorithms were implemented by the scikit-learn package (version 0.23.2), and all model training process was completed on python 3.7.1. The 10-fold cross-validation and grid search were used to find optimal hyperparameters on the training set, and then the manual fine-tuning process was executed.

Statistical Methods

The performances of all multi-classification models were quantitatively evaluated by the precision, recall, F1-score, accuracy on the training set and the independent test set:

$$Precision = \frac{TP}{TP + FN} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F_1 - score = \frac{2 \cdot recall * precision}{(recall + precision)} = \frac{2 \cdot TP}{2 \cdot TP + FN + FP} \quad (3)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

where *TP*, *TN*, *FP* and *FN* stand for true positive, true negative, false positive and false negative, respectively. And all evaluation metrics were performed in the *scikit-learn* package. The above evaluation indicators of multi-classification can be directly calculated through *python* (version 3.7.1). Other simple data recording and calculation were done using Excel 2016 (Microsoft Corp., Seattle, WA, USA). And the statistical significance of t-test was set at $p < 0.05$.

RESULTS

The Result of Patient Screening

In this study, a total of 1292 consecutive patients with pathologically confirmed lung adenocarcinoma presenting as ground glass opacity (GGO) nodules on thin-slice CT at our hospital (2016/11-2020/08) were initially collected. By inclusion criteria, 630 patients were obtained and then further screened by exclusion criteria (**Figure 3**). Finally, 346 pGGNs met the standard. All pGGNs were confirmed by experienced radiologists as AAH (n = 29), AIS (n = 80), MIA (n = 89), or IAC (n = 148).

Patients and Clinical Features Collection

The clinical features collected by the research include 4 basic clinical features from medical records and 15 conventional CT features, as shown in **Table 1**. This study used one-hot encoding to quantitatively process clinical features. One-hot encoding is a data processing method that converts qualitative disordered data into quantitative ordered data (28). The main idea is to use multiple state registers to encode multiple states, so that each state has an independent register, and only one digit is valid at any time (29). After one-hot encoding, 19 original clinical features were converted into 27 usable features. The cases in the training set and the test set do not show significant differences in all clinical features.

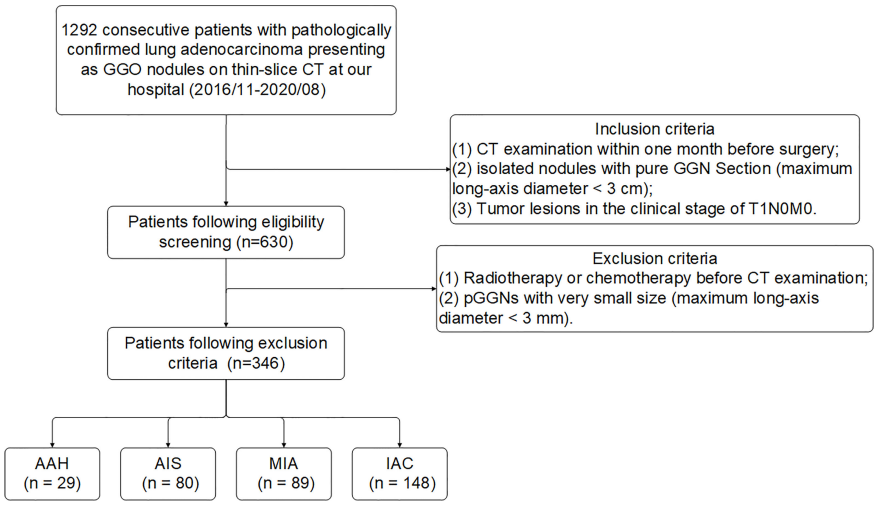


FIGURE 3 | Flowchart of patient enrollment and exclusion criteria of data set. Numbers in parentheses are the numbers of pGGNs. GGO, ground glass opacity nodule.

The Result of Feature Selection

This multi-classification research used the L2,1-norm minimization and logistic regression algorithm to perform feature selection from the 1436 hybrid clinical-radiomics features on the training set. As shown in **Figure 4**, the average accuracy and standard deviation values corresponding to the number ($1 \leq n \leq 300$) of selected features were calculated by 10-fold cross-validation. It could be seen that when the number of selected features was 166, the highest accuracy value (0.931 ± 0.026) with a small standard deviation was obtained on the training set, so these 166 features could form an effective feature set for distinguish the degree of invasion for pGGNs. The detailed results of feature selection are shown in **Supplementary Table S1**.

Analysis of Selected Features

The weight coefficients of top 10 features are shown in **Figure 5A**, and the complete weight coefficients of all 166 features are listed in **Supplementary Table S2**. The three most important features with the highest weight coefficients are wavelet-HLL_firstorder_Minimum (0.568), wavelet-HLL_ngtdm_Busyness (0.542) and square_firstorder_Kurtosis (0.476).

As shown in **Figure 5B**, the 166 selected features include 1 clinical feature (clinical short-axis diameter, ranked 23th) and 165 radiomics features. There are 4 (2%), 71 (43%) and 90 (55%) radiomics features from the tumor shape, intensity and texture features, respectively. Among the 90 tumor texture features, GLCM (n = 18), GLDM (n = 23), GLRLM (n = 18), GLSZM (n = 26) and NGTDM (n = 5) are all clearly present. We further analyze the importance of different categories of the selected 166 features through the average weight coefficient, as shown in **Figure 5C**. There is only one clinical feature, so its *p* value cannot be calculated. Among other radiomics categories, the features of intensity, texture GLDM, texture GLSZM and texture NGTDM show higher average weight coefficients than other feature categories, but no significant differences are found. Therefore, it can be considered that each feature category plays an important role for the multi-classification of invasiveness of pGGNs. The **Figure 6** shows the specific CT images of short-axis diameter with different invasion levels (AAH, AIS, MIA and IAC).

Predictive Performance of Multi-Classification Models

In this study, in order to distinguish among PIAs, MIAs and IACs, we respectively used three machine learning algorithms (LR, ET and GBRT) based on hybrid clinical-radiomics features to construct three multi-classification models. The three models were named hybrid-LR, hybrid-ET and hybrid-GBDT. We further integrated the results of three algorithms to obtain a hybrid-ensemble model through the model ensemble strategy. In addition, we also carried out the feature selection process from independent clinical features or radiomics features, as shown in **Figure S2**. Then we respectively constructed the clinical-ensemble model and radiomics-ensemble model based on the selected 20 clinical features and 275 radiomics features. Therefore, a total of 6 models were constructed, and their prediction confusion matrices on the test set are shown in **Figure 7**. It can be observed that the prediction performance of the six models for PIAs and IACs is better than MIAs, and the misclassified MIAs are more likely to be predicted as IACs than PIAs. The hybrid-ensemble model correctly classified more pGGNs on the test set compared to other five models. It could distinguish between the PIAs and IACs perfectly (There is no misclassification between the PIAs and IACs), and their wrong predictions were all classified as MIAs. For the hybrid-ensemble model, most of the misclassified cases (n = 6) of MIAs were predicted to be IACs, and only one MIA was incorrectly predicted as PIA.

For the 6 models, **Table 2** quantitatively lists their sensitivities of different invasion levels and overall classification accuracies on the training set and test set. Consistent with what is observed in **Figure 5**, the hybrid-ensemble model shows the strongest predictive

TABLE 1 | Clinical features of 346 patients on the training set and test set.

Clinical features	Total (n=346)	Training set (n=277)	Test set (n=69)	*p* value
Age (years)	55.79 ± 10.53 (24-83)	55.72 ± 10.83 (27-83)	56.06 ± 10.70 (24-76)	0.811
Gender				0.391
Female	297 (85.8)	240 (86.6)	57 (82.6)	
Male	49 (14.2)	37 (13.4)	12 (17.4)	
Smoking history				0.419
Never smoker	319 (92.2)	257 (92.8)	62 (89.9)	
Former/current smoker	27 (7.8)	20 (7.2)	7 (10.1)	
Smoking index (pack-years)	1.34 ± 6.18 (0-75)	1.27 ± 6.34 (0-75)	1.59 ± 5.50 (0-30)	0.698
Lesion involved lobe				0.764
Right upper lobe	126 (36.4)	101 (36.5)	25 (36.2)	
Right middle lobe	17 (4.9)	12 (4.3)	5 (7.2)	
Right lower lobe	62 (17.9)	51 (18.4)	11 (15.9)	
Left upper lobe	99 (28.6)	80 (28.9)	19 (27.5)	
Left lower lobe	42 (12.1)	33 (11.9)	9 (13.0)	
Maximum long-axis diameter (mm)	12.39 ± 5.60 (3.5-30)	12.37 ± 5.63 (3.5-30)	12.49 ± 5.50 (3.9-26)	0.879
Short-axis diameter (mm)	9.92 ± 4.47 (1.8-29)	9.90 ± 4.53 (1.8-29)	10.01 ± 4.23 (2.3-22)	0.856
Mean CT attenuation (HU)	-531.60 ± 138.07 (-801.5, -188)	-533.19 ± 138.15 (-790.9, -188)	-525.23 ± 138.59 (-801.5, -202.3)	0.669
SD of CT attenuation (HU)	96.72 ± 76.21 (4.5-1059)	97.88 ± 79.59 (4.5-1059)	92.07 ± 61.10 (16.1-317.1)	0.572
Nodule shape				0.980
Round or oval	165 (47.7)	132 (47.7)	33 (47.8)	
Irregular or polygonal	181 (52.3)	145 (52.3)	36 (52.2)	
Tumor-lung interface				0.448
Ill-defined	38 (11.0)	31 (11.2)	7 (10.1)	
Well-defined and smooth	207 (59.8)	168 (60.6)	39 (56.5)	
Well-defined but coarse	101 (29.2)	78 (28.2)	23 (33.3)	
Spiculation (-)	115 (33.2)	91 (32.9)	24 (34.8)	0.761
Spiculation (+)	231 (66.8)	186 (67.1)	45 (65.2)	
Lobulation (-)	128 (37.0)	106 (38.3)	22 (31.9)	0.327
Lobulation (+)	218 (63.0)	171 (61.7)	47 (68.1)	
Cavity (-)	343 (1)	274 (98.9)	69 (100)	0.387
Cavity (+)	3 (0)	3 (1.1)	0 (0)	
Vacuole sign (-)	92 (26.6)	72 (26.0)	20 (29.0)	0.616
Vacuole sign (+)	254 (73.4)	205 (74.0)	49 (71.0)	
Air bronchogram (-)	224 (64.7)	184 (66.4)	40 (58.0)	0.189
Air bronchogram (+)	122 (35.3)	93 (33.6)	29 (42.0)	
Vascular convergence (-)	155 (44.8)	127 (45.8)	28 (40.6)	0.432
Vascular convergence (+)	191 (55.2)	150 (54.2)	41 (59.4)	
Intranodular vascular anomaly				0.737
None	86 (24.9)	70 (25.3)	16 (23.2)	
Vessels entering with natural contour	99 (28.6)	79 (28.5)	20 (29.0)	
Vessels ingress into the nodule with dilated or distorted branches	161 (46.5)	128 (46.2)	33 (47.8)	
Pleural retraction sign (-)	243 (70.2)	193 (69.7)	50 (72.5)	0.651
Pleural retraction sign (+)	103 (29.8)	84 (30.3)	19 (27.5)	

The data are displayed as mean ± standard deviation (range) or number (%).
P value is derived from the t-test (two-tailed distribution, equal variance assumption) between training set and test set.

performance among all 6 models. On the training set and test set, it obtained the F1-scores of 0.917 ± 0.024 and 0.824, and its accuracy values respectively were 0.918 ± 0.022 and 0.841. That indicated that the model ensemble strategy and hybrid clinical-radiomics features are important to improve the three-classification performance.

DISCUSSION

In this study, we collected 27 clinical features and extracted 1409 radiomics features from each tumor three-dimensional VOI. After feature selection, we selected an effective feature set consisting of 166 features from the 1436 hybrid clinical-radiomics features. Based on the 166 hybrid features, we used three machine learning algorithms (LR, ET and GBDT) to construct three multi-classification models to distinguish the different invasion levels (PIA, MIA and IAC) of pGGNs. We further integrated the results of three algorithms to obtain a hybrid-ensemble model through the model ensemble strategy. Finally, we successfully constructed a multi-classification model to effectively distinguish different degrees of invasion for pGGNs. The proposed hybrid-ensemble model achieved the F1-score of 0.824 and an accuracy value of 0.841 on the independent test set, showing promising classification performance.

A precise diagnosis of the tumor invasion status is very important to guide individualized therapy in clinical practice.

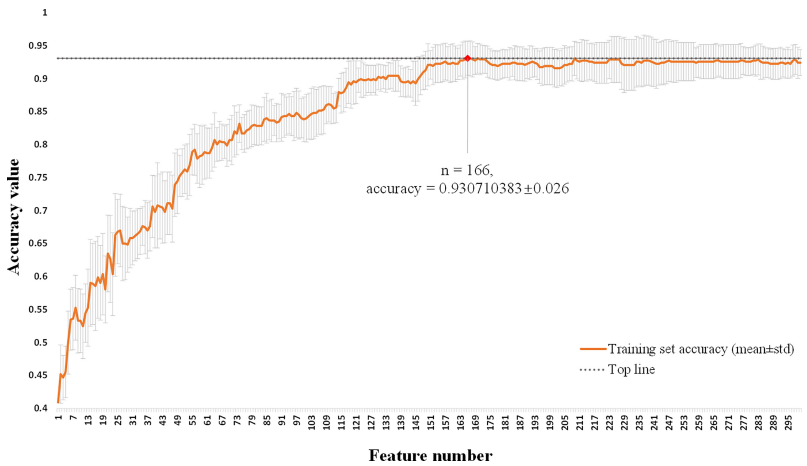


FIGURE 4 | Feature selection of hybrid clinical-radiomics model using L2,1-norm minimization and logistic regression algorithm. The horizontal axis is the number of selected features ($1 \leq n \leq 300$). The vertical axis shows the corresponding average accuracy value of 10-fold cross-validation on the training set, and the gray area is the standard deviation. When the feature number is 166, the maximum accuracy value is obtained with the small standard deviation.

Early-stage lung adenocarcinoma often presents as GGN and has atypical features, which makes the differential diagnosis of the adenocarcinoma subtypes more difficult. Therefore, auxiliary identification by radiomics is necessary for early detection and prognosis of patients. Current researches mainly predicted the invasiveness of lung adenocarcinoma as invasive or non-invasive (2, 16–20, 30–33), and the multi-classification studies were rarely conducted to distinguish the degree of invasion in more detail. Our study attempted the three-classification of aggressive pGGNs, which is more meaningful.

Through the quantitative analysis of CT images, radiomics could objectively reflect both the attenuation and dispersion of

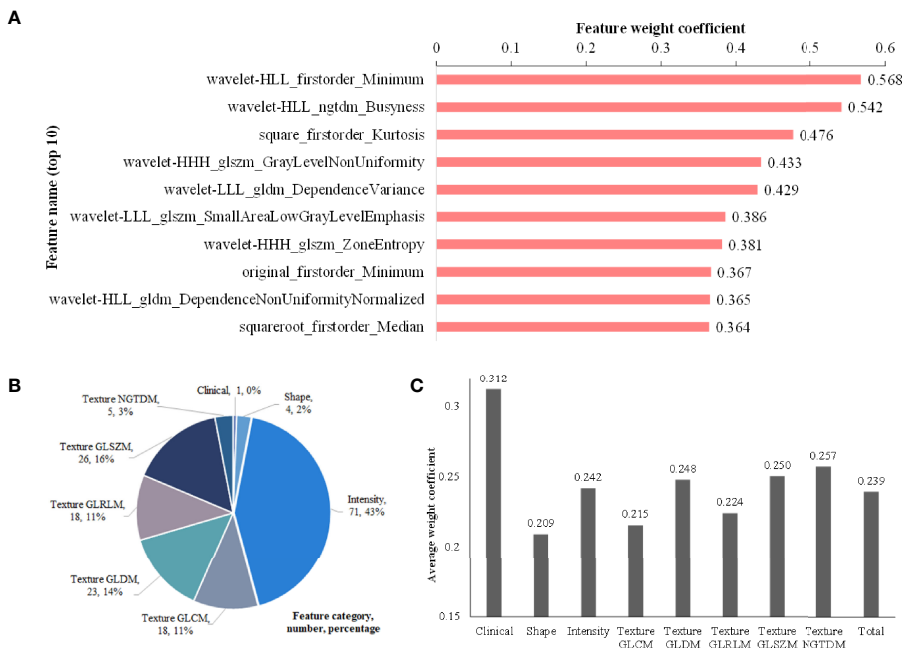


FIGURE 5 | Feature analysis. **(A)** Histogram showing the weight coefficients of top 10 features within selected 166 hybrid features; **(B)** Description about category names, numbers and percentages of the 166 features; **(C)** The average weight coefficient of every category for the selected 166 features (There is only one clinical feature, so its *p* value cannot be calculated. No significant differences are found among other categories).

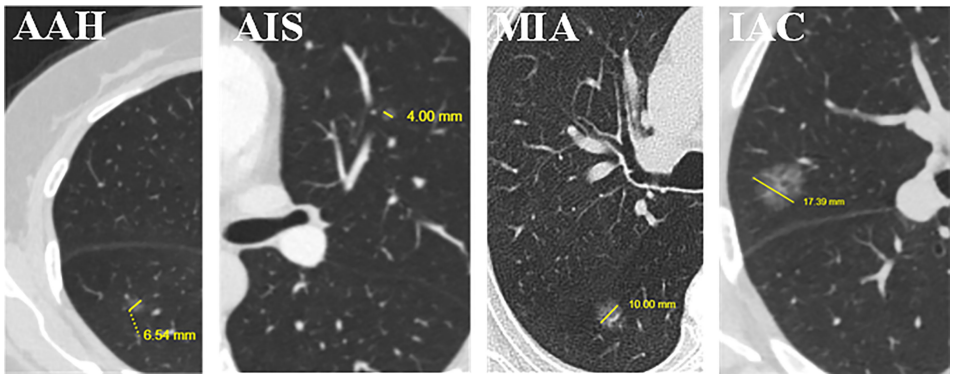


FIGURE 6 | Examples of short-axis diameter (the vertical diameter of the longest diameter of the largest cross-section) (mm) for the four levels of invasion. We found that it is the only clinical feature in the 166 selected features used by the proposed hybrid-ensemble model. From left to right: atypical adenomatous hyperplasia (AAH), 6.54 mm; adenocarcinoma *in situ* (AIS), 4.00 mm; minimally invasive adenocarcinoma (MIA), 10.00 mm; invasive adenocarcinoma (IAC), 17.39 mm.

gray level intensity, which might not be evident in direct visual assessments. Recent studies have shown that intensity and texture radiomics features are useful for predicting the invasiveness of lung adenocarcinoma presenting as GGNs (17, 18). This finding is consistent with our study, as the machine learning feature selection procedure selected 71 (43%) intensity and 90 (55%) texture features to establish the hybrid-ensemble model. In addition, in total 166 features were selected, of which only one clinical feature (short-axis diameter, ranked 23th). It meant that the short-axis diameter was the most important parameter for the invasive classification of pGGNs among the 27 clinical features. We found that in general, lung nodules with large short-axis diameter have the higher degree of invasion. Compared with the maximum long-axis diameter, the short-axis

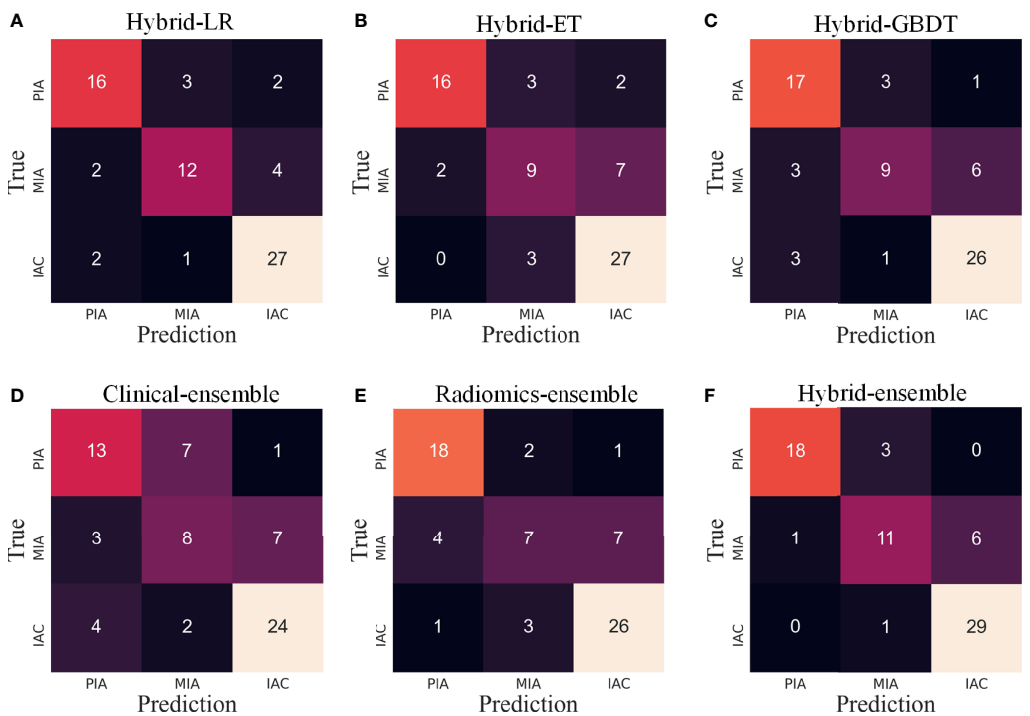


FIGURE 7 | The confusion matrices of various models on the test set. LR, logistic regression; ET, extra trees; GBDT, gradient boosting decision tree. **(A–C)** Three algorithms (LR, ET, GBDT) with hybrid clinical-radiomics features; **(D–F)** Clinical model, radiomics model and hybrid clinical-radiomics model based on model ensemble of the three algorithms.

TABLE 2 | The comparison of classification performance using different feature groups and algorithms.

Feature groups	Algorithms	Training set				Test set			
		Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score	Accuracy
Hybrid	LR	0.937 ± 0.033	0.933 ± 0.054	0.935 ± 0.041	0.931 ± 0.026	0.789	0.776	0.782	0.797
	ET	0.808 ± 0.103	0.805 ± 0.097	0.806 ± 0.100	0.807 ± 0.057	0.746	0.721	0.733	0.754
	GBDT	0.921 ± 0.057	0.908 ± 0.060	0.914 ± 0.058	0.909 ± 0.021	0.740	0.726	0.733	0.754
Clinical	Ensemble	0.795 ± 0.067	0.785 ± 0.049	0.790 ± 0.057	0.785 ± 0.029	0.624	0.621	0.622	0.652
Radiomics		0.865 ± 0.082	0.815 ± 0.037	0.839 ± 0.051	0.813 ± 0.067	0.710	0.704	0.707	0.739
Hybrid-ensemble model		0.917 ± 0.024	0.917 ± 0.024	0.917 ± 0.024	0.918 ± 0.022	0.836	0.812	0.824	0.841

LR, logistic regression; ET, extra trees; GBDT, gradient boosting decision tree. ± represents the standard deviation of the values in the 10-fold cross-validation.

diameter implies a longer diameter in the vertical direction, which represents more nodule size information. Previous studies (16, 30–32) found that the size (usually quantified by area) of the nodule is an important parameter for assessment of lung adenocarcinoma invasiveness, which is somewhat consistent with short-axis diameter. However, we believe that the short-axis diameter may be more advantageous in some respects, as it contains information about the shape of the nodule in addition to its size.

Previous studies tried hybrid clinical-radiomics features to build radiomics models, and the results showed that this is effective for more accurate classification (2, 20, 33). Our study also demonstrated this, using the joint 1436 features make the hybrid-ensemble model perform better than clinical-ensemble and radiomics-ensemble. In addition, we further introduced the model ensemble strategy, which has not been tried by researchers before, and our model comparison experiments showed that this strategy is also very effective. For the proposed hybrid-ensemble model, the classification performance of MIAs is slightly low, similar to the fact that it is more difficult for clinicians to distinguish MIAs in actual clinical diagnosis, which may be because MIAs are of the intermediate degree of invasion. We further found that most of the misclassified cases of MIAs were predicted to be IACs, which means that these two grades were more difficult to be distinguished. In addition, the hybrid-ensemble model had no misclassification to distinguish between IACs and PIAs, showing its potential clinical application value.

This study has several limitations. First of all, this is a single-center retrospective study, and a multi-center study is better to be conducted to further evaluate the model performance. Second, relying only on the radiologists to manually delineate and segment the region of interest is more time-consuming and subjective, and reliable and automatic methods are essential to simplify the complex procedures.

In conclusion, this study used the short-axis diameter parameter and 165 radiomics features to construct a multi-classification model for precisely predicting the invasiveness of lung adenocarcinoma with pGGNs. We found that short-axis diameter was the most important parameter among 27 clinical features. The hybrid-ensemble model based on hybrid clinical-radiomics features and model ensemble strategy had better predictive performance, and could have a promising clinical application value.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

FS and LS made major contributions to the research and writing of manuscripts.TX, YF, and XS assisted in the research and data analysis. PZ and TZ provided some code suggestions. ZZ assisted in data curation and processing. WS and GZ provided the supervisions of the entire study and made contributions to review and editing of manuscripts. All authors contributed to the article and approved the submitted version.

FUNDING

This work was partially supported by the Beijing Natural Science Foundation (7202102), the National Natural Science Foundation of China (61871022), the Fundamental Research Funds for Central Universities, the 111 Project (B13003), and the 2021 SKY Imaging Research Fund of Chinese International Medical Exchange Foundation (Z-2014-07-2101).

ACKNOWLEDGMENTS

We are grateful to the Chinese Academy of Medical Sciences and Peking Union Medical College for the clinical data collection and analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.800811/full#supplementary-material

REFERENCES

1. Henschke CI, Yankelevitz DF, Mirtcheva R, McGuinness G, McCauley D, Miettinen OS, et al. CT Screening for Lung Cancer: Frequency and Significance of Part-Solid and Nonsolid Nodules. *Am J Roentgenol* (2002) 178:1053–7. doi: 10.2214/ajr.178.5.1781053
2. Meng F, Guo Y, Li M, Lu X, Wang S, Zhang L, et al. Radiomics Nomogram: A Noninvasive Tool for Preoperative Evaluation of the Invasiveness of Pulmonary Adenocarcinomas Manifesting as Ground-Glass Nodules. *Transl Oncol* (2021) 14(1):100936. doi: 10.1016/j.tranon.2020.100936
3. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger K, Yatabe Y, et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Muhidiseiplinary Classification of Lung Adenocarcinoma. *J Thorac Oncol* (2011) 6(2):244–85. doi: 10.1513/pats.201107-042ST
4. Tafe L, Abreu FD, Peterson J, Finley D, Black C. Genomic Relationship Between Lung Adenocarcinoma and Synchronous AIS/AAH Lesions in the Same Lobe. *J Thorac Oncol* (2017) 12(1):S537–7. doi: 10.1016/j.jtho.2016.11.664
5. Dembitzer FR, Flores RM, Parides MK, Beasley MB. Impact of Histologic Subtyping on Outcome in Lobar vs Sublobar Resections for Lung Cancer: A Pilot Study. *Chest* (2014) 146(1):175–81. doi: 10.1378/chest.13-2506
6. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol* (2015) 10(9):1243–60. doi: 10.1097/jto.0000000000000630
7. Mei X, Rui W, Yang W, Qian F, Ye X, Zhu L, et al. Predicting Malignancy of Pulmonary Ground-Glass Nodules and Their Invasiveness by Random Forest. *J Thorac Dis* (2018) 10(1):458–63. doi: 10.21037/jtd.2018.01.88
8. Si M, Tao X, Du G, Cai L, Han H, Liang X, et al. Thin-Section Computed Tomography_Histopathologic Comparisons of Pulmonary Focal Interstitial Fibrosis, Atypical Adenomatous Hyperplasia, Adenocarcinoma in Situ, and Minimally Invasive Adenocarcinoma With Pure Ground-Glass Opacity. *Eur J Radiol* (2016) 85:1708–15. doi: 10.1016/j.ejrad.2016.07.012
9. Jin X, Zhao S, Gao J, Wang D, Wu J, Wu C, et al. CT Characteristics and Pathological Implications of Early Stage (T1N0M0) Lung Adenocarcinoma With Pure Ground Glass Opacity. *Eur Radiol* (2015) 25:2532–40. doi: 10.1007/s00330-015-3637-z
10. Son JY, Lee HY, Kim JH, Han J, Jeong J, Lee KS, et al. Quantitative CT Analysis of Pulmonary Ground-Glass Opacity Nodules for Distinguishing Invasive Adenocarcinoma From Non-Invasive or Minimally Invasive Adenocarcinoma: The Added Value of Using Iodine Mapping. *Eur Radiol* (2016) 26:43–54. doi: 10.1007/s00330-015-3816-y
11. Shikuma K, Menju T, Chen F, Kubo T, Muro S, Sumiyoshi S, et al. Is Volumetric 3-Dimensional Computed Tomography Useful to Predict Histological Tumour Invasiveness? Analysis of 211 Lesions of Ct1n0m0 Lung Adenocarcinoma. *Interact CardiovTh* (2016) 22:831–8. doi: 10.1093/icvts/ivw037
12. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Stiphout R, Granton P, et al. Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis. *Eur J Cancer* (2012) 48(4):441–6. doi: 10.1016/j.ejca.2011.11.036
13. Gillies R, Kinahan P, Hricak H. Radiomics: Images are More Than Pictures, They are Data. *Radiology* (2015) 278:563–77. doi: 10.1148/radiol.2015151169
14. Fan M, Xia P, Clarke R, Wang Y, Li L. Radiogenomic Signatures Reveal Multiscale Intratumour Heterogeneity Associated With Biological Functions and Survival in Breast Cancer. *Nat Commun* (2020) 11:4861. doi: 10.1038/s41467-020-18703-2
15. Vaidya P, Bera K, Gupta A, Wang X, Corredor G, Fu P, et al. CT Derived Radiomic Score for Predicting the Added Benefit of Adjuvant Chemotherapy Following Surgery in Stage I, II Resectable Non-Small Cell Lung Cancer: A Retrospective Multi-Cohort Study for Outcome Prediction. *Lancet Digit Health* (2020) 2:e116–28. doi: 10.1016/s2589-7500(20)30002-9
16. Lee SM, Park CM, Goo JM, Lee HJ, Wi JY, Kang CH. Invasive Pulmonary Adenocarcinomas Versus Preinvasive Lesions Appearing as Ground-Glass Nodules: Differentiation by Using CT Features. *Radiology* (2013) 268(1):265–73. doi: 10.1148/radiol.13120949
17. She Y, Zhang L, Zhu H, Dai C, Xie H, Zhang W, et al. The Predictive Value of CT-Based Radiomics in Differentiating Indolent From Invasive Lung Adenocarcinoma in Patients With Pulmonary Nodules. *Eur Radiol* (2018) 28(12):5121–8. doi: 10.1007/s00330-018-5509-9
18. Fan L, Fang M, Li Z, Tu W, Wang S, Chen W, et al. Radiomics Signature: A Biomarker for the Preoperative Discrimination of Lung Invasive Adenocarcinoma Manifesting as a Ground-Glass Nodule. *Eur Radiol* (2018) 29(2):889–97. doi: 10.1007/s00330-018-5530-z
19. Weng Q, Zhou L, Wang H, Hui J, Chen M, Pang P, et al. A Radiomics Model for Determining the Invasiveness of Solitary Pulmonary Nodules That Manifest as Part-Solid Nodules. *Clin Radiol* (2019) 74(12):933–43. doi: 10.1016/j.crad.2019.07.026
20. Song L, Xing T, Zhu Z, Han W, Fan G, Li J, et al. Hybrid Clinical-Radiomics Model for Precisely Predicting the Invasiveness of Lung Adenocarcinoma Manifesting as Pure Ground-Glass Nodule. *Acad Radiol* (2020) 44(8):1892–5. doi: 10.1016/j.acra.2020.05.004
21. Tan AC, Gilbert D. Ensemble Machine Learning on Gene Expression Data for Cancer Classification. *Bioinformatics* (2003) 2:S75–83.
22. Zhang C, Ma Y. *Ensemble Machine Learning: Methods and Applications*. New York: Springer-Verlag (2012). doi: 10.1007/978-1-4419-9326-7, .
23. Griethuysen J, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339
24. Chawla NV, Bowyer KW, Hall HO , Kegelmeyer WP. SMOTE: Synthetic Minority Over-Sampling Technique. *J Artif Intell Res* (2002) 16:321–57. doi: 10.1613/jair.953
25. Nie F, Huang H, Cai X, Ding CHQ. Efficient and Robust Feature Selection *via* Joint L2,1-Norms Minimization. In: *Proceedings of the 23rd International Conference on Neural Information Processing Systems*. United States: AAAI (2010). 1813–21. doi: 10.5555/2997046.2997098
26. Liu J, Cui J, Liu F, Yuan Y, Guo F, Zhang G. Multi-Subtype Classification Model for Non-Small Cell Lung Cancer Based on Radiomics: SLS Model. *Med Phys* (2019) 46:3091–100. doi: 10.1002/mp.13551
27. Arlot S, Celisse A. A Survey of Cross-Validation Procedures for Model Selection. *Stat Survey* (2010) 4:40–79. doi: 10.1214/09-SS054
28. Chren W. One-Hot Residue Coding for Low Delay-Power Product CMOS Design. *Circuit Syst Signal Proc* (1998) 45(3):303–13. doi: 10.1109/82.664236
29. Rabinowitz L. Mathematical Statistics and Data Analysis. *Technometrics* (1989) 31(3):390–1. doi: 10.2307/1269179
30. Wang B, Tang Y, Chen Y, Hamal P, Zhu Y, Wang T. Joint Use of the Radiomics Method and Frozen Sections Should Be Considered in the Prediction of the Final Classification of Peripheral Lung Adenocarcinoma Manifesting as Ground-Glass Nodules. *Lung Cancer* (2020) 139:103–10. doi: 10.1016/j.lungcan.2019.10.031
31. She Y, Zhao L, Dai C, Ren Y, Zha J, Xie H, et al. Preoperative Nomogram for Identifying Invasive Pulmonary Adenocarcinoma in Patients With Pure Ground-Glass Nodule: A Multi-Institutional Study. *Oncotarget* (2016) 8 (10):17229. doi: 10.18632/oncotarget.11236
32. Xue X, Yang Y, Huang Q, Cui F, Lian Y, Zhang S, et al. Use of a Radiomics Model to Predict Tumor Invasiveness of Pulmonary Adenocarcinomas Appearing as Pulmonary Ground-Glass Nodules. *BioMed Res Int* (2018) 2018:1–9. 10.1155/2018/6803971.
33. Wu YJ, Liu YC, Liao CY, Tang EK, Wu FZ. A Comparative Study to Evaluate CT-Based Semantic and Radiomic Features in Preoperative Diagnosis of Invasive Pulmonary Adenocarcinomas Manifesting as Subsolid Nodules. *Sci Rep* (2021) 11:66. doi: 10.1038/s41598-020-79690-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 06 May 2022
doi: 10.3389/fonc.2022.881467

Obesity Does Not Increase Perioperative Outcomes in Older Patients Undergoing Thoracoscopic Anatomic Lung Cancer Surgery

Chaoyang Tong[1,2†], *Tingting Li*[1†], *Yaofeng Shen*[1†], *Hongwei Zhu*[1], *Jijian Zheng*[2*] *and Jingxiang Wu*[1*]

[1] Department of Anesthesiology, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China, [2] Department of Anesthesiology, Shanghai Children's Medical Center, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

OPEN ACCESS

Edited by:
Vamsi Velcheti,
New York University, United States

Reviewed by:
Massimo Baudo,
Spedali Civili Brescia, Italy
Michael Ried,
University Hospital Regensburg, Germany

***Correspondence:**
Jingxiang Wu
wjx1132xk@163.com
Jijian Zheng
zhengjijian626@sina.com

†These authors share first authorship

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: 22 February 2022
Accepted: 07 April 2022
Published: 06 May 2022

Citation:
Tong C, Li T, Shen Y, Zhu H, Zheng J and Wu J (2022) Obesity Does Not Increase Perioperative Outcomes in Older Patients Undergoing Thoracoscopic Anatomic Lung Cancer Surgery. Front. Oncol. 12:881467. doi: 10.3389/fonc.2022.881467

Objectives: To investigate the relationship between obesity status and perioperative outcomes in elderly patients undergoing thoracoscopic anatomic lung cancer surgery.

Methods: From January 2016 to December 2018, we performed a monocentric retrospective cohort study among 4164 consecutive patients aged 65 years or older who underwent thoracoscopic anatomic lung cancer surgery at Shanghai Chest Hospital. Two groups were stratified by body mass index (BMI): nonobese (BMI<28kg/m^2) and obese status (BMI≥28kg/m^2). Using a 1:1 propensity score matching (PSM) analysis to compare perioperative outcomes between two groups.

Results: 4035 older patients were eventually enrolled, with a mean age of 69.8 years (range: 65-87), and 305 patients were eligible for obese status, with a mean BMI of 29.8 ± 1.7kg/m^2. Compared with nonobese patients, obese patients were more likely to have higher rates of intraoperative hypoxemia (1.2% vs 3.9%, P=0.001) and new-onset arrhythmia (2.3% vs 4.3%, P=0.034). The difference in intraoperative transfusion and conversion rates and postoperative outcomes regarding pulmonary complications, new-onset arrhythmia, transfusion, length of hospital stay, 30-day readmission and hospitalization costs between two groups were not significant (P>0.05). After a 1:1 PSM analysis, the difference in both intraoperative and postoperative complications among two groups were not significant (P>0.05). In subgroup analysis, patients with BMI≥30kg/m^2 had a similar incidence of perioperative complications compared to patients with BMI between 28 and 30 kg/m^2 (P>0.05).

Conclusions: Our research data support evidence for "obesity paradox" and also contribute the growing body of evidence that obesity in older patients should not exclude candidates for thoracoscopic anatomic lung cancer surgery.

Keywords: elderly, obesity, thoracoscopic surgery, lung cancer, outcomes

INTRODUCTION

Over the last few decades, the prevalence of obesity has shifted dramatically (1, 2), and obesity-related metabolic diseases such as hypertension, diabetes and hyperlipidemia have also gradually increased (3–5), becoming a worldwide health problem. Currently, obesity-related cancer rates have become a focus of concern, although the effect varies widely depending on the types of cancer (4, 6–9). Among these types, an inverse relationship between obesity and lung cancer risk has been reported in previous studies (6–9). And the biological mechanisms underlying the major role of obesity in chronic inflammation and carcinogenesis have also been extensively evaluated (4, 10–12).

Indeed, the proportion of overweight and obese lung cancer patients undergoing lung surgery has increased during the last decades (13, 14). Besides, obesity-related preoperative comorbidities increase the risk of perioperative surgical complications, mutually reinforcing (15–17). Additionally, obesity characterized by substantial girth and excess mediastinal fat has a profound impact on the technical issues during the procedure, especially in minimally invasive surgery (MIS) (18, 19). Moreover, with the aging of society, the influence of obesity on perioperative outcomes is further complicated by the specific attributes of elderly surgical patients, such as frailty, cognitive decline, impaired preoperative pulmonary function and tissue fragility (20, 21).

Recently, a retrospective study conducted by Tabatabai and colleagues showed that obesity class II-III had prognostic value in predicting increased rates of postoperative complications, prolonged length of hospital stay and no-home discharge locations in elderly patients undergoing spine, hip, and knee procedures (22). Up to now, there was no literature about the influence of obesity on perioperative outcomes in elderly patients during thoracoscopic anatomic lung cancer surgery. In the present study, by reviewing a large sample of clinical data, we aimed to evaluate the relationship of obesity to perioperative outcomes in thoracoscopic anatomic lung cancer surgery.

MATERIALS AND METHODS

Study Design and Patients

From January 2016 to December 2018, we performed a monocentric retrospective cohort study based on a prospectively collected database, including 4164 consecutive patients aged 65 years or older who underwent thoracoscopic anatomic lung cancer surgery. Excluded patients were described in the flow diagram (**Figure 1**). 4035 elderly patients were enrolled in the final analysis. The Institutional Review Board (IRB) of Shanghai Jiao tong University, Shanghai Chest Hospital (IS21119) approved this study and waived the need for informed consent.

Anesthesia Protocol

All patients were routinely monitored by electrocardiogram, pulse oximetry, non-invasive blood pressure (NIBP) and capnography. Radial artery intubation and right internal jugular central venous catheterization (CVP) were used to monitor invasive blood pressure (IBP). For thoracic paravertebral blockade (TPVB), 20 mL 0.5% ropivacaine was injected to the T4-T5 by the experienced anesthesiologist under the guidance of ultrasound before surgery. Intraoperative lung protective ventilation (LPV) strategies consisted of low-tide ventilation based on ideal body weight (≤8mL/kg), PEEP=5cmH2O, lung recruitment and maintenance of airway pressure <30cmH2O. Extubation was performed in the operating room or post anesthesia care unit (PACU) for all suitable patients. All patients treated with patient-controlled analgesia (PCA) pump, including sufentanil1.0 μg/kg + desoxocin 0.4mg/kg.

Technique of Operation

Since 2016, the high-volume center of Shanghai Chest Hospital has performed nearly 10,000 lung operations annually, of which thoracoscopic surgery accounts for more than 80 percent. All possible thoracoscopic surgery was determined by the participating surgeons based on the individual patient's preoperative evaluation. For all patients, thoracoscopic anatomic lung resection plus systematic lymph node dissection were considered as the best treatment for primary lung cancer.

Data Collection and Definition

Perioperative clinical data were prospectively pooled from our institution's electronic medical system, enrolling patient's baseline and intraoperative characteristics, and perioperative outcomes including hypoxemia, transfusion, new-onset arrhythmia, conversion to thoracotomy, pulmonary complications, length of hospital stay (LOS), 30-day readmission and hospitalization costs. Hypoxemia was defined as SpO_2<90% for more than 10 consecutive minutes. Postoperative pulmonary complications (PPCs) refer to the European Perioperative Clinical Outcome (EPCO) (23). According to the 2014 Guidelines of the American Association of Thoracic Surgeons (AATS) (24), perioperative new-onset arrhythmia included incidents of atrial fibrillation (AF) and atrial flutter. The 30-day readmission included in the analysis was an unplanned return to the hospital due to various postoperative complications.

We classified the obesity status according to the body mass index (BMI, weight in kilograms divided by height in meters squared) of the Guidelines for Prevention and Control of Overweight and Obesity in Chinese Adults (25). Two groups were stratified by BMI: nonobese (BMI<28kg/m^2) and obese status (BMI≥28kg/m^2). In subgroup analysis, the perioperative outcomes between patients with obesity class I (28kg/m^2 ≤BMI<30kg/m^2) and obesity class II (BMI≥30kg/m^2) were investigated.

Statistical Analysis

Data was tested for normal distribution with Q-Q plot. Continuous variables were compared between nonobese and obese patients using Two independent sample t-test or Mann-Whitney U test. Categorical variables were compared with Chi-square test or Fisher exact test, depending on the sample size. To reduce the selection bias and other potential confounding effects, we performed a 1:1 propensity score matching (PSM) analysis using a caliper size of

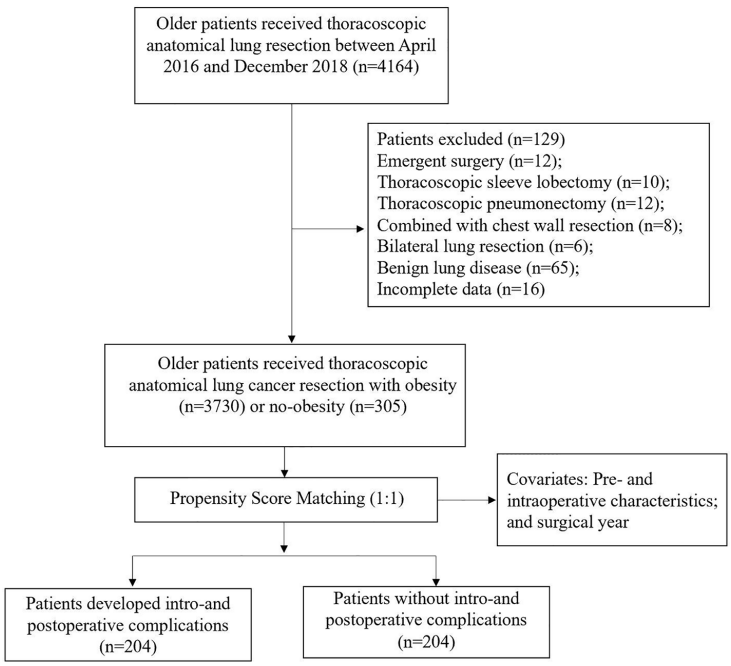


FIGURE 1 | Patient flowchart.

0.01 to compare perioperative outcomes between two cohorts. All pre-, intraoperative variables and surgical year were included in the PSM analysis. Standardized mean difference (SMD) between two cohorts on all covariables after matching was calculated, with differences of <10% indicating adequate balance in the matched cohort. Statistical analysis was performed using the SPSS 26.0 software (IBM Corp., Armonk, NY, USA). R version 4.1.2 was used with the tableone, ggplot2, reshape2, survey and Matching packages. *P*-value<0.05 was considered statistically significant.

RESULTS

Study Cohort

From January 2016 to December 2018, 4035 older patients with a mean age of 69.8 years (range: 65-87) underwent thoracoscopic anatomical lung cancer surgery, of which 17.6% (712 out of 4035) underwent segmentectomy resection and 82.4% (3323 out of 4035) underwent lobectomy resection, and 17.6% (305 out of 4035) were eligible for obese status, with a mean BMI of 29.8 ± 1.7kg/m^2 (**Figure 1**). The BMI distribution of all enrolled patients were depicted in **Figure 2**.

Patients with obese status had a higher American Society of Anesthesiologists (ASA) classification (III/IV, 23.3% vs 16.3%, P=0.007), and increased rates of hypertension (16.1% vs 11.8%, P=0.03) and diabetes mellitus (10.2% vs 6.7%, P=0.022), and better preoperative lung function (FEV_1/FVC, 105.2 ± 8.3 vs 100.9 ± 9.9, P<0.001; DLCO%, 96.7 ± 17.9 vs 90.5 ± 17.8, P<0.001) when compared with their counterparts. Additionally, patients developed with obese status required longer operative time (108.5 ± 40.1 vs 102.4 ± 39.5mins, P=0.009) compared with nonobese patients (**Tables 1**, **2**).

Perioperative Outcomes Between Two Cohorts Before and After a 1:1 PSM

Compared with nonobese patients, the rates of intraoperative hypoxemia (1.2% vs 3.9%, P=0.001) and new-onset arrhythmia (2.3% vs 4.3%, P=0.034) were higher in obese patients (**Table 3**). The differences among intraoperative transfusion, conversion rates and postoperative outcomes including pulmonary complications, new-onset arrhythmia, transfusion, LOS, 30-day readmission and hospitalization costs were not significant (**Table 3**). After a 1:1 PSM analysis, all patient characteristics were comparable in two cohorts (**Tables 4**, **5**). We investigated perioperative outcomes in 408 patients (204 pairs), the differences in both intra- and postoperative complications were not significant (**Table 6**). In subgroup analysis, the perioperative outcomes between patients with obesity class I and obesity class II were explored. When all perioperative data were comparable (**Supplementary Tables S1, S2**), we found that higher severity of obesity did not increase the incidence of perioperative complications (**Supplementary Table S3**).

DISCUSSION

305 older patients who underwent thoracoscopic anatomical lung cancer surgery were eligible for obese status. Our study found that elderly obese patients had similar rates of perioperative complications compared to nonobese patients,

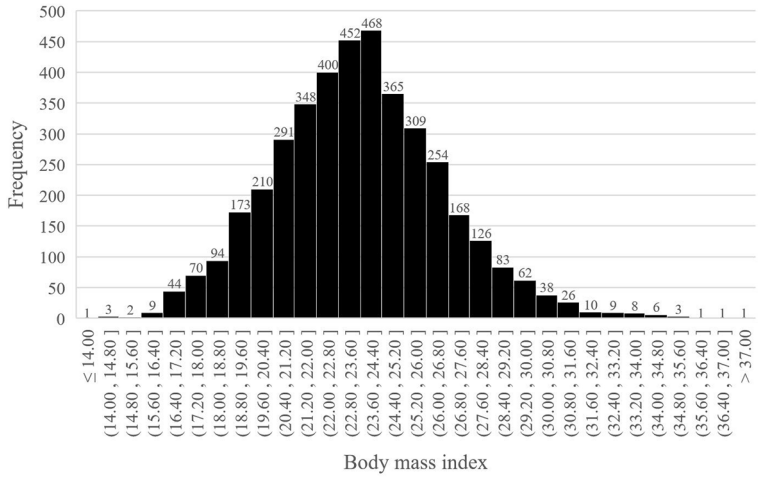


FIGURE 2 | Body mass index distribution.

TABLE 1 | Baseline characteristics stratified by BMI.

Variables[a]	BMI<28 (n = 3730)	BMI≥28 (n = 305)	P Value
Age, years	69.8 ± 4.1	69.5 ± 3.9	0.309
Sex			0.513
Male sex	1748 (46.9)	137 (44.9)	
Female sex	1982 (53.1)	168 (55.1)	
ASA classification			0.007[b]
I	108 (2.9)	8 (2.6)	
II	3014 (80.8)	226 (74.1)	
III/IV	608 (16.3)	71 (23.3)	
Comorbidity			
Hypertension	442 (11.8)	49 (16.1)	0.030[b]
Diabetes mellitus	250 (6.7)	31 (10.2)	0.022[b]
Coronary artery disease	45 (1.2)	5 (1.6)	0.585
Stroke/TIA	22 (0.6)	2 (0.7)	1.000
FEV_1/FVC, %	100.9 ± 9.9	105.2 ± 8.3	<0.001[b]
DLCO%	90.5 ± 17.8	96.7 ± 17.9	<0.001[b]
Chemoradiotherapy	4 (0.1)	1 (0.3)	0.325
Tumor size, cm	2.1 ± 1.1	2.1 ± 1.1	0.791
Clinical tumor stage			0.349
T1a	522 (14.0)	33 (10.8)	
T1b	1577 (42.3)	146 (47.9)	
T1c	1013 (27.2)	74 (24.3)	
T2a	401 (10.8)	37 (12.1)	
T2b	135 (3.6)	11 (3.6)	
T3	67 (1.8)	3 (1.0)	
T4	15 (0.4)	1 (0.3)	
Advanced clinical stage (T≥2)	618 (16.6)	52 (17.0)	0.828

[a]*Continuous data are shown as mean ± standard deviation and categoric data as number (%).*
[b]*Statistically significant (P < 0.05). BMI, Body mass index (kg/m²); ASA, American Society of Anesthesiology; TIA, Transient cerebral ischemic attack; FEV_1, Forced expiratory volume in 1 second; FVC, Forced vital capacity; DLCO, Diffusion capacity for carbon monoxide.*

TABLE 2 | Intraoperative characteristics stratified by BMI.

Variables[a]	BMI<28 (n = 3730)	BMI≥28 (n = 305)	P Value
Lymph nodes calcification	288 (7.7)	27 (8.9)	0.479
Clinical nodal involvement	255 (6.8)	24 (7.9)	0.494
Pleural adhesions	130 (3.5)	15 (4.9)	0.196
Type of resection			0.168
Segmentectomy resection	667 (17.9)	45 (14.8)	
Lobectomy resection	3063 (82.1)	260 (85.2)	
Thoracoscopic resection			0.873
Uni-portal	328 (8.8)	26 (8.5)	
Multi-portal	3402 (91.2)	279 (91.5)	
Approach			0.743
VATS	3607 (96.7)	296 (97.0)	
RATS	123 (3.3)	9 (3.0)	
Anesthesia type			0.847
GA alone	3110 (83.4)	253 (83.0)	
GA plus TPVB	620 (16.6)	52 (17.0)	
Location of resection			0.459
Left	1452 (38.9)	106 (34.8)	
Left upper	929 (24.9)	74 (24.3)	
Left lower	523 (14.0)	32 (10.5)	
Right	2278 (61.1)	199 (65.2)	
Right upper	1218 (32.7)	109 (35.7)	
Right middle	334 (9.0)	27 (8.9)	
Right lower	726 (19.5)	63 (20.7)	
Ipsilateral reoperation	3 (0.1)	1 (0.3)	0.270
Operative time, mins	102.4 ± 39.5	108.5 ± 40.1	0.009[b]

[a]*Continuous data are shown as mean ± standard deviation and categoric data as number (%).*
[b]*Statistically significant (P < 0.05). BMI, Body mass index (kg/m²); VATS, Video-assisted thoracoscopic surgery; RATS, Robotic-assisted thoracoscopic surgery; GA, General anesthesia; TPVB, Thoracic paravertebral blockade.*

and that the severity of obesity was unlikely to increase the incidence of adverse outcomes. Therefore, obesity status among elderly patients should not be a hindrance to preoperative evaluation and surgical planning during thoracoscopic anatomic lung cancer surgery.

The World Health Organization (WHO) (26) has recommended BMI cutoff points for underweight(<18.5kg/m^2), normal weight (18.5 to 24.9kg/m^2), overweight (25 to 29.9 kg/m^2) and obesity (>30kg/m^2) to predict health risks, including risk of all cancer types and non-cancer diseases. However, whether the criteria applied to Asian populations remains controversial (25,

TABLE 3 | Intra- and postoperative complications stratified by BMI.

Variables[a]	BMI<28 (n = 3730)	BMI≥28 (n = 305)	P Value
Intraoperative complications			
Hypoxemia	46 (1.2)	12 (3.9)	0.001[b]
Transfusion	18 (0.5)	1 (0.3)	1.000
New-onset arrhythmia	86 (2.3)	13 (4.3)	0.034[b]
Conversion to thoracotomy	91 (2.4)	6 (2.0)	0.605
Postoperative complications			
PPCs	1309 (35.1)	118 (38.7)	0.207
Atelectasis	29 (0.8)	4 (1.3)	0.310
Pulmonary infection	1293 (34.7)	117 (38.4)	0.193
Respiratory failure	14 (0.4)	2 (0.7)	0.627
New-onset arrhythmia	174 (4.7)	17 (5.6)	0.472
Transfusion	51 (1.4)	2 (0.7)	0.325
Length of hospital stay, day	5[4-6]	5[4-6]	0.669
30-day readmission	19 (0.5)	2 (0.7)	1.000
Hospitalization costs, USD	10057 ± 2655	10193 ± 2250	0.434

[a]*Continuous data are shown as mean ± standard deviation and categoric data as number (%); Length of hospital stay, values as median [interquartile range].*
[b]*Statistically significant (P < 0.05). BMI, Body mass index (kg/m^2); PPCs, Postoperative pulmonary complications; USD, United States dollar.*

TABLE 4 | Baseline characteristics stratified by BMI after a 1:1 PSM.

Variables[a]	BMI<28 (n = 204)	BMI≥28 (n = 204)	SDM	P Value
Age, years	69.0 ± 3.4	69.4 ± 3.8	0.100	0.282
Sex			0.069	0.484
Male sex	84 (41.2)	91 (44.6)		
Female sex	120 (58.8)	113 (55.4)		
ASA classification			0.022	0.857
I	4 (2.0)	5 (2.5)		
II	160 (78.4)	156 (76.5)		
III/IV	40 (19.6)	43 (21.1)		
Comorbidity				
Hypertension	17 (8.3)	28 (13.7)	0.072	0.082
Diabetes mellitus	19 (9.3)	18 (8.8)	0.017	0.863
Coronary artery disease	3 (1.5)	3 (1.5)	<0.001	1.000
Stroke/TIA	0 (0)	2 (1.0)	0.014	0.499
FEV_1/FVC, %	104.8 ± 8.7	104.5 ± 7.9	0.042	0.672
DLCO%	94.2 ± 16.9	95.7 ± 17.5	0.087	0.381
Chemoradiotherapy	0 (0)	1 (0.5)	0.099	1.000
Tumor size, cm	2.0 ± 1.1	2.1 ± 1.1	0.079	0.425
Clinical tumor stage			0.082	0.449
T1a	36 (17.6)	26 (12.7)		
T1b	95 (46.6)	95 (46.6)		
T1c	45 (22.1)	49 (24.0)		
T2a	15 (7.4)	24 (11.8)		
T2b	9 (4.4)	7 (3.4)		
T3	4 (2.0)	2 (1.0)		
T4	0 (0)	1 (0.5)		
Advanced clinical stage (T≥2)	28 (13.7)	34 (16.7)	0.082	0.408

[a]*Continuous data are shown as mean ± standard deviation and categoric data as number (%). BMI, Body mass index (kg/m^2); PSM, Propensity score matching; SMD, Standardized mean difference; ASA, American Society of Anesthesiology; TIA, Transient cerebral ischemic attack; FEV_1, Forced expiratory volume in 1 second; FVC, Forced vital capacity; DLCO, Diffusion capacity for carbon monoxide.*

27, 28). The BMI classification criteria for obesity in this investigation by referring to the Guidelines for Prevention and Control of Overweight and Obesity in Chinese Adults, which may be more suitable for clinical studies in Chinese population.

TABLE 5 | Intraoperative characteristics stratified by BMI after a 1:1 PSM.

Variables[a]	BMI<28 (n = 204)	BMI≥28 (n = 204)	SMD	P Value
Lymph nodes calcification	17 (8.3)	17 (8.3)	<0.001	1.000
Clinical nodal involvement	16 (7.8)	16 (7.8)	<0.001	1.000
Pleural adhesions	14 (6.9)	11 (5.4)	0.061	0.536
Type of resection			0.014	0.885
Segmentectomy resection	28 (13.7)	27 (13.2)		
Lobectomy resection	176 (86.3)	177 (86.8)		
Thoracoscopic resection			0.010	0.307
Uni-portal	16 (7.8)	22 (10.8)		
Multi-portal	188 (92.2)	182 (89.2)		
Approach			0.058	0.241
VATS	196 (96.1)	200 (98.0)		
RATS	8 (3.9)	4 (2.0)		
Anesthesia type			0.026	0.199
GA alone	180 (88.2)	171 (83.8)		
GA plus TPVB	24 (11.8)	33 (16.2)		
Location of resection			0.044	0.912
Left	73 (35.8)	74 (36.3)		
Left upper	49 (24.0)	54 (26.5)		
Left lower	24 (11.8)	20 (9.8)		
Right	131 (64.2)	130 (63.7)		
Right upper	64 (31.3)	68 (33.3)		
Right middle	23 (11.3)	20 (9.8)		
Right lower	44 (21.6)	42 (20.6)		
Ipsilateral reoperation	1 (0.5)	0 (0)	0.099	1.000
Operative time, mins	108.4 ± 37.8	108.1 ± 40.0	0.009	0.927

[a]*Continuous data are shown as mean ± standard deviation and categoric data as number (%). BMI, Body mass index (kg/m^2); PSM, Propensity score matching; SMD, Standardized mean difference; VATS, Video-assisted thoracoscopic surgery; RATS, Robotic-assisted thoracoscopic surgery; GA, General anesthesia; TPVB, Thoracic paravertebral blockade.*

MIS has been established to improve perioperative adverse outcomes, especially in patients at high risk such as elderly and obese patients, with regard to the standard open approaches (13, 29). Presumably, the proportion of elderly patients with elevated BMI undergoing thoracic surgery will constantly increase in the future (13, 14, 19). Intuitively, it seems that longer operative time, reduced mobility, impaired diaphragm movement, and obesity-related comorbidities should be associated with an increased risk of lung resection. Therefore, it is mandatory to fully understand the influence of obesity on perioperative outcomes of elderly patients.

In terms of intraoperative complications, although before a 1:1 PSM, obese patients developed higher rates of hypoxemia and new-onset arrhythmia, but had comparable transfusion, conversion rates and operative time compared with nonobese patients, and none of these differences were significant after PSM. Our previous published literature echoed these results and did not find that elevated BMI was associated with high rates of intraoperative conversion and new-onset arrhythmia (30, 31). Similarly, Guerrera has evaluated the impact of morbidly obesity on perioperative clinical outcomes after thoracoscopic lobectomy and found that obese patients did not increase conversion rates, blood loss and surgical time (32). Conversely, a few studies have shown that obesity was associated with increased operative time (33).

Our research results may also provide the evidence for the "obesity paradox" in elderly patients undergoing thoracoscopic anatomic lung cancer surgery, which has been widely reported in published documents (14–17). The postoperative complications regarding pulmonary complications, new-onset arrhythmia,

TABLE 6 | Intra- and postoperative complications stratified by BMI after a 1:1 PSM.

Variables[a]	BMI<28 (n = 204)	BMI≥28 (n = 204)	P Value
Intraoperative complications			
Hypoxemia	4 (2.0)	10 (4.9)	0.103
Transfusion	0 (0)	1 (0.5)	1.000
New-onset arrhythmia	4 (2.0)	9 (4.4)	0.159
Conversion to thoracotomy	5 (2.5)	3 (1.5)	0.724
Postoperative complications			
PPCs	74 (36.3)	79 (38.7)	0.609
Atelectasis	2 (1.0)	3 (1.5)	1.000
Pulmonary infection	73 (35.8)	78 (38.2)	0.608
Respiratory failure	0 (0)	2 (1.0)	0.499
New-onset arrhythmia	9 (4.4)	11 (5.4)	0.647
Transfusion	1 (0.5)	0 (0)	0.100
Length of hospital stay, day	5[4-7]	5[4-6]	0.104
30-day readmission	2 (1.0)	1 (0.5)	1.000
Hospitalization costs, USD	10030 ± 2516	9966 ± 2069	0.787

[a]*Continuous data are shown as mean ± standard deviation and categoric data as number (%); Length of hospital stay, values as median [interquartile range]. BMI, Body mass index (kg/m^2); PSM, Propensity score matching; PPCs, Postoperative pulmonary complications; USD, United States dollar.*

transfusion, LOS, 30-day readmission and hospitalization costs were not significant between obese and nonobese patients. Ferguson et al. showed that being overweight or obese did not increase the risk of postoperative complications in any category after major lung resection (34). Also, Thomas et al. conducted a retrospective cohort study of 19,635 patients undergoing lobectomy for primary lung cancer and concluded that obesity was not associated with increased incidence of postoperative complications, except for cardiovascular complications and had a statistical protective effect regarding surgical complications (35). Moreover, a systematic review with meta-analysis demonstrated that overall morbidity and in-hospital mortality were significantly decreased in obese patients (36).

Our investigation also assessed the effect of obesity severity on perioperative outcomes in older patients and found that the difference was not significant between the obesity class I and obesity class II. And the results were consistent with other published researches (14, 32). A retrospective study conducted by Williams using Society of Thoracic Surgeons General Thoracic Surgery Database indicated that overweight and obese class I to II patients had a lower risk of pulmonary complications and any postoperative events (14). In contrast to the results of our present study, Zogg (37) and De Oliveira's (38) research showed that obese class II and III patients experienced marginally increased odds of morbidity and an increased risk of pulmonary complications, respectively. The variability of results may be related to the fact that high BMI does not distinguish between body composition phenotypes that have important effects on surgical risk and outcomes (39). Relying on BMI augmentation categories alone to assess the surgical risk of major lung resection is fraught with challenges.

Potential shortcomings of this study include as follows. First, as a retrospective study based on a prospectively collected database, it has the inherent design biases. Besides, single-institute study has specific generalization limitations. Second, due to the limited granularity of postoperative care data, some poor outcomes, such as pain control (40), other surgical complications and mortality, could not be pooled in this study. Third, the relationship between obesity and long-term outcomes needs further investigation (41, 42).

CONCLUSIONS

By conducting a monocentric retrospective cohort study of 4035 elderly patients receiving thoracoscopic anatomic lung cancer surgery, our study found that obesity status and obesity severity dose not increase perioperative adverse outcomes. Even among older patients, these data support evidence for "obesity paradox". These data also contribute the growing body of evidence that obesity in older patients should not exclude candidates for thoracoscopic anatomic lung cancer surgery.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

This study was approved by the Institutional Review Board at Shanghai Chest Hospital (IS21119), and the informed consent was waived because of the retrospective nature of the study.

AUTHOR CONTRIBUTIONS

CT, TL and YS: study conception, design, statistic analysis and drafting of the manuscript. CT, YS and HZ: acquisition of data. CT, JZ and JW: analysis and interpretation of data. JWand JZ: critical revision. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by National Natural Science Foundation of China (82071233) and Shanghai Shen Kang Hospital Development Center Project (SHDC2020CR4063).

ACKNOWLEDGMENTS

Our research team would like to thank Professor Liu for statistical analysis and Professor Luo for study design in Shanghai Chest Hospital, Shanghai, China, for their involvement and support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.881467/full#supplementary-material

REFERENCES

1. Blüher M. Obesity: Global Epidemiology and Pathogenesis. *Nat Rev Endocrinol* (2019) 15(5):288–98.
2. Hurt RT, Frazier TH, McClave SA, Kaplan LM. Obesity Epidemic: Overview, Pathophysiology, and the Intensive Care Unit Conundrum. *JPEN J Parenter Enteral Nutr* (2011) 35(5 Suppl):4S–13S. doi: 10.1177/0148607111415110
3. Apovian CM. Obesity: Definition, Comorbidities, Causes, and Burden. *Am J Manag Care* (2016) 22(7 Suppl):s176–85.
4. Heymsfield SB, Wadden TA. Mechanisms, Pathophysiology, and Management of Obesity. *N Engl J Med* (2017) 376(3):254–66. doi: 10.1056/NEJMra1514009
5. Stefan N. Causes, Consequences, and Treatment of Metabolically Unhealthy Fat Distribution. *Lancet Diabetes Endocrinol* (2020) 8(7):616–27. doi: 10.1016/S2213-8587(20)30110-8
6. Bhaskaran K, Douglas I, Forbes H, dos-Santos-Silva I, Leon DA, Smeeth L. Body-Mass Index and Risk of 22 Specific Cancers: A Population-Based Cohort Study of 5.24 Million UK Adults. *Lancet* (2014) 384(9945):755–65. doi: 10.1016/S0140-6736(14)60892-8
7. Gupta A, Majumder K, Arora N, Mayo HG, Singh PP, Beg MS, et al. Premorbid Body Mass Index and Mortality in Patients With Lung Cancer: A Systematic Review and Meta-Analysis. *Lung Cancer* (2016) 102:49–59. doi: 10.1016/j.lungcan.2016.10.017
8. Zhou W, Liu G, Hung RJ, Haycock PC, Aldrich MC, Andrew AS, et al. Causal Relationships Between Body Mass Index, Smoking and Lung Cancer: Univariable and Multivariable Mendelian Randomization. *Int J Cancer* (2021) 148(5):1077–86. doi: 10.1002/ijc.33292
9. Petrelli F, Cortellini A, Indini A, Tomasello G, Ghidini M, Nigro O, et al. Association of Obesity With Survival Outcomes in Patients With Cancer: A Systematic Review and Meta-Analysis. *JAMA Netw Open* (2021) 4(3): e213520. doi: 10.1001/jamanetworkopen.2021.3520
10. Saltiel AR, Olefsky JM. Inflammatory Mechanisms Linking Obesity and Metabolic Disease. *J Clin Invest* (2017) 127(1):1–4. doi: 10.1172/JCI92035
11. Sarode P, Schaefer MB, Grimminger F, Seeger W, Savai R. Macrophage and Tumor Cell Cross-Talk Is Fundamental for Lung Tumor Progression: We Need to Talk. *Front Oncol* (2020) 10:324. doi: 10.3389/fonc.2020.00324
12. Hillers-Ziemer LE, Williams AE, Janquart A, Grogan C, Thompson V, Sanchez A, et al. Obesity-Activated Lung Stromal Cells Promote Myeloid Lineage Cell Accumulation and Breast Cancer Metastasis. *Cancers (Basel)* (2021) 13(5):1005. doi: 10.3390/cancers13051005
13. Ferguson MK, Vigneswaran WT. Changes in Patient Presentation and Outcomes for Major Lung Resection Over Three Decades. *Eur J Cardiothorac Surg* (2008) 33(3):497–501. doi: 10.1016/j.ejcts.2007.12.023
14. Williams T, Gulack BC, Kim S, Fernandez FG, Ferguson MK. Operative Risk for Major Lung Resection Increases at Extremes of Body Mass Index. *Ann Thorac Surg* (2017) 103(1):296–302. doi: 10.1016/j.athoracsur.2016.05.057
15. Valentijn TM, Galal W, Hoeks SE, van Gestel YR, Verhagen HJ, Stolker RJ. Impact of Obesity on Postoperative and Long-Term Outcomes in a General Surgery Population: A Retrospective Cohort Study. *World J Surg* (2013) 37 (11):2561–8. doi: 10.1007/s00268-013-2162-y
16. Mungo B, Zogg CK, Hooker CM, Yang SC, Battafarano RJ, Brock MV, et al. Does Obesity Affect the Outcomes of Pulmonary Resections for Lung Cancer? A National Surgical Quality Improvement Program Analysis. *Surgery* (2015) 157:792–800. doi: 10.1016/j.surg.2014.10.016
17. Dotan I, Shochat T, Shimon I, Akirov A. The Association Between BMI and Mortality in Surgical Patients. *World J Surg* (2021) 45(5):1390–9. doi: 10.1007/s00268-021-05961-4
18. Lang LH, Parekh K, Tsui BYK, Maze M. Perioperative Management of the Obese Surgical Patient. *Br Med Bull* (2017) 124(1):135–55. doi: 10.1093/bmb/ldx041
19. Liou DZ, Berry MF. Thoracic Surgery Considerations in Obese Patients. *Thorac Surg Clin* (2018) 28(1):27–41. doi: 10.1016/j.thorsurg.2017.09.004
20. Billmeier SE, Jaklitsch MT. Pulmonary Surgery for Malignant Disease in the Elderly. In: RA Rosenthal, ME Zenilman, MR Katlic, editors. *Principles and Practice of Geriatric Surgery*, *2nd ed*. New York, NY: Springer (2011). p. 605–16. doi: 10.1007/978-1-4419-6999-6_48
21. Alvarez-Nebreda ML, Bentov N, Urman RD, Setia S, Huang JC, Pfeifer K, et al. Recommendations for Preoperative Management of Frailty From the Society for Perioperative Assessment and Quality Improvement (SPAQI). *J Clin Anesth* (2018) 47:33–42. doi: 10.1016/j.jclinane.2018.02.011
22. Tabatabai S, Do Q, Min J, Tang CJ, Pleasants D, Sands LP, et al. Obesity and Perioperative Outcomes in Older Surgical Patients Undergoing Elective Spine and Major Arthroplasty Surgery. *J Clin Anesth* (2021) 75:110475. doi: 10.1016/j.jclinane.2021.110475
23. Jammer I, Wickboldt N, Sander M, Smith A, Schultz MJ, Pelosi P, et al. Standards for Definitions and Use of Outcome Measures for Clinical Effectiveness Research in Perioperative Medicine: European Perioperative Clinical Outcome (EPCO) Definitions: A Statement From the ESA-ESICM Joint Taskforce on Perioperative Outcome Measures. *Eur J Anaesthesiol* (2015) 32:88–105. doi: 10.1097/EJA.0000000000000118
24. Frendl G, Sodickson AC, Chung MK, Waldo AL, Gersh BJ, Tisdale JE, et al. 2014 AATS Guidelines for the Prevention and Management of Perioperative Atrial Fibrillation and Flutter for Thoracic Surgical Procedures. *Executive Summary J Thorac Cardiovasc Surg* (2014) 148(3):772–91. doi: 10.1016/j.jtcvs.2014.06.037
25. Department of Disease Control of the Ministry of Health of the People's Republic of China. *Guidelines for the Prevention and Control of Overweight and Obesity of Adult in China*. Beijing: People's Medical Publishing House (2006) p. 34–6.
26. Organization WH. Obesity: Preventing and Managing the Global Epidemic. Report of a WHO Consultation.*World Health Organ Tech Rep Ser* (2020) 894: i-xii, 1–253.
27. Yang L, Yang G, Zhou M, Smith M, Ge H, Boreham J, et al. Body Mass Index and Mortality From Lung Cancer in Smokers and Nonsmokers: A Nationally Representative Prospective Study of 220,000 Men in China. *Int J Cancer* (2009) 125(9):2136–43. doi: 10.1002/ijc.24527
28. Koh WP, Yuan JM, Wang R, Lee HP, Yu MC. Body Mass Index and Smoking – Related Lung Cancer Risk in the Singapore Chinese Health Study. *Br J Cancer* (2010) 102(3):610–4. doi: 10.1038/sj.bjc.6605496
29. Falcoz PE, Puyraveau M, Thomas PA, Decaluwe H, Hürtgen M, Petersen RH, et al. Video-Assisted Thoracoscopic Surgery Versus Open Lobectomy for Primary Non-Small-Cell Lung Cancer: A Propensity-Matched Analysis of Outcome From the European Society of Thoracic Surgeon Database. *Eur J Cardiothorac Surg* (2015) 49:602–9. doi: 10.1093/ejcts/ezv154
30. Tong C, Li T, Huang C, Ji C, Liu Y, Wu J, et al. Risk Factors and Impact of Conversion to Thoracotomy From 20,565 Cases of Thoracoscopic Lung Surgery. *Ann Thorac Surg* (2020) 109(5):1522–9. doi: 10.1016/j.athoracsur.2019.12.009
31. Tong C, Zhang Q, Liu Y, Xu M, Wu J, Cao H. Risk Factors and Outcomes of Intraoperative Atrial Fibrillation in Patients Undergoing Thoracoscopic Anatomic Lung Surgery. *Ann Transl Med* (2021) 9(7):543. doi: 10.21037/atm-20-5035
32. Guerrera F, Lyberis P, Lausi PO, Cristofori RC, Giobbe R, Molinatti M, et al; On the behalf of the Italian VATS Group. Does Morbid Obesity Influence Perioperative Outcomes After Video-Assisted Thoracic Surgery (VATS) Lobectomy for Non-Small Cell Lung Cancer? Analysis of the Italian VATS Group Registry. *Surg Endosc* (2021) 36(5):3567–73. doi: 10.1007/s00464-021-08680-y
33. St Julien JB, Aldrich MC, Sheng S, Deppen SA, Burfeind Jr WR, Putnam JB, et al. Obesity Increases Operating Room Time for Lobectomy in the Society of Thoracic Surgeons Database. *Ann Thorac Surg* (2012) 94:1841–7. doi: 10.1016/j.athoracsur.2012.08.006
34. Ferguson MK, Im HK, Watson S, Johnson E, Wigfield CH, Vigneswaran WT. Association of Body Mass Index and Outcomes After Major Lung Resection. *Eur J Cardiothorac Surg* (2014) 45(4):e94–9; discussion e99. doi: 10.1093/ejcts/ezu008
35. Thomas PA, Berbis J, Falcoz PE, Le Pimpec-Barthes F, Bernard A, Jougon J, et al; EPITHOR Group. National Perioperative Outcomes of Pulmonary Lobectomy for Cancer: The Influence of Nutritional Status. *Eur J Cardiothorac Surg* (2014) 45(4):652–9; discussion 659. doi: 10.1093/ejcts/ezt452
36. Li S, Wang Z, Huang J, Fan J, Du H, Liu L, et al. Systematic Review of Prognostic Roles of Body Mass Index for Patients Undergoing Lung Cancer Surgery: Does the 'Obesity Paradox' Really Exist? *Eur J Cardiothorac Surg* (2017) 51(5):817–28. doi: 10.1093/ejcts/ezw386

37. Zogg CK, Mungo B, Lidor AO, Stem M, Rios Diaz AJ, Haider AH, et al. Influence of Body Mass Index on Outcomes After Major Resection for Cancer. *Surgery* (2015) 158(2):472–85. doi: 10.1016/j.surg.2015.02.023
38. De Oliveira GS Jr, McCarthy RJ, Davignon K, Chen H, Panaro H, Cioffi WG. Predictors of 30-Day Pulmonary Complications After Outpatient Surgery: Relative Importance of Body Mass Index Weight Classifications in Risk Assessment. *J Am Coll Surg* (2017) 225(2):312–23.e7. doi: 10.1016/j.jamcollsurg.2017.04.013
39. Prado CM, Gonzalez MC, Heymsfield SB. Body Composition Phenotypes and Obesity Paradox. *Curr Opin Clin Nutr Metab Care* (2015) 18(6):535–51. doi: 10.1097/MCO.0000000000000216
40. Majchrzak M, Brzecka A, Daroszewski C, Błasiak P, Rzechonek A, Tarasov VV, et al. Increased Pain Sensitivity in Obese Patients After Lung Cancer Surgery. *Front Pharmacol* (2019) 10:626. doi: 10.3389/fphar.2019.00626
41. Sepesi B, Gold KA, Correa AM, Heymach JV, Vaporciyan AA, Roszik J, et al. The Influence of Body Mass Index on Overall Survival Following Surgical Resection of Non-Small Cell Lung Cancer. *J Thorac Oncol* (2017) 12(8):1280–7. doi: 10.1016/j.jtho.2017.05.010
42. Alifano M, Daffré E, Iannelli A, Brouchet L, Falcoz PE, Le Pimpec , et al. The Reality of Lung Cancer Paradox: The Impact of Body Mass Index on Long-Term Survival of Resected Lung Cancer. A French Nationwide Analysis From the Epithor Database. *Cancers (Basel)* (2021) 13(18):4574. doi: 10.3390/cancers13184574

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 13 May 2022
doi: 10.3389/fonc.2022.817372

Machine Learning for the Prediction of Synchronous Organ-Specific Metastasis in Patients With Lung Cancer

Huan Gao[1,2], Zhi-yi He[2], Xing-li Du[2], Zheng-gang Wang[2] and Li Xiang[1]**

[1] School of Medicine and Health Management, Huazhong University of Science and Technology, Wuhan, China,
[2] Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

OPEN ACCESS

Edited by:
Yutong He,
Fourth Hospital of Hebei Medical University, China

Reviewed by:
Carlo Greco,
Policlinico Universitario Campus Bio-Medico, Italy
Taner Tunç,
Ondokuz Mayıs University, Turkey

***Correspondence:**
Zheng-gang Wang
846735564@qq.com
Li Xiang
xllyf@hust.edu.cn

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: 18 November 2021
Accepted: 11 April 2022
Published: 13 May 2022

Citation:
Gao H, He Z-y, Du X-l, Wang Z-g and Xiang L (2022) Machine Learning for the Prediction of Synchronous Organ-Specific Metastasis in Patients With Lung Cancer. Front. Oncol. 12:817372. doi: 10.3389/fonc.2022.817372

Background: This study aimed to develop an artificial neural network (ANN) model for predicting synchronous organ-specific metastasis in lung cancer (LC) patients.

Methods: A total of 62,151 patients who diagnosed as LC without data missing between 2010 and 2015 were identified from Surveillance, Epidemiology, and End Results (SEER) program. The ANN model was trained and tested on an 75/25 split of the dataset. The receiver operating characteristic (ROC) curves, area under the curve (AUC) and sensitivity were used to evaluate and compare the ANN model with the random forest model.

Results: For distant metastasis in the whole cohort, the ANN model had metrics AUC = 0.759, accuracy = 0.669, sensitivity = 0.906, and specificity = 0.613, which was better than the random forest model. For organ-specific metastasis in the cohort with distant metastasis, the sensitivity in bone metastasis, brain metastasis and liver metastasis were 0.913, 0.906 and 0.925, respectively. The most important variable was separate tumor nodules with 100% importance. The second important variable was visceral pleural invasion for distant metastasis, while histology for organ-specific metastasis.

Conclusions: Our study developed a "two-step" ANN model for predicting synchronous organ-specific metastasis in LC patients. This ANN model may provide clinicians with more personalized clinical decisions, contribute to rationalize metastasis screening, and reduce the burden on patients and the health care system.

Keywords: machine learning, artificial neural network, SEER, metastasis, lung cancer

INTRODUCTION

Lung cancer (LC) is one of the most commonly diagnosed malignancy as well as the leading cause of cancer-related death both in males and females worldwide (1, 2). Approximately 30-40% of LC patients present with distant metastasis (DM) at the time of diagnosis (3–5). And distant metastasis is responsible for a large morbidity and mortality burden among LC patients (6, 7). The most common metastatic site is bone, followed by liver, brain and adrenal gland (8, 9). Distant metastasis

is closely related to treatment decisions and clinical outcomes. Therefore, it is important to identify and diagnose distant metastasis in the early period.

Computed tomography (CT), magnetic resonance imaging (MRI), single-photon emission computed tomography (SPECT) and positron emission tomography/computed tomography (PET/CT) are the common techniques to screen the distant metastasis in LC patients. However, routine DM screening to all LC patients is controversial because of low detection rate of asymptomatic patients, invasive operation, potential risk of adverse reactions, complex process and high cost (10–14). Therefore, there are strong requirements for the identification of a high-risk group with distant metastasis and the rationalization of DM screening in LC patients.

The occurrence and development of lung cancer is very complicated, and most of the clinical characteristics exhibit a multidimensional and non-linear relationship. The artificial neural network (ANN) is a complex non-linear model inspired by the working of biological neural networks (15–17). In the face of huge and complex medical data, it has the ability to discover underlying patterns and constantly adjust the algorithm to adapt to new patient information (18–20). In recent years, the ANN has been applied successfully in clinical medicine, including diagnosis, image identification and outcome prediction (16, 21–24).

In this study, we aim to develop an ANN model to predict synchronous organ-specific metastasis in LC patients. This study may provide clinicians with more personalized clinical decisions, reduce the unnecessary financial burden of patients, and allocate medical resources more rationally.

PATIENTS AND METHODS

Patient Selection and Data Collection

We obtained the research participants from the Surveillance, Epidemiology, and End Results (SEER) Program. The SEER program is supported by the US National Cancer Institute, covers cases from 18 cancer registries, and represents approximately 28-30% of the population (25). Patient data were screened *via* the SEER*Stat software (version 8.3.6). Since the data was anonymized, no additional institutional review board approval or patient informed consent was required.

We included patients diagnosed with lung cancer between 2010 and 2015. Variables of interest included age, sex, race, marital status, insurance, primary site, histology, grade, tumor size, separate tumor nodules, visceral pleural invasion, T-stage, N-stage, and organ-specific metastases. We excluded the patients whose reporting sources were "Autopsy only" or "Death certification only", as well as those who did not have complete information on all the above variables.

Model Development

A multilayer perceptron ANN was created consisting of an input, an output, and one or more hidden layers (**Figure 1**). In this research, thirteen selected demographic or clinical variables were served as the input layers neurons, and one variable (metastasis or no metastasis) was served as the output layer neuron. The number of neurons in the hidden layer was set empirically. 75% of patients was used to develop the model (the training group), while the remaining 25% was used to

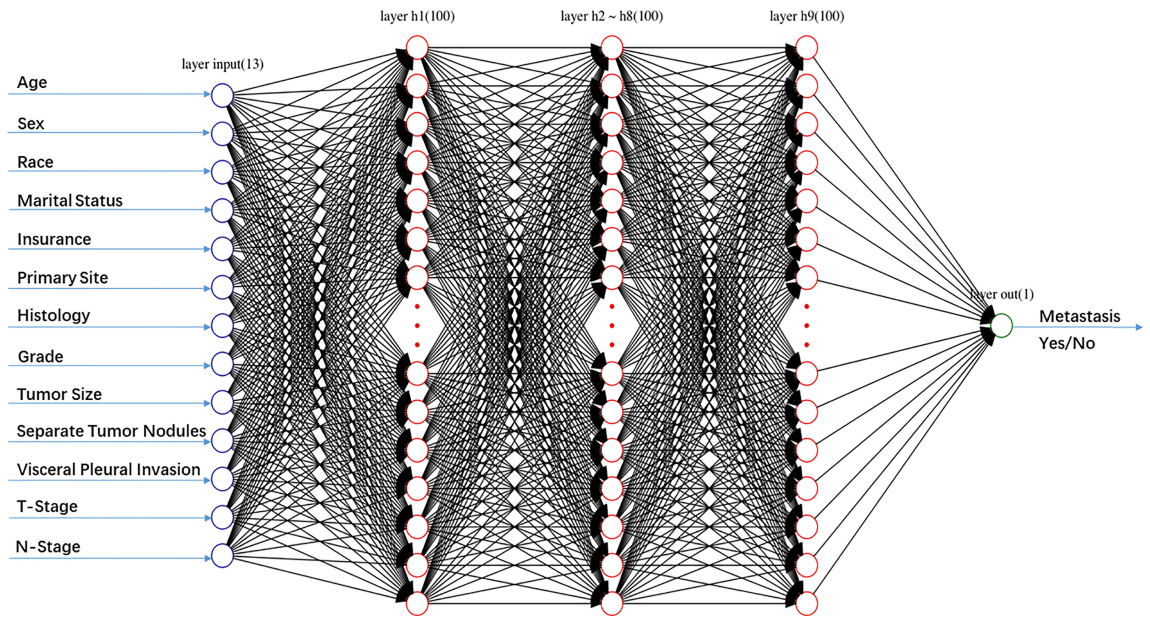


FIGURE 1 | Schematic structure of the artificial neural network (ANN) model including one input layer with 13 nodes, nine hidden layers with 100 nodes, and one output layer with 1 node.

evaluate the developed model (the testing group). A back propagation (BP) method was used to train the multilayer perceptron ANN, which modified the weight of the interneuron connections to reduce the total errors during the repeated development cycles. During the learning progresses, the errors between ANN model outputs and expected outputs were minimized (21). In this study, the number of epochs was selected from the set {10, 20, 30, 50, 100, 500}.

Statistical Analysis

Kaplan-Meier analysis was used for comparison of survival among the subgroups classified by distant metastasis. Multivariate Cox proportional hazard analyses was conducted to estimate the hazard ratio (HR), and the corresponding 95% confidence interval (CI), for the potential risk factors. The model performance was evaluated with the receiver operating characteristic (ROC) curves and areas under the curve (AUC), which is a score ranging from 0.50 to 1.0. All statistical analyses were conducted using SPSS version 21.0 and RStudio Version 1.0.153. A two-tailed P value<0.05 was considered statistically significant.

RESULTS

Patient Demographics and Clinical Characteristics

From 2010 to 2015, 62,151 patients with lung cancer were consecutively included in this study. Patient characteristics were described in **Table 1**. The population with a median age of 68 (IQR, 61-75) years and White people (n=50589, 81.4%) predominated. The distribution of male and female was almost 1:1. The most common primary site was upper lobe (n=37284, 60%) and the most common histological subtype was adenocarcinoma (n=33036, 53.2%). Of these patients, 12,182 (19.6%) developed distant metastases, including 3,982 (6.4%) with bone metastases, 3,674 (5.9%) with brain metastases, and 1,307 (2.1%) with liver metastases.

Survival Analysis

A cohort of 29,296 patients was used to analyze cancer-specific survival (CSS). The median CSS for patients with none metastasis, bone metastasis, brain metastasis, liver metastasis and two or three metastases were 10 months, 4 months, 4 months, 4 months and 3 months, respectively (**Table 2**). Kaplan-Meier analysis showed the similar trend in **Figure 2**. In addition, multivariate Cox proportional hazard analyses revealed that bone metastasis (OR=1.630, p<0.001), brain metastasis (OR=1.698, p<0.001), liver metastasis (OR=1.673, p<0.001) and two or three metastases (OR=2.025, p<0.001) were associated with poor prognosis (**Table 2**).

Construction of the ANN Model

In the training of ANN model, we manually increased the number of hidden layers starting with 5 layers. The predictive sensitivity culminated with 9 layers and adding more layer did not improve the performance but increased time of computation (**Table 3**). In the end, the ANN model was constructed with 13 neurons in the input layer, 100 neurons in each of the 9 hidden layers and 1 neuron in the output layer (**Figure 1**). Meanwhile, we compared the RF model (ntree=500) with the ANN model, and the RF model showed obvious overfitting (**Figure 3**).

Evaluation of the ANN Model

In this study, we first evaluated the model performance for predicting distant metastasis in the whole cohort (AUC: 0.759, accuracy: 0.669, sensitivity: 0.906, specificity: 0.613, false positive rate: 0.387, false negative rate: 0.094, likelihood ratio positive: 2.339, likelihood ratio negative: 0.154). Then we evaluated the model performance for predicting organ-specific metastasis in the cohort with distant metastasis (**Figure 4**; **Table 4**). The sensitivity in bone metastasis, brain metastasis and liver metastasis were 0.913, 0.906 and 0.925, respectively.

Variable Importance Measure

By applying ANN methods with variable importance measures, the importance of the 13 variables was standardized and the top 10 were showed in **Figure 5**. The most important variable was separate tumor nodules with 100% importance. The second important variable was visceral pleural invasion for distant metastasis, while histology for organ-specific metastasis. And the sex variable only appeared in bone metastases. Relatively, the race and insurance variable were less important in the whole cohort.

DISCUSSION

With the increasing incidence of distant metastasis of lung cancer, this field has gradually become one of the hot spots in clinical research (26–29). Our study suggested that distant metastasis was a risk factor for poor prognosis, and the median CSS for LC patients with bone metastasis, brain metastasis, liver metastasis and two or three metastases are 4 months, 4 months, 4 months and 3 months, respectively, which was similar to previous studies (28–32). Thus, early identification and diagnosis of distant metastasis is meaningful to improve prognosis and can assist clinicians in making therapeutic choices.

However, the cost of screening in an unselected population is considerable and the benefit is questionable, given the conflicting international screening guidelines and clinicians' possible tendency to conduct investigations in excess of the recommended stage (14, 33–35). In this study, we developed a "two-step" ANN model for predicting synchronous organ-specific metastasis in LC patients. Our ANN model has high predictive power, with sensitivity of 0.906 for distant metastasis, 0.913 for bone metastasis, 0.925 for brain metastasis and 0.906 for liver metastasis. It can help predict the possibility of organ-specific metastasis in LC patients and alert high-risk patients for further investigation, which can provide clinicians with more accurate and personalized clinical decisions.

TABLE 1 | Baseline demographic and clinical characteristics of patients with lung cancer.

Characteristics	Total patients	Patients with no metastasis	Patients with metastases	Patients with bone metastasis	Patients with brain metastasis	Patients with liver metastasis
	n=62151	n=49969	n=12182	n=3982	n=3674	n=1307
Age, year						
Mean±SD	68±11	68+10	66+11	68±11	64±10	68±10
Median	68	68	66	68	64	68
(IQR 25%-75%)	(61-75)	(61-76)	(59-74)	(60-76)	(57-72)	(61-76)
Sex						
Male	31736 (51.1%)	24926 (49.9%)	6810 (55.9%)	2333 (58.6%)	1904 (51.8%)	724 (55.4%)
Female	30415 (48.9%)	25043 (50.1%)	5372 (44.1%)	1649 (41.4%)	1770 (48.2%)	583 (44.6%)
Race						
White	50589 (81.4%)	40911 (81.9%)	9678 (79.4%)	3134 (78.7%)	2885 (78.5%)	1076 (82.3%)
Blake	6855 (11%)	5326 (10.7%)	1529 (12.6%)	526 (13.2%)	491 (13.4%)	169 (12.9%)
American Indian/Alaska Native	291 (0.5%)	241 (0.5%)	50 (0.4%)	17 (0.4%)	18 (0.5%)	4 (0.3%)
Asian or Pacific Islander	4416 (7.1%)	3491 (7%)	925 (7.6%)	305 (7.7%)	280 (7.6%)	58 (4.4%)
Marital Status						
Single (never married)	8840 (14.2%)	6834 (13.7%)	2006 (16.5%)	605 (15.2%)	700 (19.1%)	199 (15.2%)
Married (including common law)	34269 (55.1%)	27547 (55.1%)	6722 (55.2%)	2235 (56.1%)	1941 (52.8%)	683 (52.3%)
Separated	726 (1.2%)	577 (1.2%)	149 (1.2%)	49 (1.2%)	43 (1.2%)	21 (1.6%)
Divorced	8267 (13.3%)	6637 (13.3%)	1630 (13.4%)	499 (12.5%)	525 (14.3%)	179 (13.7%)
Widowed	10049 (16.2%)	8374 (16.8%)	1675 (13.7%)	594 (14.9%)	465 (12.7%)	225 (17.2%)
Insurance						
Uninsured	1602 (2.6%)	1136 (2.3%)	466 (3.8%)	105 (2.6%)	178 (4.8%)	39 (3%)
Insured/Medicaid	60549 (97.4%)	48833 (97.7%)	11716 (96.2%)	3877 (97.4%)	3496 (95.2%)	1268 (97%)
Primary Site						
Main bronchus	2036 (3.3%)	1388 (2.8%)	648 (5.3%)	196 (4.9%)	154 (4.2%)	103 (7.9%)
Upper lobe	37284 (60%)	29918 (59.9%)	7366 (60.5%)	2437 (61.2%)	2324 (63.3%)	740 (56.6%)
Middle lobe	3136 (5%)	2600 (5.2%)	536 (4.4%)	170 (4.3%)	158 (4.3%)	58 (4.4%)
Lower lobe	19008 (30.6%)	15486 (31%)	3522 (28.9%)	1146 (28.8%)	1010 (27.5%)	389 (29.8%)
Overlapping lesion of lung	687 (1.1%)	577 (1.2%)	110 (0.9%)	33 (0.8%)	28 (0.8%)	17 (1.3%)
Histology						
Squamous cell carcinoma	17973 (28.9%)	15782 (31.6%)	2191 (18%)	874 (21.9%)	515 (14%)	331 (25.3%)
Small cell carcinoma	3236 (5.2%)	1807 (3.6%)	1429 (11.7%)	244 (6.1%)	339 (9.2%)	341 (26.1%)
Adenocarcinoma	33036 (53.2%)	26471 (53%)	6565 (53.9%)	2229 (56%)	2185 (59.5%)	429 (32.8%)
Large cell carcinoma	1117 (1.8%)	830 (1.7%)	287 (2.4%)	78 (2%)	101 (2.7%)	32 (2.4%)
Adenosquamous carcinoma	5244 (8.4%)	3609 (7.2%)	1635 (13.4%)	532 (13.4%)	518 (14.1%)	161 (12.3%)
Sarcomatoid carcinoma	183 (0.3%)	146 (0.3%)	37 (0.3%)	15 (0.4%)	12 (0.3%)	1 (0.1%)
Carcinoid tumor	1362 (2.2%)	1324 (2.6%)	38 (0.3%)	10 (0.3%)	4 (0.1%)	12 (0.9%)
Grade						

(Continued)

TABLE 1 | Continued

Characteristics	Total patients	Patients with no metastasis	Patients with metastases	Patients with bone metastasis	Patients with brain metastasis	Patients with liver metastasis
	n=62151	n=49969	n=12182	n=3982	n=3674	n=1307
Well differentiated	7619 (12.3%)	7183 (14.4%)	436 (3.6%)	170 (4.3%)	111 (3%)	37 (2.8%)
Moderately differentiated	21737 (35%)	18991 (38%)	2746 (22.5%)	1072 (26.9%)	816 (22.2%)	199 (15.2%)
Poorly differentiated	29483 (47.4%)	21774 (43.6%)	7709 (63.3%)	2489 (62.5%)	2406 (65.5%)	785 (60.1%)
Undifferentiated	3312 (5.3%)	2021 (4%)	1291 (10.6%)	251 (6.3%)	341 (9.3%)	286 (21.9%)
Tumor Size, mm						
Mean±SD	42±25	39±24	52	51±25	52±25	53±26
Median	35	32	48	46	48	50
(IQR 25%-75%)	(22-56)	(20-52)	(32-69)	(32-67)	(32-68)	(33-70)
Separate Tumor Nodules						
STN0	55677 (89.6%)	47096 (94.3%)	8581 (70.4%)	2798 (70.3%)	2788 (75.9%)	945 (72.3%)
STN1	2276 (3.7%)	901 (1.8%)	1375 (11.3%)	445 (11.2%)	365 (9.9%)	145 (11.1%)
STN2	2416 (3.9%)	1187 (2.4%)	1229 (10.1%)	421 (10.6%)	312 (8.5%)	117 (9%)
STN3	1782 (2.9%)	785 (1.6%)	997 (8.2%)	318 (8%)	209 (5.7%)	100 (7.7%)
Visceral Pleural Invasion						
PL0	21565 (34.7%)	20633 (41.3%)	932 (7.7%)	278 (7%)	338 (9.2%)	101 (7.7%)
PL1	1758 (2.8%)	1715 (3.4%)	43 (0.4%)	5 (0.1%)	26 (0.7%)	4 (0.3%)
PL2	1513 (2.4%)	1455 (2.9%)	58 (0.5%)	15 (0.4%)	30 (0.8%)	6 (0.5%)
PL3	686 (1.1%)	648 (1.3%)	38 (0.3%)	18 (0.5%)	12 (0.3%)	2 (0.2%)
PLX	36629 (58.9%)	25518 (51.1%)	11111 (91.2%)	3666 (92.1%)	3268 (88.9%)	1194 (91.4%)
T-Stage						
T1a	11271 (18.1%)	10696 (21.4%)	575 (4.7%)	183 (4.6%)	214 (5.8%)	69 (5.3%)
T1b	8238 (13.3%)	7397 (14.8%)	841 (6.9%)	288 (7.2%)	267 (7.3%)	86 (6.6%)
T2a	17176 (27.6%)	14653 (29.3%)	2523 (20.7%)	832 (20.9%)	840 (22.9%)	264 (20.2%)
T2b	5989 (9.6%)	4615 (9.2%)	1374 (11.3%)	400 (10%)	485 (13.2%)	143 (10.9%)
T3	9616 (15.5%)	6763 (13.5%)	2853 (23.4%)	951 (23.9%)	869 (23.7%)	293 (22.4%)
T4	9861 (15.9%)	5845 (11.7%)	4016 (33%)	1328 (33.4%)	999 (27.2%)	452 (34.6%)
N-Stage						
NX	626 (1%)	346 (0.7%)	280 (2.3%)	93 (2.3%)	83 (2.3%)	32 (2.4%)
N0	32972 (53.1%)	30260 (60.6%)	2712 (22.3%)	863 (21.7%)	1066 (29%)	281 (21.5%)
N1	6262 (10.1%)	5116 (10.2%)	1146 (9.4%)	386 (9.7%)	386 (10.5%)	120 (9.2%)
N2	17174 (27.6%)	11319 (22.7%)	5855 (48.1%)	1885 (47.3%)	1641 (44.7%)	642 (49.1%)
N3	5117 (8.2%)	2928 (5.9%)	2189 (18%)	755 (19%)	498 (13.6%)	232 (17.8%)

SD, standard deviation; IQR, interquartile range; STN0, no separate tumor nodules noted; STN1, separate tumor nodules in ipsilateral lung, same lobe; STN2, separate tumor nodules in ipsilateral lung, different lobe; STN3, separate tumor nodules, ipsilateral lung, same and different lobe.

TABLE 2 | Cancer-specific survival and multivariate analysis for patients with lung cancer.

Site	No. (%)	Cancer-specific survival			Multivariate analysis	
		Median	Mean	SD	HR (95% CI)	P-value
None	19139 (65.3)	10	13.4	12.761	1	
Bone	3262 (11.1)	4	6.97	8.061	1.630 (1.568-1.695)	<0.001
Brain	2974 (10.2)	4	7.22	8.4	1.698 (1.631-1.768)	<0.001
Liver	1126 (3.8)	4	6.46	7.63	1.673 (1.573-1.778)	<0.001
Two or Three	2795 (9.5)	3	5.48	7.075	2.025 (1.941-2.112)	<0.001
Total	29296	7	11.03	11.769		

SD, standard deviation; HR, hazard ratio; CI, confidence interval.

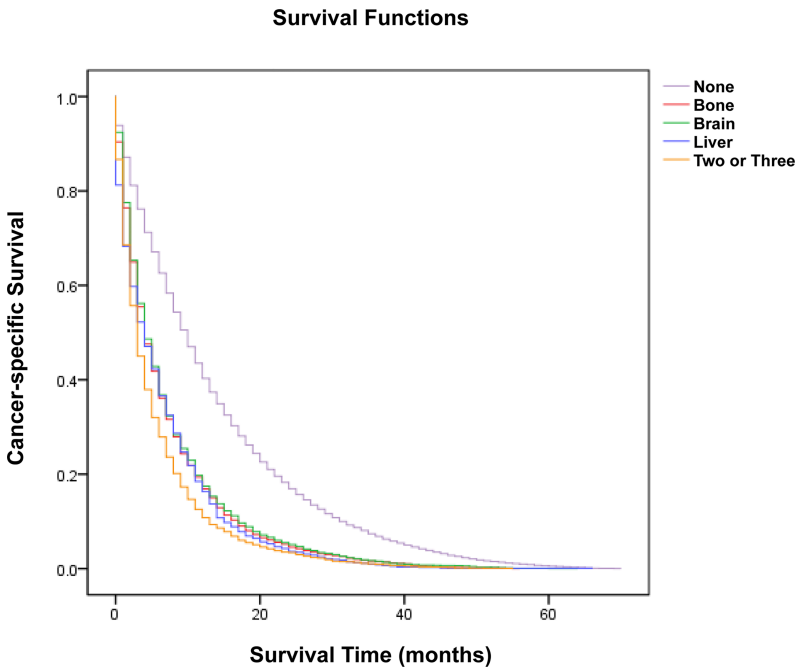


FIGURE 2 | Kaplan-Meier analysis of cancer-specific survival for patients with lung cancer stratified by organ-specific metastasis.

TABLE 3 | Performance of the artificial neural network (ANN) model with increasing layers for predicting distant metastasis.

Number of the hidden layer	AUC	Sensitivity	Specificity	Accuracy	FPR	FNR	LRP	LRN
5	0.737	0.776	0.697	0.713	0.303	0.224	2.565	0.321
6	0.747	0.815	0.679	0.705	0.321	0.185	2.536	0.273
7	0.748	0.837	0.660	0.691	0.340	0.163	2.460	0.247
8	0.759	0.889	0.629	0.679	0.371	0.111	2.398	0.176
9	0.759	0.906	0.613	0.669	0.387	0.094	2.339	0.154
10	0.761	0.902	0.620	0.674	0.380	0.098	2.371	0.158
11	0.756	0.896	0.609	0.665	0.391	0.104	2.293	0.170

AUC, area under curve; FPR, false positive rate; FNR, false negative rate; LRP, likelihood ratio positive; LRN, likelihood ratio negative.

Previously, Zhou et al. used machine learning methods to analyze the distant metastasis possibility of lung cancer based on clinical and radiomic features (36). In this study, if only the features extracted from the CT image were used, the AUC was 72.84%. After combined with the patients' clinical features, 89.09% could be achieved. The authors did not utilize ANN and included radiomic features, limiting direct comparison with our model. Recently, Liu et al. constructed a nomogram to predict bone metastasis of small cell lung cancer (SCL), which had a c-index of 0.745 in the internal validation set (30).

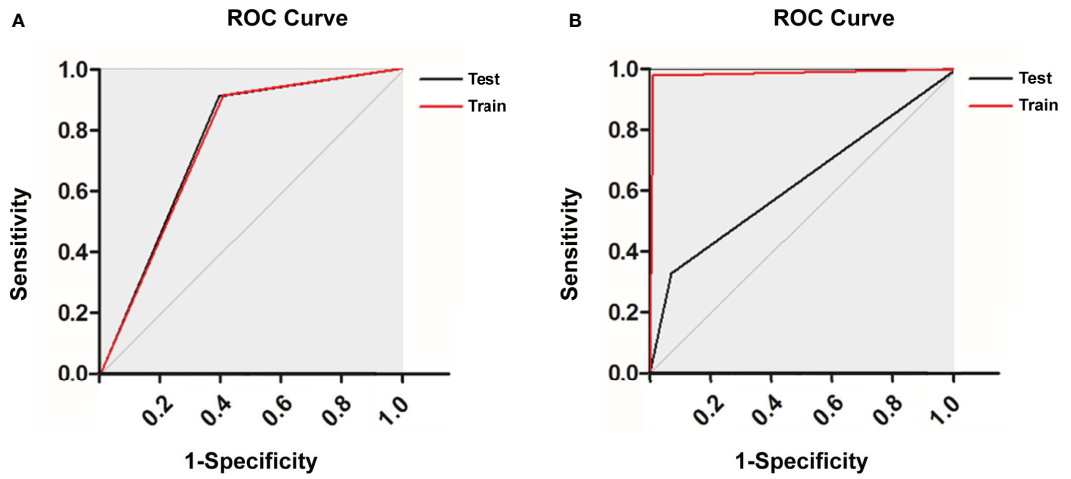


FIGURE 3 | Receiver operating characteristic curve of **(A)** the artificial neural network (ANN) model and **(B)** the random forest (RF) model.

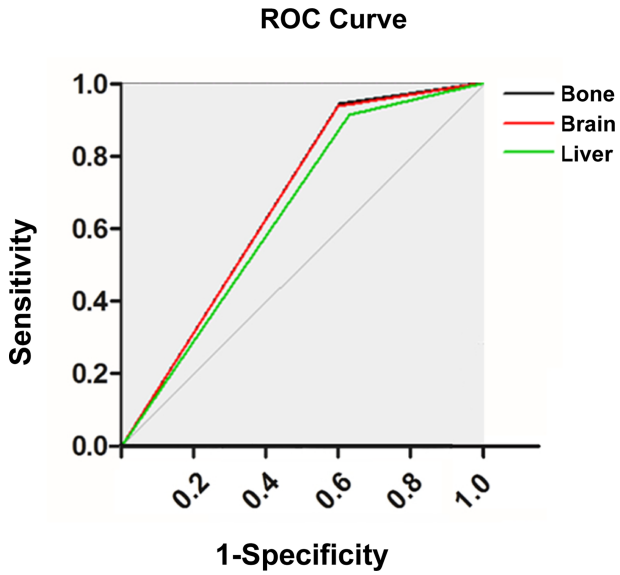


FIGURE 4 | Receiver operating characteristic curve of the artificial neural network (ANN) model for predicting organ-specific metastasis.

Meanwhile, a multivariate model developed by Cacho-Díaz et al. was used to predict brain metastases of non-small cell lung cancer (NSCLC) and showed a predictive sensibility of 72% (27). Although the random forest classifier showed a good performance in predicting overall survival and the early response during radiotherapy in NSCLC, it performed unsatisfactorily in the predictions of our study (37, 38). Therefore, compared with traditional statistical models, our ANN model has superior performance.

In this study, we identified important features in the ANN model, with the top five including separate tumor nodules, visceral pleural invasion, histology, N-stage and tumor size, which were in line with the previous studies (27, 28, 30, 32, 36, 39, 40). Similar to our study, sex and N-stage were reported to be related to the occurrence of bone metastases (30, 32, 40). Interestingly, the correlation between larger tumor size and a higher risk of bone metastasis was uncertain (30, 39). And it was reported that age, sex, T-stage were independent predictors of brain metastasis (27, 28, 31, 41). Although the carcinoembryonic antigen (CEA) levels and epidermal growth factor receptor gene (EGFR) mutation status were associated with brain metastasis in patients with newly diagnosed NSCLC, we did not include these variables because they were not provided in the SEER database (27, 41).

This study should be considered in the context of several limitations. First, the study does not include an independent external cohort to validate the model, which is an important focus of future research. Nonetheless, we hope that the use of

TABLE 4 | Performance of the artificial neural network (ANN) model for predicting organ-specific metastasis.

Site of the organ-specific metastasis	AUC	Sensitivity	Specificity	Accuracy	FPR	FNR	LRP	LRN
Bone	0.688	0.913	0.443	0.539	0.557	0.087	1.638	0.197
Brain	0.686	0.906	0.449	0.525	0.551	0.094	1.646	0.209
Liver	0.664	0.925	0.403	0.453	0.597	0.075	1.548	0.187

AUC, area under curve; FPR, false positive rate; FNR, false negative rate; LRP, likelihood ratio positive; LRN, likelihood ratio negative.

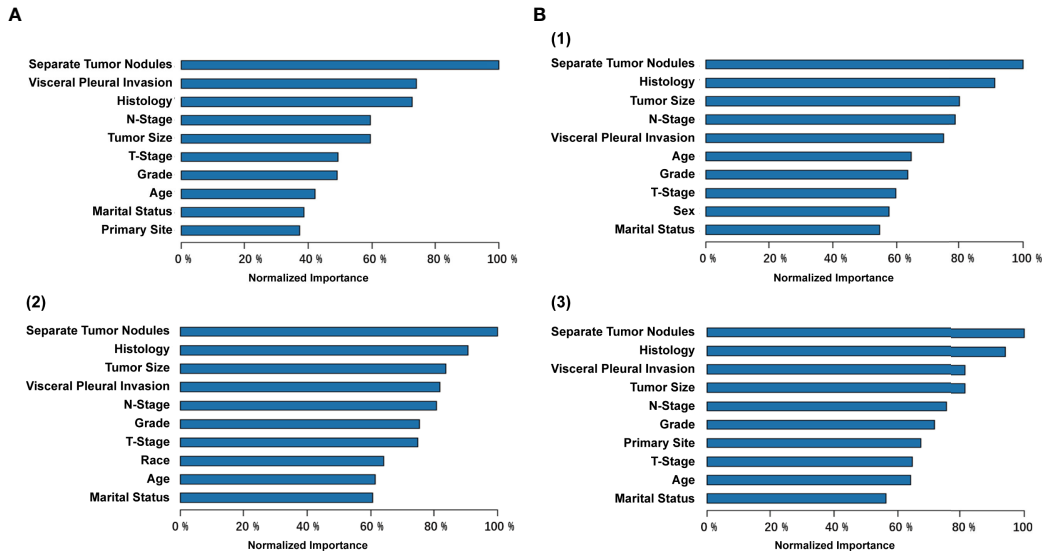


FIGURE 5 | Variable importance from the artificial neural network (ANN) model for predicting **(A)** distant metastasis and **(B)** organ-specific metastasis [(1) bone, (2) brain, and (3) liver].

the SEER database, which accounts for about 28% of the United States population, will improve generalizability. Second, due to retrospective studies, the excluded missing data may lead to selection bias. Therefore, 25% of patients were randomly assigned to the testing group, which allowed for pseudo-prospective evaluation of our model and thus reduced bias.

In conclusion, despite the limitations, we developed and validated a novel ANN model for the prediction of synchronous organ-specific metastasis in patients with lung cancer. This ANN model may help clinicians to make individualized prediction and rational metastasis screening.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

Conception and design: HG, Z-GW and LX. Administrative support: LX and XLD. Provision of study materials or patients: Z-YH and HG. Collection and assembly of data: Z-YH and HG. Data analysis and interpretation: Z-GW and HG. Manuscript writing: All authors. Final approval of manuscript: All authors.

FUNDING

This study was supported by the National Natural Science Foundation of China (grant number 71874058).

ACKNOWLEDGMENTS

The authors acknowledge Guanglian Xiong, Jia Song and Yinghui Li for their advice on the modification of the model parameters. The authors acknowledge the efforts of the National Cancer Institute.

REFERENCES

1. 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 Cancer J Clin* (2018) 68 (6):394–424. doi: 10.3322/caac.21492
2. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. *CA Cancer J Clin* (2017) 67(1):7–30. doi: 10.3322/caac.21387
3. Rosti G, Bevilacqua G, Bidoli P, Portalone L, Santo A, Genestreti G. Small Cell Lung Cancer. *Ann Oncol* (2006) 17 Suppl 2:ii5–10. doi: 10.1093/annonc/mdj910
4. Quint LE. Staging Non-Small Cell Lung Cancer. *Cancer Imaging* (2007) 7 (1):148–59. doi: 10.1102/1470-7330.2007.0026
5. He YY, Zhang XC, Yang JJ, Niu FY, Zeng Z, Yan HH, et al. Prognostic Significance of Genotype and Number of Metastatic Sites in Advanced Non-Small-Cell Lung Cancer. *Clin Lung Cancer* (2014) 15(6):441–7. doi: 10.1016/j.cllc.2014.06.006
6. Wang X, Adjei AA. Lung Cancer and Metastasis: New Opportunities and Challenges. *Cancer Metastasis Rev* (2015) 34(2):169–71. doi: 10.1007/s10555-015-9562-4
7. Altorki NK, Markowitz GJ, Gao D, Port JL, Saxena A, Stiles B, et al. The Lung Microenvironment: An Important Regulator of Tumour Growth and Metastasis. *Nat Rev Cancer* (2019) 19(1):9–31. doi: 10.1038/s41568-018-0081-9
8. Tamura T, Kurishima K, Nakazawa K, Kagohashi K, Ishikawa H, Satoh H, et al. Specific Organ Metastases and Survival in Metastatic Non-Small-Cell Lung Cancer. *Mol Clin Oncol* (2015) 3(1):217–21. doi: 10.3892/mco.2014.410

9. Nakazawa K, Kurishima K, Tamura T, Kagohashi K, Ishikawa H, Satoh H, et al. Specific Organ Metastases and Survival in Small Cell Lung Cancer. *Oncol Lett* (2012) 4(4):617–20. doi: 10.3892/ol.2012.792
10. Schoenmaekers JJAO, Dingemans AC, Hendriks LEL. Brain Imaging in Early Stage Non-Small Cell Lung Cancer: Still a Controversial Topic? *J Thorac Dis* (2018) 10(Suppl 18):S2168–71. doi: 10.21037/jtd.2018.06.68
11. Tang C, Liu Y, Qin H, Li X, Guo W, Li J, et al. Clinical Significance of Serum BAP, TRACP 5b and ICTP as Bone Metabolic Markers for Bone Metastasis Screening in Lung Cancer Patients. *Clin Chim Acta* (2013) 426:102–7. doi: 10.1016/j.cca.2013.09.011
12. Barak A, Neudorfer M, Heilweil G, Merimsky O, Lowenstein A, Inbar M, et al. Decreased Prevalence of Asymptomatic Choroidal Metastasis in Disseminated Breast and Lung Cancer: Argument Against Screening. *Br J Ophthalmol* (2007) 91(1):74–5. doi: 10.1136/bjo.2006.099416
13. Chai X, Yinwang E, Wang Z, Wang Z, Xue Y, Li B, et al. Predictive and Prognostic Biomarkers for Lung Cancer Bone Metastasis and Their Therapeutic Value. *Front Oncol* (2021) 11:692788. doi: 10.3389/fonc.2021.692788
14. Vinod SK. Should We Screen for Brain Metastases in Non-Small Cell Lung Cancer? *J Med Imaging Radiat Oncol* (2018) 62(3):380–2. doi: 10.1111/1754-9485.12743
15. Sargent DJ. Comparison of Artificial Neural Networks With Other Statistical Approaches: Results From Medical Data Sets. *Cancer* (2001) 91(8 Suppl):1636–42. doi: 10.1002/1097-0142(20010415)91:8+<1636::aid-cncr1176>3.0.co;2-d
16. Kononenko I. Machine Learning for Medical Diagnosis: History, State of the Art and Perspective. *Artif Intell Med* (2001) 23(1):89–109. doi: 10.1016/s0933-3657(01)00077-x
17. Dayhoff JE, DeLeo JM. Artificial Neural Networks: Opening the Black Box. *Cancer* (2001) 91(8 Suppl):1615–35. doi: 10.1002/1097-0142(20010415)91:8+<1615::aid-cncr1175>3.0.co;2-l
18. Patel JL, Goyal RK. Applications of Artificial Neural Networks in Medical Science. *Curr Clin Pharmacol* (2007) 2(3):217–26. doi: 10.2174/157488407781668811
19. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine Learning Applications in Cancer Prognosis and Prediction. *Comput Struct Biotechnol J* (2014) 13:8–17. doi: 10.1016/j.csbj.2014.11.005
20. Cruz JA, Wishart DS. Applications of Machine Learning in Cancer Prediction and Prognosis. *Cancer Inform* (2007) 2:59–77. doi: 10.1177/117693510600200030
21. Hu X, Cammann H, Meyer HA, Miller K, Jung K, Stephan C. Artificial Neural Networks and Prostate Cancer–Tools for Diagnosis and Management. *Nat Rev Urol* (2013) 10(3):174–82. doi: 10.1038/nrurol.2013.9
22. Bhambhvani HP, Zamora A, Shkolyar E, Prado K, Greenberg DR, Kasman AM, et al. Development of Robust Artificial Neural Networks for Prediction of 5-Year Survival in Bladder Cancer. *Urol Oncol* (2021) 39(3):193.e7–193.e12. doi: 10.1016/j.urolonc.2020.05.009
23. Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep Learning in Cancer Diagnosis, Prognosis and Treatment Selection. *Genome Med* (2021) 13(1):152. doi: 10.1186/s13073-021-00968-x
24. Khanagar SB, Naik S, Al Kheraif AA, Vishwanathaiah S, Maganur PC, Alhazmi Y, et al. Application and Performance of Artificial Intelligence Technology in Oral Cancer Diagnosis and Prediction of Prognosis: A Systematic Review. *Diagn (Basel)* (2021) 11(6):1004. doi: 10.3390/diagnostics11061004
25. Wang ZG, He ZY, Chen YY, Huan G, Du XL. Incidence and Survival Outcomes of Secondary Liver Cancer: A Surveillance Epidemiology and End Results Database Analysis. *Transl Cancer Res* (2021) 10(3):1273–83. doi: 10.21037/tcr-20-3319
26. Liu W, Wu J. Lung Cancer With Bone Metastases in the United States: An Analysis From the Surveillance, Epidemiologic, and End Results Database. *Clin Exp Metastasis* (2018) 35(8):753–61. doi: 10.1007/s10585-018-9943-5
27. Cacho-Díaz B, Cuapaténcatl LD, Rodríguez JA, Garcilazo-Reyes YJ, Reynoso-Noverón N, Arrieta O. Identification of a High-Risk Group for Brain Metastases in Non-Small Cell Lung Cancer Patients. *J Neurooncol* (2021) 155(1):101–6. doi: 10.1007/s11060-021-03849-w
28. Reddy SP, Dowell JE, Pan E. Predictors of Prognosis of Synchronous Brain Metastases in Small-Cell Lung Cancer Patients. *Clin Exp Metastasis* (2020) 37(4):531–9. doi: 10.1007/s10585-020-10040-4
29. Cai H, Wang H, Li Z, Lin J, Yu J. The Prognostic Analysis of Different Metastatic Patterns in Extensive-Stage Small-Cell Lung Cancer Patients: A Large Population-Based Study. *Future Oncol* (2018) 14(14):1397–407. doi: 10.2217/fon-2017-0706
30. Liu C, Yi J, Jia J. Diagnostic and Prognostic Nomograms for Bone Metastasis in Small Cell Lung Cancer. *J Int Med Res* (2021) 49(10):3000605211050735. doi: 10.1177/03000605211050735
31. Zhu H, Zhou L, Guo Y, Yang G, Dong Q, Zhang Z, et al. Factors for Incidence Risk and Prognosis in Non-Small-Cell Lung Cancer Patients With Synchronous Brain Metastasis: A Population-Based Study. *Future Oncol* (2021) 17(19):2461–73. doi: 10.2217/fon-2021-0103
32. Zhang C, Mao M, Guo X, Cui P, Zhang L, Xu Y, et al. Nomogram Based on Homogeneous and Heterogeneous Associated Factors for Predicting Bone Metastases in Patients With Different Histological Types of Lung Cancer. *BMC Cancer* (2019) 19(1):238. doi: 10.1186/s12885-019-5445-3
33. Diaz ME, Debowski M, Hukins C, Fielding D, Fong KM, Bettington CS. Non-Small Cell Lung Cancer Brain Metastasis Screening in the Era of Positron Emission Tomography-CT Staging: Current Practice and Outcomes. *J Med Imaging Radiat Oncol* (2018) 62(3):383–8. doi: 10.1111/1754-9485.12732
34. Hudson BJ, Crawford MB, Curtin JJ. Brain Imaging in Lung Cancer Patients Without Symptoms of Brain Metastases: A National Survey of Current Practice in England. *Clin Radiol* (2015) 70(6):610–3. doi: 10.1016/j.crad.2015.02.007
35. Vernon J, Andruszkiewicz N, Schneider L, Schieman C, Finley CJ, Shargall Y, et al. Comprehensive Clinical Staging for Resectable Lung Cancer: Clinicopathological Correlations and the Role of Brain MRI. *J Thorac Oncol* (2016) 11(11):1970–5. doi: 10.1016/j.jtho.2016.06.003
36. Zhou H, Dong D, Chen B, Fang M, Cheng Y, Gan Y, et al. Diagnosis of Distant Metastasis of Lung Cancer: Based on Clinical and Radiomic Features. *Transl Oncol* (2018) 11(1):31–6. doi: 10.1016/j.tranon.2017.10.010
37. D'Amico NC, Sicilia R, Cordelli E, Tronchin L, Greco C, Fiore M, et al. Radiomics-Based Prediction of Overall Survival in Lung Cancer Using Different Volumes-of-Interest. *Appl Sci* (2020) 10:6425. doi: 10.3390/app10186425
38. Ramella S, Fiore M, Greco C, Cordelli E, Sicilia R, Merone M, et al. A Radiomic Approach for Adaptive Radiotherapy in Non-Small Cell Lung Cancer Patients. *PloS One* (2018) 13(11):e0207455. doi: 10.1371/journal.pone.0207455
39. Li J, Liu F, Yu H, Zhao C, Li Z, Wang H. Different Distant Metastasis Patterns Based on Tumor Size Could be Found in Extensive-Stage Small Cell Lung Cancer Patients: A Large, Population-Based SEER Study. *PeerJ* (2019) 7:e8163. doi: 10.7717/peerj.8163
40. Song Q, Shang J, Zhang C, Zhang L, Wu X. Impact of the Homogeneous and Heterogeneous Risk Factors on the Incidence and Survival Outcome of Bone Metastasis in NSCLC Patients. *J Cancer Res Clin Oncol* (2019) 145(3):737–46. doi: 10.1007/s00432-018-02826-7
41. Park S, Lee SM, Ahn Y, Kim M, Suh CH, Do KH, et al. Identification of Predictors for Brain Metastasis in Newly Diagnosed Non-Small Cell Lung Cancer: A Single-Center Cohort Study. *Eur Radiol* (2021) 32:990–1001. doi: 10.1007/s00330-021-08215-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 30 May 2022
doi: 10.3389/fonc.2022.882292

Comprehensive Analysis of the Function, Immune Profiles, and Clinical Implication of m1A Regulators in Lung Adenocarcinoma

Guangyao Bao[1], Tian Li[2], Xiaojiao Guan[3], Yao Yao[1], Jie Liang[1], Yifang Xiang[1] and Xinwen Zhong[1]**

[1] Department of Thoracic Surgery, First Affiliated Hospital, China Medical University, Shenyang, China, [2] School of Basic Medicine, Fourth Military Medical University, Xi'an, China, [3] Department of Pathology, Shengjing Hospital, China Medical University, Shenyang, China

OPEN ACCESS

Edited by:
Yutong He,
Fourth Hospital of Hebei Medical University, China

Reviewed by:
Ling Zhao,
Harbin Medical University Cancer Hospital, China
Zhongping Tang,
The First People's Hospital of Shuangliu District, China

***Correspondence:**
Xinwen Zhong
xwzhong@cmu.edu.cn
Tian Li
tian@fmmu.edu.cn

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: 23 February 2022
Accepted: 25 April 2022
Published: 30 May 2022

Citation:
Bao G, Li T, Guan X, Yao Y, Liang J, Xiang Y and Zhong X (2022) Comprehensive Analysis of the Function, Immune Profiles, and Clinical Implication of m1A Regulators in Lung Adenocarcinoma. Front. Oncol. 12:882292. doi: 10.3389/fonc.2022.882292

Background: Previous studies have demonstrated that transcriptional RNA methyladenosine modification significantly affects tumor initiation and progression. However, clinical implications of N1-methyladenosine (m1A) regulators and their effect on tumor immunity in lung adenocarcinoma (LUAD) are still poorly elucidated.

Methods: Herein, the characteristics of somatic mutation, copy number variation (CNV), DNA methylation, and expression levels of m1A regulators were thoroughly analyzed. We classified 955 lung adenocarcinoma patients into different m1A modification patterns based on an unsupervised consensus clustering algorithm. We then calculated the differences in gene expression, prognosis outcomes, and immune profiles among different m1A clusters. Subsequently, we screened differently expressed genes (DEGs) related to prognosis among different m1A clusters. We identified m1A related gene clusters according to the prognosis-related different expressed genes. We further constructed a scoring standard named the m1A score and comprehensively analyzed the survival outcomes, clinical-pathological features, immune microenvironment, treatment responses of immunotherapy, and drug susceptibility in different m1A score groups.

Results: In total, three different m1A modification patterns were identified, which contained cluster A, B, and C. Among them, cluster A processed the poorest clinical outcomes, the lowest immune cell infiltration rate, and the highest tumor purity score. Then, three m1A gene clusters (gene cluster A, B, C) were speculated. Subsequently, we combined m1A modification patterns and m1A gene cluster to classify lung adenocarcinoma patients into high and low m1A score groups. The low m1A score group was accompanied by higher mortality, higher tumor mutation burden (TMB) and genome mutation frequency, and lower programmed cell death-Ligand 1 (PD-L1) expression and tumor immune dysfunction and exclusion (TIDE) expression. Moreover, the m1A score exhibited positive correlation with almost all immune cells. Finally, common

chemotherapeutic and targeted therapy agents exhibited obvious differences in drug susceptibility in different m1A score groups.

Conclusions: Collectively, we explored the potential value of m1A regulators in the prognosis and treatment of lung adenocarcinoma in multiple dimensions and provided some preliminary basis for the follow-up study of m1A regulators in lung adenocarcinoma.

Keywords: m1A, lung adenocarcinoma, prognosis, immune microenvironment, immunotherapy

INTRODUCTION

As the cancer with the highest incidence, lung cancer also causes the most cancer-related deaths (1). According to reports, 85% of the total number of new lung cancer each year is non-small cell lung cancer (NSCLC) (2). Currently, lung adenocarcinoma (LUAD), the major type of NSCLC, shows an increasing incidence in young women and non-smokers (3). LUAD is often accompanied by the characteristics of not obvious early symptoms, and prone to hematogenous metastasis and local infiltration. Moreover, patients with advanced lung adenocarcinoma are often accompanied by poor long-term prognosis. In addition, chemotherapy, targeted therapy, and immunotherapy are facing challenges in treatment effectiveness due to the high drug resistance of LUAD (4–6). Therefore, the discovery of molecular markers for early diagnosis and therapeutic efficacy targets of lung adenocarcinoma is an effective way to improve the survival rate of LUAD.

RNA chemical modifications play crucial roles in regulating important cellular processes at the RNA level, including cell differentiation, key cellular signaling pathways, and cell metabolism (7–9). RNA methylation is the major component of RNA chemical modification, which contains N1-methyladenosine (m1A), N3-methylcytosine (m3C), 5-methylcytosine (m5C), and N6-methyladenosine (m6A) (10–12). Among them, N1-methyladenosine (m1A) has been demonstrated to be involved in stabilizing RNA structural, splicing, cell proliferation, and cell aptosis (13, 14). Common m1A regulators contain "writers" (TRMT10C, TRMT61B, and TRMT6/61A), "readers" (YTHDF1, YTHDF2, YTHDF3, and YTHDC1), and "erasers" (ALKBH1 and ALKBH3), which play an essential role in the m1A methylation process (15–17). In general, the "writer" and "eraser" are involved in regulating the state of m1A, while the "reader" acts as m1A binding proteins to access m1A modification information and further identify and combine with methylation sites. The "writer" acts as a methyltransferase complex. Growing evidence indicates that dysregulation of genomic mutation of m1A regulators can influence the process of transcription and translation, resulting in aberrant cell proliferation and tumor initiation (18–21). Moreover, downregulation of ALKBH3 promoted m1A levels and weakened RNA translation levels associated with the accumulation of methylated RNA in the PANC-1 cell line (22). ALKBH3 and ALKBH1 were upregulated in HNSCC and resulted in tumor development (23). However, studies on m1A regulators in LUAD are lacking. Therefore, a multi-dimensional comprehensive assessment of m1A methylation regulators will enhance our understanding of tumorigenesis and the immune microenvironment in LUAD.

In this research, we first investigated the differences in somatic mutation, CNV, DNA methylation, and expression levels of m1A regulators. Further analysis identified three m1A modification patterns and accessed the correlation with tumor microenvironment (TME). Subsequently, the m1A score was developed and used to qualify the m1A modification pattern of a single LUAD patient. Finally, we comprehensively evaluated the prognosis and treatment efficacy of LUAD based on the m1A score system.

Abbreviations: NSCLC, non-small cell lung cancer; LUAD, lung adenocarcinoma; m1A, N1-methyladenosine; TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; TPM, transcripts per kilobase million; CNV, copy number variation; IC50, half-maximal inhibitory concentration; ssGSEA, single-sample gene-set enrichment analysis; ESTIMATE, Estimation of Stromal and Immune Cells in Malignant Tumors using Expression Data; PCA, principal component analysis; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; TMB, tumor mutation burden; PD-L1, programmed cell death-Ligand 1; TIDE, tumor immune dysfunction and exclusion; TIME, tumor immune microenvironment; ICI, immune checkpoint inhibitor; FPKM, fragments per kilobase of transcript per million mapped reads.

MATERIAL AND METHODS

Data Collection and Analysis

Nine previously published m1A regulators were included in our research (16, 24–28). Somatic mutation data of LUAD were enrolled from the Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/) and visualized by utilizing maftool R package. Subsequently, sequencing data of CNV and DNA methylation were extracted from the Xena database (https://xenabrowser.net/). Transcriptome data and corresponding clinicopathologic characteristics of LUAD were retrospectively curated from TCGA database. Then, three datasets (GSE72094, GSE37745, GSE50081) with clinical information of Gene Expression Omnibus (GEO) were enrolled using GEOquery R package (29), among which GSE37745 and GSE50081 were all RNA sequencing data from the Affymetrix Human Genome U133 Plus 2.0 Array platform. Therefore, we integrated two datasets as a meta-cohort for an independent validation dataset using sva R package for removal of batch effects (30), which contained 235 LUAD samples. Next, RNA expression data of TPM format of the TCGA database and GSE72094 were also combined as a training dataset with sva R package (30), which contained 955 LUAD and 59

normal samples. Furthermore, our study included the anti-PD-L1 treatment cohort IMvigor 210, which contained gene transcriptomic data and clinical information of advanced cancer patients followed by anti-PD-L1 antibody treatment, to further assess the association between m1A modulators and tumor immunity therapy (31).

Unsupervised Consensus Clustering of Nine m1A Regulators in LUAD

Nine m1A regulators were collected to construct m1A modification patterns, including TRMT6, TRMT61A, TRMT10C, YTHDF1, YTHDF2, YTHDF3, YTHDC1, ALKBH1, and ALKBH3. Unsupervised consensus clustering was performed to identify specific m1A modification patterns. According to gene expression levels of m1A regulators, ConsensusClusterPlus R package was enrolled to clustered 955 LUAD patients into subgroups (32). We set the following clustering parameters: number of cycles = 1000; pItem = 0.8; pFeature = 0.8, and k-means was selected as the clustering algorithm. The clusters that expected the most significant difference in survival were taken into consideration.

Identification of Immune Cell Infiltration Among Different m1A Modification Patterns

Immune cell infiltration abundance of different m1A cluster groups was identified by a single-sample gene-set enrichment analysis (ssGSEA) algorithm of the GSVA R package (33). Subsequently, enriched pathways for each cluster were also determined.

Analysis of Differentially Expressed Genes (DEGs) in m1A Cluster Groups

There were 955 lung adenocarcinoma patients classified into three clusters. Then, limma R package was enrolled to identify DEGs in three m1A regulator clusters and adjust p value <0.05 was considered as DEGs.

Construction of m1A-Related Gene Signatures

Univariate Cox regression analysis of overlapped DEGs among the three m1A regulator clusters was performed to select prognosis-related genes for further analysis using survival R package, with p <0.05 as the threshold. Next, based on prognostic-related genes, an unsupervised consensus clustering algorithm was conducted to classify LUAD patients into different m1A gene clusters. Finally, we performed principal component analysis based on prognosis-related gene expression profiles and identified principal components 1 and 2 as the characteristic scores of each patient. This method mainly includes the scores of gene modules with the most significant positive or negative correlations. In view of this, we established the m1A gene signatures of patients with LUAD based on this formula from previous research: m1Ascore = Σ(PC1i + PC2i), where i represented expression level of prognosis-related gene in different m1A gene clusters.

Estimation of Drug Sensitivity

Half-maximal inhibitory concentration (IC50) of paclitaxel, gefitinib, vinblastine, and erlotinib were quantified with the pRRophetic R package by ridge regression analysis (34, 35). IC50 indicated the response to the above-mentioned chemotherapy drugs in the TCGA cohort.

Statistical Analysis

Spearman's correlation analysis was conducted to estimate composition differences. Wilcoxon signed rank test was applied for comparisons between the two groups. Kaplan-Meier survival curve was implemented for evaluating the survival differences between groups. Statistical analysis was achieved utilizing R software (version 4.02). P <0.05 was taken into consideration statistically.

RESULTS

Multi-Omic Landscapes of m1A Regulators in LUAD

We first screened the mutation frequency of nine m1A regulators in LUAD. Our results showed that 37 of 561 LUAD samples (6.6%) contained m1A regulators-related mutation, which ranged from 0 to 2% (**Figure 1A**). Further analysis revealed that CNV events occurred frequently in nine m1A regulators. YTHDF1, YTHDF3, TRMT10C, YTHDC1, and ALKBH3 all displayed widespread copy number amplification. Conversely, TRMT6, TRMT61A, YTHDF2, and ALKBH1 exhibited prevalent copy number deletion (**Figure 1B**). Then, the CNV alternation positions of m1A regulators in human chromosome were visualized (**Supplementary Figure S1A**). The differences in the DNA methylation levels of nine m1A regulators in LUAD were subsequently revealed (**Figure 1C**). The results showed TRMT61A, TRMT10C, YTHDF1, YTHDF3, and ALKBH3 were accompanied with higher DNA methylation levels in LUAD (**Supplementary Figure S1B**). Furthermore, the expression levels of YTHDF1, YTHDF2 TRMT6, TRMT61A, TRMT10C, and ALKBH1 were significantly different compared to normal patients (**Figure 1D**). Finally, a comprehensive survival analysis of nine m1A regulators was listed (**Supplementary Figures S1C–F**).

Identification of Specific m1A Modification Patterns

We first investigated and visualized the interaction network between the nine m1A regulators (**Figure 2A**). A significant interaction network indicated that correlation among different m1A regulators may act as mutually complementary roles in initiation and development of LUAD. According to expression levels of nine m1A regulators, 955 LUAD patients in TCGA and GSE72094 datasets were enrolled in unsupervised clusters for classifying the different m1A modification patterns. We finally determined three different m1A modification patterns: m1Acluster A (222 samples), m1Acluster B (395 samples), and m1Acluster C (338 samples) (**Figure 2B**). With corresponding clinical information, we performed survival analysis among

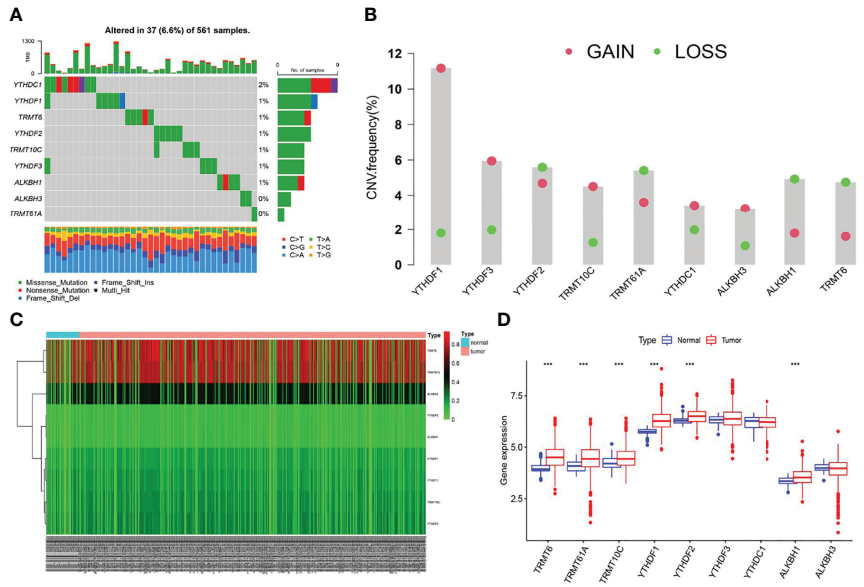


FIGURE 1 | The multi-omic landscapes of nine m1A regulators in LUAD. **(A)** Somatic mutations of nine m1A regulators in TCGA-LUAD. **(B)** The CNV features of m1A regulators in TCGA-LUAD. **(C)** DNA methylation levels of nine m1A regulators in TCGA-LUAD and normal patients (Tumor: 563; Normal: 53). **(D)** Gene expression levels of nine m1A regulators in TCGA-LUAD and normal patients (***P < 0.001).

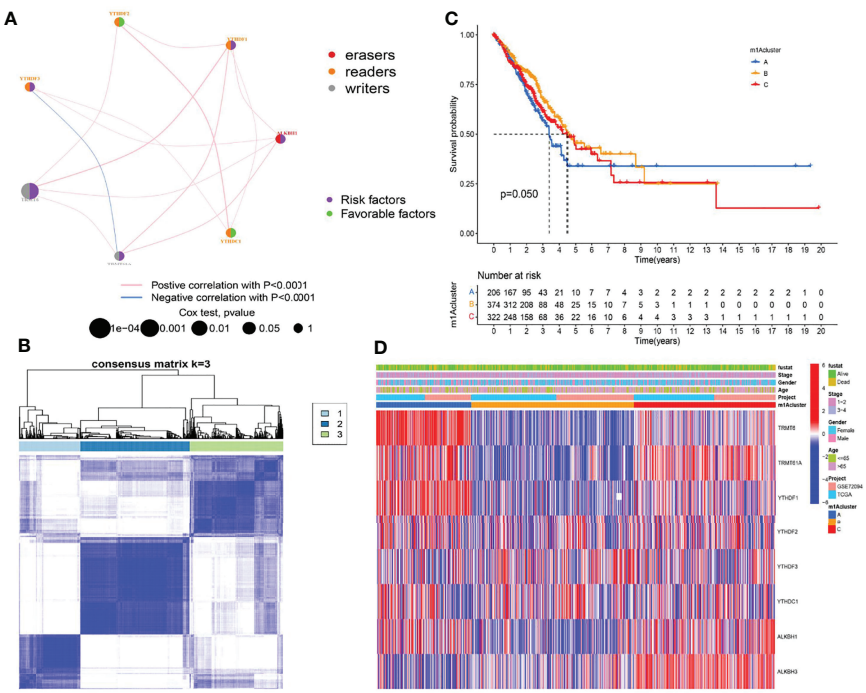


FIGURE 2 | Establishment of three m1A modification patterns. **(A)** The correlation of nine m1A regulators in LUAD. **(B)** Consensus clustering of m1A clusters for LUAD patients in the training cohort. **(C)** Differences in survival outcomes of three m1A clusters in the training cohort. **(D)** Heatmap of three m1A modification patterns in the training cohort.

different m1A clusters, and the result showed patients in cluster A exhibited the poorest clinical outcome (**Figure 2C**). Finally, a heatmap systemically depicted the difference in expression levels and clinical pathological features among 3 m1A clusters (**Figure 2D**).

The Immune Landscape of Different m1A Modification Patterns

The GSVA algorithm was performed to investigate specific biological pathways within the different m1A modification patterns (**Figures 3A–C**). The results of GSVA revealed the m1A cluster A mainly enriched in basal transcription factors, RNA degradation, and cell cycle. The m1A cluster B was strongly associated with complement and coagulation cascades, cell adhesion molecules cams, and cytokine receptor interaction. Moreover, the m1A cluster C exhibited high correlation with cell metabolism, RNA polymerase, and splicesome. Then, we calculated the levels of immune and stromal components across LUAD tissues through the ESTIMATE algorithm. Accordingly, m1A cluster A was accompanied by the highest tumor purity score and the lowest estimate, immune, and stromal score, whereas the m1A cluster B was characterized by the lowest tumor purity score and highest estimate, immune, and stromal score (**Figures 3D–G**). Finally, we systematically qualified the distribution landscape of immune cell infiltration among different m1A modification patterns, with the result indicating that the m1A cluster B displayed the most abundant distribution of adaptive and innate immune cells (**Figure 3H**).

Investigation of m1A-Related DEGs in LUAD

The above research fully clarified the effects of different m1A modification patterns on the immune microenvironment and clinical outcomes of LUAD patients. To further investigate the underlying impact of m1A regulators in LUAD, we first performed principal component analysis (PCA) based on m1A gene expression and clustering data and revealed that the m1A modification patterns could well reflect the heterogeneity of LUAD patients (**Supplementary Figure S2**). Subsequently, DEGs that intersected between the three m1A clusters were screened and 2986 DEGs were finally uncovered (**Figure 4A**). Then, 1787 prognosis-related DEGs were uncovered by utilizing univariate Cox regression analysis. Gene ontology (GO) enrichment analysis suggested that prognosis-related DEGs were mainly enriched in DNA replication, chromosome segregation, and ATPase activity (**Figure 4B**). Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis uncovered that prognosis-related DEGs exhibited strong association with DNA replication, cell cycle, and cell adhesion molecules (**Figure 4C**). These results uncovered that m1A-related genes participated in vital cellular pathways and predicted poor clinical outcomes, which may lead to the occurrence and progression of LUAD. Next, we performed unsupervised consensus clusters based on the expression profiling data of prognosis-related DEGs. Then, three m1A-related gene clusters were identified (gene cluster A, B, C) (**Figure 4D**). Subsequently, we found that the LUAD patients divided into m1A gene cluster A were highly correlated with worse survival outcomes (**Figure 4E**). Then, the heatmap comprehensively depicted the clinicopathological characteristics and differences of these subgroups (**Figure 4F**). Finally, we screened the differential expression of m1A regulators among m1A gene clusters. The results revealed that TRMT6, TRMT61A, YTHDF1, and ALKBH1 were significantly upregulated in m1A gene cluster A, whereas YTHDC1 and ALKBH3 were upregulated in m1A gene cluster B (**Figure 4G**).

Construction of m1A-Related Gene Signatures

The above studies were based on the different m1A classifications of LUAD patients, so it is far from accurate to qualify the impact of m1A modification patterns on specific patient samples.

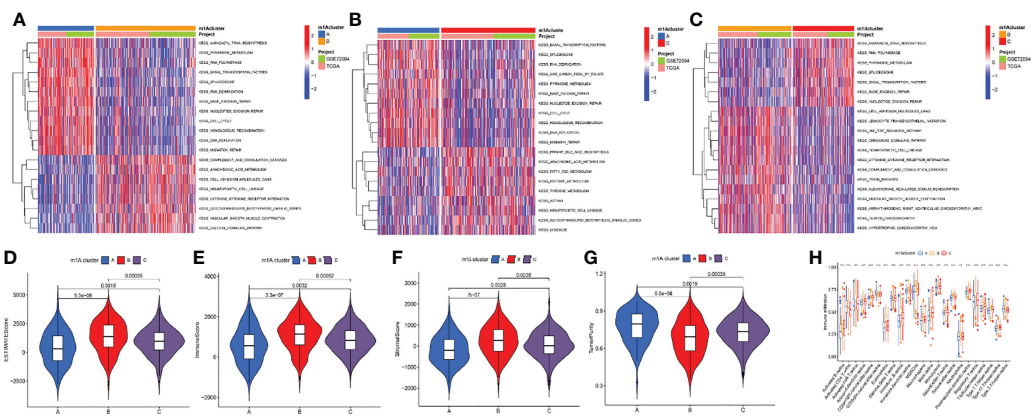


FIGURE 3 | Differences of biological features and immune profiles of three m1A clusters in LUAD. **(A–C)** GSVA results revealed specific biological pathways of three m1A modification patterns. **(A)** Cluster A vs. cluster B; **(B)** cluster A vs. cluster C; **(C)** cluster B vs. cluster C. **(D–G)** Violin plots depicted the distribution of ESTIMATE, immune, and stroma scores as well as tumor purity in three m1A modification patterns. **(H)** Differences in abundance of 23 TME-infiltrating cells under three m1A modification patterns. ** indicates $p < 0.01$; *** indicates $p < 0.001$.

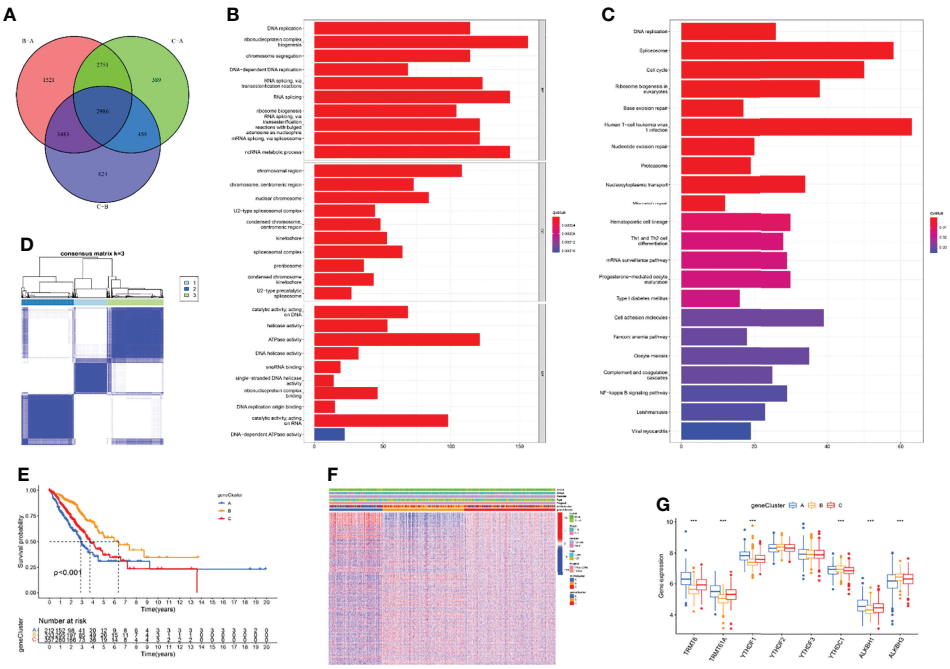


FIGURE 4 | Identification of three m1A gene clusters. **(A)** 2986 DEGs of three m1A modification patterns were shown by Venn diagram. **(B, C)** GO and KEGG results revealed the potential function of 1787 prognosis-related DEGs. **(D)** Consensus clustering of m1A gene clusters for LUAD patients in the training cohort. **(E)** Survival outcome prediction of three m1A gene clusters in the training cohort. **(F)** Heatmap of consensus clustering of 1787 prognosis-related DEGs. **(G)** The different expression levels of m1A regulators in three m1A gene clusters (***P < 0.001).

Therefore, we scored LUAD patients according to the m1A modification patterns and gene clusters, named m1A score. We then classified LUAD patients into high and low m1A score groups based on the median m1A score (median m1A score = 4.88). Alluvial plots were performed to depict the correspondence in m1A clusters, m1A gene clusters, m1A score, and the survival status of patients (**Figure 5A**). Then, we noticed that LUAD patients who were surviving showed a higher m1A score than dead LUAD patients (**Figure 5B**). We also uncovered that low m1A score in LUAD patients who were diagnosed with pathological stage I and II often indicated poor prognostic outcomes, whereas the effects of m1A score on predicting prognostic outcomes in LUAD patients with stage III and stage IV showed no statistical difference (**Figures 5C, D**). In addition, based on the information of 889 LUAD patients with clinicopathological stage, we found significant differences in m1A scores among LUAD patients with different pathological stages (**Figure 5E**). Moreover, m1A scores in different m1A modification patterns and gene clusters also showed significant differences (**Figures 5F, G**). Next, we comprehensively screened the differences in somatic mutations in tumor genomes based on the grouping of m1A scores, which indicated that the low m1A score group exhibited the higher mutation frequency (**Figures 5H, I**). Further, we identified that low m1A score subpopulations were often accompanied by high mortality, which was consistent with the result of the meta-cohort (**Figures 6A, B**). Subsequently, univariate and multivariate analysis uncovered m1A scores possessed the potential in independently predicting LUAD prognosis (**Figures 6C, E**), which was also validated in the meta-cohort (**Figures 6D, F**).

Identifying and Comparing the Immune Profiles of Different m1A Score Groups

We first determined the association of m1A score with immune cells. The result uncovered that m1A score was positively correlated with almost all immune cells except CD56 natural killer cell (**Figure 7A**). These findings revealed that m1Ascore could be used to effectively evaluate m1A modification patterns and the differences in immune cell infiltration in a single LUAD patient. We also investigated the relationship between m1A score and tumor burden mutation (TMB) and noticed that the low m1A score group exhibited high-level of TMB (**Figure 7B**). Moreover, the m1A score was negatively correlated with TMB (**Supplementary Figure S3**). Further analysis uncovered that LUAD patients in low m1A score and low TMB group displayed the worst survival outcomes (**Figure 7C**). Next, the interaction of m1A score and PD-L1 expression was investigated. Patients with LUAD in the low m1A score group were accompanied by low levels of PD-L1 expression (**Figure 7D**). We comprehensively evaluated the immunotherapy response of patients with LUAD based on the m1A score. High m1A score group patients expressed therapeutic advantage to CTLA-4 and PD-1 monotherapy (**Figures 7E, F**). Similarly, LUAD patients with high m1A score showed treatment advantages to CTLA-4 and PD-1 combined treatment (**Figure 7G**). In addition, we determined

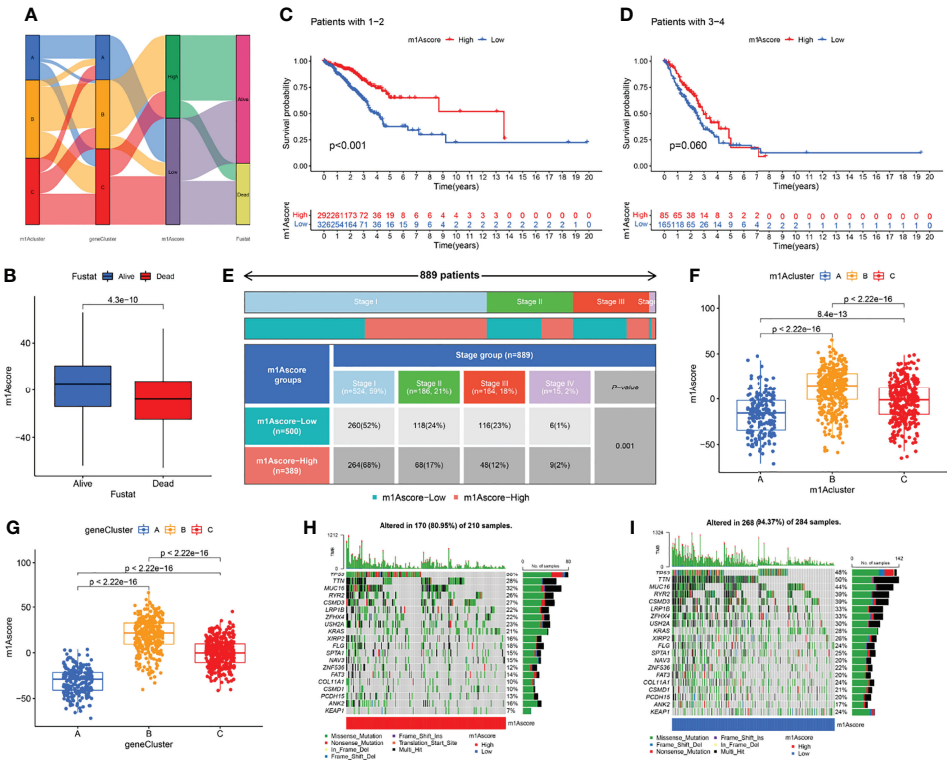


FIGURE 5 | Clinical features in different m1A score groups. **(A)** Alluvial plots depicted the correspondence in m1A clusters, m1A gene clusters, m1A score, and the survival status of patients in LUAD. **(B)** The differences of fustat in different m1A score groups. **(C–E)** Kaplan-Meier curves uncovered differences in m1A scores among LUAD patients at different pathological stages. **(F, G)** The differences of m1A score among three m1A modification patterns and three m1A gen clusters. **(H, I)** The differences of somatic mutation frequency in tumor genomes based on the grouping of m1A scores.

immunotherapy efficacy in patients with different m1A score groups based on the TIDE database. The results indicated that the TIDE expression showed significant difference in the high and low m1A score groups (**Figure 7H**). Further, we noticed that LUAD patients in high m1A score group were accompanied by higher tumor dysfunction score and lower tumor exclusion score (**Figures 7I, J**). Finally, we speculated the ability of the m1A score in predicting patients' immune effects based on the IMvigor210 immunotherapy cohort. Interestingly, we noticed that the immunotherapy efficacy in different m1A score groups exhibited no statistical difference (**Supplementary Figure S4**).

Drug Susceptibility Prediction in Different m1A Score Groups

Chemotherapy and targeted therapy were gradually applied in treatments for patients with advanced LUAD. It is of great significance to evaluate the responses of certain drugs in different subpopulations. Herein, we identified the treatment responses of some drugs that were widely used in the treatment of LUAD. As shown in **Figure 8**, the high m1A score group possessed prominently high IC50 values of erlotinib and paclitaxel, indicating that this subpopulation showed higher sensitivity to these therapeutic agents, whereas patients in the low m1A score group showed therapeutic superiority to gefitinib and vinblastine. The above research results provided more reference values for formulating personalized treatment strategies for LUAD patients.

DISCUSSION

Previous studies have confirmed that m1A methylation modification significantly affects the occurrence and development of tumors (36, 37). However, there are few studies exploring the role of m1A modification in the tumorigenesis of LUAD. Herein, we first revealed the underlying role of m1A modification in LUAD from multiple perspectives. Then, we identified the differences of TME cells infiltration among three m1A modification patterns. Subsequently, the m1A score system was constructed and used to qualify the m1A modification pattern of a single LUAD patient. Finally, we comprehensively evaluated the prognosis and treatment efficacy of LUAD based on the m1A score system.

In this study, the characteristics of somatic mutation, copy number variation (CNV), DNA methylation, and gene expression levels of nine m1A regulators in TCGA-LUAD cohort were screened. We found that 37 of 561 LUAD samples (6.6%) contained m1A regulators-related mutation, with mutation frequencies ranging from

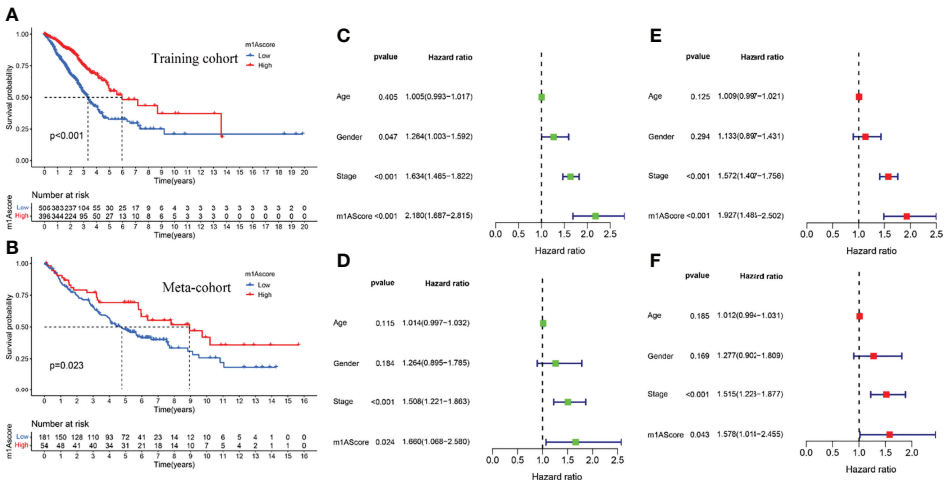


FIGURE 6 | Construction of m1A score and validation. **(A)** Survival outcomes prediction of the m1A score in the training cohort. **(B)** Survival outcomes prediction of the m1A score in the meta-cohort. **(C, E)** Univariate and multivariate analyses revealed the prognostic value of m1A score in the training cohort. **(D, F)** Univariate and multivariate analyses revealed the prognostic value of m1A score in the meta cohort.

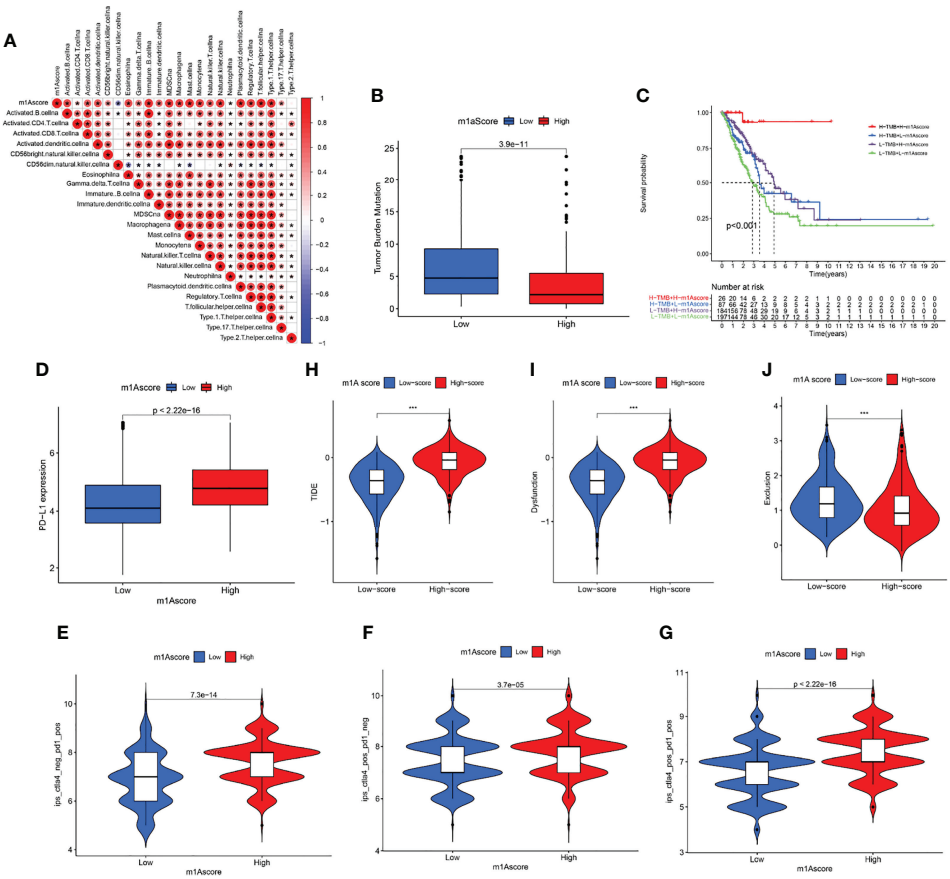


FIGURE 7 | Profile of differences in the immune microenvironment in different m1A score groups. **(A)** Correlation between m1Ascore and immune-related cellular components. Blue indicates negative correlation; red indicates positive correlation; * indicates P < 0.05. **(B)** Comparisons of TMB score in different m1A score groups. **(C)** Overall survival analysis of different m1A score and TMB score groups. **(D)** PD-L1 expression levels in different m1A score groups. **(E–G)** Treatment effects of CTLA-4 or PD-1 and combined CTLA-4 and PD-1 were evaluated in patients with high and low m1A scores. **(H–J)** TIDE, dysfunction, and exclusion scores in different m1A score groups.

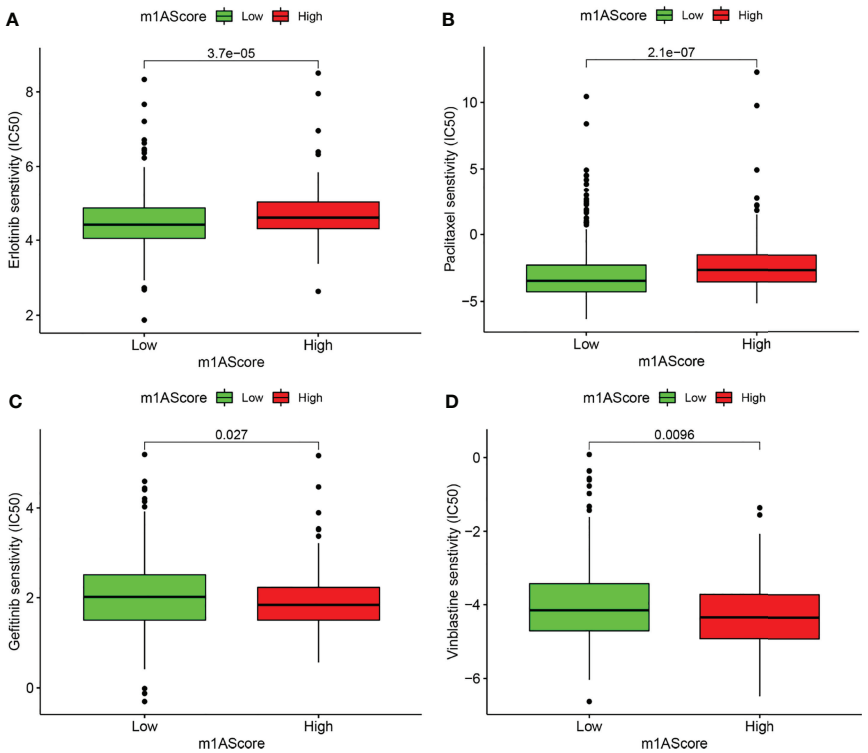


FIGURE 8 | Drug susceptibility prediction in different m1A score groups. **(A–D)** Comparisons of sensitivity to erlotinib, paclitaxel, gefitinib, and vinblastine in different m1A score groups.

0 to 2%, and YTHDC1 occupied the top mutation frequency. Previous research suggested that YTHDC1 deficiency could significantly increase the level of alternative splicing defects in mouse oocytes (38). Further, six m1A regulators displayed significantly high expression levels in LUAD patients. Moreover, high expression of ALKBH1, TRMT6, and TRMT61A was found to be indicative of poor clinical outcomes, which were consistent with their high expression in LUAD. Further studies on the CNV signature of nine m1A regulators showed that YTHDF3, TRMT10C, YTHDC1, YTHDF1, and ALKBH3 displayed copy number amplification, while TRMT6, TRMT61A, YTHDF2, and ALKBH1 exhibited copy number deletion. Genomic alternations of the m1A regulators in LUAD could be due to abnormal gene expression, which contributed to tumor development.

Then, three m1A modification patterns were revealed, named m1A clusters A, B, C. Within these modification patterns, cluster A displayed the poorest long-term survival outcomes. Meanwhile, cluster A was accompanied by the lowest estimate, immune, and stromal score, and the highest tumor purity. Further analysis revealed cluster A showed lower infiltration of immune cells. The purity of the tumor depends on the proportion of tumor cells in TME, and its level affects the prognosis of cancer patients (39, 40). Hence, we speculated that the poor long-term prognosis of LUAD patients in m1A cluster A may be related to high tumor purity and suppression of immune function. GSVA analysis revealed the m1A cluster A mainly enriched in basal transcription factors, RNA degradation, and cell cycle, which may be involved in the progression of LUAD (41, 42). In view of this, we believe that different m1A modification patterns may shape the tumor immune microenvironment (TIME) of LUAD, thereby potentially affecting the prognosis of LUAD.

Based on prognostic-related DEGs among different m1A modification patterns, we revealed three m1A gene clusters. Similarly, the m1A gene cluster exhibited the poorest clinical outcome. The m1A scoring system was subsequently constructed to assess the impact of m1A methylation on single LUAD patients. Survival analysis revealed that LUAD patients with low m1A score suggested high mortality. In addition, we found LUAD patients classified into m1A cluster A and m1A gene cluster A groups were accompanied by the lowest m1A score. In view of this, we noticed that LUAD patients showed different clinical prognosis with different grouping methods based on m1A modification patterns, which revealed that m1A had a clear prognostic value for LUAD patients. TMB was characterized as an effective indicator for prediction of clinical response to immunotherapy (43). Our data suggested that low m1A score group showed a high-level of TMB, while LUAD patients with both low m1A score and low TMB exhibited poor clinical prognosis. Moreover, m1A score was positively correlated with almost all immune cells. The above findings suggested m1A score may have the ability to predict prognosis of LUAD and evaluate

the tumor immune microenvironment and immune response of LUAD.

Immunotherapy is gradually becoming an important treatment for advanced LUAD. PD-L1 has become a powerful biomarker to assess the efficacy of immune checkpoint inhibitor (ICI) in LUAD patients recently (44). High expression of PD-L1 often predicts better treatment response to ICIs (6, 44, 45). Here, we noticed that high m1A score group exhibited high expression level of PD-L1, which uncovered LUAD patients with high m1A score may occupy a higher priority for anti-PD-L1 therapy. The above results indicated that predicting anti-PD-L1 efficacy based on m1A score required more clinical trials to verify. Moreover, we found that high m1A score patients presented higher sensitivity to erlotinib and paclitaxel, and patients with low m1A score had higher priority to gefitinib and vinblastine, providing a reference for the choice of the optimal chemotherapeutic or targeted therapeutic regimen.

CONCLUSION

To sum up, our study characterizes m1A regulators in LUAD from multiple dimensions and qualified its significant role in predicting prognosis value and immune performance. Further analysis revealed the interaction between m1A score and immune microenvironment. Importantly, we provided some preliminary basis for the follow-up study of m1A regulators in lung adenocarcinoma. Nevertheless, their potential significance as prognostic indicators and therapeutic guidance value of LUAD is worthy of further study.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: TCGA-LUAD, GSE72094, GSE37745, and GSE50081.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics committee of China Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

GYB performed the statistical analyses and wrote the manuscript. GYB completed all of the data entry and provided assistance for the data analysis. GYB, XJG, YY, JL and XWZ were responsible for the diagnosis and clinical assessment of the participants. XW Z and TL designed and wrote the study protocol and reviewed the manuscript. XJG and YX participated the revision of this manuscript. In addition, YY and JL offered many constructive opinions on this study and provided a critical revision of the manuscript for important intellectual content. All authors contributed to and approved the final manuscript.

FUNDING

This work was supported by Wu Jieping Medical Foundation [grant number 320.6750.2020-17-7].

ACKNOWLEDGMENTS

The authors would like to express their sincere thanks to all who supported and helped them conduct this research, as well as Editage for language editing services.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.882292/full#supplementary-material

Supplementary Figure 1 | Comprehensive analysis of nine m1A regulators in lung adenocarcinoma multi-omics. **(A)** The CNV alternation positions in human chromosome. **(B)** The DNA methylation levels in TCGA-LUAD and normal patients (*P < 0.05; **P < 0.01; ***P < 0.001). **(C–F)** Survival outcome prediction of ALKBH1, TRMT6, TRMT61A, and YTHDF2 in the training cohort.

Supplementary Figure 2 | Principal component analysis revealed that the m1A modification patterns could well reflect the heterogeneity of LUAD patients.

Supplementary Figure 3 | TMB exhibited a significantly negative correlation with m1A score (R = -0.37, p < 2.2e-16).

Supplementary Figure 4 | Immune responses of different m1A score groups exhibited no statistical difference in IMvigor210 immunotherapy cohort (p = 0.78).

REFERENCES

1. Siegel R, Miller K, Jemal A. Cancer Statistics, 2020. *CA* (2020) 70:7–30. doi: 10.3322/caac.21590
2. Oudkerk M, Liu S, Heuvelmans M, Walter J, Field J. Lung Cancer LDCT Screening and Mortality Reduction - Evidence, Pitfalls and Future Perspectives. *Nat Rev Clin Oncol* (2021) 18(3):135–51. doi: 10.1038/s41571-020-00432-6
3. Zhu X, Chen L, Liu L, Niu X. EMT-Mediated Acquired EGFR-TKI Resistance in NSCLC: Mechanisms and Strategies. *Front Oncol* (2019) 9:1044. doi: 10.3389/fonc.2019.01044
4. Sequist L, Martins R, Spigel D, Grunberg S, Spira A, Jänne P, et al. First-Line Gefitinib in Patients With Advanced non-Small-Cell Lung Cancer Harboring Somatic EGFR Mutations. *J Clin Oncol* (2008) 26(15):2442–9. doi: 10.1200/JCO.2007.14.8494
5. Wang Y, Yang N, Zhang Y, Li L, Han R, Zhu M, et al. Effective Treatment of Lung Adenocarcinoma Harboring EGFR-Activating Mutation, T790M, and Cis-C797S Triple Mutations by Brigatinib and Cetuximab Combination Therapy. *J Thorac Oncol* (2020) 15(8):1369–75. doi: 10.1016/j.jtho.2020.04.014
6. Garon E, Rizvi N, Hui R, Leighl N, Balmanoukian A, Eder J, et al. Pembrolizumab for the Treatment of non-Small-Cell Lung Cancer. *N Engl J Med* (2015) 372(21):2018–28. doi: 10.1056/NEJMoa1501824

7. Shi H, Chai P, Jia R, Fan X. Novel Insight Into the Regulatory Roles of Diverse RNA Modifications: Re-Defining the Bridge Between Transcription and Translation. *Mol Cancer* (2020) 19(1):78. doi: 10.1186/s12943-020-01194-6
8. Li S, Mason C. The Pivotal Regulatory Landscape of RNA Modifications. *Annu Rev Genomics Hum Genet* (2014) 15:127–50. doi: 10.1146/annurev-genom-090413-025405
9. Barbieri I, Kouzarides T. Role of RNA Modifications in Cancer. *Nat Rev Cancer* (2020) 20:303–22. doi: 10.1038/s41568-020-0253-2
10. Desrosiers R, Friderici K, Rottman F. Identification of Methylated Nucleosides in Messenger RNA From Novikoff Hepatoma Cells. *Proc Natl Acad Sci USA* (1974) 71:3971–5. doi: 10.1073/pnas.71.10.3971
11. Adams J, Cory S. Modified Nucleosides and Bizarre 5'-Termini in Mouse Myeloma mRNA. *Nature* (1975) 255:28–33. doi: 10.1038/255028a0
12. Wei C, Gershowitz A, Moss B. Methylated Nucleotides Block 5' Terminus of HeLa Cell Messenger RNA. *Cell* (1975) 4:379–86. doi: 10.1016/0092-8674(75)90158-0
13. Dominissini D, Nachtergaele S, Moshitch-Moshkovitz S, Peer E, Kol N, Ben-Haim M, et al. The Dynamic N(1)-Methyladenosine Methylome in Eukaryotic Messenger RNA. *Nature* (2016) 530(7591):441–6. doi: 10.1038/nature16998
14. Li X, Xiong X, Zhang M, Wang K, Chen Y, Zhou J, et al. Base-Resolution Mapping Reveals Distinct mA Methylome in Nuclear- and Mitochondrial-Encoded Transcripts. *Mol Cell* (2017) 68:993–1005.e9. doi: 10.1016/j.molcel.2017.10.019
15. Zhang C, Jia G. Reversible RNA Modification N-Methyladenosine (Ma) in mRNA and tRNA. *Genomics Proteomics Bioinf* (2018) 16:155–61. doi: 10.1016/j.gpb.2018.03.003
16. Safra M, Sas-Chen A, Nir R, Winkler R, Nachshon A, Bar-Yaacov D, et al. The M1a Landscape on Cytosolic and Mitochondrial mRNA at Single-Base Resolution. *Nature* (2017) 551(7679):251–5. doi: 10.1038/nature24456
17. Anreiter I, Mir Q, Simpson J, Janga S, Soller M. New Twists in Detecting mRNA Modification Dynamics. *Trends Biotechnol* (2021) 39:72–89. doi: 10.1016/j.tibtech.2020.06.002
18. Waku T, Nakajima Y, Yokoyama W, Nomura N, Kako K, Kobayashi A, et al. NML-Mediated rRNA Base Methylation Links Ribosomal Subunit Formation to Cell Proliferation in a P53-Dependent Manner. *J Cell Sci* (2016) 129(12):2382–93. doi: 10.1242/jcs.183723
19. Engel M, Chen A. The Emerging Role of mRNA Methylation in Normal and Pathological Behavior. *Genes Brain Behav* (2018) 17:e12428. doi: 10.1111/gbb.12428
20. Woo H, Chambers S. Human ALKBH3-Induced mA Demethylation Increases the CSF-1 mRNA Stability in Breast and Ovarian Cancer Cells. *Biochim Biophys Acta Gene Regul Mech* (2019) 1862:35–46. doi: 10.1016/j.bbagrm.2018.10.008
21. Shi Q, Xue C, Yuan X, He Y, Yu Z. Gene Signatures and Prognostic Values of M1a-Related Regulatory Genes in Hepatocellular Carcinoma. *Sci Rep* (2020) 10(1):15083. doi: 10.1038/s41598-020-72178-1
22. Ueda Y, Ooshio I, Fusamae Y, Kitae K, Kawaguchi M, Jingushi K, et al. AlkB Homolog 3-Mediated tRNA Demethylation Promotes Protein Synthesis in Cancer Cells. *Sci Rep* (2017) 7:42271. doi: 10.1038/srep42271
23. Pilžys T, Marcinkowski M, Kukwa W, Garbicz D, Dylewska M, Ferenc K, et al. ALKBH Overexpression in Head and Neck Cancer: Potential Target for Novel Anticancer Therapy. *Sci Rep* (2019) 9(1):13249. doi: 10.1038/s41598-019-49550-x
24. Chujo T, Suzuki T. Trmt61B is a Methyltransferase Responsible for 1-Methyladenosine at Position 58 of Human Mitochondrial tRNAs. *RNA (New York NY)* (2012) 18:2269–76. doi: 10.1261/rna.035600.112
25. Liu F, Clark W, Luo G, Wang X, Fu Y, Wei J, et al. ALKBH1-Mediated tRNA Demethylation Regulates Translation. *Cell* (2016) 167(7):1897. doi: 10.1016/j.cell.2016.11.045
26. Li X, Xiong X, Wang K, Wang L, Shu X, Ma S, et al. Transcriptome-Wide Mapping Reveals Reversible and Dynamic N(1)-Methyladenosine Methylome. *Nat Chem Biol* (2016) 12(5):311–6. doi: 10.1038/nchembio.2040
27. Dai X, Wang T, Gonzalez G, Wang Y. Identification of YTH Domain-Containing Proteins as the Readers for N1-Methyladenosine in RNA. *Anal Chem* (2018) 90(11):6380–4. doi: 10.1021/acs.analchem.8b01703
28. Chen Z, Qi M, Shen B, Luo G, Wu Y, Li J, et al. Transfer RNA Demethylase ALKBH3 Promotes Cancer Progression *via* Induction of tRNA-Derived Small RNAs. *Nucleic Acids Res* (2019) 47(5):2533–45. doi: 10.1093/nar/gky1250
29. Davis S, Meltzer P. GEOquery: A Bridge Between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinf (Oxford England)* (2007) 23:1846–7. doi: 10.1093/bioinformatics/btm254
30. Leek J, Johnson W, Parker H, Jaffe A, Storey J. The Sva Package for Removing Batch Effects and Other Unwanted Variation in High-Throughput Experiments. *Bioinf (Oxford England)* (2012) 28(6):882–3. doi: 10.1093/bioinformatics/bts034
31. Mariathasan S, Turley S, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. Tgfβ Attenuates Tumour Response to PD-L1 Blockade by Contributing to Exclusion of T Cells. *Nature* (2018) 554(7693):544–8. doi: 10.1038/nature25501
32. Wilkerson M, Hayes D. ConsensusClusterPlus: A Class Discovery Tool With Confidence Assessments and Item Tracking. *Bioinf (Oxford England)* (2010) 26:1572–3. doi: 10.1093/bioinformatics/btq170
33. Hänzelmann S, Castelo R, Guinney J. GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinf* (2013) 14:7. doi: 10.1186/1471-2105-14-7
34. Geeleher P, Cox N, Huang R. Clinical Drug Response can be Predicted Using Baseline Gene Expression Levels and *In Vitro* Drug Sensitivity in Cell Lines. *Genome Biol* (2014) 15:R47. doi: 10.1186/gb-2014-15-3-r47
35. Geeleher P, Cox N, Huang R. Prrophetic: An R Package for Prediction of Clinical Chemotherapeutic Response From Tumor Gene Expression Levels. *PloS One* (2014) 9:e107468. doi: 10.1371/journal.pone.0107468
36. Shi L, Yang X, Tang D, Liu G, Yuan P, Yang Y, et al. Expression and Significance of M1a Transmethylase, Htrm6p/Htrm61p and its Related Gene Htrm6/Htrm61 in Bladder Urothelial Carcinoma. *Am J Cancer Res* (2015) 5(7):2169–79.
37. Zhao Y, Zhao Q, Kaboli P, Shen J, Li M, Wu X, et al. M1a Regulated Genes Modulate PI3K/AKT/mTOR and ErbB Pathways in Gastrointestinal Cancer. *Trans Oncol* (2019) 12:1323–33. doi: 10.1016/j.tranon.2019.06.007
38. Kasowitz S, Ma J, Anderson S, Leu N, Xu Y, Gregory B, et al. Nuclear M6a Reader YTHDC1 Regulates Alternative Polyadenylation and Splicing During Mouse Oocyte Development. *PloS Genet* (2018) 14(5):e1007412. doi: 10.1371/journal.pgen.1007412
39. Rhee J, Jung Y, Kim K, Yoo J, Kim J, Lee Y, et al. Impact of Tumor Purity on Immune Gene Expression and Clustering Analyses Across Multiple Cancer Types. *Cancer Immunol Res* (2018) 6(1):87–97. doi: 10.1158/2326-6066.CIR-17-0201
40. Aran D, Sirota M, Butte A. Systematic Pan-Cancer Analysis of Tumour Purity. *Nat Commun* (2015) 6:8971. doi: 10.1038/ncomms9971
41. Icard P, Fournel L, Wu Z, Alifano M, Lincet H. Interconnection Between Metabolism and Cell Cycle in Cancer. *Trends Biochem Sci* (2019) 44(6):490–501. doi: 10.1016/j.tibs.2018.12.007
42. Beishline K, Azizkhan-Clifford J. Sp1 and the 'Hallmarks of Cancer'. *FEBS J* (2015) 282:224–58. doi: 10.1111/febs.13148
43. Richard C, Fumet J, Chevrier S, Derangère V, Ledys F, Lagrange A, et al. Exome Analysis Reveals Genomic Markers Associated With Better Efficacy of Nivolumab in Lung Cancer Patients. *Clin Cancer Res* (2019) 25(3):957–66. doi: 10.1158/1078-0432.CCR-18-1940
44. Reck M, Rodríguez-Abreu D, Robinson A, Hui R, Csőszi T, Fülöp A, et al. Pembrolizumab Versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. *N Engl J Med* (2016) 375(19):1823–33. doi: 10.1056/NEJMoa1606774
45. Mok T, Wu Y, Kudaba I, Kowalski D, Cho B, Turna H, et al. Pembrolizumab Versus Chemotherapy for Previously Untreated, PD-L1-Expressing, Locally Advanced or Metastatic non-Small-Cell Lung Cancer (KEYNOTE-042): A Randomised, Open-Label, Controlled, Phase 3 Trial. *Lancet (London England)* (2019) 393(10183):1819–30.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 31 May 2022
doi: 10.3389/fonc.2022.855305

A Germline Mutation in ATR Is Associated With Lung Adenocarcinoma in Asian Patients

Guangyao Bao[1†], Xiaojiao Guan[2†], Jie Liang[1], Yao Yao[1], Yifan Xiang[1], Tian Li[3] and Xinwen Zhong[1*]*

[1] Department of Thoracic Surgery, First Affiliated Hospital, China Medical University, Shenyang, China, [2] Department of Pathology, Shengjing Hospital, China Medical University, Shenyang, China, [3] School of Basic Medicine, Fourth Military Medical University, Xi'an, China

OPEN ACCESS

Edited by:
Yutong He,
Fourth Hospital of Hebei Medical University, China

Reviewed by:
Ling Zhao,
Harbin Medical University Cancer Hospital, China
Wenguo Jiang,
Binzhou Medical University, China

***Correspondence:**
Xinwen Zhong
xwzhong@cmu.edu.cn
Tian Li
fmmult@foxmail.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: 15 January 2022
Accepted: 26 April 2022
Published: 31 May 2022

Citation:
Bao G, Guan X, Liang J, Yao Y, Xiang Y, Li T and Zhong X (2022) A Germline Mutation in ATR Is Associated With Lung Adenocarcinoma in Asian Patients. Front. Oncol. 12:855305. doi: 10.3389/fonc.2022.855305

Background: Familial lung cancer (FLC) accounts for 8% of lung adenocarcinoma. It is known that a few germline mutations are associated with risk increasing and may provide new screening and treatment option. The goal of this study is to identify an FLC gene among three members of an FLC family.

Methods: To uncover somatic and embryonic mutations linked with familial lung cancer, whole exome sequencing was done on surgical tissues and peripheral blood from three sisters in a family diagnosed with pulmonary lung adenocarcinoma (LUAD). At the same time, single-cell RNA sequencing (scRNA-seq) and bulk RNA sequencing data in public databases were enrolled to identify specific gene expression level.

Results: Ataxia Telangiectasia and Rad3-Related Protein (ATR) gene C.7667C >G (p.T2556S) mutation were found in 3 patients with familial lung cancer. Whole-genome sequencing revealed that the three sisters exhibited similar somatic mutation patterns. Besides ATR mutations, common mutated genes (BRCA1, EGFR, and ROS1) that characterize LUAD were also found in 5 tumor samples. Analysis for the ATR expression in LUAD patients by single-cell sequencing data, we found ATR expression of tumor patients at high level in immune cells when compared with normal patients, but the expression of ATR in stromal cells has the opposite result.

Conclusion: We found a germline mutation in the ATR gene in three sisters of a Chinese family affected by familial lung cancer, which may be a genetic factor for lung cancer susceptibility.

Keywords: familiar lung cancer, ATR, sequencing, germline mutation, lung adenocarcinoma

INTRODUCTION

Lung adenocarcinoma is the primary cause of cancer-related death worldwide, with an estimated 1.8 million fatalities each year (1). Researchers are still considering possible links involving lung cancer risk. Although smoking is still the leading cause of lung cancer, accounting for 80 to 90 percent of all cases, genetic susceptibility may play a role in lung cancer in some situations (2). Previous studies

have demonstrated that genetic susceptibility was thought to be closely related to the development of familial lung cancer (3, 4). However, the risk factors for hereditary lung cancer are unknown at this time. Several research has looked into the function of germline mutations, primarily selected somatic mutations, in lung cancer susceptibility (5–7).

Ataxia Telangiectasia and Rad3 related (ATR) kinase plays a crucial role in the repair of replication-associated DNA damage (8). The study of cell cycle checkpoint signaling through ATR, as well as the related pathways implicated in oncogenesis and cancer progression, has resulted in the discovery and development of effective and selective ATR inhibitors (ATRi) (9). According to recent research of inherited germline mutations in a Chinese population with lung cancer assessed using an NGS, the majority of the discovered germline mutations (85.5 percent) were implicated in DNA damage repair (DDR) pathways (10). There have previously been reports of rare events in lung cancer patients, which may be related to the genetic susceptibility to lung cancer (11). As a high-throughput method, next-generation sequencing (NGS) has become routine in clinical practice and has altered the clinical management of lung cancer (12, 13). This method enables for massively parallel characterization of thousands of cells at the transcriptome level. In this study, we identified an interesting ATR mutation by whole-exome sequencing in three sisters of this family with lung cancers.

MATERIAL AND METHODS

Patients Data

The three patients were sisters and were treated successively in the Thoracic Surgery Department of the First Hospital of China Medical University due to lung nodules. We have drawn the pedigree of the lung adenocarcinoma family (**Figure 1**). The first woman was a 61-year-old female with 40 years of smoking history (II -2), who was diagnosed with micro-invasive adenocarcinoma and invasive adenocarcinoma of two nodules in the upper right lobe (**Figures 2A, B**). Another 56-year-old female, non-smoker, with 2 nodules in the upper right lobe (II -4), was pathologically confirmed to be adenocarcinoma *in situ* and invasive adenocarcinoma after surgery (**Figures 2C, D**). Invasive adenocarcinoma of the right lower lobe (II-5) was diagnosed in a 47-year-old non-smoker sister of the family (**Figure 2E**). The other members of this family have no history of lung tumor. This study was approved by the ethics committees of the First Hospital of China Medical University. The patient signed written informed consent, and we kept the patient's identity protection for the duration of the study.

Whole-Exome Sequencing (WES)

In total, peripheral blood, cancer, para-cancerous samples, and clinical data of three patients were collected for whole-exome sequencing (WES) in this study. NEBNext dsDNA Fragmentase (NEB, Ipswich, MA, USA) was used to extract and fragment genomic DNA from blood and cancer or para-cancer (normal tissues adjacent to cancer) tissues, followed by DNA end mending. After being detailed, end-repaired DNA segments were ligated with the NEBNext adaptor (NEB, Ipswich, MA, USA). Biotinylated RNA library baits and magnetic beads were coupled with the barcoded library to detect particular areas using the SureSelect Human All Exon V6 Kit (Agilent Technologies, Palo Alto, Calif.). On an Illumina X-ten system, the acquired sequences were amplified further for 150bp paired-end sequencing (Illumina, San Diego, CA, USA). Readings of high quality that passed the Illumina filter were kept for further processing.

Mutation Analysis

Using fastp, raw reads would be processed to make high-quality clean reads (14). To match the clean reads from each sample to the reference genome, the Burrows-Wheeler Aligner (BWA) (15) was employed (GRCh38.p12). The Genome Analysis Toolkit (GATK) (16) was used to call variants for multi-sample analyses. For multi-sample variant calling, local realignment, and base quality score recalibration, Unified Genotyper was employed. To

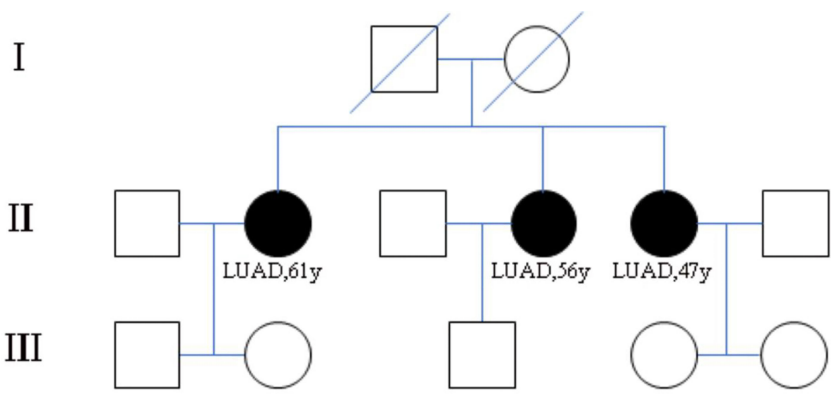


FIGURE 1 | Pedigree of family (case 3) with multiple cases of lung adenocarcinoma. II-2, II-4, II-5 are sisters and have been diagnosed as lung adenocarcinoma. I-1 and I-2 are their parents and have passed away. The other members of this family have no history of lung tumor.

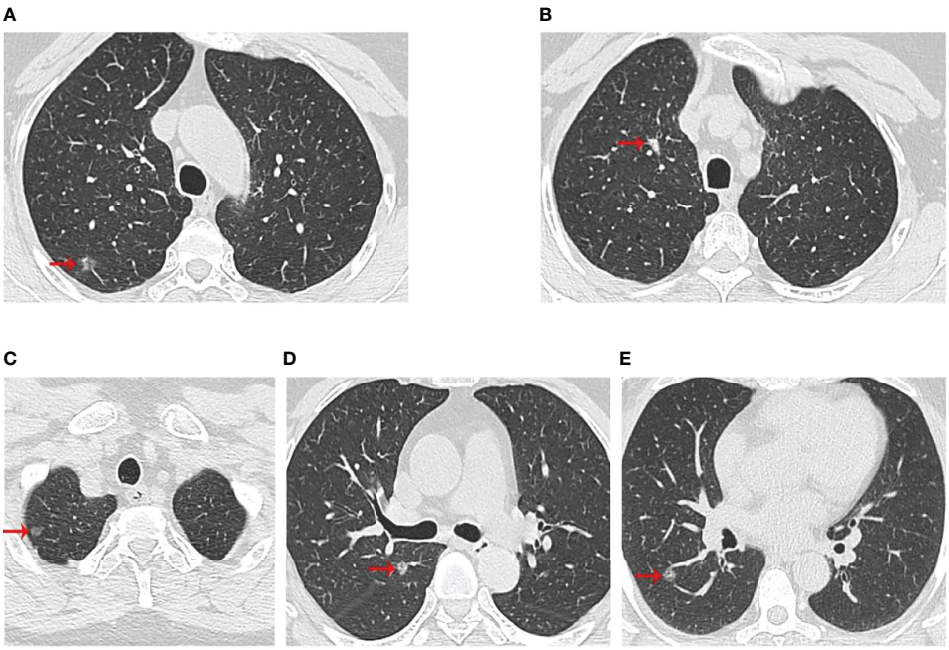


FIGURE 2 | Computed tomographic scan of chest at diagnosis of three patients. **(A, B)** were II-2 CT scan results, which showed two nodules in the upper right lobe; **(C, D)** were II-4 CT scan results, which showed two nodules in the upper lobe of the right lung; **(E)** was II-5 CT scan results with a right lower lobe nodule. The red arrows indicated the location of the lesions.

assess the frequency of each SNP, the software package ANNOVAR (17) was used to align and annotate SNPs or InDels.Variants related to diseases were annotated with Clinvar database (http://www.clinvar.com/).

Bioinformatic Analysis

Gene expression profiles of 483 tumor samples and 347 normal samples in LUAD were conducted by GEPIA database (18). RNA-seq data and its relevant clinical information and mutation data of 561 LUAD samples were received from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/). Survival analysis was conducted by "survival" R package. The "maftools" R package was used to find high-frequency mutation genes. The GEO database (http://www.ncbi.nlm.nih.gov/geo/) was used to download raw UMI counts per gene, as well as the sample and cluster annotations of 11 human LUAD and 11 human normal lung samples from Kim et al's research (19). The raw counts were normalized using the "scran" R program, and then cluster annotation was performed using the "SingleR" R tool.

RESULTS

Molecular Characterization of Genetic ATR Mutations

We performed whole-exome sequencing on surgical samples and peripheral blood to discover variants associated with familial lung cancer. Then, we applied the following filters for the variants called by whole exome sequencing to identify the most significant germline mutations: (1) all affected family members must have mutations; (2) variations in the SNP databases with a minor allele frequency (MAF) greater than 0.5 percent were eliminated from further research, cause they are more likely to be functional neutral polymorphisms (the SNP databases that we used for filtering are Clinvar from Complete Genomics); (3) the variations must be missense or nonsense mutations or indels that cause in alterations or a shortened protein.; (4) bioinformatics algorithms such as PolyPhen-2 (20), SIFT (21), or MutationTaster must forecast that these mutations are functionally "non-benign" or "intolerable" and are linked to cancer (22). We summarized and annotated the mutation data of three patients (**Table 1**), and seven candidate variants were discovered in three patients of this family, only the ATR c.7667C>G (p.T2556S) mutation is the most important germline mutation in this family.

Analysis Somatic Mutations in FLC

Then, the results of somatic mutations in five tumor lesions of three patients were thoroughly studied. The somatic SNV/INDEL mutation spectrum and CNV mutation spectrum of the three lung cancer patients in this investigation were mapped (**Supplementary Figure S1**). Moreover, we searched the reported driver gene data to screen out the known driver genes in the tumor sample based on Cancer Gene Census (http://cancer.sanger.ac.uk/census), MDG125 (23), SMG127 (24), CDG291 (25) database and we noticed that somatic mutations frequently found in this family, such as EGFR, BRCA1 and ROS1

TABLE 1 | Characteristics of the seven candidate variants for familial lung cancer.

Gene	Genomic Position	Genomic Mutation	dbSNP	Protein Alteration	Function	ClinVar Assessment
ATR	Chr3: 142453222	c.7667C>G	rs200490116	p.Thr2556Ser	Missense	Uncertain significance
SMG5	Chr1: 156263493	c.1933A>G	rs200093957	p.Lys645Glu	Missense	Not listed
RAB3GAP2	Chr1: 220164744	c.3143A>G	rs151244742	p.His1048Arg	Missense	Likely benign
EI24	Chr11: 125575342	c.122G>A	rs559933286	p.Arg41His	Missense	Not listed
XDH	Chr2: 31350061	:c.2794G>A	rs141291583	p.Ala932Thr	Missense	Not listed
PTPRA	Chr20: 3035681	c.2017G>A	rs61742029	p.Val673Ile	Missense	Not listed
GSTZ1	Chr14:77326864	c.94G>A	rs7975	p.Glu32Lys	Missense	Stop Gained

** indicates nucleotide number and translation termination (stop) codon.*

were also observed (**Table 2**). Furthermore, new candidate driver genes of lung adenocarcinoma, like PNCK and IP6K2 could also be found in this family (**Table 3**). Finally, we used MuSiC software (26) to find genes with higher mutation frequencies of the three patients, we found the top thirty high-frequency mutation genes (**Figure 3A**) and we summarized the mutation information of these genes in the TCGA mutation data, and we found that AMT, BICRA, ARMCX4 are high-frequency mutation genes unique to these three patients (**Figure 3B**).

ATR Expression in LUAD Specimen

ATR is required for maintaining genome integrity during DNA replication and plays a key role in avoiding replication stress at toxic levels, according to previous research. Because of these critical functions, cancers infrequently involve loss-of-function mutations in the ATR pathway, and a subset of tumors with particular mutations is more vulnerable to ATR pathway inhibition than normal cells (27). The underlying mechanism of ATR expression in lung adenocarcinoma requires further study, hence we analyzed ATR related expression and prognosis in cases of lung adenocarcinoma. As demonstrated in **Figure 4**, the expression of ATR was downregulated in LUAD patients as compared to normal individuals, and the downregulation of ATR expression was related to poor prognosis. Moreover, we used single-cell sequencing data to further explore the expression of ATR in designated cell types of lung adenocarcinoma and found the expression of ATR in stromal cells (Fibroblasts, Epithelial cells, Endothelial cells) of tumor patients is at a lower level when compared with normal patients, but the expression of ATR in the immune cells (Myeloid cells, B lymphocytes, MAST cells) of tumor patients is at a high level, except for T/NK cells (**Figure 5**).

Analysis of Common Germline Genes in Chinese LUAD Patients

Next, we analyzed the whole-exome sequencing data of 2706 patients to uncover the germline gene mutations in the Chinese LUAD cohort. We then screened out 27 germline gene mutations and listed their mutation frequencies and mutation sites (**Table 4**). Among the relatively common mutant genes, including MLH1 (1.03474%), BRCA (0.77605%), STK11 (0.62823%), CHEK2 (0.44346%), TP53 (0.44346%), have been reported related to lung cancer. Meanwhile, multiple mutation sites of these germline gene mutations were also revealed. For example, we have detected 3 different mutation sites in the germline mutation of MLH1, including c.649C>T (p.R217C), p.Q701K (c. C2101A), and p.V384D (c.T1151A).

DISCUSSION

Carcinogenesis is a multi-factor and multi-stage process and gene mutation is an important cause of cancer occurrence and development (28). Germline mutations including somatic mutation and germline mutations. Germline mutations which

TABLE 2 | Known somatic mutation patterns in three patients.

Gene	Chr	Start	End	Variation Type	Ref	Mut	Patients
BRCA1	17	43104073	43104089	Deletion	AAAAAAAAGAAAAGAAG	–	II-4 Tumor B
EGFR	7	55173052	55173052	SNP	G	T	II-2 Tumor A
EGFR	7	55202802	55202802	SNP	G	A	II-2 Tumor B
EGFR	7	55174774	55174788	Deletion	AATTAAGAGAAGCAA	–	II-4 Tumor A
EGFR	7	55174773	55174787	Deletion	GAATTAAGAGAAGCA	–	II-4 Tumor A
ROS1	6	117356812	117356812	SNP	G	T	II-2 Tumor A
ROS1	6	117341369	117341369	SNP	C	A	II-5

TABLE 3 | Predict Driver Genes in Three FLC Patients.

PNCK	IP6K2	CLEC18C	PARD3	TTN	CMTM3	TMPRSS6
PCBP4	CNOT10	RAB43	DZANK1	C1orf43	KIF9	INPP5B
SUOX	ASIC3	PLCD4	NPAS3	SPTBN2	CDC25B	ARHGAP33
SLC6A2	LTF	SLC26A11	DNAH10	MUC17	AP4M1	GLG1
GSN	AMT	CHRNA2	RUBCNL	KCNH6	ANK2	CYTH2
NAV2	ADORA1	MYO7A	ARHGAP25	LTBP3	MASP1	PRSS54
AURKB	KLHDC2	WIPF1	FBLN7	TRAIP	PHF12	MVK
CRY2	ITGA7	CLCN2	ARHGEF3	KLHL13	EHMT1	PTK7
SPHK2	SLC4A11	SLFN13	CCDC120	TIMMDC1	LOXL3	SYNGAP1
HLA-G	ARHGAP27	SYNE2	SORL1	RPH3A	EYA2	RAB5C
LIMK1	EPN2	SEC16A	FARSA	TMEM267	COL6A3	IPO5
ZSWIM8	CNGB1	GRIN1	CAPN1	WSCD1	PCGF2	TM7SF2
LMCD1	RBBP7	SEMA4A	PTPRA	KDM4C	ZDHHC8	FDXR
ZAN	RASA4	PSMD13	HOMER3	TPI1	HLA-C	SKIL
ATP2B2	PRPF31	DLGAP4	SLC8B1	TPM2	TADA2B	STARD3NL
ABCC12	TRIM2	RIC8B	GABRB2	CXCR2	GRB10	TBC1D16
MAP3K19	TAF1C	SORBS1	DTNB	CACNA1A	POLDIP3	CES2
OBSCN	EPS8L2	SYVN1	BSDC1	RNF32	STK25	KCNT1
ATXN2L	TLE2	MVB12A				

occur in the reproductive cells, it can be passed on to the next generation, affecting every cell in the next generation. However, most studies on tumor mutations have focused on somatic mutations. The mining and identification of tumor germline mutations are very limited, with only a few cases reported in lung adenocarcinoma. As a result, the identification of novel germline mutations is critical for both fundamental research and clinical cancer treatment (29).

In previous studies, most of the genes with germline mutations in lung adenocarcinoma were typical oncogenes and proto-oncogenes, which included BRCA2, CHECK2, CDKN2, BAP1, EGFR (30–32). The genomic alternations of these genes were thought to be important in familial lung cancer. Then, with the development of high-throughput sequencing data, germline altered variants in other genes (MAST1, CENPE, LCT) (33–35) were identified to exhibit an essential role in the development of lung adenocarcinoma, and they were suspected to be related to the development of lung cancer through whole-exome or genome sequencing of samples from many lung cancer patients.

In this study, we continue to focus on the germline mutation of genes in lung adenocarcinoma and we have identified seven different variants (SMG5, RAB3GAF2, EI24, XDH, PTPRA, ATR, GSTZ1) in three sibling patients suffering from lung cancer. The most significant is a rare somatic mutation in the ATR gene named c.7667C>G (p.T2556S), and this ATR gene variation has never been observed in genome and exome research within combined populations. Then, we analyzed the expression and prognosis of ATR in lung adenocarcinoma. Compared with normal patients, the expression of ATR was down-regulated in LUAD, and the down-regulated expression of ATR was associated with poor prognosis. Furthermore, we employed single-cell sequencing data to investigate the expression of ATR in lung adenocarcinoma cell types. Compared with normal patients, we found that the expression level of ATR in the stromal cells (fibroblasts, epithelial cells, and

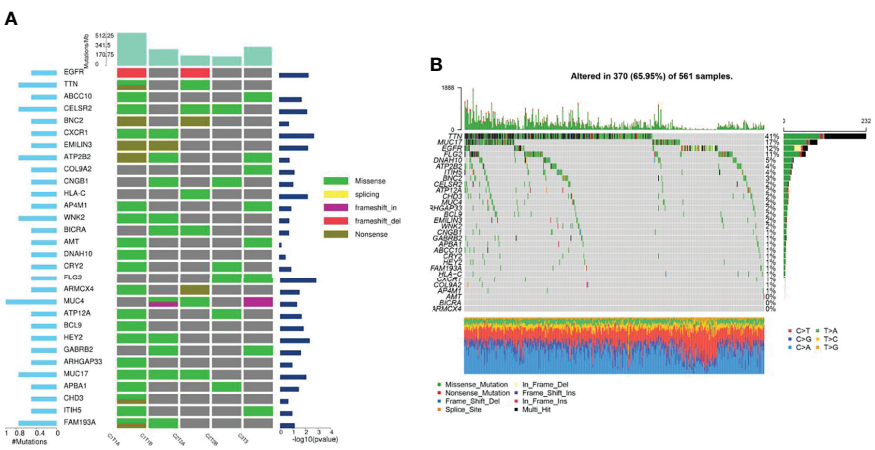


FIGURE 3 | Gene mutation characteristics of three patients and in TCGA database. **(A)** High-frequency mutation genes of three lung adenocarcinoma patients; **(B)** Mutation spectrum of high-frequency mutation genes in TCGA mutation data.

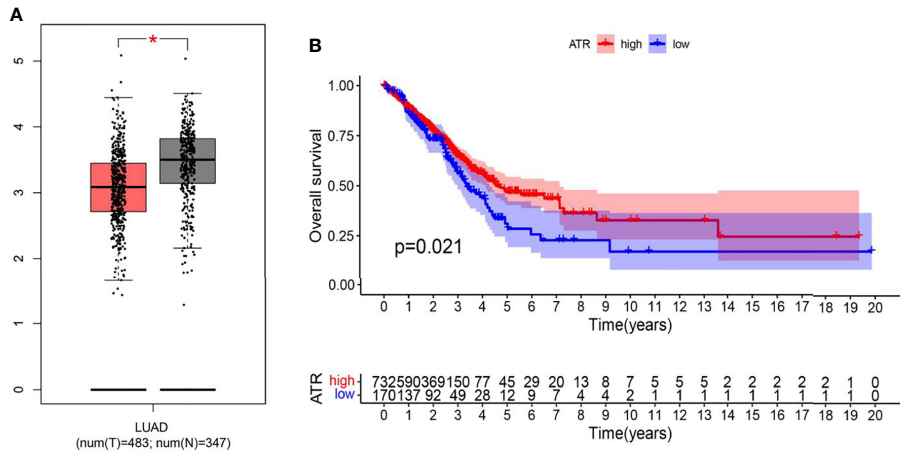


FIGURE 4 | The expression and prognostic differences of ATR in TCGA database. **(A)** The expression of ATR in LUAD and normal tissues in GEPIA database (Tumor: 483; Normal: 347) (* indicates $p < 0.05$); **(B)** Down regulation of ATR is associated with harmful outcome in TCGA database (Tumor: 535; Normal: 59) ($p = 0.021$).

endothelial cells) of tumor patients was low. In addition to T/NK cells, the expression level of ATR in immune cells (myeloid cells, B lymphocytes, MAST cells) of tumor patients is higher. In view of this, we have reason to believe that germline ATR c.7667C>G (p.T2556S) mutation may lead to familial clusters of lung adenocarcinoma and differentially expressed ATR may play an essential role in the occurrence and development of lung adenocarcinoma. Moreover, based on somatic mutation data, we found common driver gene mutations such as EGFR, BRCA1, ROS1, etc. We predicted 136 new driver gene mutations such as PNCK, IP6K2, CLEC18C, etc. through OncodriveCLUST software (36). These gene mutations may be secondary somatic mutations based on germline ATR c.7667C>G (p.T2556S) mutations, which eventually lead to the occurrence and progression of lung adenocarcinoma.

ATR, as a necessary gene in the DNA damage repair pathway, is involved in the coordination of cell-cycle transitions, DNA replication, DNA repair, and apoptosis (37). Previous studies have shown that germline gene mutations that occur in the DNA damage repair pathway often irreversibly lead to tumors (38, 39). Furthermore, germline abnormalities in DNA repair genes lead to advanced solid tumor (such as prostate, ovarian, pancreatic, and breast cancer) susceptibility to poly ADP ribose polymerase (PARP) inhibitors (40–42). To our knowledge, this is the first such case report confirming the presence of c.7667C>G (p.T2556S), a germline ATR mutation, in a family with three occurrences of lung adenocarcinoma The resulting amino acid mutation, p.Thr2556Ser, was detected in the ATR's conserved Phosphoinositide 3- and 4-kinase domain, which is involved in biological processes such as cell growth, proliferation, differentiation, motility, survival, and intracellular trafficking (43).

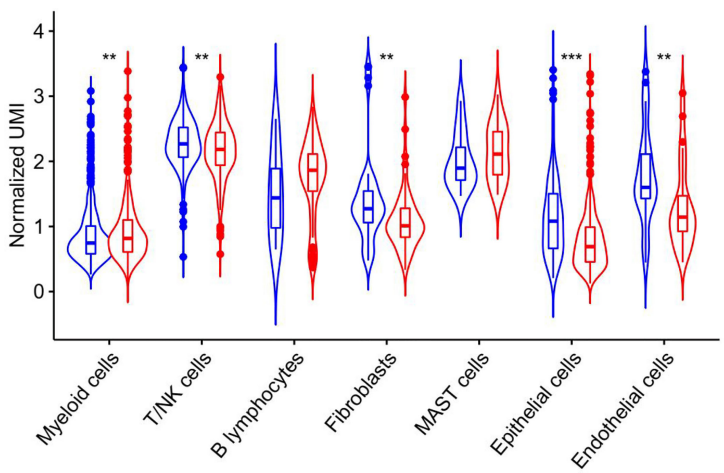


FIGURE 5 | ATR expression in unique molecular identifiers in the indicated cell types from single-cell RNA-Seq data derived from human LUAD specimen. Blue represents the expression of ATR in each component of normal tissue, and red represents the expression of ATR in each component of tumor tissue.

TABLE 4 | Common germline gene mutations of 2706 Chinese LUAD patients.

Germline Mutation Gene	Mutation Frequency	Mutation Sites
MLH1	1.03474%	c.649C>T(p.R217C), p.Q701K (c.C2101A), p.V384D (c.T1151A)
BRCA2	0.77605%	c.2830A>T(p.K944*), c.671-1G>A, c.956dupA(p.N319Kfs*8)
STK11	0.62823%	p.F354L (c.C1062G)
CHEK2	0.44346%	p.H371Y (c.C1111T), c.1245dup(p.K416Qfs*22)
TP53	0.44346%	c.91G>A(p.V31I), p.V31I (c.G91A), c.7516-1G>A
ATM	0.29564%	c.3602_3603delTT(p.F1201Wfs*3), p.S1042R (c.A3124C)
BRCA1	0.25868%	c.671-1G>A
CDH1	0.25868%	p.T340A (c.A1018G)
MSH6	0.25868%	c.1406A>G(p.Y469C)
MUTYH	0.18477%	c.55C>T(p.R19*)
PALB2	0.18477%	c.2480_2481del(p.T827Mfs*6)
PMS2	0.18477%	c.1A>G(p.M1)?
RAD51D	0.18477%	c.270_271dup(p.K91Ifs*13)
ATR	0.11086%	c.4681C>T(p.Q1561*)
BRIP1	0.11086%	c.918+1G>A
MSH2	0.11086%	p.L390F (c.C1168T)
NBN	0.11086%	c.1651dupA(p.R551Kfs*5)
EPCAM	0.07391%	c.556-14A>G
FANCA	0.07391%	c.1900+1G>T
POLE	0.07391%	c.G4952+1T
RET	0.07391%	p.V804M (c.G2410A)
SDHA	0.07391%	c.1A>T(p.M1)?
VHL	0.07391%	p.W8X (c.G23A)
APC	0.03695%	c.3374T>C(p.V1125A)
FANCI	0.03695%	c.504-1G>A
MEN1	0.03695%	c.1A>G(p.M1)?
SDHB	0.03695%	c.200+1G>C

** indicates nucleotide number and translation termination (stop) codon.*

The population frequency in the entire population is 0.0399361 percent, suggesting that this ATR germline mutation may be a genetic susceptibility factor for lung adenocarcinoma. Our findings contribute to the understanding of the genesis and mechanism of lung cancer and may give clues for oncology treatment strategies and cancer prevention.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: NCBI [accession: PRJNA800743].

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by ethics committees of the First Hospital of China Medical University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

GB conceived and designed the study, and drafted the manuscript. XG, YX, and YY collected, analyzed and interpreted the experimental data. JL, TL, and XZ revised the manuscript for important intellectual content. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by Wu Jieping Medical Foundation (320.6750.2020-17-7).

ACKNOWLEDGMENTS

We are grateful to Nanjing Shihe Gene Bio-Technology Co., Ltd. for providing the whole exome sequencing data of 2706 Chinese LUAD patients. We thanked Home for Researcher for language editing service.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.855305/full#supplementary-material

Supplementary Figure 1 | Somatic mutation spectrum of five lesions in three patients. **(A, B)** Somatic SNV/INDEL mutation spectrum; **(C)** Somatic CNV mutation spectrum. C1T1A: II-2 tumor A; C1T2A: II-2 tumor B; C2T2A: II-4 tumor A; C2T2B: II-4 tumor B; C3T3: II-5.

REFERENCES

1. Sung H, Ferlay J, Siegel R, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: Cancer J Clin* (2021) 71:209–49. doi: 10.3322/caac.21660
2. Alberg A, Brock M, Ford J, Samet J, Spivack S. Epidemiology of Lung Cancer: Diagnosis and Management of Lung Cancer, 3rd Ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* (2013) 143:e1S–e29S. doi: 10.1378/chest.12-2345
3. Bailey-Wilson J, Amos C, Pinney S, Petersen G, De Andrade M, Wiest J, et al. A Major Lung Cancer Susceptibility Locus Maps to Chromosome 6q23-25. *Am J Hum Genet* (2004) 75:460–74. doi: 10.1086/423857
4. Schwartz A, Ruckdeschel J. Familial Lung Cancer: Genetic Susceptibility and Relationship to Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* (2006) 173:16–22. doi: 10.1164/rccm.200502-235PP
5. Tibaldi C, Giovannetti E, Vasile E, Boldrini L, Gallegos-Ruiz M, Bernardini I, et al. Inherited Germline T790M Mutation and Somatic Epidermal Growth Factor Receptor Mutations in non-Small Cell Lung Cancer Patients. *J Thorac Oncol* (2011) 6:395–6. doi: 10.1097/JTO.0b013e3182059a6f
6. Oxnard G, Nguyen K, Costa D. Germline Mutations in Driver Oncogenes and Inherited Lung Cancer Risk Independent of Smoking History. *J Natl Cancer Institute* (2014) 106:djt361. doi: 10.1093/jnci/djt361
7. Lu S, Yu Y, Li Z, Yu R, Wu X, Bao H, et al. EGFR and ERBB2 Germline Mutations in Chinese Lung Cancer Patients and Their Roles in Genetic Susceptibility to Cancer. *J Thorac Oncol* (2019) 14:732–6. doi: 10.1016/j.jtho.2018.12.006
8. Jackson S, Bartek J. The DNA-damage Response in Human Biology and Disease. *Nature* (2009) 461:1071–8. doi: 10.1038/nature08467
9. Bradbury A, Hall S, Curtin N, Drew Y. Targeting ATR as Cancer Therapy: A New Era for Synthetic Lethality and Synergistic Combinations? *Pharmacol Ther* (2020) 207:107450. doi: 10.1016/j.pharmthera.2019.107450
10. Tian P, Cheng X, Zhao Z, Zhang Y, Bao C, Wang Y, et al. Spectrum of Pathogenic Germline Mutations in Chinese Lung Cancer Patients Through Next-Generation Sequencing. *Pathol Oncol Res POR* (2020) 26:109–14. doi: 10.1007/s12253-019-00771-5
11. Zhang Y, Zhang L, Li R, Chang D, Ye Y, Minna J, et al. Genetic Variations in Cancer-Related Significantly Mutated Genes and Lung Cancer Susceptibility. *Ann Oncol* (2017) 28:1625–30. doi: 10.1093/annonc/mdx161
12. Puram S, Tirosh I, Parikh A, Patel A, Yizhak K, Gillespie S, et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* (2017) 171:1611–1624.e1624. doi: 10.1016/j.cell.2017.10.044
13. Azizi E, Carr A, Plitas G, Cornish A, Konopacki C, Prabhakaran S, et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell* (2018) 174:1293–1308.e1236. doi: 10.1016/j.cell.2018.05.060
14. Chen S, Zhou Y, Chen Y, Gu J. Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor. *Bioinf (Oxford England)* (2018) 34:i884–90. doi: 10.1093/bioinformatics/bty560
15. Li H, Durbin R. Fast and Accurate Short Read Alignment With Burrows-Wheeler Transform. *Bioinf (Oxford England)* (2009) 25:1754–60. doi: 10.1093/bioinformatics/btp324
16. Mckenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res* (2010) 20:1297–303. doi: 10.1101/gr.107524.110
17. Wang K, Li M, Hakonarson H. ANNOVAR: Functional Annotation of Genetic Variants From High-Throughput Sequencing Data. *Nucleic Acids Res* (2010) 38:e164. doi: 10.1093/nar/gkq603
18. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: A Web Server for Cancer and Normal Gene Expression Profiling and Interactive Analyses. *Nucleic Acids Res* (2017) 45(1):W98–102. doi: 10.1093/nar/gkx247
19. Kim N, Kim H, Lee K, Hong Y, Cho J, Choi J, et al. Single-Cell RNA Sequencing Demonstrates the Molecular and Cellular Reprogramming of Metastatic Lung Adenocarcinoma. *Nat Commun* (2020) 11:2285. doi: 10.1038/s41467-020-16164-1
20. Adzhubei I, Schmidt S, Peshkin L, Ramensky V, Gerasimova A, Bork P, et al. A Method and Server for Predicting Damaging Missense Mutations. *Nat Methods* (2010) 7:248–9. doi: 10.1038/nmeth0410-248
21. Ng P, Henikoff S. Sift: Predicting Amino Acid Changes That Affect Protein Function. *Nucleic Acids Res* (2003) 31:3812–4. doi: 10.1093/nar/gkg509
22. Schwarz J, Rödelsperger C, Schuelke M, Seelow D. MutationTaster Evaluates Disease-Causing Potential of Sequence Alterations. *Nat Methods* (2010) 7:575–6. doi: 10.1038/nmeth0810-575
23. Vogelstein B, Papadopoulos N, Velculescu V, Zhou S, Diaz L, Kinzler K. Cancer Genome Landscapes. *Sci (N Y NY)* (2013) 339:1546–58. doi: 10.1126/science.1235122
24. Kandoth C, Mclellan M, Vandin F, Ye K, Niu B, Lu C, et al. Mutational Landscape and Significance Across 12 Major Cancer Types. *Nature* (2013) 502:333–9. doi: 10.1038/nature12634
25. Tamborero D, Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Kandoth C, Reimand J, et al. Comprehensive Identification of Mutational Cancer Driver Genes Across 12 Tumor Types. *Sci Rep* (2013) 3:2650. doi: 10.1038/srep02650
26. Dees N, Zhang Q, Kandoth C, Wendl M, Schierding W, Koboldt D, et al. MuSiC: Identifying Mutational Significance in Cancer Genomes. *Genome Res* (2012) 22:1589–98. doi: 10.1101/gr.134635.111
27. Karnitz L, Zou L. Molecular Pathways: Targeting ATR in Cancer Therapy. *Clin Cancer Res Off J Am Assoc Cancer Res* (2015) 21:4780–5. doi: 10.1158/1078-0432.CCR-15-0479
28. Curtius K, Wright N, Graham T. An Evolutionary Perspective on Field Cancerization. *Nat Rev Cancer* (2018) 18:19–32. doi: 10.1038/nrc.2017.102
29. Przytycki P, Singh M. Differential Analysis Between Somatic Mutation and Germline Variation Profiles Reveals Cancer-Related Genes. *Genome Med* (2017) 9:79. doi: 10.1186/s13073-017-0465-6
30. Abdel-Rahman M, Pilarski R, Cebulla C, Massengill J, Christopher B, Boru G, et al. Germline BAP1 Mutation Predisposes to Uveal Melanoma, Lung Adenocarcinoma, Meningioma, and Other Cancers. *J Med Genet* (2011) 48:856–9. doi: 10.1136/jmedgenet-2011-100156
31. Thomas A, Xi L, Carter C, Rajan A, Khozin S, Szabo E, et al. Concurrent Molecular Alterations in Tumors With Germ Line Epidermal Growth Factor Receptor T790M Mutations. *Clin Lung Cancer* (2013) 14:452–6. doi: 10.1016/j.cllc.2013.01.005
32. Wang Y, Mckay J, Rafnar T, Wang Z, Timofeeva M, Broderick P, et al. Rare Variants of Large Effect in BRCA2 and CHEK2 Affect Risk of Lung Cancer. *Nat Genet* (2014) 46:736–41. doi: 10.1038/ng.3002
33. Renieri A, Mencarelli M, Cetta F, Baldassarri M, Mari F, Furini S, et al. Oligogenic Germline Mutations Identified in Early non-Smokers Lung Adenocarcinoma Patients. *Lung Cancer (Amsterdam Netherlands)* (2014) 85:168–74. doi: 10.1016/j.lungcan.2014.05.020
34. Clamon G, Bossler A, Abu Hejleh T, Furqan M. Germline Mutations Predisposing to non-Small Cell Lung Cancer. *Familial Cancer* (2015) 14:463–9. doi: 10.1007/s10689-015-9796-x
35. Tomoshige K, Matsumoto K, Tsuchiya T, Oikawa M, Miyazaki T, Yamasaki N, et al. Germline Mutations Causing Familial Lung Cancer. *J Hum Genet* (2015) 60:597–603. doi: 10.1038/jhg.2015.75
36. Tamborero D, Gonzalez-Perez A, Lopez-Bigas N. OncodriveCLUST: Exploiting the Positional Clustering of Somatic Mutations to Identify Cancer Genes. *Bioinf (Oxford England)* (2013) 29:2238–44. doi: 10.1093/bioinformatics/btt395
37. Cimprich K, Cortez D. ATR: An Essential Regulator of Genome Integrity. *Nat Rev Mol Cell Biol* (2008) 9:616–27. doi: 10.1038/nrm2450
38. Parry E, Gable D, Stanley S, Khalil S, Antonescu V, Florea L, et al. Germline Mutations in DNA Repair Genes in Lung Adenocarcinoma. *J Thorac Oncol* (2017) 12:1673–8. doi: 10.1016/j.jtho.2017.08.011
39. Wei Y, Wu J, Gu W, Qin X, Dai B, Lin G, et al. Germline DNA Repair Gene Mutation Landscape in Chinese Prostate Cancer Patients. *Eur Urol* (2019) 76:280–3. doi: 10.1016/j.eururo.2019.06.004
40. Fong P, Boss D, Yap T, Tutt A, Wu P, Mergui-Roelvink M, et al. Inhibition of Poly(ADP-Ribose) Polymerase in Tumors From BRCA Mutation Carriers. *New Engl J Med* (2009) 361:123–34. doi: 10.1056/NEJMoa0900212

41. Kaufman B, Shapira-Frommer R, Schmutzler R, Audeh M, Friedlander M, Balmaña J, et al. Olaparib Monotherapy in Patients With Advanced Cancer and a Germline BRCA1/2 Mutation. *J Clin Oncol* (2015) 33:244–50. doi: 10.1200/JCO.2014.56.2728
42. Mateo J, Carreira S, Sandhu S, Miranda S, Mossop H, Perez-Lopez R, et al. Dna-Repair Defects and Olaparib in Metastatic Prostate Cancer. *New Engl J Med* (2015) 373:1697–708. doi: 10.1056/NEJMoa1506859
43. Lecona E, Fernandez-Capetillo O. Targeting ATR in Cancer. *Nat Rev Cancer* (2018) 18:586–95. doi: 10.1038/s41568-018-0034-3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 27 June 2022
doi: 10.3389/fonc.2022.880478

Development of a Prognostic Alternative Splicing Signature Associated With Tumor Microenvironment Immune Profiles in Lung Adenocarcinoma

Guangyao Bao[1], Tian Li[2], Xiaojiao Guan[3], Yao Yao[1], Jie Liang[1], Yifan Xiang[1] and Xinwen Zhong[1]**

[1] Department of Thoracic Surgery, First Affiliated Hospital, China Medical University, Shenyang, China, [2] School of Basic Medicine, Fourth Military Medical University, Xi'an, China, [3] Department of Pathology, Shengjing Hospital, China Medical University, Shenyang, China

OPEN ACCESS

Edited by:
Yutong He,
Fourth Hospital of Hebei Medical University, China

Reviewed by:
Wei He,
University of Texas MD Anderson Cancer Center, United States
Tao Han,
Northern Theater General Hospital, China
Wenguo Jiang,
Binzhou Medical University, China

***Correspondence:**
Xinwen Zhong
xwzhong@cmu.edu.cn
Tian Li
tian@fmmu.edu.cn

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: *21 February 2022*
Accepted: *24 May 2022*
Published: *27 June 2022*

Citation:
Bao G, Li T, Guan X, Yao Y, Liang J, Xiang Y and Zhong X (2022) Development of a Prognostic Alternative Splicing Signature Associated With Tumor Microenvironment Immune Profiles in Lung Adenocarcinoma. Front. Oncol. 12:880478. doi: 10.3389/fonc.2022.880478

Background: Alternative splicing (AS), a pivotal post-transcriptional process across more than 95% of human transcripts, is involved in transcript structural variations and protein complexity. Clinical implications of AS events and their interaction with tumor immunity were systematically analyzed in lung adenocarcinoma (LUAD).

Methods: Transcriptome profiling as well as AS data of LUAD were retrospectively curated. Then, the network of the overall survival (OS)-relevant AS events with splicing factors was established. After screening OS-relevant AS events, a LASSO prognostic model was conducted and evaluated with ROC curves. A nomogram that integrated independent prognostic indicators was created. Immune response and immune cell infiltration were estimated with ESTIMATE, CIBERSORT, and ssGSEA algorithms. Drug sensitivity was inferred with pRRophetic package.

Results: In total, 2415 OS-relevant AS events were identified across LUAD patients. The interaction network of splicing factors with OS-relevant AS events uncovered the underlying regulatory mechanisms of AS events in LUAD. Thereafter, a prognostic model containing 12 AS events was developed, which acted as a reliable and independent prognostic indicator following verification. A nomogram that constituted stage and risk score displayed great effectiveness in evaluating the survival likelihood. Moreover, the AS-based prognostic model was in relation to immune response and immune cell infiltration. Patients with a high-risk score displayed therapeutic superiority to cisplatin, erlotinib, gefitinib, and gemcitabine. Finally, three AS-relevant genes (CDKN2A, TTC39C, and PKIB) were identified as prognostic markers.

Conclusion: Collectively, our findings developed an AS event signature with powerful prognostic predictive efficacy in LUAD.

Keywords: alternative splicing, lung adenocarcinoma, prognosis, immune response, immune microenvironment

INTRODUCTION

As the highest incidence cancer type, lung cancer also causes the most cancer-related deaths (1, 2). According to reports, 85% of all new lung cancers each year are non-small cell lung cancer (NSCLC), with a dismal 5-year survival rate of < 16% (3). Currently, lung adenocarcinoma (LUAD) accounts for the leading pathological subtype of NSCLC, which exhibits rising morbidity among young women and non-smokers (4). Moreover, patients with advanced lung adenocarcinoma are often accompanied by poor long-term prognosis. Currently, surgical resection plus radio- or chemotherapy represents the first choice and main therapeutic means against LUAD (5, 6). Despite recent advances in immunotherapeutic strategies, LUAD patients display diverse responses to immune-based therapies (6–8). Few schemes to prevent and early treat LUAD are developed mainly because of the few characteristic targets upon molecular pathogenesis (9).

Alternative splicing (AS), a pervasive cellular process, exerts a critical function in the post-transcriptional process where a variety of transcripts from the same gene are generated, contributing to proteome complexity (10). More than 95% of human genes incur AS events during physiological process (11). AS events are remarkedly modulated with tissue and developmental stage-specific manners, which are often deregulated in diverse cancer types (12). Abnormal RNA splicing drives tumor initiation and progression through affecting metabolic reprogramming, proliferation, metastases, and resistance of tumor cells and microenvironment (13–16). Moreover, deregulated splice variants produce effects on the therapeutic responses to targeted therapy, radio-, chemo- and immunotherapies (17). Thus, it is mostly important to ascertain pathological splicing isoforms regarding the development of novel practical markers and clarifying the mechanisms involving in deregulated AS events, eventually expounding the influences on cancers, and offering more effective treatment schemes. To date, accumulated evidence uncovers the biological relevance as well as clinical implications of AS events during lung tumorigenesis (18–21). Lung carcinogenesis principally evolves by sequential genetic changes and genomic deregulation, which is also influenced by tumor microenvironment. LUAD exhibits interpatient and intratumor heterogeneity in tumor cells and microenvironment (22). Nevertheless, the underlying relations of AS events with tumor microenvironment of LUAD remain ill-defined.

Herein, our research conducted comprehensive analyses upon AS events across LUAD and identified LUAD-specific AS events for developing novel prognostic markers. Moreover, our findings provided novel thinking about the interactions between AS events and immunity in LUAD.

Abbreviations: NSCLC, non-small cell lung cancer; LUAD, lung adenocarcinoma; AS, alternative splicing; PSI, Percent Spliced In; AD, Alternate Donor site; AA, Alternate Acceptor site; AT, Alternate Terminator; AP, Alternate Promoter; ME, Mutually Exclusive Exons; ES, Exon Skip; RI, Retained Intron; OS, overall survival; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic; ssGSEA, single-sample gene set enrichment analysis; CIBERSORT, Cell type Identification By Estimating Relative Subsets Of RNA Transcripts; ESTIMATE, Estimation of Stromal and Immune Cells in Malignant Tumors using Expression Data; TMB, Tumor mutational burden; IC50, half-maximal inhibitory concentration; AUC, area under the curve.

MATERIALS AND METHODS

Data Retrieval

Transcriptome profiling and clinicopathologic characteristics of 522 LUAD specimens were retrospectively curated from the Cancer Genome Atlas (TCGA) project utilizing TCGAbiolinks R package (23). **Table 1** lists clinicopathological data of 522 LUAD patients. AS data were curated from TCGA SpliceSeq (https://bioinformatics.mdanderson.org/TCGASpliceSeq) (24). Then, Percent Spliced In (PSI) values that ranged from 0 to 1 were determined for AS events across transcripts. AS events were classified into seven forms, containing Alternate Donor site (AD), Alternate Acceptor site (AA), Alternate Terminator (AT), Alternate Promoter (AP), Mutually Exclusive Exons (ME), Exon Skip (ES), and Retained Intron (RI). AS events with PSI value ≥ 75%, and average PSI value ≥ 0.05 were enrolled for subsequent analysis. UpSetR package was employed for visualizing the distribution of AS events in LUAD (25).

Screening OS-Relevant AS Events in LUAD

OS-relevant AS events were selected across LUAD patients through the survival R package utilizing univariate regression analyses following the criteria of p-value < 0.05. In addition, UpSet and volcano plot were adopted for describing the distribution of OS-relevant AS events. Thereafter, the first 20 AS events in different types of AS were visualized into bubble plots.

Establishment of an OS-Relevant Splicing Factor-AS Interaction Network

SpliceAid project was employed to curate specific splicing factors (26). Furthermore, Pearson correlation test was adopted for analyzing the interactions of splicing factors with OS-relevant

TABLE 1 | Clinicopathological characteristics of 522 LUAD patients from TCGA cohort.

Characteristics	Type	n	Proportion (%)
Age	≤65	241	46.2
	> 65	262	50.2
	unknown	19	3.6
Gender	Female	280	53.6
	Male	242	46.4
Stage	I-II	403	77.2
	III-IV	111	21.3
	unknown	8	1.5
T stage	T1-2	453	86.8
	T3-4	66	12.6
	unknown	3	0.6
N stage	N0-1	433	83.0
	N2-3	77	14.8
	unknown	12	2.2
M stage	M0	353	67.6
	M1	25	4.8
	unknown	144	27.6

AS events. The Cytoscape (version 3.8.0) was utilized for visualizing this interactional network of splicing factors with OS-relevant AS events and correlation coefficient > 0.6 as well as $p < 0.05$ as the filtering criteria (27).

Construction and Validation of Predictive Models Based on AS Events

The glmnet R package was adopted to establish a least absolute shrinkage and selection operator (LASSO) prognostic model based on OS-relevant AS events across LUAD patients (28). The prognostic scoring formula was conducted with this formula: risk score= PSI value of AS event1 × Coef1 + PSI value of AS event2 × Coef2 … + PSI value of AS eventn × Coefn, in which Coefn represented the regression coefficient. Then, we stratified LUAD patients into different risk subpopulations according to median risk score. The receiver operating characteristic (ROC) curve was generated utilizing timeROC R package for showing the specificity and sensitivity of risk score in evaluating prognosis of LUAD. The Kaplan-Meier curves were applied to assess the differences in OS rate with the survival R package. Additionally, Cox regression models were conducted for analyzing the interactions of age, gender, tumor stage, and risk score with OS outcomes.

Construction of a Prognostic Nomogram

In order to evaluate OS outcomes, a prognostic nomogram comprised of independently prognostic indicators AS-relevant risk signature as well as stage was conducted for estimating 1-, 2-, and 3-year OS probabilities with the rms R package. Subsequently, calibration curves which showed the survival implications of this nomogram were depicted. The calibration curve close to 45° was considered as an excellent indicator in this nomogram.

Immune Cell Infiltrations Estimated *via* Deconvolution Algorithm and Single-Sample Gene Set Enrichment Analysis (ssGSEA)

The cell type identification by Estimating Relative Subsets Of RNA Transcripts (CIBERSORT) deconvolution algorithm was adopted to estimate the abundances of 22 diverse leukocyte subsets (29). CIBERSORT results for samples with $p < 0.05$ indicated that the estimated abundances of leukocyte subsets were reliable, which were eligible for subsequent analysis. For each specimen, estimations were standardized to sum up to 1, thereby being interpreted directly as cellular fraction. The ssGSEA from Gene Set Variation Analysis (GSVA) was employed for quantification of the relative abundances of 29 immune cells as well as functions following the special feature gene panels across LUAD specimens (30). The ssGSEA enrichment score was indicative of the relative abundance, which was standardized to range from 0 to 1.

Identifying and Comparing the Immune Profiles of Different Risk Groups

The Estimation of Stromal and Immune Cells in Malignant Tumors using Expression Data (ESTIMATE) R package possesses the significant advantage in estimating the specific features of transcriptome profiles (31). The gene sets of immune checkpoints were downloaded from recent research (32, 33). The mRNA expression of immune checkpoints was quantified across LUAD specimens. Tumor mutational burden (TMB) was employed to predict clinical response to immunotherapy (34). TMB was calculated according to the formula: (entire counts of variants)/(the entire lengths of exons) in line with the variants of LUAD specimens that were extracted from the mutational profiles.

Estimation of Drug Sensitivity

Half-maximal inhibitory concentration (IC50) values for cisplatin, gemcitabine, gefitinib, and erlotinib were estimated with the pRRophetic R package by ridge regression analysis (35, 36). IC50 indicated the treatment response to above chemotherapeutic agents in TCGA cohort.

Statistical Analysis

Spearman's correlation analysis was conducted to estimate composition differences. Wilcoxon signed rank test was applied for comparisons in two groups. Kaplan-Meier survival curve was implemented for evaluating the survival differences between groups. Cox regression analysis was conducted for verifying the associations of certain indicators with LUAD prognosis. To evaluate the performance of prognosis prediction, time-independent ROC curves were conducted and area under the curve (AUC) was calculated with timeROC R package. Statistical analysis was achieved utilizing R software (version 4.02). $P < 0.05$ was taken into consideration statistically.

RESULTS

Identification of OS-Relevant AS Events in LUAD

In total, 43,945 AS events were identified across 522 LUAD patients (**Figure 1A**). ES accounted for the most frequent AS signature, followed by AT and AP. Univariate analyses were presented to qualify the impact of each AS event on patients' OS. Subsequently, 2415 AS events displayed remarked associations with survival outcomes of LUAD patients, in which 1356 were protective factors and 1059 were risk factors (**Figure 1B**). Notably, one gene may possess two or more OS-relevant AS events across LUAD patients, as shown in the UpSet plots (**Figure 1C**). The first 20 significant OS-relevant genes of AS events are separately shown in **Figures 1D–J**, which indicated that the seven alternative splicing modes exhibit great variability.

Construction of an OS-Relevant Splicing Factor-AS Interaction Network in LUAD

Splicing factors act as dominant regulators of AS events, which may affect the splicing of oncogenes as well as tumor suppressors (37). For exploring the underlying interactions of the expressions of splicing factors with AS events, we visualized the splicing-

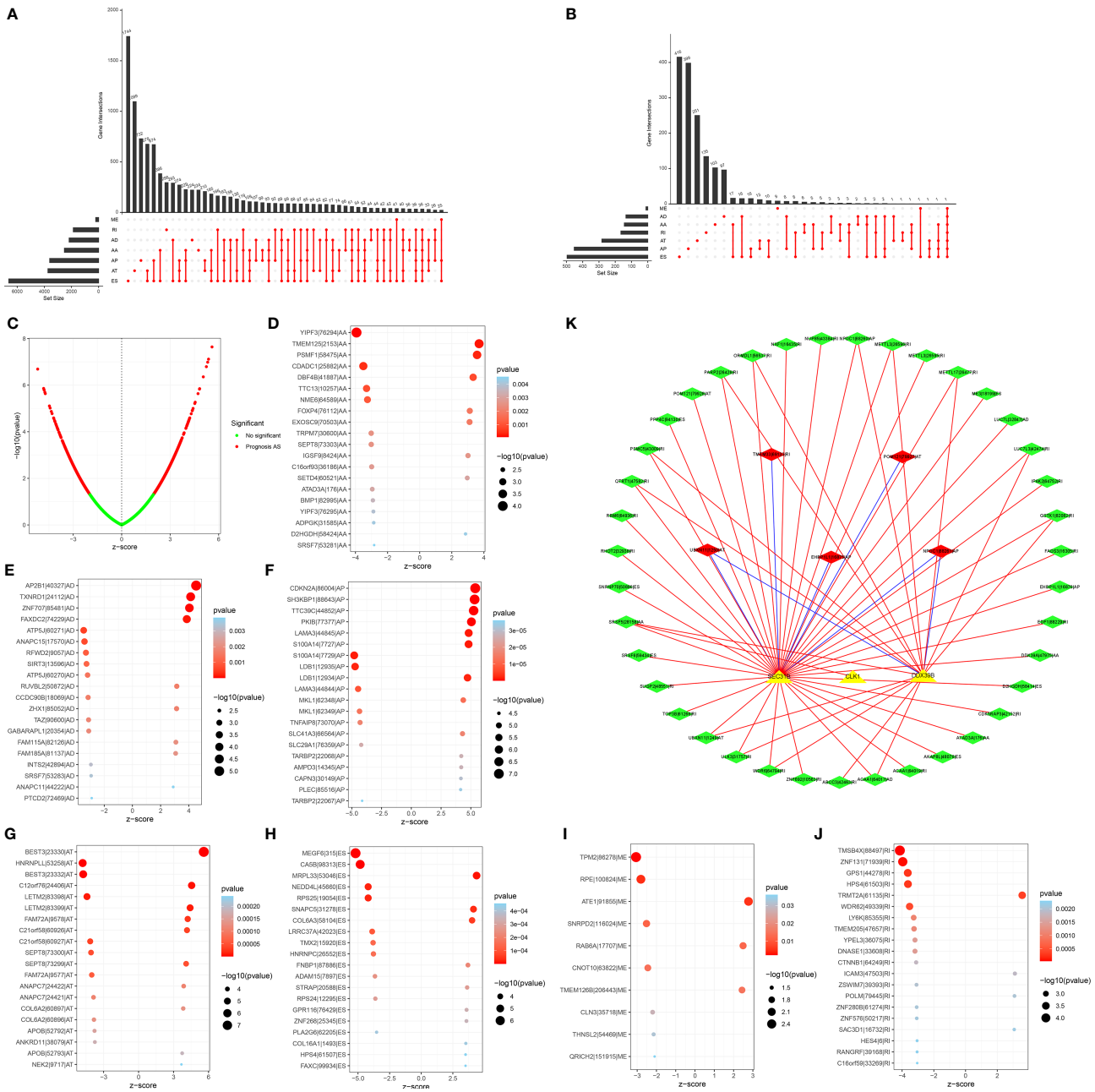


FIGURE 1 | Prognosis-relevant AS events and their interactions with splicing factors across LUAD patients. **(A)** UpSet plot showing numbers and percentages of AS events as well as their interactions across LUAD specimens. **(B)** Volcano plots of OS-relevant AS events in LUAD. Red dots represented AS events that were distinctly correlated to OS, but green dots did not affect patients' OS. **(C)** UpSet showing numbers and percentages of seven types of OS-relevant AS events and their interactions in LUAD. **(D–J)** Bubble plots of the distribution of the first 20 most significant AS events in LUAD, which indicated that the seven alternative splicing modes exhibit great variability. **(K)** The OS-relevant splicing factor-AS interaction network in LUAD. Triangle bubbles indicated splicing factors and diamond bubbles indicated AS events. The red and blue line separately indicated positive and negative connection in splicing factors and AS events. Red and green diamond bubbles separately meant adverse and favorable prognosis-relevant AS events.

regulatory network, as depicted in **Figure 1K**. In total, three splicing factors (including SEC31B, CLK1, and DDX39B) displayed prominent associations with 44 OS-relevant AS events. Furthermore, most favorable AS events exhibited positive interactions with the expression of splicing factors and the three splicing factors were in relation to multiple AS events. Thus, splicing factors may act as an indispensable role in modulating AS events during lung carcinogenesis.

Development of a Reliable Prognostic AS Event-Based Signature in LUAD

For avoiding over-fitting, LASSO Cox analysis was adopted for developing a prognostic model of LUAD on the basis of OS-relevant AS events. Through cross-verification, the optimal parameters were selected (**Figure 2A**) and the coefficients in LASSO regression model were determined (**Figure 2B**). Ultimately, 12 OS-relevant AS events (BEST3|23330|AT, CDKN2A|86004|AP, TTC39C|44852|AP, MEGF6|315|ES, PKIB|77377|AP, CA5B|98313|ES, HNRNPLL|53258|AT, LDB1|12935|AP, C12orf76|24406|AT, AP2B1|40327|AD, LETM2|83398|AT, MRPL33|53046|ES) were identified (**Table 2**). In line with the regression coefficients and PSI value of 12 OS-relevant AS events, we calculated risk scores of LUAD patients. Thereafter, LUAD patients were classified into different groups with median risk score of 0.8834 (**Figure 2C**). Moreover, we noticed that high-risk subpopulations were often accompanied by high mortality (**Figure 2D**). Heatmap depicted the heterogeneity in PSI values of 12 OS-relevant AS events (**Figure 2E**). Prognostic analyses uncovered that high-risk subpopulations exhibited remarkedly dismal OS outcomes (**Figure 2F**). The validity of the prognostic model in prognosis prediction was verified through ROC analysis. The AUC values at 1-, 3-, and 5-year OS were separately 0.762, 0.770, and 0.725, showing the good effectiveness of this model in prognosis prediction (**Figure 2G**).

Associations of the Prognostic Model With Clinicopathological Characteristics of LUAD

Through ROC analysis, we presented the comparisons of AUC values and noticed that risk score displayed the higher AUC values under 1-, 3-, and 5-year survival compared with clinicopathological features (age, gender, and stage; **Figures 3A–C**). Additionally, the differences in risk score between distinct clinicopathological features were compared among LUAD patients. No significant differences were observed between age ≤65 and > 65 (**Figure 3D**) as well as between non-metastasis (M0) and metastasis (M1; **Figure 3E**). Increased risk score was investigated in male and female patients

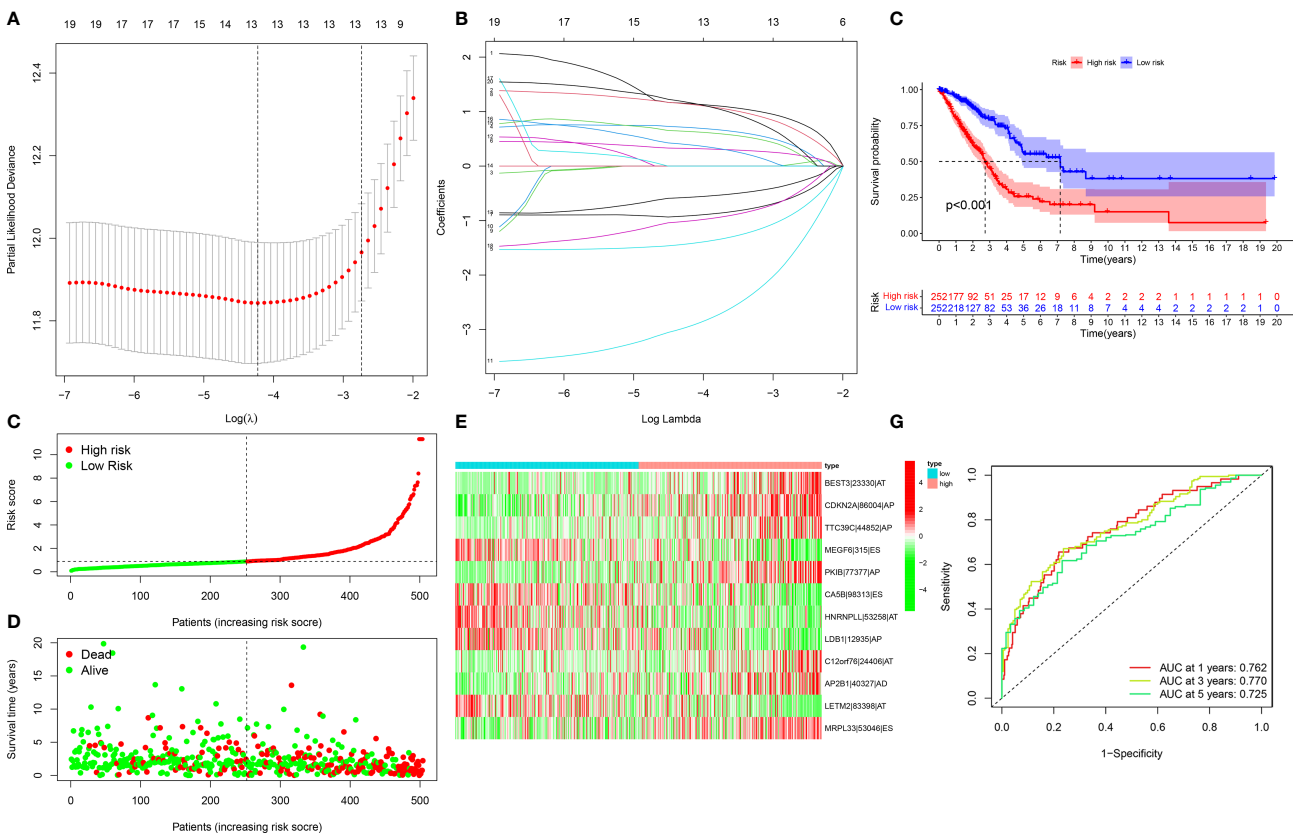


FIGURE 2 | Development of a reliable prognostic model for LUAD patients. **(A)** The distribution of partial likelihood deviance corresponding to λ-logarithm value. **(B)** LASSO coefficient profiling of OS-relevant AS events. The lines stood for OS-relevant AS events and candidates AS events were chosen utilizing ten-fold cross-verification with minimum criteria. **(C)** The distribution of risk score across LUAD patients. Red dots meant high-risk patients while green dots meant low-risk patients. **(D)** Scatter plots depicted distribution of LUAD patients' survival time and status. Red dots denoted patients who were dead, whereas green dots denoted patients who were alive. **(E)** Heatmap displayed the distribution of PSI values for the established prognostic model. **(F)** Kaplan–Meier survival curves of high- and low-risk LUAD patients. **(G)** ROC curves at 1, 3, and 5 years of the prognostic model for LUAD patients.

TABLE 2 | Twelve OS-relevant AS events in the LASSO prognostic model.

AS events	Coefficient	HR	HR.95L	HR.95H	P-value
BEST3|23330|AT	1.23	3.43	1.07	10.93	0.038
CDKN2A|86004|AP	1.27	3.55	1.68	7.52	< 0.001
TTC39C|44852|AP	0.78	2.19	0.86	5.59	0.101
MEGF6|315|ES	-1.57	0.21	0.08	0.52	< 0.001
PKIB|77377|AP	0.35	1.42	0.75	2.71	0.281
CA5B|98313|ES	-0.99	0.37	0.14	0.98	0.045
HNRNPLL|53258|AT	-3.38	0.03	0.004	0.30	0.002
LDB1|12935|AP	-0.65	0.52	0.16	1.67	0.275
C12orf76|24406|AT	0.70	2.01	0.39	10.38	0.402
AP2B1|40327|AD	0.51	1.68	0.38	7.51	0.497
LETM2|83398|AT	-1.13	0.32	0.11	0.93	0.036
MRPL33|53046|ES	1.40	4.06	0.57	29.21	0.163

(**Figure 3F**). As T, N, and stage increased, risk score was gradually elevated (**Figures 3G–I**), indicating that the prognostic model contributed to LUAD progression.

The Prognostic Model Acts as an Independently Prognostic Indicator of LUAD

We further verified the prognostic value of clinical characteristics and risk score and found that risk score and stage possessed the potential to independently predict LUAD prognosis (**Figures 4A, B**). Thereafter, a prognostic nomogram containing independent prognostic indicators risk score as well as clinicopathological stage was conducted for forecasting patients' outcomes (**Figure 4C**). Calibration curves were indicative of the powerful prognostic predictive capacity of this nomogram in 1-, 3-, and 5-year OS (**Figures 4D–F**).

Development of a Prognostic Nomogram Containing the Prognostic Model and Stage

For further applying our findings to clinical practice, this study constructed a nomogram prognostic score system in the prediction of 1-, 3-, and 5-year OS outcomes of LUAD patients (**Figure 4C**). The scoring system included the prognostic model and stage. Thereafter, for verifying the reliability of the prognostic nomogram, calibration plots were conducted and confirmed the practical significance of the model. As depicted in **Figures 4D–F**, the model possessed the potential in determining survival outcomes with a high predicted accuracy.

Associations of the Prognostic Model With Tumor Immunity

We firstly estimated infiltration levels of immune and stromal cells across LUAD patients. Accordingly, patients with a high-risk score displayed reduced immune score and stromal score (**Figures 5A, B**). Nevertheless, higher tumor purity was investigated in high-risk patients (**Figure 5C**). Then, we determined ESTIMATE and noticed the prominently decreased ESTIMATE score in the high-risk group (**Figure 5D**). Thus, low-risk tumors were accompanied by abundant infiltrations of immune and stromal cells. Then, we systematically investigated the immune cell infiltration landscape across LUAD with CIBERSORT algorithm. We noticed that the low-risk group displayed high infiltration levels of B cells naïve, T cells CD4 memory resting, monocytes, and mast cells resting (**Figure 5E**). Oppositely, the high-risk group exhibited increased infiltration levels in T cells CD4 memory activated, T cells follicular helper, T cells regulatory (Tregs), NK cells resting, macrophages M0, and macrophages M1. **Supplementary Figure 1** displays interactions of risk score signature with above immune cell infiltrations. Subsequently, we revealed the activities of immune functions and immune cell infiltrations across LUAD ssGSEA method. In **Figures 5F, G**, higher abundance levels of aDCs, B cells, HLA, iDCs, mast cells, neutrophils, T helper cells, TIL, and type II IFN response were investigated in the low-risk group while MHC class I and NK cells exhibited higher abundance levels in the high-risk group. We also evaluated the interactions of the prognostic model with immune checkpoints across LUAD. As depicted in **Figure 5H**, this prognostic model possessed a positive association with CD274. Moreover, we observed that the low-risk group was characterized by increased expression of most immune checkpoint-related genes (**Figure 5I**). Thus, low-risk patients were indicative of higher immune response as well as immune cell infiltration.

Associations of the Prognostic Model With TMB and Drug Responses

The interaction of the prognostic model with TMB was also observed across LUAD. As shown in **Figure 6A**, the high-risk group exhibited a remarkedly increased TMB score. Moreover, we presented survival analysis among diverse subgroups. We noticed that subpopulations possessing an elevated TMB score as well as a reduced risk score displayed the most favorable survival outcomes while those with a low TMB score and high-risk score exhibited the poorest survival outcomes (**Figure 6B**). Chemotherapy and targeted therapy were gradually applied in treatments for patients with advanced lung adenocarcinoma. It is of great significance to evaluate the responses of certain drugs in different risk subpopulations. Herein, we identified the treatment responses of some drugs that were widely used in the treatment of LUAD. As shown in **Figures 6C–F**, the high-risk group possessed prominently lowered IC50 values of cisplatin, erlotinib, gefitinib, and gemcitabine, indicating that this

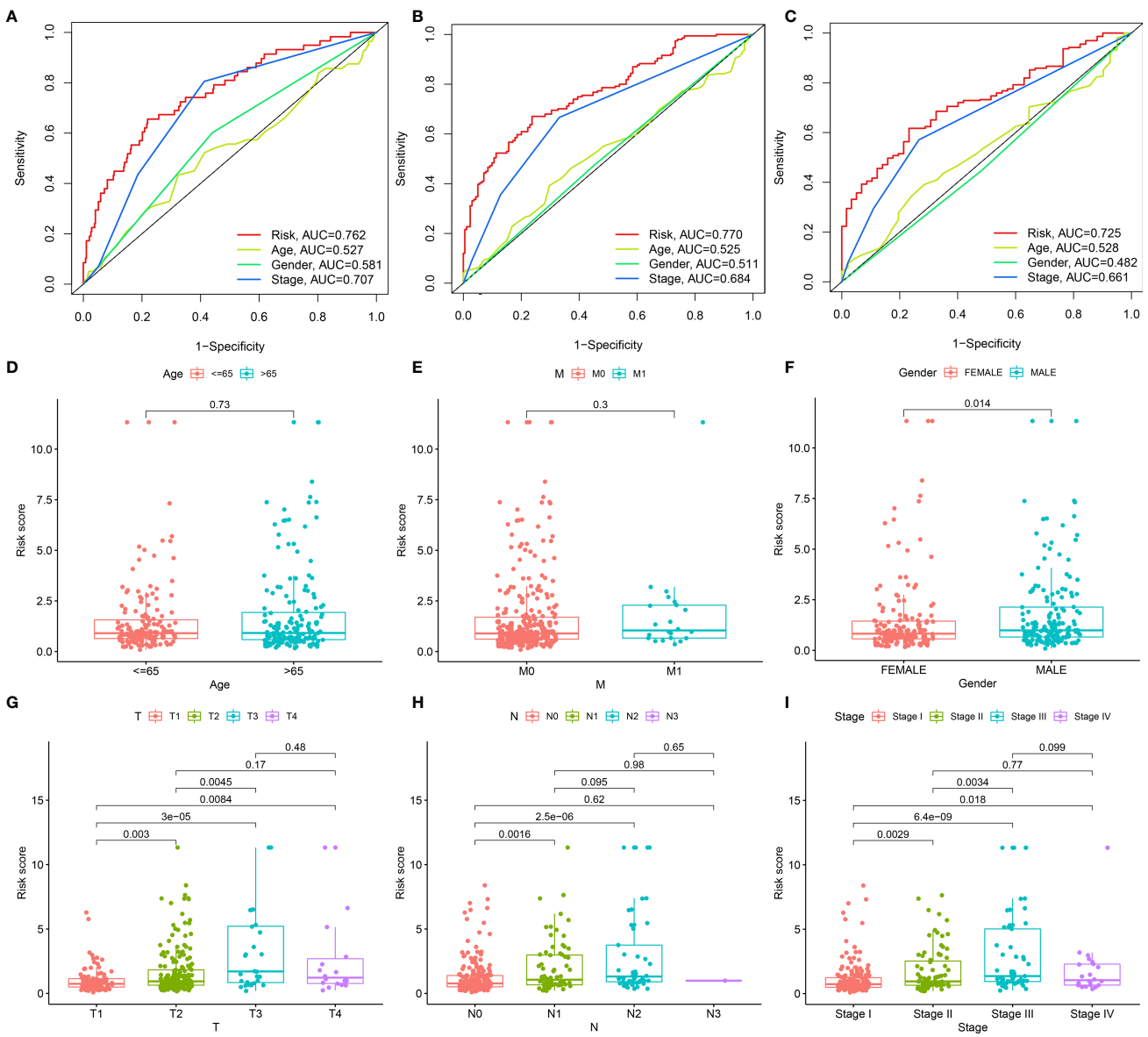


FIGURE 3 | Associations of the prognostic model with clinicopathological characteristics of LUAD. **(A–C)** Comparisons of AUC at 1-, 3-, and 5-year survival estimated by risk score and clinicopathological characteristics through ROC analysis. **(D–I)** Box plots showing the distribution of risk scores in distinct clinicopathological characteristics, containing **(D)** age (≤65 vs. > 65), **(E)** M stage (M0 vs. M1), **(F)** gender (female vs. male), **(G)** T stage (T1 vs. T2 vs. T3 vs. T4), **(H)** N stage (N0 vs. N1 vs. N2 vs. N3) and **(I)** stage (stage I vs. stage II vs. stage III vs. stage IV).

subpopulation possessed higher sensitivity to these therapeutic agents. The above findings provide more clues for individualized treatment strategies in LUAD patients.

Identification of Prognostic AS Events-Related Genes

We found that CDKN2A, PKIB, and TTC39C exhibited a higher expression in LUAD than normal tissues among the 12 AS events-relevant genes in the prognostic models (**Figures 7A–C**). Moreover, survival analysis uncovered that highly expressed CDKN2A and PKIB were in relation to more dismal survival probabilities of LUAD (**Figures 7D, E**). In contrast, high TTC39C expression was indicative of the marked survival advantage (**Figure 7F**).

Associations of Prognostic AS Events-Related Genes With Immune Microenvironment

We further investigated the interactions of the three prognostics AS events-related genes (CDKN2A, PKIB, and TTC39C) with immune response and immune cell infiltration across LUAD. We found that deregulated CDKN2A did not affect estimate,

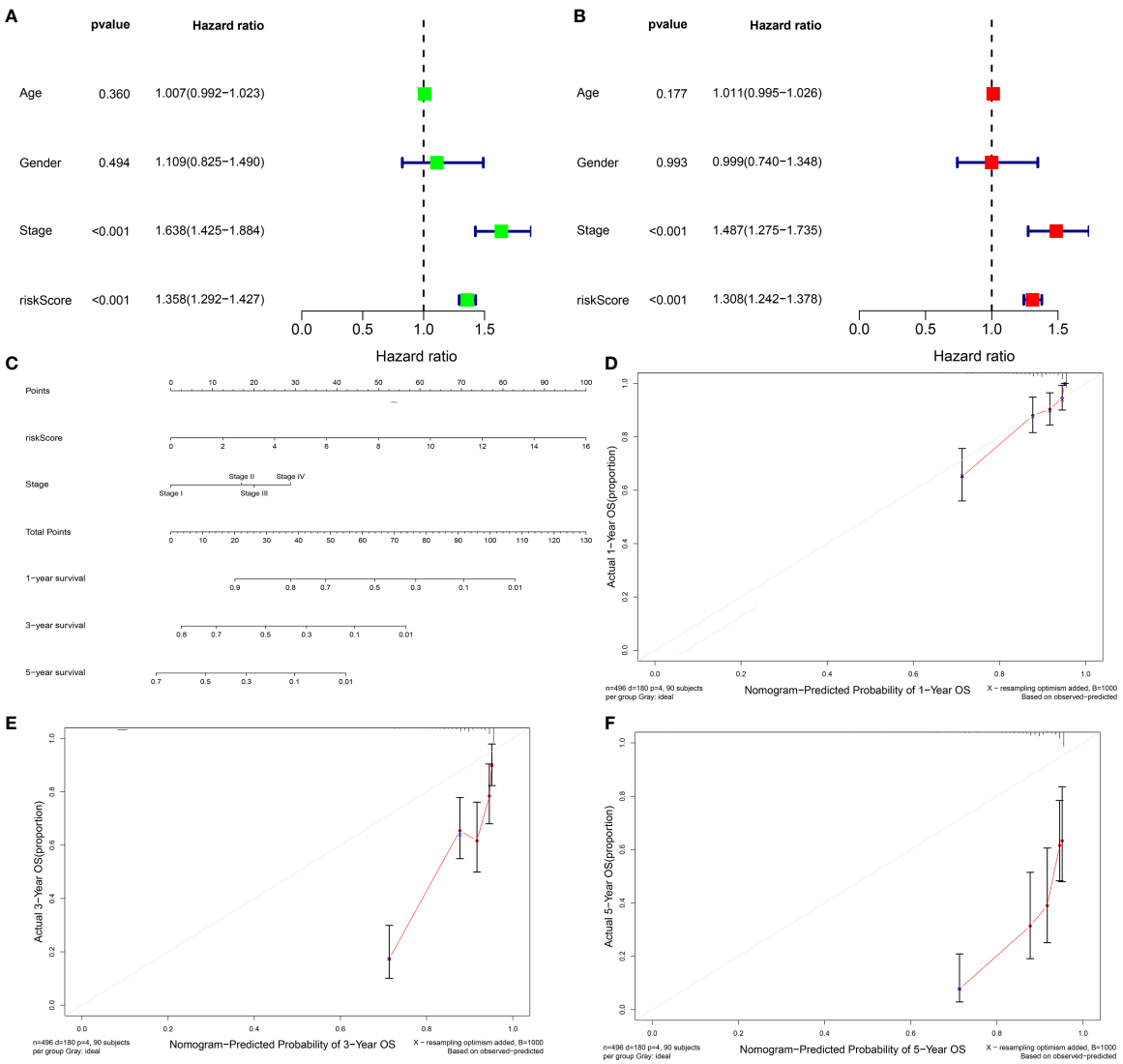


FIGURE 4 | Evaluation of the independence of prognostic model in prognostic prediction and construction of prognostic nomograms for LUAD. **(A, B)** Univariate and multivariate Cox analysis of risk score and clinicopathological features with LUAD prognosis. **(C)** The nomogram of risk score signature and stage for prediction of 1-, 3-, and 5-year OS of LUAD. **(D–F)** Calibration curves used to compare nomogram estimated 1-, 3-, and 5-year survival probabilities with actual survival time.

immune, and stromal score as well as tumor purity (**Figures 8A–D**). For **Figures 8E–H**, high PKIB expression was characterized by increased estimate, immune, and stromal score but reduced tumor purity. Moreover, high TTC39C expression displayed remarkedly decreased estimate, immune, and stromal score but elevated tumor purity (**Figures 8I–L**). In **Figure 8M**, CDKN2A upregulation was in relation to increased infiltration levels of T cells CD8, T cells CD4 memory activated, and macrophages M1. PKIB deregulation was in relation to infiltrations of B cell naïve, B cells memory, plasma cells, T cells CD8, macrophages M1, macrophages M2, dendritic cells resting and mast cells resting (**Figure 8N**). B cells native, plasma cells, T cells follicular helper, and NK cells activated exhibited the increased infiltration levels in high TTC39C expression group (**Figure 8O**). The ssGSEA results uncovered the increased infiltrations of APC co-inhibition, CD8+ T cells, inflammation-promoting, MHC class I, NK cells, T cell co-stimulation, Tfh, and Th1 cells in high CDKN2A expression group (**Figure 8P**). PKIB upregulation was in relation to most immune functions and immune cell infiltrations (**Figure 8Q**). In **Figure 8R**, we noticed the prominent interactions of high TTC39C expression with activation of most immune functions and immune cell infiltrations. We also estimated the associations of CDKN2A, PKIB, and TTC39C with immune checkpoint molecules. Most immune checkpoint molecules exhibited positive interactions with CDKN2A, PKIB, and TTC39C (**Figures 8S–U**).

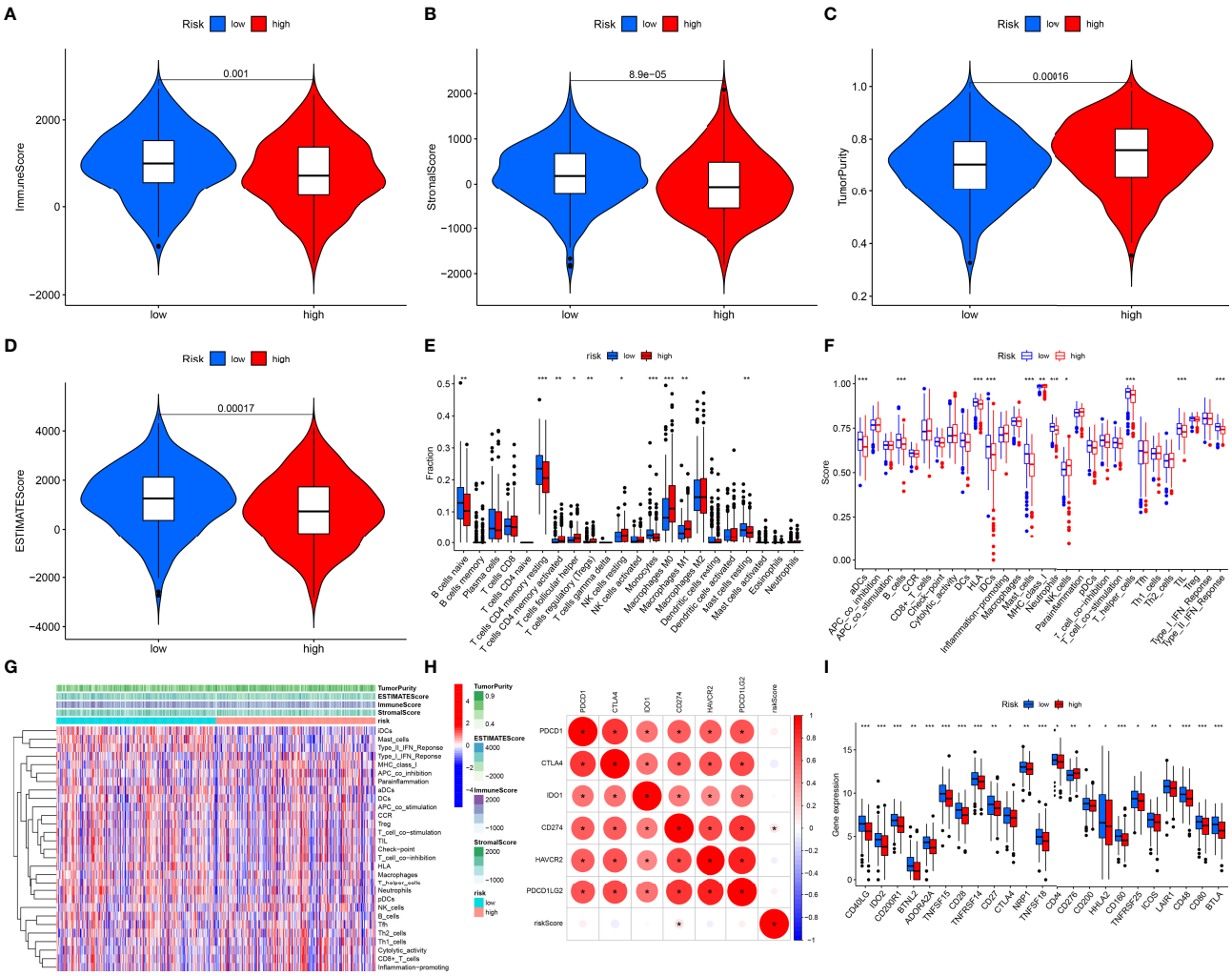


FIGURE 5 | Associations of prognostic model with immune microenvironment across LUAD patients. **(A–D)** Distribution of estimate, immune, and stromal score and tumor purity in different risk groups. **(E)** Comparisons of the levels of tumor immune infiltration in different risk subpopulations with CIBERSORT algorithm. **(F)** Comparisons of the abundance levels of immune cell infiltrations and immune functions in different risk subpopulations utilizing ssGSEA algorithm. **(G)** Heatmap visualizing the distribution of the abundance levels of immune cell infiltrations and immune functions. **(H)** Associations of risk score signature and common immune checkpoint molecules across LUAD. **(I)** Comparisons of the expressions of immune checkpoint molecules in different risk subpopulations. *P < 0.05; **p < 0.01; ***p < 0.001.

DISCUSSION

AS, a crucial post-transcriptional modification, can produce diverse mRNA variants, which results in structural transcription variation and protein diversity (38, 39). Emerging evidence suggests the functions of AS events in lung carcinogenesis (40). For instance, diverse splicing types of regulators of cell apoptosis may affect NSCLC progression through modulating the imbalance between pro-apoptosis and apoptosis (41–43). Herein, we systematically uncovered the prognostic implications and immunity of AS events in LUAD.

Herein, in total, 43,945 AS events were identified across LUAD, indicating that AS might be a common modification in LUAD. Following survival analysis, we observed 2415 OS-related AS events as well as distinct splicing types had specific splicing preferences, which might assist in formulating more effective treatment regimens. Previous studies have shown that the binding of splicing factors to specific RNA sequences in genome determines precise regulation of RNA splicing (44). Thus, an integrative analysis was conducted for addressing the underlying mechanisms involving them during lung tumorigenesis. The OS-relevant splicing factor-AS interaction network showed the prominent interactions of 44 OS-relevant AS events with three splicing factors (SEC31B, CLK1, and DDX39B). Previously, CLK1 could modulate the chemoresistance of glioma cells *via* glycolytic signaling mediated by AMPK/mTOR/HIF-1α (45) as well as participating in modulating the splicing process of gastric cancer, serving as an underlying therapeutic target against this

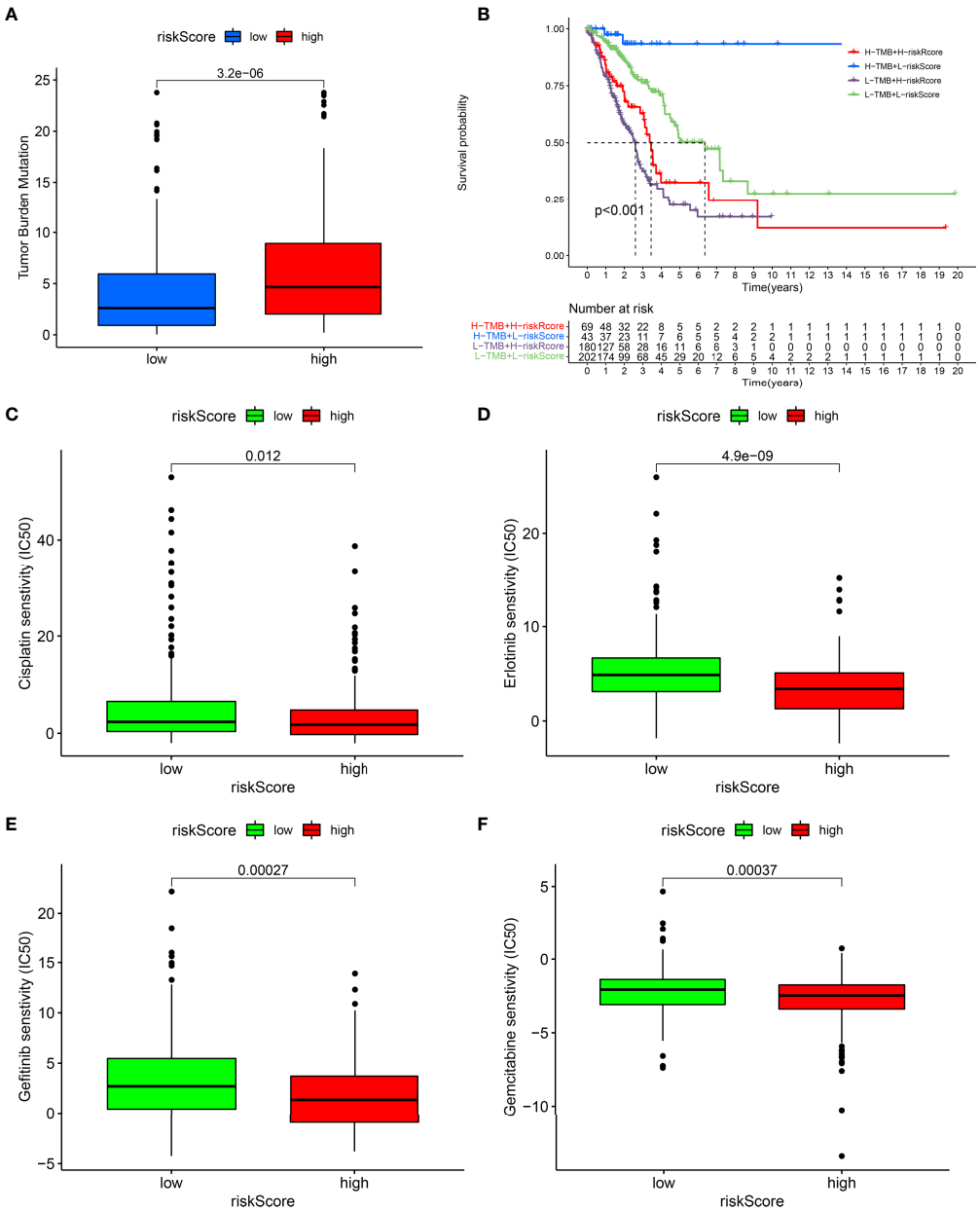


FIGURE 6 | Associations of prognostic model with TMB and drug responses. **(A)** Comparisons of TMB score between high- and low-risk groups. **(B)** Survival analysis of different TMB score and risk score groups. **(C–F)** Comparisons of sensitivity to cisplatin, erlotinib, gefitinib, and gemcitabine between high- and low-risk groups.

malignancy (46). Chemical suppression of CLK1 may disrupt the recruitment of internal kinetochores as well as impair cell cycle progression, contributing to unprogrammed cell death (47). Moreover, inhibition of DDX39B triggers sensitivity of BRCA1-mutant ovarian cancer cells to chemotherapy drugs such as platinum and PARPi (48). Our data indicate that splicing factors and AS events were not only one-to-one coordination or antagonistic regulatory interactions, revealing the complexity of their regulatory network.

With the LASSO method, we established an AS event-based prognostic model (BEST3|23330|AT, CDKN2A|86004|AP, TTC39C|44852|AP, MEGF6|315|ES, PKIB|77377|AP, CA5B|98313|ES, HNRNPLL|53258|AT, LDB1|12935|AP, C12orf76|24406|AT, AP2B1|40327|AD, LETM2|83398|AT, MRPL33|53046|ES) in LUAD. In-depth analysis verified that this model could accurately indicate outcomes of LUAD patients. Accumulated evidence suggests that AS events are in relation to the remodeling of the tumor microenvironment (15, 49).

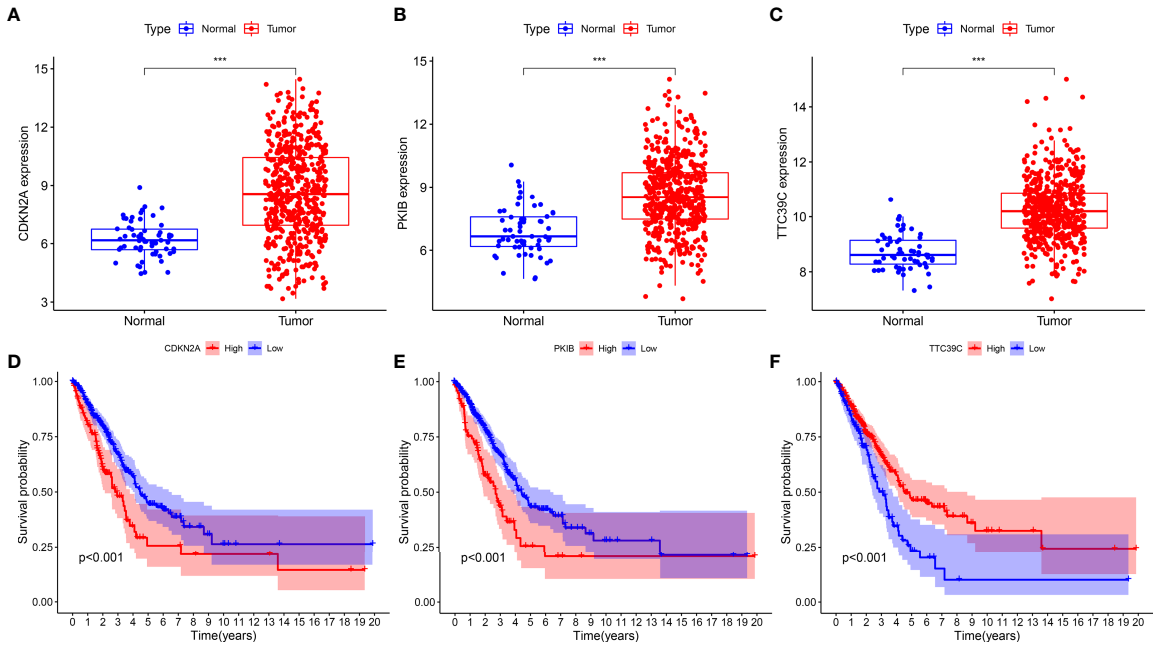


FIGURE 7 | Identification of prognostic AS events-related genes. **(A–C)** The expression patterns of CDKN2A, PKIB, and TTC39C in LUAD and normal tissues. **(D–F)** Kaplan-Meier plots of different expression of CDKN2A, PKIB, and TTC39C. ***p < 0.001.

Herein, our data uncovered the high-risk group presented the features of decreased infiltrations of immune and stromal cells as well as increased tumor purity. Additionally, LUAD patients with a high risk presented worse immune reactivity, which might contribute to shorter survival duration as well as higher degree of malignancy. TMB was characterized as an effective indicator for prediction of clinical response to immunotherapy (34, 50, 51). Our data indicated that the high-risk group presented higher TMB score, which revealed that patients in high-risk groups may experience better outcomes with immunotherapy. Subgroup analysis uncovered those patients with reduced TMB score and increased risk score tended to exhibit more malignant clinical outcomes and shorter survival duration. Moreover, we noticed that patients with a high-risk score presented higher priority to cisplatin, gemcitabine, erlotinib, and gefitinib, providing a reference for the choice of the optimal chemotherapeutic or targeted therapeutic regimen.

Previous research revealed the parental genes of AS events displayed deregulation owing to abnormal AS events (52). Therefore, we identified 12 AS-relevant genes (BEST, CDKN2A, TTC39C, MEGF6, PKIB, CA5B, HNRNPLL, LDB1, C12orf76, AP2B1, LETM2, MRPL33) in the AS event-based prognostic model. Further, we investigated the upregulation of CDKN2A, TTC39C, and PKIB expressions in LUAD as well as their upregulation was indicative of dismal outcomes in LUAD. Further analysis uncovered that highly expressed PKIB was related to increased infiltrations of immune and stromal cells and opposite findings were investigated for TTC39C. Additionally, CDKN2A, TTC39C, and PKIB exhibited positive associations with most immune checkpoint molecules across LUAD. The data indicated that CDKN2A, TTC39C, and PKIB exerted critical functions in modulating tumor immunity of LUAD. Previously, CDKN2A was shown to be associated with polymorphism of GSTs genes in esophageal squamous cell carcinoma (53). PKIB facilitates breast and lung carcinogenesis through modulating Akt signaling (54). To date, TTC39C has no relevant literature reports on its role in tumorigenesis. Several limitations have been pointed out in our study. Firstly, the AS event-based prognostic model was developed based on a retrospective cohort. The predictive power of this model needs to be validated in more prospective cohorts. Moreover, the limited evidence is not enough to fully explain the specific roles of these genes in lung tumorigenesis. In follow-up studies, we will conduct further experiments to validate our findings.

CONCLUSION

Collectively, our research presented systematic analyses of AS events across LUAD, and finally developed a reliable and independent prognostic model on the basis of AS events. Our in-depth analyses revealed the interactions of AS events with immune response and immune cell infiltrations. Finally, we identified three prognostic AS-event-related genes that might play a non-negligible role in lung carcinogenesis. Nevertheless, their potential significance as prognostic indicators and therapeutic targets in clinical applications deserve further study.

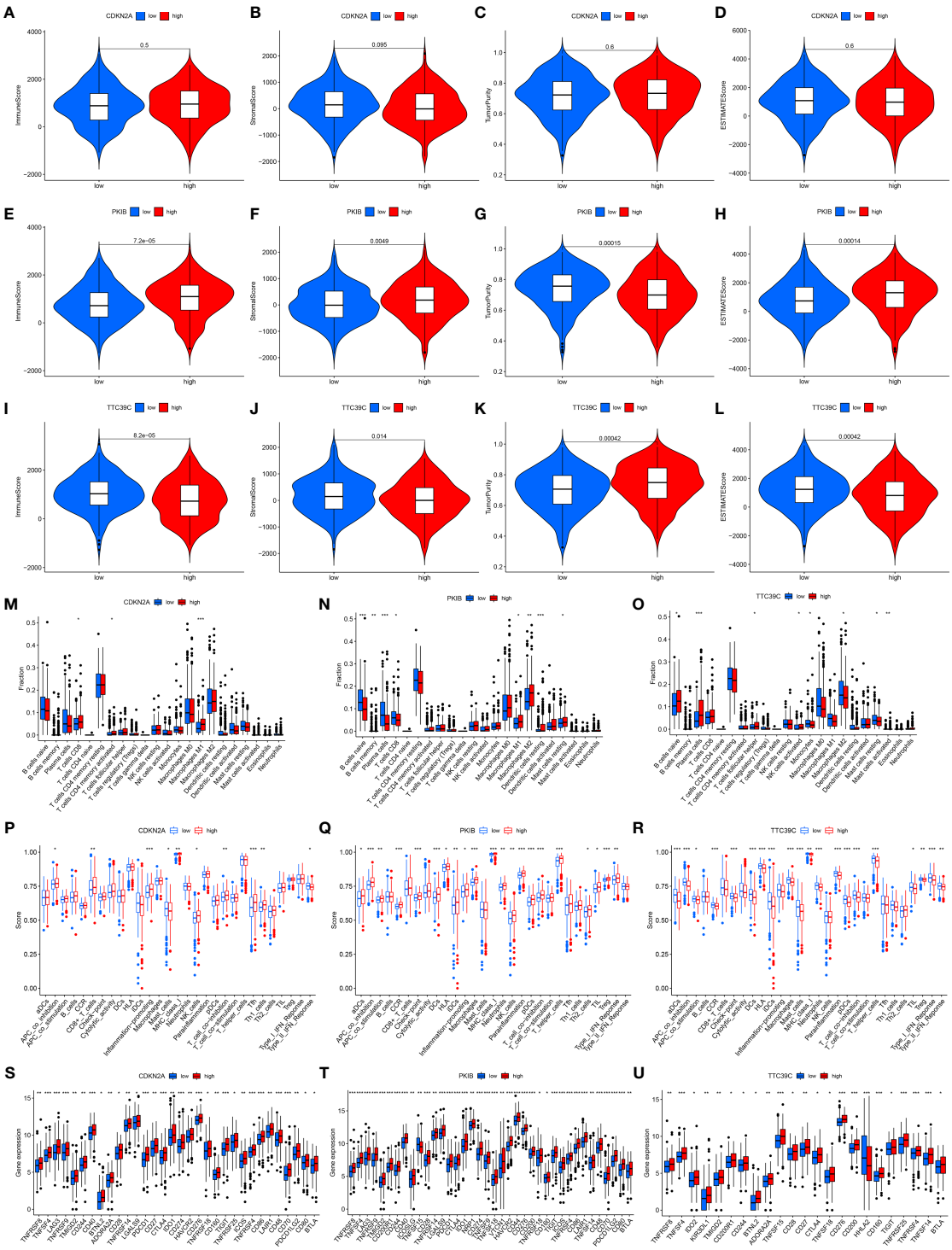


FIGURE 8 | Associations of prognostic AS events-related genes with immune response and immune cell infiltration. **(A–D)** Violin plots depicted the distribution of estimate, immune, and stromal score as well as tumor purity in high and low CDKN2A groups. **(E–H)** Violin plots depicted the distribution of estimate, immune, and stromal score as well as tumor purity in high and low PKIB groups. **(I–L)** Violin plots depicted the distribution of estimate, immune, and stromal score as well as tumor purity in high and low TTC39C groups. **(M–O)** The distribution of the abundance levels of tumor-infiltrating immune subpopulations in high and low expression of CDKN2A, PKIB, and TTC39C groups. **(P–R)** The distribution of the abundance levels of immune cell infiltrations and immune functions in high and low expression of CDKN2A, PKIB, and TTC39C groups. **(S–U)** Expression levels of immune checkpoint related genes in high and low CDKN2A, PKIB, and TTC39C subpopulations. *P < 0.05; **p < 0.01; ***p < 0.001.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics committee of China Medical Uni. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

Author GB and YX performed the statistical analyses and wrote the manuscript. Author GB completed all of the data entry and provided assistance for the data analysis. Author GB, XG, YY, JL, and XZ were responsible for the diagnosis and clinical assessment of the participants. Author XZ and TL designed and wrote the study protocol and reviewed the manuscript. Author XG participated the revision of this manuscript. In addition, author YY, YX, and JL offered many constructive opinions on this study and provided a critical revision of the manuscript for important intellectual content. All authors contributed to and approved the final manuscript.

FUNDING

This work was supported by Wu Jieping Medical Foundation (320.6750.2020-17-7).

ACKNOWLEDGMENTS

Authors would like to profusely thank all individuals who supported and helped them to conduct this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.880478/full#supplementary-material

Supplementary Figure 1 | Correlations between prognostic model and immune cell infiltrations across LUAD, including **(A)** B cells naïve, **(B)** dendritic cells resting, **(C)** mast cells resting, **(D)** monocytes, **(E)** T cells CD4 memory resting, **(F)** macrophage M0, **(G)** macrophage M1, **(H)** NK cells resting, **(I)** T cells CD4 memory activated, **(J)** T cells follicular helper, and **(K)** Tregs.

REFERENCES

1. 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 Cancer J Clin* (2018) 68 (6):394–424. doi: 10.3322/caac.21492
2. Woodman C, Vundu G, George A, Wilson C. Applications and Strategies in Nanodiagnosis and Nanotherapy in Lung Cancer. *Semin Cancer Biol* (2021) 69:349–64. doi: 10.1016/j.semcancer.2020.02.009
3. Oudkerk M, Liu S, Heuvelmans MA, Walter JE, Field JK. Lung Cancer LDCT Screening and Mortality Reduction - Evidence, Pitfalls and Future Perspectives. *Nat Rev Clin Oncol* (2021) 18(3):135–51. doi: 10.1038/s41571-020-00432-6
4. Zhu X, Chen L, Liu L, Niu X. EMT-Mediated Acquired EGFR-TKI Resistance in NSCLC: Mechanisms and Strategies. *Front Oncol* (2019) 9:1044. doi: 10.3389/fonc.2019.01044
5. Liu H, Qin Y, Zhou N, Ma D, Wang Y. ZNF280A Promotes Lung Adenocarcinoma Development by Regulating the Expression of EIF3C. *Cell Death Dis* (2021) 12(1):39. doi: 10.1038/s41419-020-03309-9
6. Wu Y, Biswas D, Swanton C. Impact of Cancer Evolution on Immune Surveillance and Checkpoint Inhibitor Response. *Semin Cancer Biol* (2021) S1044-579X(21):00042-0.. doi: 10.1016/j.semcancer.2021.02.013
7. Mansfield AS, Każarnowicz A, Karaseva N, Sánchez A., De Boer R, Andric Z, et al. Safety and Patient-Reported Outcomes of Atezolizumab, Carboplatin, and Etoposide in Extensive-Stage Small-Cell Lung Cancer (IMpower133): A Randomized Phase I/III Trial. *Ann Oncol* (2020) 31(2):310–7. doi: 10.1016/j.annonc.2019.10.021
8. Garassino MC, Gadgeel S, Esteban E, Felip E, Speranza G, Domine M, et al. Patient-Reported Outcomes Following Pembrolizumab or Placebo Plus Pemetrexed and Platinum in Patients With Previously Untreated, Metastatic, non-Squamous non-Small-Cell Lung Cancer (KEYNOTE-189): A Multicentre, Double-Blind, Randomised, Placebo-Controlled, Phase 3 Trial. *Lancet Oncol* (2020) 21(3):387–97. doi: 10.1016/S1470-2045(19)30801-0
9. Borczuk AC. Therapeutic Interception of Early Lung Adenocarcinoma Progression: Not Just How, But When? *Am J Respir Crit Care Med* (2021) 203:8–9. doi: 10.1164/rccm.202008-3087ED
10. Di Matteo A, Belloni E, Pradella D, Cappelletto A, Volf N, Zacchigna S, et al. Alternative Splicing in Endothelial Cells: Novel Therapeutic Opportunities in Cancer Angiogenesis. *J Exp Clin Cancer Res* (2020) 39(1):275. doi: 10.1186/s13046-020-01753-1
11. Montes M, Sanford BL, Comiskey DF, Chandler DS. RNA Splicing and Disease: Animal Models to Therapies. *Trends Genet* (2019) 35:68–87. doi: 10.1016/j.tig.2018.10.002
12. Bessa C, Matos P, Jordan P, Gonçalves V. Alternative Splicing: Expanding the Landscape of Cancer Biomarkers and Therapeutics. *Int J Mol Sci* (2020) 21 (23). doi: 10.3390/ijms21239032
13. Sciarrillo R, Wojtuszkiewicz A, Assaraf YG, Jansen G, Kaspers GJL, Giovannetti E, et al. The Role of Alternative Splicing in Cancer: From Oncogenesis to Drug Resistance. *Drug Resist Update* (2020) 53:100728. doi: 10.1016/j.drup.2020.100728
14. Ule J, Blencowe BJ. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. *Mol Cell* (2019) 76:329–45. doi: 10.1016/j.molcel.2019.09.017
15. Bonnal SC, López-Oreja I, Valcárcel J. Roles and Mechanisms of Alternative Splicing in Cancer - Implications for Care. *Nat Rev Clin Oncol* (2020) 17:457–74. doi: 10.1038/s41571-020-0350-x
16. Wang S, Sun Z, Lei Z, Zhang H. RNA-Binding Proteins and Cancer Metastasis. *Semin Cancer Biol* (2022). doi: 10.1016/j.semcancer.2022.03.018
17. Cherry S, Lynch KW. Alternative Splicing and Cancer: Insights, Opportunities, and Challenges From an Expanding View of the Transcriptome. *Genes Dev* (2020) 34:1005–16. doi: 10.1101/gad.338962.120

18. Cai Q, He B, Zhang P, Zhao Z, Peng X, Zhang Y, et al. Exploration of Predictive and Prognostic Alternative Splicing Signatures in Lung Adenocarcinoma Using Machine Learning Methods. *J Transl Med* (2020) 18(1):463. doi: 10.1186/s12967-020-02635-y
19. Mao S, Li Y, Lu Z, Che Y, Huang J, Lei Y, et al. PHD Finger Protein 5A Promoted Lung Adenocarcinoma Progression *via* Alternative Splicing. *Cancer Med* (2019) 8(5):2429–41. doi: 10.1002/cam4.2115
20. Zhang S, Bao Y, Shen X, Pan Y, Sun Y, Xiao M, et al. RNA Binding Motif Protein 10 Suppresses Lung Cancer Progression by Controlling Alternative Splicing of Eukaryotic Translation Initiation Factor 4H. *EBioMedicine* (2020) 61:103067. doi: 10.1016/j.ebiom.2020.103067
21. Khan F, Bhat B, Sheikh B, Tariq L, Padmanabhan R, Verma J, et al. Microbiome Dysbiosis and Epigenetic Modulations in Lung Cancer: From Pathogenesis to Therapy. *Semin Cancer Biol* (2021). doi: 10.1016/j.semcancer.2021.07.005
22. He D, Wang D, Lu P, Yang N, Xue Z, Zhu X, et al. Single-Cell RNA Sequencing Reveals Heterogeneous Tumor and Immune Cell Populations in Early-Stage Lung Adenocarcinomas Harboring EGFR Mutations. *Oncogene* (2021) 40(2):355–68. doi: 10.1038/s41388-020-01528-0
23. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAbiolinks: An R/Bioconductor Package for Integrative Analysis of TCGA Data. *Nucleic Acids Res* (2016) 44(8):e71. doi: 10.1093/nar/gkv1507
24. Ryan M, Wong WC, Brown R, Akbani R, Su X, Broom B, et al. TCGASpliceSeq a Compendium of Alternative mRNA Splicing in Cancer. *Nucleic Acids Res* (2016) 44(D1):D1018–1022. doi: 10.1093/nar/gkv1288
25. Conway JR, Lex A, Gehlenborg N. UpSetR: An R Package for the Visualization of Intersecting Sets and Their Properties. *Bioinformatics* (2017) 33:2938–40. doi: 10.1093/bioinformatics/btx364
26. Piva F, Giulietti M, Nocchi L, Principato G. SpliceAid: A Database of Experimental RNA Target Motifs Bound by Splicing Proteins in Humans. *Bioinformatics* (2009) 25:1211–3. doi: 10.1093/bioinformatics/btp124
27. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J Proteome Res* (2019) 18:623–32. doi: 10.1021/acs.jproteome.8b00702
28. Engebretsen S, Bohlin J. Statistical Predictions With Glmnet. *Clin Epigenet* (2019) 11:123. doi: 10.1186/s13148-019-0730-1
29. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust Enumeration of Cell Subsets From Tissue Expression Profiles. *Nat Methods* (2015) 12(5):453–7. doi: 10.1038/nmeth.3337
30. Hänzelmann S, Castelo R, Guinney J. GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinf* (2013) 14:7. doi: 10.1186/1471-2105-14-7
31. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring Tumour Purity and Stromal and Immune Cell Admixture From Expression Data. *Nat Commun* (2013) 4:2612. doi: 10.1038/ncomms3612
32. Burugu S, Dancsok A, Nielsen T. Emerging Targets in Cancer Immunotherapy. *Semin Cancer Biol* (2018) 52:39–52. doi: 10.1016/j.semcancer.2017.10.001
33. Gaikwad S, Agrawal M, Kaushik I, Ramachandran S., Srivastava S.. Immune Checkpoint Proteins: Signaling Mechanisms and Molecular Interactions in Cancer Immunotherapy. *Semin Cancer Biol* (2022). doi: 10.1016/j.semcancer.2022.03.014
34. Richard C, Fumet JD, Chevrier S, Derangère V, Ledys F, Lagrange A, et al. Exome Analysis Reveals Genomic Markers Associated With Better Efficacy of Nivolumab in Lung Cancer Patients. *Clin Cancer Res* (2019) 25(3):957–66. doi: 10.1158/1078-0432.CCR-18-1940
35. Geeleher P, Cox N, Huang R. Clinical Drug Response can be Predicted Using Baseline Gene Expression Levels and *In Vitro* Drug Sensitivity in Cell Lines. *Genome Biol* (2014) 15:R47. doi: 10.1186/gb-2014-15-3-r47
36. Geeleher P, Cox N, Huang R. Prrophetic: An R Package for Prediction of Clinical Chemotherapeutic Response From Tumor Gene Expression Levels. *PloS One* (2014) 9:e107468. doi: 10.1371/journal.pone.0107468
37. Chang YS, Tu SJ, Chiang HS, Yen JC, Lee YT, Fang HY, et al. Genome-Wide Analysis of Prognostic Alternative Splicing Signature and Splicing Factors in Lung Adenocarcinoma. *Genes (Basel)* (2020) 11(11). doi: 10.3390/genes11111300
38. Blencowe B. The Relationship Between Alternative Splicing and Proteomic Complexity. *Trends Biochem Sci* (2017) 42:407–8. doi: 10.1016/j.tibs.2017.04.001
39. Wang E, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, et al. Alternative Isoform Regulation in Human Tissue Transcriptomes. *Nature* (2008) 456(7221):470–6. doi: 10.1038/nature07509
40. Pio R, Montuenga L. Alternative Splicing in Lung Cancer. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer* (2009) 4:674–8. doi: 10.1097/JTO.0b013e3181a520dc
41. Taylor J, Zhang Q, Wyatt J, Dean N. Induction of Endogenous Bcl-xS Through the Control of Bcl-X pre-mRNA Splicing by Antisense Oligonucleotides. *Nat Biotechnol* (1999) 17:1097–100. doi: 10.1038/15079
42. Bauman J, Li S, Yang A, Huang L, Kole R. Anti-Tumor Activity of Splice-Switching Oligonucleotides. *Nucleic Acids Res* (2010) 38:8348–56. doi: 10.1093/nar/gkq731
43. Gautschi O, Tschopp S, Olie R, Leech S, Simões-Wüst A, Ziegler A, et al. Activity of a Novel Bcl-2/bcl-xL-Bispecific Antisense Oligonucleotide Against Tumors of Diverse Histologic Origins. *J Natl Cancer Institute* (2001) 93(6):463–71. doi: 10.1093/jnci/93.6.463
44. Zhu L, Zhu Y, Dai D, Wang X, Jin H. Epigenetic Regulation of Alternative Splicing. *Am J Cancer Res* (2018) 8(12):2346–58.
45. Zhang L, Yang H, Zhang W, Liang Z, Huang Q, Xu G, et al. Clk1-Regulated Aerobic Glycolysis is Involved in Glioma Chemoresistance. *J neurochemistry* (2017) 142(4):574–88. doi: 10.1111/jnc.14096
46. Babu N, Pinto S, Biswas M, Subbannayya T, Rajappa M, Mohan S, et al. Phosphoproteomic Analysis Identifies CLK1 as a Novel Therapeutic Target in Gastric Cancer. *Gastric Cancer* (2020) 23(5):796–810. doi: 10.1007/s10120-020-01062-8
47. Saldivia M, Fang E, Ma X, Myburgh E, Carnielli J, Bower-Lepts C, et al. Targeting the Trypanosome Kinetochore With CLK1 Protein Kinase Inhibitors. *Nat Microbiol* (2020) 5(10):1207–16. doi: 10.1038/s41564-020-0745-6
48. Xu Z, Li X, Li H, Nie C, Liu W, Li S, et al. Suppression of DDX39B Sensitizes Ovarian Cancer Cells to DNA-Damaging Chemotherapeutic Agents *via* Destabilizing BRCA1 mRNA. *Oncogene* (2020) 39(47):7051–62. doi: 10.1038/s41388-020-01482-x
49. Silva A, Faria M, Matos P. Inflammatory Microenvironment Modulation of Alternative Splicing in Cancer: A Way to Adapt. *Adv Exp Med Biol* (2020) 1219:243–58. doi: 10.1007/978-3-030-34025-4_13
50. Anderson T, Wooster A, Piersall S, Okpalanwaka I, Lowe D.. Disrupting Cancer Angiogenesis and Immune Checkpoint Networks for Improved Tumor Immunity. *Semin Cancer Biol* (2022). doi: 10.1016/j.semcancer.2022.02.009
51. Liu H, Zhao H, Sun Y. Tumor Microenvironment and Cellular Senescence: Understanding Therapeutic Resistance and Harnessing Strategies. *Semin Cancer Biol* (2021). doi: 10.1016/j.semcancer.2021.11.004
52. Zhou R, Moshgabadi N, Adams K. Extensive Changes to Alternative Splicing Patterns Following Allopolyploidy in Natural and Resynthesized Polyploids. *Proc Natl Acad Sci USA* (2011) 108:16122–7. doi: 10.1073/pnas.1109551108
53. Forghanifard M, Aarabi A, Nasiri Aghdam M, Memar B, Hasanzadeh Khayat M, Dadkhah E, et al. GSTs Polymorphisms are Associated With Epigenetic Silencing of CDKN2A Gene in Esophageal Squamous Cell Carcinoma. *Environ Sci pollut Res Int* (2020) 27(25):31269–77. doi: 10.1007/s11356-020-09408-6
54. Dabanaka K, Chung S, Nakagawa H, Nakamura Y, Okabayashi T, Sugimoto T, et al. PKIB Expression Strongly Correlated With Phosphorylated Akt Expression in Breast Cancers and Also With Triple-Negative Breast Cancer Subtype. *Med Mol morphology* (2012) 45(4):229–33. doi: 10.1007/s00795-011-0565-0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 20 July 2022
doi: 10.3389/fonc.2022.900110

Electrical Impedance Analysis for Lung Cancer: A Prospective, Multicenter, Blind Validation Study

Dawei Yang[1,2,3,4†], Chuanjia Gu[5,6†], Ye Gu[7], Xiaodong Zhang[8], Di Ge[9], Yong Zhang[1], Ningfang Wang[1], Xiaoxuan Zheng[5,6], Hao Wang[7], Li Yang[7], Saihua Chen[8], Pengfei Xie[8], Deng Chen[10], Jinming Yu[10], Jiayuan Sun[5,6,11] and Chunxue Bai[1,2,3,4*]*

[1] Department of Pulmonary Medicine and Critical Care Medicine, Zhongshan Hospital, Fudan University, Shanghai, China, [2] Shanghai Respiratory Research Institution, Shanghai, China, [3] Chinese Alliance Against Lung Cancer, Shanghai, China, [4] Shanghai Engineer & Technology Research Center of Internet of Things for Respiratory Medicine, Shanghai, China, [5] Department of Respiratory Endoscopy, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China, [6] Department of Respiratory and Critical Care Medicine, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China, [7] Department of Thoracic Surgery, Shanghai Pulmonary Hospital, Tongji University, Shanghai, China, [8] Department of Pulmonary Medicine, Nantong Tumor Hospital, Nantong, China, [9] Department of Thoracic Surgery, Zhongshan Hospital, Fudan University, Shanghai, China, [10] Key Laboratory of Public Health Safety, Ministry of Education, School of Public Health, Fudan University, Shanghai, China, [11] Shanghai Engineering Research Center of Respiratory Endoscopy, Shanghai, China

OPEN ACCESS

Edited by:
Yutong He,
Fourth Hospital of Hebei Medical University, China

Reviewed by:
Beatrice Aramini,
University of Bologna, Italy
Qi Wang,
Second Affiliated Hospital of Dalian Medical University, China

***Correspondence:**
Chunxue Bai
bai.chunxue@zs-hospital.sh.cn
Jiayuan Sun
xkyyjysun@163.com

[†]These authors share first authorship

Specialty section:
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

Received: 20 March 2022
Accepted: 17 June 2022
Published: 20 July 2022

Citation:
Yang D, Gu C, Gu Y, Zhang X, Ge D, Zhang Y, Wang N, Zheng X, Wang H, Yang L, Chen S, Xie P, Chen D, Yu J, Sun J and Bai C (2022) Electrical Impedance Analysis for Lung Cancer: A Prospective, Multicenter, Blind Validation Study. Front. Oncol. 12:900110. doi: 10.3389/fonc.2022.900110

Hypothesis: Patients with cancer have different impedances or conductances than patients with benign normal tissue; thus, we can apply electrical impedance analysis (EIA) to identify patients with cancer.

Method: To evaluate EIA's efficacy and safety profile in diagnosing pulmonary lesions, we conducted a prospective, multicenter study among patients with pulmonary lesions recruited from 4 clinical centers (Zhongshan Hospital Ethics Committee, Approval No. 2015-16R and 2017-035(3). They underwent EIA to obtain an Algorithm Composite Score or 'Prolung Index,' PI. The classification threshold of 29 was first tested in an analytical validation set of 144 patients and independently validated in a clinical validation set of 418 patients. The subject's final diagnosis depended on histology and a 2-year follow-up.

Results: In total, 418 patients completed the entire protocol for clinical validation, with 186 true positives, 145 true negatives, 52 false positives, and 35 false negatives. The sensitivity, specificity, and diagnostic yield were 84% (95% CI 79.3%-89.0%), 74% (95% CI 67.4%-79.8%), and 79% (95%CI 75.3%-83.1%), respectively, and did not differ according to age, sex, smoking history, body mass index, or lesion types. The sensitivity of small lesions was comparable to that of large lesions (p = 0.13). Four hundred eighty-four patients who underwent the analysis received a safety evaluation. No adverse events were considered to be related to the test.

Conclusion: Electrical impedance analysis is a safe and efficient tool for risk stratification of pulmonary lesions, especially for patients with a suspicious lung lesion.

Keywords: electrical impedance, lung cancer, pulmonary nodules, diagnosis, prospective

HIGHLIGHTS

- What is the key question?

Could we apply electrical impedance analysis (EIA) to identify patients with cancer?

- What is the bottom line?

EIA is a safe and efficient tool for risk stratification of pulmonary lesions, especially for patients presenting with a suspicious lung lesion.

- Why read on?

As a non-invasive test, EIA can be adjunctively incorporated with CT screening to both avoid overdiagnosis and missed diagnosis.

INTRODUCTION

Lung cancer has become the most common incident cancer and the leading cause of cancer death in China (1). Low-dose computed tomography (LDCT) screening has reduced the mortality in high-risk populations (2, 3), but since LDCT involves the use of radiation, asymptomatic people are reluctant to undergo routine screening. Consequently, there is an unmet need for a non-invasive, radiation-free, and easy-to-use diagnostic tool for lung cancer risk stratification. Electrical impedance analysis (EIA) can be used to obtain impedance information related to human physiological and pathological conditions. In EIA, an electrode probe is placed on the body, and a small amount of current is passed through the body, allowing analysis of its impedance. The application scope of EIA includes monitoring pulmonary function, constructing functional brain imaging, evaluating body composition and nutrition status, and identifying tumors (4).

Bioimpedance has been valuable in detecting various cancers, including skin, thyroid, liver, cervix, and breast cancers (5–12), with breast cancer being the most extensively studied bioimpedance technology (9–12). It is well known that cancerous tissue has distinctly different electrical properties from non-cancerous tissue (13). This has been attributed to the high water and sodium content within cancerous tissues, altered membrane composition, the nucleus-to-cytoplasm ratio, and cellular composition and density (14–16). To construct an impedance analytical system for lung cancer detection, Kimura et al. (17) inserted an electric probe into the pulmonary mass during thoracotomy among 53 patients (17). They diagnosed 9 patients with intrathoracic lesions using the analytical system confirmed by a needle biopsy, resulting in no false negatives and only one false positive (17). Transcutaneous measurement of dermal impedance has been developed as an indication of internal organ pathologies, which was confirmed to be helpful in the diagnosis of lung cancer (18, 19). Current devices on the market to characterize biopsied tissue utilize what is generally referred to as Electrical Impedance Spectroscopy (EIS). Additional devices that monitor or image the various electrical properties of tissues are typically referred to as using Electrical Impedance Tomography (EIT). While the ProLung device shares some common aspects of these other technologies, it is unique in that the device does not measure or image the tumor nodule directly, but rather measures the systemic changes (bulk resistive changes to the interstitial fluids within the extracellular matrix and lymph system due to the presence of cancer in the body), which are significant and measurable due to the presence of cancer in the body (14, 20–26). Therefore, its accuracy is less affected by the size of the lesion, unlike traditional imaging technologies such as CT and PET scans.

In our previous study, we developed an EIA approach using 31 bilateral points on the skin surface (27). We achieved 89.7% sensitivity and 91.7% specificity in distinguishing between a cohort of lung cancer patients and healthy volunteers. This study optimized the detection sites to 20 skin surface points and updated the algorithm to compute an Algorithm Composite Score combining the impedance results from all detection sites. Based on our previous study, we enrolled more participants in this prospective, multicenter, blind validation study aiming to confirm EIA's efficacy and safety profile in lung cancer diagnosis.

MATERIALS AND METHODS

Participants

From June 2015 to August 2019, we recruited consecutive patients with pulmonary lesions suspected to be lung cancer from 4 clinical centers in China: Zhongshan Hospital Fudan University, Shanghai Chest Hospital, Shanghai Pulmonary Hospital, and Nantong Tumor Hospital. Patients recruited between June 2015 and June 2016 were enrolled in an analytical validation set for testing of the previously reported method and threshold. Patients recruited between October 2017 and August 2019 were enrolled in a clinical validation set for independent validation.

Inclusion criteria were as follows (1). Age 18–80 years (2). Presence of a pulmonary lesion 4–50 mm in diameter (3). Provision of a CT/PET (positron emission tomography) within 30 days. Exclusion criteria were as follows (1). Benign tumors of the central nervous system or other known malignancies, except for non-melanoma skin cancer, during the past five years (2). Confounding factors affecting thoracic impedance, including apparent pulmonary inflammation, tuberculosis, pleural effusion, thoracic anatomy abnormality, thoracic interventional therapy, implanted electric devices, dermatosis, and thoracic radiation or chemotherapy within the last 30 days (3). Other factors affecting thoracic anatomy and conductivity properties, such as strenuous exercise within the last 24 hours (4). Pregnancy or lactation (5). Presence of an unusually low conductivity, such as an Algorithm Composite Score of <20 when measured between the two hands following a 5-minute dwell time.

All patients provided written informed consent and agreed to undergo histological diagnosis. The ethics committee approved the protocol at each hospital. The trial was registered at www.

clinicaltrials.gov (NCT02726633). All procedures involving human participants were as per the Declaration of Helsinki.

Procedure

The EIA (BSP-E2-1000-A, Prolung Biotech Wuxi Co, Wuxi, China) comprised a host computer, a reference electrode, and an impedance detector integrated with an electric probe (**Figure 1**). The electric probe passes a weak current (≤25 μA) through the body, forming a series circuit with the reference electrode. By sending a standard induction voltage to the interrogation location, the voltage difference across the body is detected, thus allowing the bioimpedance to be calculated. The operator measured 20 given sites on the surface of the skin. All data were delivered automatically to the host computer, and the algorithm generated an Algorithm Composite Score (or 'Prolung Index,' PI) for each point measured. The PI reveals the overall conductance property of the participants, which is the inverse of bioimpedance. Previous studies have shown that EIA can effectively distinguish malignancies from benign conditions; patients with Algorithm Composite Score ≥29 have a high risk for malignancy. Those with Algorithm Composite Score <29 are at low risk (27, 28). The same threshold was tested and validated in this study.

Follow-Up

A final diagnosis was established by a comprehensive analysis of pathology and clinical follow-up. When the diagnosis was pathologically confirmed by surgery or biopsy, no follow-up

FIGURE 1 | Electrical impedance analysis (EIA) platform.

was needed for newly discovered pulmonary lesions. A definitive result confirmed by biopsy should be a conclusive malignancy or a specific benign condition, such as a granuloma, fibrosis, or clear microbiological evidence.

In cases where a histological diagnosis was not performed, or the histology was indeterminate by biopsy, the patient underwent a 2-year clinical follow-up until further intervention was performed and a definitive histological diagnosis was established. Solid lesions that remained stable after two years were recorded as benign. Subsolid lesions were further discussed by a multi-disciplinary team consisting of two radiologists and one respiratory physician to determine whether they were benign. Histological examination was recommended when a follow-up CT showed morphological changes defined by the Clinical Practice Consensus Guidelines (29). Additional details on this method are provided in the online data supplement.

Statistical Analysis

Statistical Analysis Software (SAS) (version 9.4) was used for statistical analyses. Descriptive statistics were recorded as frequency, percentage, range, and mean ± standard deviation. Categorical variables were compared using a chi-square test or Fisher's exact probability test. All tests were performed bilaterally. A p value less than 0.05 was considered to be statistically significant.

RESULTS

Study Cohorts

One hundred sixty-three participants with CT-detected pulmonary lesions were prospectively enrolled in the analytical validation set. Of these, 19 patients dropped out without a definitive diagnosis, resulting in 144 eligible patients for threshold testing. The baseline characteristics of the patients are shown in **Table 1**. Of the 144 cases, 60 had surgery, 68 had a nonsurgical bronchoscopic biopsy, and 16 had a nonsurgical fine-needle aspiration (FNA) biopsy. All nonsurgical patients were followed up for two years. The final diagnoses revealed 25 benign lesions and 119 lung cancers, including 20 squamous cell carcinomas, 88 adenocarcinomas, 3 other types of non-small-cell lung cancer, 4 small cell lung cancers, and 4 malignancies not otherwise specified.

Four hundred eighty-four participants were prospectively enrolled in the study to validate the diagnostic performance of EIA (**Figure 2**). Of these, 42 patients did not meet the inclusion criteria and were excluded. Another 24 patients dropped out without a definitive diagnosis, resulting in 418 patients in the final validation set. The baseline characteristics of patients are shown in **Table 1**. Of the 418 cases in the clinical validation set, 183 had surgery, 94 had a nonsurgical bronchoscopic biopsy or FNA biopsy, and 141 had at least two years of clinical follow-up. The final diagnoses revealed 197 benign lesions and 221 malignancies, including 19 squamous cell carcinomas, 190 adenocarcinomas, 4 small cell lung cancers, and 6 malignancies not otherwise specified, as well as 1 non-small-cell lung cancer confirmed by pathology at Shanghai Chest Hospital and 1 combined adenocarcinoma and large cell neuroendocrine carcinoma at Shanghai Pulmonary Hospital.

Diagnostic Efficacy and Ease of Operation of EIA

EIA was first performed in the analytical validation set. There were 98 true positives, 18 true negatives, 7 false positives, and 21 false negatives in the EIA analysis (**Table 2**). The overall sensitivity, specificity, positive predictive value, negative predictive value, and diagnostic yield were 82% (95% CI 75.5%-89.2%), 72% (95% CI 54.4%-89.6%), 93% (95% CI 88.6%-98.1%), 46% (95% CI 30.5%-61.8%), and 81% (95% CI 74.1%-87.0%), respectively. These results indicated that the Algorithm Composite Score threshold previously developed using a North American cohort was sufficiently applicable for Chinese cohorts.

In the independent clinical validation set, there were 186 true positives, 145 true negatives, 52 false positives, and 35 false negatives in the EIA analysis (**Table 2**). The overall sensitivity, specificity, positive predictive value, negative predictive value, and diagnostic yield were 84% (95% CI 79.3%-89.0%), 74% (95% CI 67.4%-79.8%), 78% (95% CI 72.9%-83.4%), 81% (95% CI 74.8%-86.3%), and 79% (95% CI 75.3%-83.1%), respectively. The sensitivity, specificity, and diagnostic yield were statistically comparable for the four clinical centers. However, the positive (p = 0.005) and negative predictive values (p = 0.004) differed between the four clinical centers, potentially due to differences in their benign/malignant case ratios. The overall kappa value was 0.58 (95% CI 0.50–0.65), indicating moderate consistency.

Influence of Clinical Variables on the Diagnostic Yield

The sensitivity, specificity, and diagnostic yield of EIA did not differ according to age, sex, body mass index, or patients' smoking history (**Tables 3**, **4**). However, patients with malignant pulmonary lesions (85%) yielded a higher sensitivity than the specificity of patients with benign lesions (56%, $p < 0.01$, **Table 4**). For lesions <10 mm, 10–30 mm, and 30–50 mm, respectively, the diagnostic yield was 82% (95% CI 70.0%-94.1%), 77% (95% CI 70.5%-82.6%), and 84% (95% CI 73.8%-94.2%) (p = 0.45), with sensitivity of 88% (95% CI 76.2%-100.0%), 82% (95% CI 76.3%-88.4%), 90% (95% CI 81.2%-99.3%) (p = 0.35), and specificity of 69% (95% CI 44.1%-94.3%), 51% (95% CI 34.9%-67.0%), 56% (95% CI 23.1%-88.0%) (p = 0.55). The specificity decreased when evaluating subjects with large pulmonary lesions (>10 mm). The sensitivity, specificity and diagnostic yield did not vary among different lesions types, including solid lesions, pure GGOs, mixed GGOs, and patchy shadows.

Because a 2-year clinical follow-up was insufficient to establish a definitive diagnosis for subsolid lesions, we tabulated the diagnostic results of EIA. We then compared them with the findings obtained from pathology and clinical

TABLE 1 | Baseline characteristics of the patients.

Characteristics	Analytical Validation Dataset (N=144)	Clinical Validation Dataset				
		Total (N=418)	Center A (n=137)	Center B (n=149)	Center C (n=79)	Center D (n=53)
Age, year						
Mean (SD)	60.2 (9.41)	57.8 (11.6)	57.3 (12.2)	59.0 (11.4)	53.4 (11.2)	62.3 (8.62)
Median [Min, Max]	61.0 [31.0, 79.0]	60 [20, 80]	60 [20, 80]	61 [32, 80]	56 [26, 73]	64 [45, 77]
Gender, n (%)						
Male	90 (62.5)	185 (44.3)	59 (43.1)	63 (42.3)	35 (44.3)	28 (52.8)
Female	54 (37.5)	233 (55.7)	78 (56.9)	86 (57.7)	44 (55.7)	25 (47.2)
BMI						
Mean (SD)	23.7 (3.20)	23.3 (3.27)	23.2 (3.03)	23.2 (3.50)	23.4 (2.88)	24.0 (3.74)
Median [Min, Max]	23.6 [15.4, 31.1]	23.1 [14.5, 37.9]	23.4 [16.9, 37.9]	22.8 [14.5, 31.2]	23.3 [18.0, 31.9]	23.1 [17.6, 32.0]
Smoke, n (%)						
No	101 (70.1)	305 (73.0)	101 (73.7)	106 (71.1)	63 (79.7)	35 (66.0)
Yes	43 (29.9)	113 (27.0)	36 (26.3)	43 (28.9)	16 (20.3)	18 (34.0)
Lesion Size, mm						
Mean (SD)	28.0 (13.1)	16.0 (10.8)	14.6 (10.2)	17.6 (10.6)	11.3 (8.96)	22.5 (11.6)
Median [Min, Max]	26.0 [4.90, 50.0]	13.0 [4.00, 50.0]	11.0 [4.00, 48.3]	15.0 [4.00, 46.0]	8.0 [4.00, 46.0]	21.0 [4.00, 50.0]
Lobe Location, n (%)						
RLL/LLL	56 (38.9)	117 (28.0)	33 (24.1)	42 (28.2)	16 (20.3)	26 (49.1)
RML	19 (13.2)	47 (11.2)	18 (13.1)	10 (6.71)	13 (16.5)	6 (11.3)
RUL/LUL	69 (47.9)	254 (60.8)	86 (62.8)	97 (65.1)	50 (63.3)	21 (39.6)
Lesion Type, n (%)						
MGGO	7 (4.9)	110 (26.3)	28 (20.4)	30 (20.1)	16 (20.3)	36 (67.9)
PGGO	21 (14.6)	146 (34.9)	55 (40.1)	40 (26.8)	45 (57.0)	6 (11.3)
Solid	111 (77.1)	161 (38.5)	54 (39.4)	79 (53.0)	18 (22.8)	10 (18.9)
Other	5 (3.5)	1 (0.2)	0 (0)	0 (0)	0 (0)	1 (1.9)
Final Diagnosis, n (%)						
Benign	25 (17.4)	197 (47.1)	72 (52.6)	62 (41.6)	54 (68.4)	9 (17.0)
Malignant	119 (82.6)	221 (52.9)	65 (47.4)	87 (58.4)	25 (31.6)	44 (83.0)
SQ	20 (13.9)	19 (4.5)	3 (2.2)	10 (6.7)	2 (2.5)	4 (7.5)
Ad	88 (61.1)	190 (45.5)	61 (44.5)	69 (46.3)	22 (27.8)	38 (71.7)
SCLC	4 (2.8)	4 (1.0)	1 (0.7)	1 (0.7)	0 (0)	2 (3.8)
NOS	4 (2.8)	6 (1.4)	0 (0)	6 (4.0)	0 (0)	0 (0)
Other malignancy	0 (0)	2 (0.5)	0 (0)	1 (0.7)	1 (1.3)	0 (0)
Diagnostic Method, n (%)						
Biopsy & follow-up	84 (58.3)	84 (20.1)	14 (10.2)	69 (46.3)	0 (0)	1 (1.9)
Surgery	60 (41.7)	193 (46.2)	76 (55.5)	36 (24.2)	31 (39.2)	50 (94.3)
Follow-up only		141 (33.7)	47 (34.3)	44 (29.5)	48 (60.8)	2 (3.8)

AD, adenocarcinoma; Center A, Zhongshan Hospital Fudan University; Center B, Shanghai Chest Hospital; Center C, Shanghai Pulmonary Hospital; Center D, Nantong Tumor Hospital; FNA, fine needle aspiration; LLL, left lower lobe; LUL, left upper lobe; MGGO, mixed ground-glass opacity; NOS, not otherwise specified; NSCLC, non-small cell lung cancer; Another malignancy refers to 1 non-small-cell lung cancer confirmed by pathology in Shanghai Chest Hospital and one combined adenocarcinoma and large cell neuroendocrine carcinoma in Shanghai Pulmonary Hospital; PGGO, pure ground-glass opacity; RLL, right lower lobe; RML, right middle lobe; RUL, right upper lobe; SD, standard deviation; SQ, squamous cell carcinoma; SCLC, small cell lung cancer.

follow-up (**Table 5**). Of the pure GGOs, 44.52% (65/146) were diagnosed by pathology, resulting in a sensitivity, specificity, and diagnostic yield of 78% (45/58, 95% CI 66.9%-88.3%), 57% (4/7, 95% CI 20.5%-93.8%), 75% (49/65, 95% CI 64.9%-85.9%), respectively. A total of 81 pure GGOs were established through a final diagnosis by clinical follow-up, resulting in a sensitivity, specificity, and diagnostic yield of 100% (1/1), 80% (64/80, 95% CI 71.2%-88.8%), and 80% (65/81, 95% CI 71.6%-88.9%), respectively. Of the mixed GGOs, 81% (89/110) were diagnosed by pathology, resulting in a sensitivity, specificity, and diagnostic yield of 84% (64/76, 95% CI 76.0%-92.4%), 54% (7/13, 95% CI 26.7%-80.9%), 80% (71/89, 95% CI 71.4%-88.1%), respectively. A total of 21 mixed GGOs were established through a final diagnosis by clinical follow-up, resulting in a sensitivity, specificity, and diagnostic yield of 100% (1/1), 75% (15/20, 95% CI 56.0%-94.0%), and 76% (16/21, 95% CI 58.0%-94.4%), respectively.

Complications

The safety evaluation was conducted on 484 patients who were evaluated by EIA. No patient discomfort related to the measurement procedure was reported during the course of, or within 24 hours of, the operation. However, one patient died of lung cancer progression during clinical follow-up.

DISCUSSION

In this prospective, multicenter study assessing the use of EIA as a diagnostic tool for Chinese patients with pulmonary lesions,

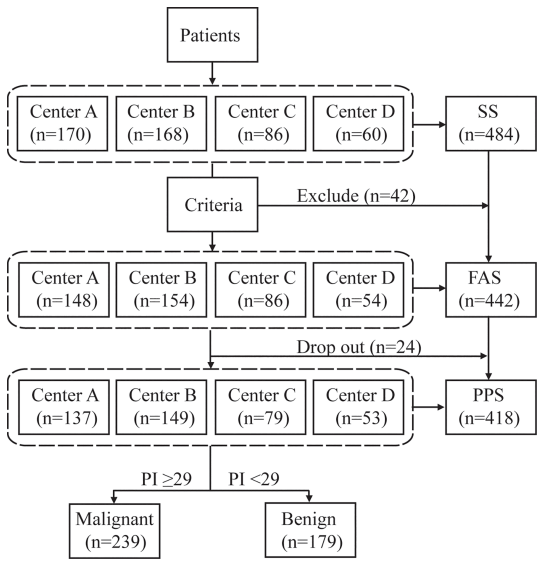


FIGURE 2 | Study design flow chart. The safety dataset (SS) was composed of all subjects who underwent EIA and had at least one safety evaluation. The full analysis set (FAS) was made up of all eligible participants who underwent EIA. The per protocol set (PPS) consisted of eligible patients who completed the whole study and excluded severe violations of the protocol. Forty-two patients who were not suitable for the test were excluded, of whom 21 did not have eligible CT, 9 had Algorithm Composite Score < 20, 4 had pulmonary mass >50 mm, 3 had nodules <4 mm, 2 received thoracic intervention therapy, 2 had tuberculosis, 1 had an implanted steel plate in the thorax, one was aged >80 years, 1 had cancer history within five years. Twenty-four patients dropped out of the study, 23 due to the absence of compliance, and one died of cancer progression.

EIA was shown to be capable of safely and accurately discriminating between malignant and benign lesions with high sensitivity, specificity, and diagnostic yield. This accuracy was seen to be unaffected by patient demographics and clinical characteristics such as age, sex, smoking history, body mass index, or lesion types. Of note, the sensitivity associated with small lesions was comparable to that for large lesions.

EIA has long been used for electrocardiographs and electroencephalograms. It has also been utilized in skin cancer identification, thyroid nodule differentiation, and breast cancer risk stratification and screening (30–34). Stojadinovic et al. conducted breast cancer screening among 1,103 young women using EIA and established a sensitivity of 50% and a specificity of 90% (32, 33). No large-scale clinical trials have been conducted to evaluate EIA's utility in pulmonary lesion risk stratification.

The American National Lung Screening Trial (NLST) demonstrated a relative reduction in lung cancer mortality of 20% following implementation of screening using three annual low-dose CT scans compared with planar chest radiographs among a high-risk population (3). However, using the NLST data, Bach et al. (35) estimated that approximately one radiation-associated cancer death would result per 2,500 people screened (35). Unlike CT, EIA does not employ ionizing radiation and so would be expected to have a lower risk of screening associated malignancies. CT screening is a trade-off that benefits people with potential high risk (aged 55–74 years, smoking history ≥30 pack-years). But for the low-risk population, the harm may outweigh the benefits (35, 36).

Following CT screening, false negatives result in late diagnosis and poor prognosis. False-positive CT readings lead to psychological distress, more frequent follow-up exposure to ionizing radiation (full dose CTs and PET-CT scans), and potentially unnecessary and harmful invasive procedures. Individuals with indeterminate pulmonary nodules >8 mm are recommended to undergo diagnostic procedures, such as PET, nonsurgical biopsy, surgery, and CT surveillance (37). Although the non-invasive PET has a sensitivity of 72–94% for malignant lesions, it is not very effective in diagnosing small nodules (<8–10 mm), pure GGOs, and mixed GGOs with a solid component ≤8 mm (37, 38), for which EIA has better performance than PET. The diagnostic yield of EIA for pure GGOs (78%) was not inferior to that for solid nodules (80%), mixed GGOs (79%), and patchy shadows (100%; p = 0.37), as was its sensitivity (p = 0.39) and specificity (p = 0.32). Additionally, EIA has a very impressive sensitivity of 85% for nodules <10 mm, which allows it to identify lung cancer at a very early stage. Based upon combined results from EIA and CT, clinicians could recommend

TABLE 2 | The diagnostic efficacy and ease of operation evaluation of electrical impedance analysis by different clinical centers.

Indicators	Analytical Validation Dataset (N=144)	Clinical Validation Dataset (N=418)	*P* Value	Clinical Validation Dataset				
				Center A (n=137)	Center B (n=149)	Center C (n=79)	Center D (n=53)	*P* value
ACC	0.81 (116/144)	0.79 (331/418)	0.817	0.77 (105/137)	0.78 (116/149)	0.82 (65/79)	0.85 (45/53)	0.530
Sens	0.82 (98/119)	0.84 (186/221)	0.783	0.85 (55/65)	0.83 (72/87)	0.88 (22/25)	0.84 (37/44)	0.944
Spec	0.72 (18/25)	0.74 (145/197)	1.000	0.69 (50/72)	0.71 (44/62)	0.80 (43/54)	0.89 (8/9)	0.394
PPV	0.93 (98/105)	0.78 (186/238)	0.001	0.71 (55/77)	0.80 (72/90)	0.67 (22/33)	0.97 (37/38)	0.005
NPV	0.46 (18/39)	0.81 (145/180)	<0.001	0.83 (50/60)	0.75 (44/59)	0.93 (43/46)	0.53 (8/15)	0.004
Kappa Value	0.445	0.580		0.535	0.541	0.623	0.577	
Easy Operation	100%	100%		100%	100%	100%	100%	

Center A, Zhongshan Hospital Fudan University; Center B, Zhongshan Hospital Fudan University; Center C, Shanghai Pulmonary Hospital; Center D, Nantong Tumor Hospital; ACC, accuracy; Sens, sensitivity; Spec, specificity; PPV, positive predictive value; NPV, negative predictive value.

When kappa value < 0, the consistency intensity is extremely poor; 0 ~ 0.20, faint; 0.21 ~ 0.40, weak; 0.41 ~ 0.60, moderate; 0.61 ~ 0.80, high; kappa > 0.81, extremely strong.

TABLE 3 | Diagnostic yield in the analytical validation dataset by different variables.

Variable	Diagnostic yield	*P-Value*	Sensitivity	*P-Value*	Specificity	*P-Value*
Age, year						
18~44	0.50 (4/8)	0.092	0.75 (3/4)	0.907	0.25 (1/4)	0.054
45~69	0.82 (91/111)		0.82 (75/92)		0.84 (16/19)	
≥70	0.84 (21/25)		0.87 (20/23)		0.50 (1/2)	
Gender						
Male	0.79 (71/90)	0.664	0.82 (59/72)	1.000	0.67 (12/18)	0.626
Female	0.83 (45/54)		0.83 (39/47)		0.86 (6/7)	
BMI						
<24	0.86 (66/77)	0.143	0.86 (57/66)	0.299	0.82 (9/11)	0.407
≥24	0.75 (50/67)		0.77 (41/53)		0.64 (9/14)	
Smoke						
No	0.84 (85/101)	0.149	0.85 (72/85)	0.425	0.81 (13/16)	0.205
Yes	0.72 (31/43)		0.76 (26/34)		0.56 (5/9)	
Lesion Size, mm						
<10	0.70 (7/10)	0.017	0.71 (5/7)	0.026	0.67 (2/3)	1.000
10~30	0.74 (56/76)		0.75 (47/63)		0.69 (9/13)	
30~50	0.91 (53/58)		0.94 (46/49)		0.78 (7/9)	
Lobe Location						
RLL/LLL	0.82 (46/56)	0.374	0.86 (37/43)	0.268	0.69 (9/13)	1.000
RML	0.68 (13/19)		0.69 (11/16)		0.67 (2/3)	
RUL/LUL	0.83 (57/69)		0.83 (50/60)		0.78 (7/9)	
Lesion Type						
MGGO	1 (7/7)	0.337	1 (5/5)	0.585	1 (2/2)	0.718
PGGO	0.76 (16/21)		0.76 (13/17)		0.75 (3/4)	
Solid	0.79 (88/111)		0.82 (77/94)		0.65 (11/17)	
Patchy shadows	1 (5/5)		1 (3/3)		1 (2/2)	
Final Diagnosis						
Benign	0.72 (18/25)	0.262			0.72 (18/25)	
Malignant	0.82 (98/119)		0.82 (98/119)			
Diagnostic Method						
Biopsy	0.83 (70/84)	0.434	0.87 (58/67)	0.260	0.71 (12/17)	1.000
Surgery	0.77 (46/60)		0.77 (40/52)		0.75 (6/8)	

AD, adenocarcinoma; Center A, Zhongshan Hospital Fudan University; Center B, Shanghai Chest Hospital; Center C, Shanghai Pulmonary Hospital; Center D, Nantong Tumor Hospital; FNA, fine needle aspiration; LLL, left lower lobe; LUL, left upper lobe; MGGO, mixed ground-glass opacity; NOS, not otherwise specified; NSCLC, non-small cell lung cancer; PGGO, pure ground-glass opacity; RLL, right lower lobe; RML, right middle lobe; RUL, right upper lobe; SD, standard deviation; SQ, squamous cell carcinoma; SCLC, small cell lung cancer.

a relatively conservative or a more aggressive intervention according to the test results.

In addition, an easy-to-use risk stratification method for use after CT screening is needed to avoid repeated radiation exposure in low-risk populations. As EIA had a good sensitivity (84%) and a better specificity (74%) than CT, it could be sufficiently accurate for use as a risk stratification tool. The positive (78%) and negative (81%) predictive values suggest that EIA could be used as a valid "rule out" test while effectively capturing early cancers. Unlike CT screening, EIA provides a direct and immediate conclusion after the test without requiring further interpretation by experts. When an EIA test result suggests a high malignancy risk, physicians could then advise the patients to undergo the necessary CT examinations or more invasive biopsies. EIA could also be beneficial for large-scale lung cancer risk stratification initiatives, especially in less developed geographical areas with limited access to medical professionals and advanced healthcare facilities, for people who are prone to psychological distress due to suspected illness, and for patients who cannot afford an annual physical examination.

We evaluated multiple variables that might affect the diagnostic efficacy of EIA. Since many participants had multiple pulmonary lesions, we only focused on lesions suspected to be malignant while regarding other lesions as normal. The sensitivity, specificity, and diagnostic yield were not affected by the patients' age, sex, or body mass index. This confirms that EIA can be generally used in a wide range of cases, especially when there is suspicion of lung cancer, as it was more sensitive for malignant lesions (84%) than benign lesions (74%; $p = 0.01$). For lesions <10 mm, 10–30 mm, and 30–50 mm, the sensitivity was not significantly different (85%, 82%, and 90%, respectively; $p = 0.43$). The specificity, however decreased significantly for lesions of these sizes (82%, 61%, and 50%, respectively; $p = 0.004$). With the high sensitivity and specificity in nodules smaller than 10 mm, EIA may be more effective in the risk stratification of small pulmonary lesions. At the same time, certain clinical considerations may be necessary for excluding false-positive cases when identifying large pulmonary lesions. Because of the low reported specificity of CT, patients with sub-centimeter nodules, which have lower cancer risk than larger lesions, may be required to undergo successive annual CT follow-ups. Using EIA risk stratification, patients with small nodules could obtain a cancer risk assessment immediately after CT detection to help decide the optimal frequency of subsequent CT follow-ups. In addition, EIA could be used as an adjunctive follow-up test so that patients only need

TABLE 4 | Diagnostic yield in the clinical validation dataset by different variables.

Variable	Diagnostic yield	*P-Value*	Sensitivity	*P-Value*	Specificity	*P-Value*
Age, year						
18-44	0.78 (29/37)	0.28	0.79 (22/28)	0.07	0.78 (7/9)	0.35
45-69	0.81 (156/193)		0.88 (134/152)		0.54 (22/41)	
≥70	0.70 (33/47)		0.75 (30/40)		0.43 (3/7)	
Gender						
Male	0.82 (106/130)	1.00	0.87 (91/105)	0.52	0.60 (15/25)	0.80
Female	0.81 (112/138)		0.83 (95/115)		0.74 (17/32)	
BMI						
<24	0.75 (123/163)	0.15	0.82 (105/128)	0.30	0.51 (18/35)	0.53
≥24	0.83 (95/114)		0.88 (81/92)		0.64 (14/22)	
Smoke						
No	0.78 (143/184)	0.68	0.83 (119/144)	0.38	0.60 (24/40)	0.54
Yes	0.81 (75/93)		0.88 (67/76)		0.47 (8/17)	
Lesion size, mm						
<10	0.82 (32/39)	0.45	0.88 (23/26)	0.35	0.69 (9/13)	0.55
10~30	0.77 (144/188)		0.82 (126/153)		0.51 (18/35)	
30~50	0.84 (42/50)		0.90 (37/41)		0.56 (5/9)	
Lobe location						
RLL/LLL	0.80 (83/104)	0.26	0.83 (66/80)	0.01	0.71 (17/24)	0.03
RML	0.65 (15/23)		0.61 (11/18)		0.80 (4/5)	
RUL/LUL	0.80 (120/150)		0.89 (109/122)		0.39 (11/28)	
Lesion type						
MGGO	0.79 (59/75)	0.15	0.81 (51/63)	0.03	0.67 (8/12)	0.66
PGGO	0.80 (83/104)		0.85 (73/86)		0.56 (10/18)	
Solid	0.66 (25/38)		0.73 (22/30)		0.38 (3/8)	
Patchy shadows	0.85 (51/60)		0.98 (40/41)		0.58 (11/19)	
Final diagnosis						
Benign	0.56 (32/57)	<0.01			0.56 (32/57)	
Malignant	0.85 (186/220)		0.85 (186/220)			
Diagnostic method						
Biopsy	0.75 (63/84)	0.40	0.85 (51/60)	1.00	0.50 (12/24)	0.60
Surgery	0.80 (155/193)		0.84 (135/160)		0.61 (20/33)	

AD, adenocarcinoma; Center A, Zhongshan Hospital Fudan University; Center B, Shanghai Chest Hospital; Center C, Shanghai Pulmonary Hospital; Center D, Nantong Tumor Hospital; FNA, fine needle aspiration; LLL, left lower lobe; LUL, left upper lobe; MGGO, mixed ground-glass opacity; NOS, not otherwise specified; NSCLC, non-small cell lung cancer; Another malignancy refers to 1 combined adenocarcinoma and large cell neuroendocrine carcinoma in Shanghai Pulmonary Hospital; PGGO, pure ground-glass opacity; RLL, right lower lobe; RML, right middle lobe; RUL, right upper lobe; SD, standard deviation; SQ, squamous cell carcinoma; SCLC, small cell lung cancer.

to receive CT detection when EIA indicates an increased risk. The distinct yet complementary diagnostic capabilities of EIA compared to CT make it a powerful tool to supplement, rather than challenge, the comprehensive analysis provided by CT imaging.

The sensitivity of EIA for lesions in the right middle lobe (56%) was lower than bilateral upper lobes (90%) and bilateral lower lobes (82%; $p = 0.002$), which may be because there were very few subjects with malignant lesions located in the right middle lobe (18/418). A single false negative would result in a 6% drop in sensitivity. EIA could be utilized for various lesion types, including solid lesions, pure GGOs, mixed GGOs, and patchy shadows that are suspected to be malignant. Of note, most patients with pure GGOs require a long period of clinical follow-up. However, EIA offers a risk stratification with a single test, making it easier for physicians to decide whether the pure GGOs should undergo invasive intervention.

There were some limitations in this study. First, The EIA suggests an overall but not the individual pathological status of a particular lesion. For patients with multiple pulmonary lesions, it cannot indicate which one needs further intervention. Second, the study enrolled patients with single and multiple pulmonary lesions, making it challenging to eliminate confounding factors caused by lesions other than the target lesions when analyzing

TABLE 5 | Diagnostic result of Electrical impedance analysis (EIA) compared with pathology and follow-up in subsolid lesions.

Lesion type	EIA	Pathology		Follow-up	
		+	-	+	-
Pure GGO	+	45	3	1	16
	–	13	4	0	64
Mixed GGO	+	64	6	1	5
	–	12	7	0	15

the factors that affected diagnostic efficacy. Third, other variables such as lesion depth and tumor stage of the patients should be evaluated to see whether they affect the diagnostic effectiveness in order to learn more about the scope of EIA application. Finally, follow-up was limited to two years, which resulted in the inability to receive conclusive diagnoses for some pure GGOs. However, a two-year follow-up is sufficient to prove whether the lesion is stable and has a low risk of progression in the short term. Additionally, providing a conclusive pathological diagnosis for begin lesions arose since they were diagnosed by clinical follow-up.

In conclusion, Electrical Impedance Analysis (EIA) is a sufficiently accurate diagnostic tool that effectively detects the overall pathological conditions of pulmonary lesions. As a non-invasive test, it is very safe and easy to use. It can be adjunctively incorporated with CT screening to both avoid overdiagnosis and missed diagnosis.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because of confidentiality issues. Requests to access the datasets should be directed to CB, bai.chunxue@zs-hospital.sh.cn.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Zhongshan Hospital Fudan University Ethics Committee, Approval No. 2015-16R and 2017-035(3). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Conceptualization and design: DY, JS, CB. Data curation, analysis, and interpretation: DY, CG, YG, XZ (4th Author), DG, YZ, NW, XZ (8th Author), HW, LY, SC, PX. Manuscript drafting: DY, CG, DC, JS, CB. Manuscript revision: DY, JY, MG, JS, CB. All authors have approved the version to be published and agree to be accountable for all aspects of the work.

FUNDING

National Natural Science Foundation of China (82170110), Shanghai Pujiang Program (20PJ1402400), Science and Technology Commission of Shanghai Municipality (20DZ2254400, 21DZ2200600, 20DZ2261200) and Shanghai Municipal Key Clinical Specialty(shslczdzk02201).

REFERENCES

1. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer Statistics in China, 2015. *CA Cancer J Clin* (2016) 66:115–32. doi: 10.3322/caac.21338
2. de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, et al. Reduced Lung-Cancer Mortality With Volume CT Screening in a Randomized Trial. *N Engl J Med* (2020) 382:503–13. doi: 10.1056/NEJMoa1911793
3. National Lung Screening Trial Research T, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced Lung Cancer Mortality With Low-Dose Computed Tomographic Screening. *N Engl J Med* (2011) 365:395–409. doi: 10.1056/NEJMoa1102873
4. Bayford R, Tizzard A. Bioimpedance Imaging: An Overview of Potential Clinical Applications. *Analyst* (2012) 137:4635–43. doi: 10.1039/c2an35874c
5. Braun RP, Mangana J, Goldinger S, French L, Dummer R, Marghoob AA. Electrical Impedance Spectroscopy in Skin Cancer Diagnosis. *Dermatol Clin* (2017) 35(4):489–93. doi: 10.1016/j.det.2017.06.009
6. Stojadinovic A, Fields SI, Shriver CD, Lenington S, Ginor R, Peoples GE, et al. Electrical Impedance Scanning of Thyroid Nodules Before Thyroid Surgery: A Prospective Study. *Ann Surg Oncol* (2005) 12(2):152–60. doi: 10.1245/ASO.2005.03.062
7. Laufer S, Ivorra A, Reuter VE, Rubinsky B, Solomon SB. Electrical Impedance Characterization of Normal and Cancerous Human Hepatic Tissue. *Physiol Meas* (2010) 31(7):995–1009. doi: 10.1088/0967-3334/31/7/009
8. Das L, Das S, Chatterjee J. Electrical Bioimpedance Analysis: A New Method in Cervical Cancer Screening. *J Med Eng* (2015) 2015:636075. doi: 10.1155/2015/636075
9. Zou Y, Guo Z. A Review of Electrical Impedance Techniques for Breast Cancer Detection. *Med Eng Phys* (2003) 25(2):79–90. doi: 10.1016/S1350-4533(02)00194-7
10. Akhtari-Zavare M, Latiff LA. Electrical Impedance Tomography as a Primary Screening Technique for Breast Cancer Detection. *Asian Pac J Cancer Prev* (2015) 16(14):5595–7. doi: 10.7314/APJCP.2015.16.14.5595
11. Zuluaga-Gomez J, Zerhouni N, Al Masry Z, Devalland C, Varnier C. A Survey of Breast Cancer Screening Techniques: Thermography and Electrical Impedance Tomography. *J Med Eng Technol* (2019) 43(5):305–22. doi: 10.1080/03091902.2019.1664672
12. Murillo-Ortiz B, Hernández-Ramírez A, Rivera-Villanueva T, Suárez-García D, Murguía-Pérez M, Martínez-Garza S, et al. Monofrequency Electrical Impedance Mammography (EIM) Diagnostic System in Breast Cancer Screening. *BMC Cancer* (2020) 20(1):876. doi: 10.1186/s12885-020-07283-5
13. Pathiraja AA, Weerakkody RA, von Roon AC, Ziprin P, Bayford R. The Clinical Application of Electrical Impedance Technology in the Detection of Malignant Neoplasms: A Systematic Review. *J Transl Med* (2020) 18 (1):227. doi: 10.1186/s12967-020-02395-9
14. Mustafa BAD, Scott PF, Brackenbury WJ. *In Vivo* Evidence for Voltage-Gated Sodium Channel Expression in Carcinomas and Potentiation of Metastasis. *Cancers (Basel)* (2019) 11:1675. doi: 10.3390/cancers11111675
15. Ward LC, Müller MJ. Bioelectrical Impedance Analysis. *Eur J Clin Nutr* (2013) 67(Suppl 1):S1. doi: 10.1038/ejcn.2012.148
16. Khalil SF, Mohktar MS, Ibrahim F. The Theory and Fundamentals of Bioimpedance Analysis in Clinical Status Monitoring and Diagnosis of Diseases. *Sensors (Basel)* (2014) 14(6):10895–928. doi: 10.3390/s140610895
17. Kimura S, Morimoto T, Uyama T, Monden Y, Kinouchi Y, Iritani T. Application of Electrical Impedance Analysis to Diagnose a Pulmonary Mass. *Chest* (1994) 105:1679–82. doi: 10.1378/chest.105.6.1679
18. Weitzen R, Epstein N, Shoenfeld Y, Zimlichman E. Diagnosing Diseases by Measuring Electrical Skin Impedance: A Novel Technique. *Ann N Y Acad Sci* (2007) 1109:185–92. doi: 10.1196/annals.1398.022
19. Zimlichman E, Lahad A, Aron-Maor A, Kanevsky A, Shoenfeld Y. Measurement of Electrical Skin Impedance of Dermal-Visceral Zones as a Diagnostic Tool for Inner Organ Pathologies: A Blinded Preliminary Evaluation of a New Technique. *Isr Med Assoc J* (2005) 7:631–4.
20. Helge W, Swartz MA. Interstitial Fluid and Lymph Formation and Transport: Physiological Regulation and Roles in Inflammation and Cancer. *Physiol Rev* (2012) 92:1005–60. doi: 10.1152/physrev.00037.2011

21. Maria ASB, Maillat L, . Clement CC, Bordry N, Corthesy P, Auger A, et al. Tumor-Associated Factors are Enriched in Lymphatic Exudate Compared to Plasma in Metastatic Melanoma Patients. *J Exp Med* (2019) 216:1091–107. doi: 10.1084/jem.20181618
22. Kirk CH, D'Alessandro A, Clement CC, Santambrogio L. Lymph Formation, Composition and Circulation: A Proteomics Perspective. *Int Immunol* (2015) 27:219–27. doi: 10.1093/intimm/dxv012
23. Mitchell HR, Dalkin AC. Electrolyte Disorders Associated With Cancer. *Adv Chronic Kidney Dis* (2014) 21:7–17. doi: 10.1053/j.ackd.2013.05.005
24. Meghan JO, Lund AW, Thomas SN. The Biophysics of Lymphatic Transport: Engineering Tools and Immunological Consequences. *iScience* (2019) 22:28–43. doi: 10.1016/j.isci.2019.11.005
25. Helge W, Schröder A, Neuhofer W, Jantsch J, Kopp C, Karlsen TV, et al. Immune Cells Control Skin Lymphatic Electrolyte Homeostasis and Blood Pressure. *J Clin Invest* (2013) 123:2803–15. doi: 10.1172/JCI60113
26. Petros CB, Wells RG, Sackey-Aboagye B, Klavan H, Reidy J, Buonocore D, et al. Structure and Distribution of an Unrecognized Interstitium in Human Tissues. *Sci Rep* (2018) 8:4947. doi: 10.1038/s41598-018-23062-6
27. Yang D, Yung R, Li J, Wang N, Gong X, Lu J, et al. Evaluate the Utility of the Computed Bioconductance Measurement in the Diagnosis of Lung Cancer. *J Thorac Oncol* (2017) 12:S1098. doi: 10.1016/j.jtho.2016.11.1535
28. Yung RC, Zeng MY, Stoddard GJ, Garff M, Callahan K. Transcutaneous Computed Bioconductance Measurement in Lung Cancer: A Treatment Enabling Technology Useful for Adjunctive Risk Stratification in the Evaluation of Suspicious Pulmonary Lesions. *J Thorac Oncol* (2012) 7:681–9. doi: 10.1097/JTO.0b013e31824a8dcd
29. Bai C, Choi CM, Chu CM, Anantham D, Chung-Man Ho J, Khan AZ, et al. Evaluation of Pulmonary Nodules: Clinical Practice Consensus Guidelines for Asia. *Chest* (2016) 150:877–93. doi: 10.1016/j.chest.2016.02.650
30. Aberg P, Nicander I, Hansson J, Geladi P, Holmgren U, Ollmar S. Skin Cancer Identification Using Multifrequency Electrical Impedance–A Potential Screening Tool. *IEEE T BIO-MED Eng* (2004) 51:2097–102. doi: 10.1109/TBME.2004.836523
31. Nissan A, Peoples GE, Abu-Wasel B, Adair CF, Prus D, Howard RS, et al. Prospective Trial Evaluating Electrical Impedance Scanning of Thyroid Nodules Before Thyroidectomy: Final Results. *Ann Surg* (2008) 247:843–53. doi: 10.1097/SLA.0b013e318165c757
32. Stojadinovic A, Nissan A, Gallimidi Z, Lenington S, Logan W, Zuley M, et al. Electrical Impedance Scanning for the Early Detection of Breast Cancer in Young Women: Preliminary Results of a Multicenter Prospective Clinical Trial. *J Clin Oncol* (2005) 23:2703–15. doi: 10.1200/JCO.2005.06.155
33. Stojadinovic A, Nissan A, Shriver CD, Mittendorf EA, Akin MD, Dickerson V, et al. Electrical Impedance Scanning as a New Breast Cancer Risk Stratification Tool for Young Women. *J Surg Oncol* (2008) 97:112–20. doi: 10.1002/jso.20931
34. Prasad SN, Houserkova D, Campbell J. Breast Imaging Using 3D Electrical Impedence Tomography. *BioMed Pap Med Fac Univ Palacky Olomouc Czech Repub* (2008) 152:151–4. doi: 10.5507/bp.2008.024
35. Bach PB, Mirkin JN, Oliver TK, Azzoli CG, Berry DA, Brawley OW, et al. Benefits and Harms of CT Screening for Lung Cancer: A Systematic Review. *JAMA* (2012) 307:2418–29. doi: 10.1001/jama.2012.5521
36. Tanoue LT, Tanner NT, Gould MK, Silvestri GA. Lung Cancer Screening. *Am J Respir Crit Care Med* (2015) 191:19–33. doi: 10.1164/rccm.201410-1777CI
37. Gould MK, Donington J, Lynch WR, Mazzone PJ, Midthun DE, Naidich DP, et al. Evaluation of Individuals With Pulmonary Nodules: When Is It Lung Cancer? Diagnosis and Management of Lung Cancer, 3rd Ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* (2013) 143:e93S–e120S. doi: 10.1378/chest.12-2351
38. Chun EJ, Lee HJ, Kang WJ, Kim KG, Goo JM, Park CM, et al. Differentiation Between Malignancy and Inflammation in Pulmonary Ground-Glass Nodules: The Feasibility of Integrated (18)F-FDG PET/Ct. *Lung Cancer* (2009) 65:180–6. doi: 10.1016/j.lungcan.2008.11.015

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 05 August 2022
DOI 10.3389/fonc.2022.892571

OPEN ACCESS

EDITED BY
H Dean Hosgood,
Albert Einstein College of Medicine,
United States

REVIEWED BY
Chun Ian Soo,
University Malaya Medical Centre, Malaysia
Chiara Colarusso,
University of Salerno, Italy

*CORRESPONDENCE
Yutong He
heyutong@hebmu.edu.cn

SPECIALTY SECTION
This article was submitted to Cancer Epidemiology and Prevention, a section of the journal Frontiers in Oncology

RECEIVED 09 March 2022
ACCEPTED 07 July 2022
PUBLISHED 05 August 2022

CITATION
Li D, Shi J, Dong X, Liang D, Jin J and He Y (2022) Epidemiological characteristics and risk factors of lung adenocarcinoma: A retrospective observational study from North China. *Front. Oncol.* 12:892571. doi: 10.3389/fonc.2022.892571

Epidemiological characteristics and risk factors of lung adenocarcinoma: A retrospective observational study from North China

Daojuan Li, Jin Shi, Xiaoping Dong, Di Liang, Jing Jin and Yutong He*

Cancer Institute, The Fourth Hospital of Hebei Medical University, Shijiazhuang, China

Background: The main aim of the study was to determine the risk factors of lung adenocarcinoma and to analyze the variations in the incidence of lung adenocarcinoma according to time, sex, and smoking status in North China.

Methods: Patients with lung cancer in local household registries diagnosed and treated for the first time in the investigating hospital were enrolled from 11 cities in North China between 2010 and 2017. Baseline characteristics and tumor-related information were extracted from the patients' hospital medical record, clinical course records, and clinical examination. Some of the variables, such as smoking, alcohol consumption, medical history, and family history of cancer, were obtained from interviews with the enrolled patients. The statistical method used were the chi-square test and multi-factor logistic regression analysis. The time trend was statistically analyzed using Joinpoint regression models, and *p* values were calculated.

Results: A total of 23,674 lung cancer cases were enrolled. People in severely polluted cities were at higher risk for lung adenocarcinoma ($p < 0.001$). Most patients with lung adenocarcinoma had no history of lung-related diseases ($p = 0.001$). Anatomically, lung adenocarcinoma was more likely to occur in the right lung ($p < 0.001$). Non-manual labor workers were more likely to develop from lung adenocarcinoma than manual workers ($p = 0.015$). Notably, non-smokers were more likely to develop lung adenocarcinoma than smokers ($p < 0.001$). The proportion of lung adenocarcinoma increased significantly in Hebei Province ($p < 0.001$). Among non-smokers, the proportion of lung adenocarcinoma showed a higher rise than in smokers ($p < 0.001$).

Conclusions: Lung adenocarcinoma is the most common histological type of lung cancer in North China (Hebei Province), and the proportion of lung adenocarcinoma is increasing, especially among non-smokers. Lung adenocarcinoma is more common in women, severely polluted cities, individuals with no history of lung-related diseases, in the right lung, and in non-smokers. These can serve as a great guide in determining the accuracy of

lung adenocarcinoma high-risk groups and lung cancer risk assessment models.

KEYWORDS

lung adenocarcinoma, epidemiology, risk factors, never-smokers, North China

Background

According to GLOBOCAN 2018, there were 2.09 million cases of lung cancer worldwide in 2018, accounting for 11.6% of all cancer incidences, and approximately 1.76 million lung cancer deaths, accounting for 18.4% of all cancer deaths, and lung cancer ranked first among all malignant tumor incidences and deaths. Moreover, there were 1.225 million new cases of lung cancer in Asia with 1.069 million lung cancer deaths, accounting for 58.5% of the incidence and 60.7% of the mortality of lung cancer worldwide (1). In China, lung cancer was the leading cancer in men and the second leading cancer in women with 550,000 and 278,000 new lung cancer cases, respectively. Lung cancer was the leading cancer in both men and women with 455,000 and 202,000 new lung cancer cases, respectively (2). Approximately 85% of lung cancers were non-small cell lung cancers, and 15% were small cell lung cancers. The following histological subtypes of non-small cell lung cancer were distinguished: adenocarcinoma, accounting for 38.5% of all lung cancers, and squamous cell carcinoma, accounting for 20% (3). A lung cancer study in the United States showed that the incidence of adenocarcinoma in men and women from 2004 to 2009 was 24.5/100,000 and 20.0/100,000, the incidence of squamous cell carcinoma was 18.8/100,000 and 8.5/100,000, and the incidence of small cell lung cancer was 9.8/100,000 and 7.9/100,000, respectively (4). Lung cancer was also the leading cause of cancer incidence and mortality in North China (Hebei Province) (5). Lung adenocarcinoma, the most common histological type of non-small cell lung cancer, has been on the rise in most countries over the past few decades (6), with a 5-year relative survival of only 12.8% (7). The increasing incidence of lung adenocarcinoma has been a subject of global interest. However, there is currently no report on the risk factors and epidemiological characteristics of lung adenocarcinoma in North China (Hebei Province). Therefore, this study aimed to determine the risk factors of lung adenocarcinoma and analyze the variations in the incidence of lung adenocarcinoma according to time, sex, and smoking status based on the distribution of lung adenocarcinoma in Hebei Province.

Methods

Study design and participants

Lung cancer clinical diagnosis and treatment information was collected from 133 hospitals in 11 cities of Hebei Province from 2010 to 2017: Shijiazhuang City, Baoding City, Tangshan City, Handan City, Xingtai City, Cangzhou City, Hengshui City, Langfang City, Qinhuangdao City, Chengde City, and Zhangjiakou City. Among the cities, Qinhuangdao City, Chengde City, and Zhangjiakou City were defined as the lightly polluted cities, while the other cities were severely polluted cities (8). Relevant variables were extracted from the patients' hospital medical records, clinical course records, clinical examination, and clinical imaging studies. Baseline characteristics included the following: sex, age at diagnosis, marital status, occupation, height and weight at admission, blood type, smoking history, alcohol consumption, history of lung-related diseases, and family history of cancer. Tumor-related information was as follows: pathological type, diagnosis basis, grade, and sub-site of lung cancer patients. The inclusion criteria of study participants were as follows (1): admitted to hospital from 1 January 2010 to 31 December 2017; (2) new cases that were first diagnosed or treated in the investigating hospital and had not undergone surgery, radiotherapy, or chemotherapy in the previous hospital; and (3) cases of local household registration. The exclusion criterion included multiple primary or metastatic cancer cases.

Some of the variables, such as smoking, alcohol consumption, medical history, and family history of cancer, were obtained by interviewing the enrolled patients. Other variables were obtained by extracting information from the patients' medical records.

The occupation classification in the questionnaire included non-manual labor and manual labor. Non-manual labor included persons in charge of party organs, mass organizations, social organizations, enterprises, and institutions; professional and technical personnel; and office and related personnel. Manual labor workers included social production service and life service personnel; agriculture, forestry, animal husbandry and fishery production, and auxiliary personnel; production and related personnel; soldiers; and other personnel who cannot be classified.

Lung cancer was mainly diagnosed by the following methods. Lung cancers were diagnosed by primary histology; that is, the pathological type of lung cancer was diagnosed by surgery or outpatient biopsy. Cytology and blood sample diagnosis referred to the diagnosis of lung cancer by sputum exfoliated cytology. Biochemical and immunological diagnosis referred to the diagnosis of lung cancer by biochemical,

immunological, and tumor marker tests. Clinical diagnosis referred to the diagnosis of lung cancer by collecting medical history data through the patients' clinical symptoms and signs and by asking patients about their subjective symptoms. Secondary histology referred to the examination and diagnosis of patients with discomfort caused by a secondary metastatic cancer and tracking their primary cancer as lung cancer. Related disease history included pulmonary tuberculosis, chronic bronchitis, emphysema, asthma, silicosis/pneumoconiosis, others, and unknown.

Patient's height and weight were from the medical record system. Body mass index (BMI) was calculated with the formula of BMI = weight (kg) ÷ height (m^2). BMI was categorized into <18.5 kg/m^2 (underweight), 18.5–24.9 kg/m^2 (normal weight), and 25–29.9 kg/m^2 (overweight) (9).

When comparing lung adenocarcinoma with other pathological types, cases with unknown information on relevant variables were excluded from the study.

Data collection and quality control

An expert group including experts in various fields such as clinical medicine, epidemiology, health statistics, cancer registration, and specialized persons who had long been engaged in clinical data collection was established. An expert seminar was held to present the study design and conduct technical guidance and quality control, followed by the training meeting. Investigators were selected by the project site according to the workload, and all the project team underwent the same technical training. All investigators followed the standardized operating procedures strictly. Experts from the project team visited the project sites at least once a year, and the supervision included (1) supervision of the collection of lung cancer clinical diagnosis data, (2) data storage check, (3) checking whether the collected data met the inclusion and exclusion criteria, and (4) checking data completeness and extracting 5% of the newly collected data in the current year to verify the information by querying the original data again.

Statistical analysis

The statistical methods used were the chi-square test and multi-factor logistic regression analysis. According to well-known statistical methods, univariate analysis was performed for factor analysis. Through univariate analysis, a variable with statistical significance was screened out. Univariate analysis was performed on the variables collected in the questionnaire of this study, including sex, age, smoking status, alcohol intake, occupation, related disease history, family history of cancer, BMI, areas of residence, marriage, blood type, position, and morphology. After univariate analysis, we selected statistically significant variables for multivariate analysis, including sex, age, smoking status, areas, related disease history, occupation, and position. The time trend was statistically analyzed using Joinpoint regression models, and *p*-values were calculated. All statistical analysis and data quality control were performed using the Statistical Package for Social Science (SPSS) (version 20, SPSS Inc., Chicago, IL, USA), SAS software (Version 9.3), and the Joinpoint software (version 4.8.0.1; National Cancer Institute, Rockville, MD, US). $p \leq 0.05$, established on two-sided probabilities, was considered statistically significant.

Results

Distribution of characteristics in patients with lung cancer in Hebei Province, 2010–2017

This study enrolled 23,674 patients diagnosed with lung cancer for the first time in 11 cities of Hebei Province. Among the lung cancer patients, the male-to-female ratio was 1.94:1. The average age of all lung cancer patients was 68.91 ± 11.21 years old, and the largest number of lung cancer cases was observed in the 65–69 age group.

Among all the lung cancer cases, regarding the smoking situation, the number of non-smokers was 10,913 (46.2%), current smokers was 7,533 (31.8%), former smokers (those who had quit smoking) was 2,732 (11.5%), and the number of patients with an unknown smoking status was 2,496 (10.5%). Moreover, 14,976 patients (63.3%) never drank, 5,881 patients (24.8%) drank frequently, and the drinking status of 2,817 (11.9%) patients was unknown. There were 2,850 non-manual labor (12.0%) and 9,724 (41.1%) manual workers. Among the lung cancer cases, 19,357 (81.8%) had no history of related diseases, 1,643 (6.9%) had a history of related diseases, and 2,674 had unknown information. Moreover, there were 18,378 (77.7%) patients without family history of cancer, 1,735 (7.3%) with family history of cancer, and 3,561 with unknown family history of cancer. Furthermore, 727 (3.1%) patients had a BMI index < 18.5 kg/m^2, 6,444 (27.2%) had a BMI between 18.5 and 24.9 kg/m^2, and 3,135 (13.2%) had a BMI ≥ 25 kg/m^2. Among the lung cancer patients, 17.8% lived in areas with light air pollution, while 82.2% lived in areas with heavy air pollution. Regarding marital status, the married proportion was the highest at 82.7%, and the other proportions were all lower than 15%. Regarding blood type, blood type A accounted for 6.3%, blood type B accounted for 8.4%, blood type O accounted for 7.1%, and blood type AB accounted for 2.9%, and the proportion of unknown blood type was high (75.3%). Regarding the distribution of lung subsites, 1.2% occurred in both lungs, 38.9% occurred in the left lung, and 50.5% occurred in the right lung. There were 8,121 cases (34.3%) of adenocarcinoma, 3,443 cases (14.5%) of squamous cell carcinoma, 2,843 cases (12.1%) of small cell

carcinoma, 956 cases (4.0%) of other cancer types, and 8,311 cases (35.1%) of unknown type with no pathology. Regarding the diagnosis, 13,785 lung cancer cases (58.2%) were diagnosed by primary histology, 892 cases (3.8%) were diagnosed by cytology and blood samples, 642 cases (2.7%) were diagnosed by biochemistry and immunology, and 1,822 cases (7.7%) were clinically diagnosed. Moreover, 215 cases (0.9%) were diagnosed by secondary histology, 5,198 cases (22.0%) were diagnosed by other special examinations (x-ray, ultrasound, CT, MRI, endoscope, etc.), and 1,120 cases (4.7%) had an unknown diagnosis (Table 1).

Lung adenocarcinoma was compared with other pathological types after excluding variables with unknown information. Through univariate analysis, variables with t statistical significance were obtained. Variables such as sex, age, smoking status, alcohol intake situation, occupation, related disease history, family history of cancer, BMI, air pollution level in areas, and subsites were included in the multivariate logistic regression analysis model. The multivariate regression findings suggested that women were at higher risk for lung adenocarcinoma than men ($p < 0.001$). Individuals over 65 years old had a lower risk of developing lung adenocarcinoma than those younger than 65 years old ($p < 0.001$). Notably, non-smokers were more likely to develop lung adenocarcinoma than smokers ($p < 0.001$). Non-manual labor workers were more likely to develop lung adenocarcinoma than manual workers ($p = 0.015$). The risk of lung adenocarcinoma was higher for people without a history of lung-related diseases ($p = 0.001$). Compared with people living in lightly polluted cities, those living in severely polluted cities were at higher risk for lung adenocarcinoma ($p < 0.001$). Moreover, lung adenocarcinoma occurred more frequently in the right lung than in the whole lung ($p < 0.001$) (Table 2 and Figure 1).

Distribution changes of different histological types of lung cancer

The incidence trends of lung cancer in Hebei Province from 2010 to 2017 were analyzed by histological type. The proportion of lung adenocarcinoma (38.90%) in 2017 was 1.98 times higher than in 2010 (19.60%) ($p < 0.001$). However, the changes in the proportion of small cell carcinoma and squamous cell carcinoma were not statistically significant in both sexes ($p = 0.100$). For male patients, the proportions of lung adenocarcinoma and small cell carcinoma increased by 93.67% and 125.34%, respectively, in 2017 compared to 2010 ($p < 0.001$ and $p < 0.001$), whereas the proportion of lung squamous cell carcinoma decreased by 29.06% from 2010 to 2017 ($p < 0.001$) (Figure 2A). For female patients, between 2010 and 2017, the proportion of lung adenocarcinoma increased by 80.36% ($p < 0.001$), and the proportion of lung squamous cell carcinoma decreased by 28.41% ($p < 0.001$) (Figure 2B).

The histological types of lung cancer were analyzed by stratifying according to smoking status. The study suggested that there was a rise in the proportion of lung adenocarcinoma of 80.22% in never-smokers during 2010–2017 ($p < 0.001$). Moreover, the change in the proportion of small cell carcinoma was not statistically significant ($p = 0.200$), and there was a decline in the proportion of lung squamous cell carcinoma of 58.54% ($p < 0.001$) (Figure 2C). Among smokers, the proportion of lung adenocarcinoma similarly increased by 56.37% from 2010 to 2017 ($p < 0.001$), while the proportion of squamous cell carcinoma decreased by 17.75% ($p < 0.001$) (Figure 2D).

Discussion

Lung cancer is the leading cause of cancer morbidity and mortality, accounting for one-fifth of all cancer deaths worldwide (1). Over the past two decades, the incidence of lung adenocarcinoma has been on the rise worldwide, and lung adenocarcinoma has become the most common lung cancer subtype (10). Investigating the risk factors of lung adenocarcinoma and analyzing how the incidence of lung adenocarcinoma varies according to time, sex, and smoking status are particularly important. This study focused on the analysis of the risk factors and epidemiological trends of lung adenocarcinoma in North China (Hebei Province) to provide an important reference for the screening and precise prevention of lung adenocarcinoma in high-risk groups.

Overall, the proportion of adenocarcinoma, squamous cell carcinoma, and small cell lung carcinomas in Hebei Province was consistent with that in the United States, Korea, and Australia (11–13). The risk factors of lung adenocarcinoma, the most pathological type of lung cancer in Hebei Province, was analyzed. The multivariate regression analysis findings revealed that lung adenocarcinoma was more common in women, individuals younger than 65 years old, those living in severely polluted cities, non-smokers, those with no history of lung-related diseases, non-manual labor workers, and the right lung. Many lung cancer studies worldwide found that among lung cancer cases in Canada, Denmark, Germany, New Zealand, the Netherlands, the United States, (Lima) Peru, and Southern Sweden, adenocarcinoma was more common in women and in younger patients (14–16).

A study from nine European countries manifested that the risk of lung cancer increased by 18% with PM2.5 increasing every 5 mg/m^3, which may be due to lung adenocarcinoma (HR: 1.55, 95% CI: 1.05–2.29) (17, 18). In Canada, researchers found that a 0.01 mg/m^3 increase in PM2.5 was inversely related to lung cancer (HR: 1.34; 95% CI: 1.10–1.65), with the strongest associations with adenocarcinoma (HR: 1.44; 95% CI: 1.06–1.97) (19). One study in the US indicated that for every 0.01 mg/m^3 increase in PM2.5 level, the hazard rate of lung adenocarcinoma

10.3389/fonc.2022.892571

TABLE 1 Distribution of clinical characteristics in patients with lung cancer in Hebei Province, 2010–2017.

Factor		Cases	%
Sex	Male	15,611	65.9
	Female	8,063	34.1
Age	<45	396	1.7
	45–64	7,639	32.2
	≥65	15,639	66.1
Smoking status	Never	10,913	46.2
	Current	7,533	31.8
	Former	2,732	11.5
	Unknown	2,496	10.5
Alcohol intake	Never	14,976	63.3
	Current	5,881	24.8
	Unknown	2,817	11.9
Occupation	Non-manual labor	2,850	12.0
	Manual labor	9,724	41.1
	Unknown	11,100	46.9
Related disease history	No	19,357	81.8
	Yes	1,643	6.9
	Unknown	2,674	11.3
Family history of cancer	No	18,378	77.7
	Yes	1,735	7.3
	Unknown	3,561	15.0
BMI	<18.5	727	3.1
	18.5–24.9	6,444	27.2
	≥25	3,135	13.2
	Unknown	13,368	56.5
Areas	Light pollution	4,222	17.8
	Severe pollution	19,452	82.2
Marriage	Unmarried	295	1.2
	Married	19,568	82.7
	Widowed	536	2.3
	Divorced	199	0.8
	Unknown	3,076	13.0
Blood type	A	1,494	6.3
	B	1,999	8.4
	O	1,676	7.1
	AB	681	2.9
	Unknown	17,824	75.3
	Whole lung	301	1.2
Position	Left lung	9,205	38.9
	Right lung	11,947	50.5
	Unknown	2,221	9.4
	Adenocarcinoma	8,121	34.3
Morphology	Small cell carcinoma	2,843	12.1
	Squamous cell carcinoma	3,443	14.5
	Others	956	4.0
	Unknown (no pathology)	8,311	35.1

TABLE 2 Distribution of clinical characteristics in patients with lung adenocarcinoma (%) in Hebei Province, 2010–2017.

Characteristic	Adenocarcinoma	Others	X^2	*p*
Sex				
Male	4,641 (57.1)	5,502 (76.0)	604.774	<0.001*
Female	3,480 (42.9)	1,740 (24.0)		
Age				
0–44	192 (2.4)	119 (1.6)	80.504	<0.001*
45–64	3,207 (39.5)	2,409 (33.3)		
≥ 65	4,722 (58.1)	4,714 (65.1)		
Smoking status				
Never	4,604 (60.2)	2,688 (39.5)	623.968	<0.001*
Current	2,259 (30.0)	2,965 (43.6)		
Former	752 (9.8)	1,149 (16.9)		
Alcohol intake				
Never	5,649 (74.8)	4,391 (65.9)	134.88	<0.001*
Current	1,902 (25.2)	2,270 (34.1)		
Occupation				
Non-manual labor	902 (21.9)	802 (19.1)	10.07	0.002*
Manual labor	3,218 (78.1)	3,473 (80.9)		
Related disease history				
No	7,137 (94.6)	6,114 (91.3)	62.595	<0.001*
Yes	403 (5.3)	584 (8.7)		
Family history of cancer				
No	6,700 (90.9)	5,804 (90.8)	0.029	0.866
Yes	672 (9.1)	588 (9.2)		
BMI				
<18.5	204 (5.8)	223 (5.8)	45	0.461
18.5–24.9	2,162 (61.4)	2,416 (62.7)		
≥25	1,157 (32.8)	1,214 (31.5)		
Areas				
Light pollution	1,120 (13.8)	1,645 (22.7)	206.538	<0.001*
Severe pollution	7,001 (86.2)	5,597 (77.3)		
Position				
Whole lung	97 (1.3)	89 (1.3)	16.199	<0.001*
Left lung	3,191 (41.3)	3,088 (44.6)		
Right lung	4,430 (57.4)	3,745 (54.1)		

* P<0.01, was considered statistically significant.

(HR: 1.31, 95% CI: 0.87–1.97) also increased (20). Moreover, in one meta-analysis, lung adenocarcinoma was more strongly associated with PM2.5 and PM10 (RR per 10 μg/m^3, 1.40 [95% CI: 1.07–1.83] and 1.29 [95% CI: 1.02–1.63]) compared with other pathological types (21). A study in Taiwan Province of China also concluded that changes in PM2.5 levels can affect the incidence of lung adenocarcinoma and patient survival (22). Air pollution and the PM2.5 concentration were severely high in Hebei Province (23–25). As in our study, they may increase the risk of lung adenocarcinoma.

Although ever smoking strongly correlated with increased risk for lung adenocarcinoma, lung adenocarcinoma was the most common subtype among never smokers (26, 27). In this study, the result found that non-smokers were more likely to develop lung adenocarcinoma than smokers. An earlier study in Singapore showed that approximately 70% of never-smoking lung cancer patients were diagnosed with adenocarcinoma (28). The World Trade Center Environmental Health Center conducted a study among people who were exposed and possibly inhaled dust and fumes from the destruction of the World Trade Center towers in 2001. The results showed that lung adenocarcinoma was more common in never-smokers compared to ever-smokers (72% *vs.* 65%) and more common in women compared to men (70% *vs.* 65%) (29). One study in China showed that though the proportion of lung adenocarcinoma increased in smokers and non-smokers, lung adenocarcinoma was more common in non-smokers (30). However, the result from a national health examination

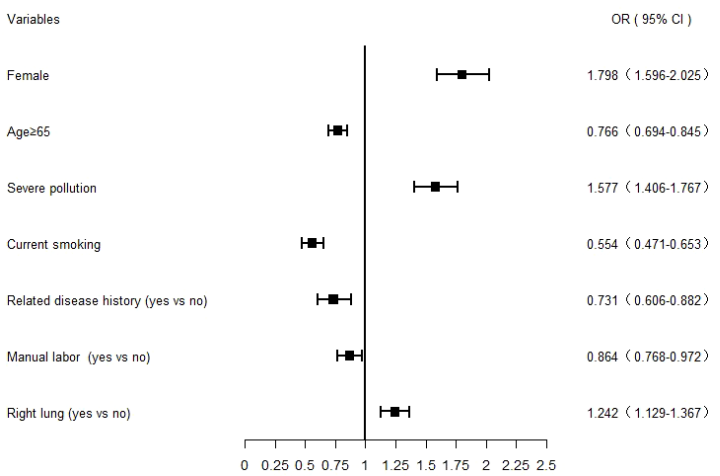


FIGURE 1
Forest plot of lung adenocarcinoma risk factors analyzed using a multivariate logistic regression model. The squares and bars represent the odds ratio (OR) with a 95% confidence interval (CI).

program in Korea was inconsistent with that of this study. The hazard ratios for lung adenocarcinoma were significantly higher in male current smoker than in never-smokers, while they were not different for female patients (31). Despite this, non-smoker lung cancer had been classified as an independent disease entity, which is different from smoker lung cancer. Gene sequencing technology has enabled us to truly understand the difference in lung adenocarcinoma between smokers and non-smokers at the

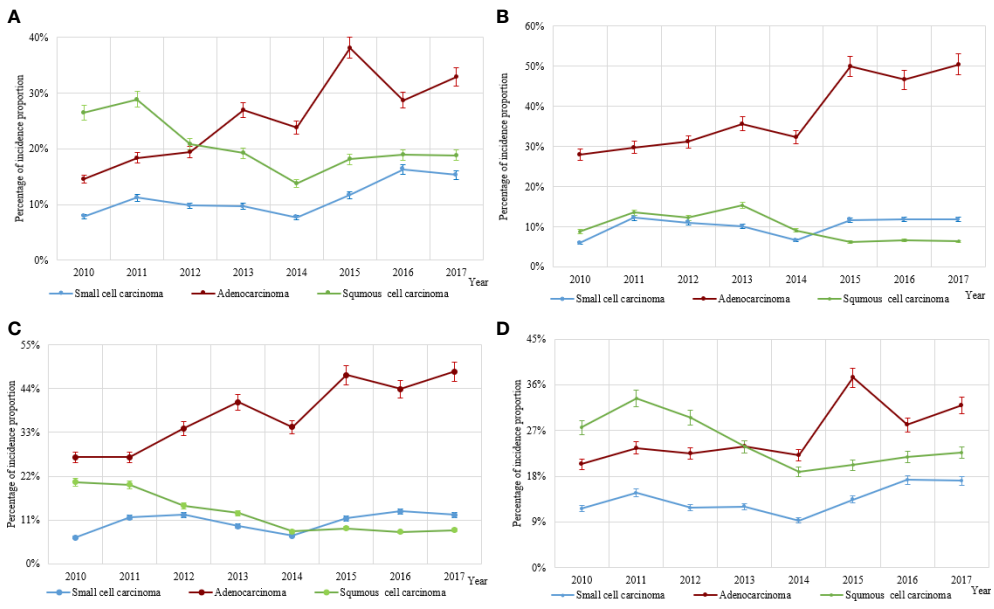


FIGURE 2
The incidence distribution of the three main histological types of lung cancer (lung adenocarcinoma, squamous cell carcinoma, and small cell carcinoma). **(A)** The incidence distribution of the three main histological types of lung cancer in men. **(B)** The incidence distribution of three main histological types of lung cancer in women. **(C)** The incidence distribution of three main histological types of lung cancer in patients with never-smoking lung cancer. **(D)** The incidence distribution of three main histological types of lung cancer in patients with smoking lung cancer.

microscopic molecular level. Most of the driver gene alterations were identified in lung adenocarcinoma in never-smokers. However, no such available molecular targets exist for smoker lung cancer (32, 33). This study found that non-smokers were more likely to develop lung adenocarcinoma, and because the mechanisms of lung adenocarcinoma in smokers and non-smokers were completely different, we should pay more attention to the screening of lung adenocarcinoma among non-smokers at high risk of lung cancer. It was of great significance to strengthen the screening of lung adenocarcinoma among non-smokers at high risk of lung cancer for related genes and to detect and diagnose lung adenocarcinoma early, improving the chances of surgery for patients and prolonging the survival time.

This study also found that most patients with lung adenocarcinoma had no history of lung-related diseases. Anatomically, lung adenocarcinoma was more likely to occur in the right lung. In addition, this study revealed that non-manual labor workers had a greater risk of lung adenocarcinoma compared with manual workers. These new findings will have a significant impact on the prevention of lung adenocarcinoma. We should strengthen the screening of lung adenocarcinoma for non-manual workers and people without a history of lung-related diseases to achieve early detection, early diagnosis, and early treatment. For designated lung adenocarcinoma high-risk groups, attention should be paid to the examination of the right lung. However, these findings may require more research to confirm.

Overall, the proportion of lung adenocarcinoma increased significantly in Hebei Province, but the proportion of lung squamous cell carcinoma decreased during 2010–2017. Among non-smokers, the proportion of lung adenocarcinoma showed a higher rise than in smokers. This implied that lung adenocarcinoma is still the main histological type of lung cancer.

This study showed that lung adenocarcinoma, one of the types of non–small cell lung carcinoma, was the most common form of lung cancer among non–smokers (34). In 10 years, lung cancer characteristics have changed: more women, more never-smokers, and more adenocarcinomas in France (35). In a study among Chinese women, the proportion of never-smokers with adenocarcinoma increased significantly compared with smokers (36). The prevention of lung adenocarcinoma has to be given more prominence, especially among non-smokers. The incidence of adenocarcinoma increased in the United States between 2006 and 2010, but there was a greater reduction in the rate of decline of squamous cell carcinoma than before (37). One study proved that the adenocarcinoma incidence trends were consistent with smoking trends; however, the relative risk with smoking for adenocarcinoma was lower than that for squamous cell carcinoma and small cell carcinoma (38). Our study showed that smoking had the greatest effect on squamous cell carcinomas, a lesser effect on lung adenocarcinomas (Supplementary Tables 1, 2), and no statistically significant effect on small cell carcinomas (Supplementary Table 3). In China, from 1990 to 2015, the age-standardized prevalence of daily smoking decreased significantly by 22.4% (95% UI: 20.7%–24.0%) and 48.4% (95% UI: 41.2%–55.1%) for men and women, respectively (39). Declining smoking rates in China may be one of the main reasons for this result.

In this study, we found that the incidence of lung adenocarcinoma was increasing, especially in non-smokers, revealing the changing trends in the pathological types of lung cancer. At present, few studies have reached this conclusion, which can have a great guiding effect on the selection of high-risk lung cancer groups for screening. In addition, we also found that lung adenocarcinoma was more common in severely polluted cities, non-smokers, individuals with no history of pulmonary diseases, non-manual labor workers, and the right lung. These conclusions have important implications. First, the conclusions have a key guiding significance for determining the accuracy of the lung cancer risk assessment model. In addition, because cancer screening programs do not currently cover the whole country, this can be more beneficial to high-risk populations through lung cancer screening, improve the detection rate, and reduce healthy individuals from receiving numerous unnecessary tests, especially those that can cause some pain and harm.

The present study also had some shortcomings. The study enrolled 23,674 lung cancer patients. For smoking status, the number of unknowns accounted for 10.5%. In the future, we need to pay attention to the integrity of variable information collection or choose an adequate method to supplement missing data; however, we currently only presented the results of real data. In this study, the main inclusion criterion for lung cancer patients was patients who were first diagnosed or treated in the hospital. Most patients with good economic conditions who were later found to have space occupancy by other means instead of a pathological diagnosis chose provincial or national specialized oncology hospitals to perform further examinations and make a pathological diagnosis. This led to the local hospital not being able to obtain the histological subtypes for the first diagnosis. Moreover, among the 35.1% of the total cases with unknown histological diagnosis, some of them may have had clear pathological types, but the local hospitals could not obtain them, which is also a limitation of this study.

Conclusion

Lung adenocarcinoma is one of the most common histological types of lung cancer in North China (Hebei Province), and the incidence of lung adenocarcinoma is increasing, especially in non-smokers. Lung adenocarcinoma is more common in women, individuals younger than 65 years old, those living in severely polluted cities, non-smokers, those with no history of lung-related diseases, non-manual labor workers, and the right lung. These can play a great guiding role in determining the accuracy of lung adenocarcinoma diagnosis in high-risk groups and lung cancer risk assessment models.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Ethics statement

This study was reviewed and approved by the Ethics Committee of Hebei Medical University Fourth Hospital. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Author contributions

DJL and YTH were involved in the design and data analysis of this study and drafted the manuscript. YTH was responsible for the oversight of the whole study and edited the manuscript. DJL conceived the review and edited the manuscript. DJL, JS, XPD, DL, and JJ collected and analyzed the data. All authors contributed to the article and approved the submitted version.

Funding

This study was approved and supported by Department of Disease Control and Prevention, Hebei Provincial Health Commission Letter (2018) No. 5.

Acknowledgments

We gratefully acknowledge the cooperation of all the population-based cancer registries in providing cancer statistics, data collection, sorting, verification, and database creation.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.892571/full#supplementary-material

References

1. 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 Cancer J Clin* (2018) 68:394–424. doi: 10.3322/caac.21492

2. Zheng R, Zhang S, Zeng H, Wang SM, Sun KX, Chen R, et al. Cancer incidence and mortality in China, 2016. *J Natl Cancer Cen* (2022) 2:1–9. doi: 10.1016/j.jncc.2022.02.002

3. Skřičková J, Kadlec B, Venclíček O, Merta Z. Lung cancer. *Cas Lek Cesk* (2018) 157:226–36.

4. Houston KA, Henley SJ, Li J, White MC, Richards TB. Patterns in lung cancer incidence rates and trends by histologic type in the united states, 2004-2009. *Lung Cancer* (2014) 86:22–8. doi: 10.1016/j.lungcan.2014.08.001

5. He Y, Liang D, Li D, Shi J, Jin J, Zhai J, et al. Cancer incidence and mortality in hebei province, 2013. *Med (Baltimore)* (2017) 96:e7293. doi: 10.1097/MD.0000000000007293

6. Nakamura H, Saji H. Worldwide trend of increasing primary adenocarcinoma of the lung. *Surg Today* (2014) 44:1004–12. doi: 10.1007/s00595-013-0636-z

7. Gedvilaitė V, Danila E, Cicėnas S, Smailytė G. Lung cancer survival in Lithuania: changes by histology, age, and sex from 2003-2007 to 2008-2012. *Cancer Control* (2019) 26:1073274819836085. doi: 10.1177/1073274819836085

8. Xu H, Xiao Z, Chen K, Tang M, Zheng N, Li P, et al. Spatial and temporal distribution, chemical characteristics, and sources of ambient particulate matter in the Beijing-Tianjin-Hebei region. *Sci Total Environ* (2019) 658:280–93. doi: 10.1016/j.scitotenv.2018.12.164

9. Gallagher D, Heymsfield SB, Heo M, Jebb SA, Murgatroyd PR, Sakamoto Y. Healthy percentage body fat ranges: an approach for developing guidelines based on body mass index. *Am J Clin Nutr* (2000) 72:694–701. doi: 10.1093/ajcn/72.3.694

10. B'chir F, Laouani A, Ksibi S, Arnaud MJ, Saguem S. Cigarette filter and the incidence of lung adenocarcinoma among Tunisian population. *Lung Cancer* (2007) 57:26–33. doi: 10.1016/j.lungcan.2007.01.034

11. AM Noone, N Howlader, M Krapcho, D Miller, A Brest, M Yu, et al eds. *SEER cancer statistics review, 1975–2015*. Bethesda, MD: Natl Cancer Inst (2018).

12. Shin A, Oh CM, Kim BW, Woo H, Won YJ, Lee JS, et al. Lung cancer epidemiology in Korea. *Cancer Res Treat* (2017) 20170 49:616–26. doi: 10.4143/crt.2016.178

13. John T, Cooper WA, Wright G, Siva S, Solomon B, Marshall HM, et al. Lung cancer in Australia. *J Thorac Oncol* (2020) 15:1809–14. doi: 10.1016/j.jtho.2020.09.005

14. Fidler-Benaoudia MM, Torre LA, Bray F, Ferlay J, Jemal A. Lung cancer incidence in young women vs. young men: a systematic analysis in 40 countries.k. *Int J Cancer* (2020) 147:811–19. doi: 10.1002/ijc.32809

15. Galvez–Nino M, Ruiz R, Pinto JA, Roque K, Mantilla R, Raez LE, et al. Lung cancer in the young. *Lung* (2020) 198:195–200. doi: 10.1007/s00408-019-00294-5

16. Fritz I, Olsson H. Lung cancer in young women in southern Sweden: a descriptive study. *Clin Respir J* (2018) 12:1565–71. doi: 10.1111/crj.12712

17. *Air pollution and cancer, IARC scientific publication*. Available at: http://publications.iarc.fr/_publications/media/download/3692/e67ee17589f8759dc6d811c03c7b061a11827ee3.pdf (Accessed September 20, 2018).

18. Raaschou-Nielsen O, Andersen ZJ, Beelen R, Samoli E, Stafoggia M, Weinmayr G, et al. Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European study of cohorts for air pollution effects (ESCAPE). *Lancet Oncol* (2013) 14:813–22. doi: 10.1016/S1470-2045(13)70279-1

19. Tomczak A, Miller AB, Weichenthal SA, To T, Wall C, van Donkelaar A, et al. Long-term exposure to fine particulate matter air pollution and the risk of lung cancer among participants of the Canadian national breast screening study. *Int J Cancer* (2016) 139:1958–66. doi: 10.1002/ijc.30255

20. Gharibvand L, Beeson WL, Shavlik D, Knutsen R, Ghamsary M, Soret S, et al. The association between ambient fine particulate matter and incident adenocarcinoma subtype of lung cancer. *Environ Health* (2017) 16:71. doi: 10.1186/s12940-017-0268-7

21. Turner MC, Andersen ZJ , Baccarelli A, Diver WR, Gapstur SM, Pope CA 3rd, et al. Outdoor air pollution and cancer: an overview of the current evidence and public health recommendations. *CA Cancer J Clin* (2020) 70(6):460–79. doi: 10.3322/caac.21632. Online ahead of print

22. Tseng CH, Tsuang BJ, Chiang CJ, Ku KC, Tseng JS, Yang TY, et al. The relationship between air pollution and lung cancer in nonsmokers in Taiwan. *J Thorac Oncol* (2019) 14:784–92. doi: 10.1016/j.jtho.2018.12.033

23. Shao J, Ge J, Feng X, Zhao C. Study on the relationship between PM2.5 concentration and intensive land use in hebei province based on a spatial regression model. *PLoS One* (2020) 15(9):e0238547. doi: 10.1371/journal.pone.0238547

24. Zhao N, Wang G, Li G, Lang J, Zhang H. Air pollution episodes during the COVID-19 outbreak in the Beijing-Tianjin-Hebei region of China: an insight into the transport pathways and source distribution. *Environ pollut* (2020) 267:115617. doi: 10.1016/j.envpol.2020.115617

25. Li X, Yan C, Wang C, Ma J, Li W, Liu J, et al. PM2.5-bound elements in hebei province, China: pollution levels, source apportionment and health risks. *Sci Total Environ* (2022) 806:150440. doi: 10.1016/j.scitotenv.2021.150440

26. Li C, Lu H. Adenosquamous carcinoma of the lung. *Onco Targets Ther* (2018) 11:4829–35. doi: 10.2147/OTT.S164574

27. Corrales L, Rosell R, Cardona AF, Martín C, Zatarain-Barrón ZL, Arrieta O, et al. Lung cancer in never smokers: the role of different risk factors other than tobacco smoking. *Crit Rev Oncol Hematol* (2020) 148:102895. doi: 10.1016/j.critrevonc.2020.102895

28. Chapman AM, Sun KY, Ruestow P, Cowan DM, Madl AK. Lung cancer mutation profile of EGFR, ALK, and KRAS: Meta-analysis and comparison of never and ever smokers. *Lung Cancer* (2016) 102:122–34. doi: 10.1016/j.lungcan.2016.10.010

29. Durmus N, Pehlivan S, Zhang Y, Shao Y, Arslan AA, Corona R, et al. Lung cancer characteristics in the world trade center environmental health center. *Int J Environ Res Public Health* (2021) 18:2689. doi: 10.3390/ijerph18052689

30. Zeng Q, Vogtmann E, Jia MM, Parascandola M, Li JB, Wu YL, et al. Tobacco smoking and trends in histological subtypes of female lung cancer at the cancer hospital of the Chinese academy of medical sciences over 13 years. *Thorac Cancer* (2019) 10:1717–24. doi: 10.1111/1759-7714.13141

31. Yun YD, Back JH, Ghang H, Jee SH, Kim Y, Lee SM, et al. Hazard ratio of smoking on lung cancer in Korea according to histological type and gender. *Lung* (2016) 194:281–89. doi: 10.1007/s00408-015-9836-1

32. Gou LY, Niu FY, Wu YL, Zhong WZ. Differences in driver genes between smoking-related and non-smoking-related lung cancer in the Chinese population. *Cancer* (2015) 17:3069–79. doi: 10.1002/cncr.29531

33. Lee JJ, Park S, Park H, Kim S, Lee J, Lee J, et al. Tracing oncogene rearrangements in the mutational history of lung adenocarcinoma. *Cell* (2019) 177:1842–57.e21. doi: 10.1016/j.cell.2019.05.013

34. Li G, Mei Y, Yang F, Yi S, Wang L. Identification of genome variations in patients with lung adenocarcinoma using whole genome re−sequencing. *Mol Med Rep* (2017) 16:9464–72. doi: 10.3892/mmr.2017.7805

35. Locher C, Debieuvre D, Coëtmeur D, Goupil F, Molinier O, Collon T, et al. Major changes in lung cancer over the last ten years in France: the KBP-CPHG studies. *Lung Cancer* (2013) 81:32–8. doi: 10.1016/j.lungcan.2013.03.001

36. Huang C, Qu X, Du J. Proportion of lung adenocarcinoma in female never-smokers has increased dramatically over the past 28 years. *J Thorac Dis* (2019) 11:2685–8. doi: 10.21037/jtd.2019.07.08

37. Lewis DR, Check DP, Caporaso NE, Travis WD, Devesa SS. US Lung cancer trends by histologic type. *Cancer* (2014) 120:2883–92. doi: 10.1002/cncr.28749

38. Jiang X, de Groh M, Liu S, Liang H, Morrison H. Rising incidence of adenocarcinoma of the lung in Canada. *Lung Cancer* (2012) 78:16–22. doi: 10.1016/j.lungcan.2012.06.002

39. GBD 2015 Tobacco Collaborators. Smoking prevalence and attributable disease burden in 195 countries and territories, 1990-2015: a systematic analysis from the global burden of disease study 2015. *Lancet* (2017) 389:1885–906. doi: 10.1016/S0140-6736(17)30819-X

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 11 August 2022
DOI 10.3389/fonc.2022.900049

OPEN ACCESS

EDITED BY
Xinming Zhao,
The Fourth Hospital of Hebei Medical University, China

REVIEWED BY
Shouliang Qi,
Northeastern University, China
Aamer Chughtai,
Case Western Reserve University, United States

*CORRESPONDENCE
Quan Zhu
zhuquan@njmu.edu.cn
Mei Yuan
njmu_yuanmei@163.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 19 March 2022
ACCEPTED 20 July 2022
PUBLISHED 11 August 2022

CITATION
Zhang T, Zhang C, Zhong Y, Sun Y, Wang H, Li H, Yang G, Zhu Q and Yuan M (2022) A radiomics nomogram for invasiveness prediction in lung adenocarcinoma manifesting as part-solid nodules with solid components smaller than 6 mm.
Front. Oncol. 12:900049.
doi: 10.3389/fonc.2022.900049

A radiomics nomogram for invasiveness prediction in lung adenocarcinoma manifesting as part-solid nodules with solid components smaller than 6 mm

Teng Zhang[1†], Chengxiu Zhang[2†], Yan Zhong[1], Yingli Sun[3], Haijie Wang[2], Hai Li[4], Guang Yang[2], Quan Zhu[5*] and Mei Yuan[1*]

[1]Department of Radiology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China, [2]Shanghai Key Laboratory of Magnetic Resonance, East China Normal University, Shanghai, China, [3]Department of Radiology, Huadong Hospital Affiliated to Fudan University, Shanghai, China, [4]Department of Pathology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China, [5]Department of Thoracic Surgery, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China

Objective: To investigate whether radiomics can help radiologists and thoracic surgeons accurately predict invasive adenocarcinoma (IAC) manifesting as part-solid nodules (PSNs) with solid components <6 mm and provide a basis for rational clinical decision-making.

Materials and Methods: In total, 1,210 patients (mean age ± standard deviation: 54.28 ± 11.38 years, 374 men and 836 women) from our hospital and another hospital with 1,248 PSNs pathologically diagnosed with adenocarcinoma *in situ* (AIS), minimally invasive adenocarcinoma (MIA), or IAC were enrolled in this study. Among them, 1,050 cases from our hospital were randomly divided into a derivation set (n = 735) and an internal validation set (n = 315), 198 cases from another hospital were used for external validation. Each labeled nodule was segmented, and 105 radiomics features were extracted. Least absolute shrinkage and selection operator (LASSO) was used to calculate Rad-score and build the radiomics model. Multivariable logistic regression was conducted to identify the clinicoradiological predictors and establish the clinical-radiographic model. The combined model and predictive nomogram were developed based on identified clinicoradiological independent predictors and Rad-score using multivariable logistic regression analysis. The predictive performances of the three models were compared *via* receiver operating characteristic (ROC) curve analysis. Decision curve analysis (DCA) was performed on both the internal and external validation sets to evaluate the clinical utility of the nomogram.

Results: The radiomics model showed superior predictive performance than the clinical-radiographic model in both internal and external validation sets (Az values, 0.884 vs. 0.810, $p = 0.001$; 0.924 vs. 0.855, $p < 0.001$, respectively). The

combined model showed comparable predictive performance to the radiomics model (Az values, 0.887 vs. 0.884, p = 0.398; 0.917 vs. 0.924, p = 0.271, respectively). The clinical application value of the nomogram developed based on the Rad-score, maximum diameter, and lesion shape was confirmed, and DCA demonstrated that application of the Rad-score would be beneficial for radiologists predicting invasive lesions.

Conclusions: Radiomics has the potential as an independent diagnostic tool to predict the invasiveness of PSNs with solid components <6 mm.

KEYWORDS

radiomics, nomogram, adenocarcinoma of lung, neoplasm invasiveness, tomography, X-ray computed

Introduction

With the increasing use of low-dose computed tomography (LDCT) in the screening of high-risk populations for lung cancer, the detection rate of part-solid nodules (PSNs) has been increasing, especially in Asian women and non-smokers (1–4). Previous research has shown that persistent PSNs are highly correlated with early-stage lung adenocarcinoma, including invasive adenocarcinoma (IAC), minimally invasive adenocarcinoma (MIA), and adenocarcinoma *in situ* (AIS) (5–7).

Accurate differentiation between IAC and AIS/MIA appearing as PSNs is critical and can determine the patient's treatment options. Unlike IAC, AIS/MIA can be resected by limited wedge resection or segmentectomy rather than lobectomy to maximize the preservation of functional pulmonary parenchyma. Moreover, lymph node exploration is not required for AIS/MIA (8, 9). For clinical management, as long as the lesions suspected of AIS/MIA remain stable, the strategy of conservative periodic follow-up with CT, with surgical resection in case of lesion growth, has been widely accepted by clinicians, ultimately avoiding unnecessary surgery for patients (4, 10, 11).

One of the key factors to predict the invasiveness of PSNs is the assessment of the size of the solid component within the nodules, which is highly correlated with the pathologically invasive foci of adenocarcinomas (12–14). For PSNs, a size criterion of solid component diameter ≥6 mm is widely accepted to discriminate IAC from AIS/MIA on CT, which is also the newly revised threshold standard for T-factor staging of adenocarcinoma. Most IACs commonly manifest as PSNs with solid components ≥6 mm and can be easily and accurately diagnosed by radiologists and thoracic surgeons (4, 15, 16). A statement from the Fleischner Society suggested that surgical resection should be considered if solid components were ≥6 mm in PSNs, while yearly surveillance CT is recommended for PSNs with solid components <6 mm (4, 17). However, due to the insufficient CT resolution, the ground-glass components of PSNs may contain invasive foci that cannot be recognized by the naked eyes; a large number of IAC cases also present as PSNs with solid component <6 mm or even non-solid nodules (NSNs), which were difficult to distinguish from AIS/MIA, and closer follow-up is needed for these lesions (18). Ahn et al. (19) indicated that the sizes of the solid component measured on CT images commonly underestimate the real size of invasive foci on pathology. Therefore, it is a great challenge for radiologists to predict the invasiveness of PSNs with a solid component <6 mm and NSNs because of their pathological diversity.

Many studies regarding the invasiveness prediction of PSNs have been reported with different methods, including radiographic feature evaluation and quantitative analysis (20–22). However, these studies have limitations: 1) there are no limits set on the size of solid components in PSNs. If too many PSNs with a solid component ≥6 mm (the pathological type is mostly IAC) are included in a study, the predictive performance may be exaggerated; 2) many evaluated radiographic features show an overlap between IAC and AIS/MIA; 3) previous studies used varied quantitative parameters and reported different results, making it unclear whether these studies can help radiologists improve the prediction performance.

Radiomics has been widely used to establish diagnosis and prediction models for tumor grading and staging, treatment outcome evaluation, and prognosis prediction by extracting and selecting predefined subtle image features (23, 24). Sun et al. (25) and Yuan et al. (26) have successfully used radiomics to predict the invasiveness of NSNs and PSNs, respectively. However, few studies have focused on the prediction of invasiveness of PSNs with a solid component <6 mm, whose diagnosis remains challenging for radiologists. Therefore, the purpose of our

study was to investigate whether radiomics is able to help radiologists accurately predict IAC manifesting as PSNs with solid components <6 mm and provide a basis for rational clinical decision-making.

Materials and methods

The ethical committee of our hospital approved this retrospective study and waived the informed consent for the patients (approval number: 2021-SR-053).

Data source and patient selection

We searched the institution's database and collected the clinical and imaging data of 3,326 patients who underwent surgery in our hospital and were pathologically diagnosed with AIS, MIA, or IAC from January 2015 to December 2020. The dataset of our hospital included 1,012 patients (mean age ± standard deviation, 54.16 ± 11.21 years, 316 men and 696 women) with 1,050 PSNs satisfying the following inclusion criteria: 1) the lesions manifested as PSNs on CT imaging, with the maximum diameter of the solid components <6 mm (excluding bronchi and vessels); 2) the maximum diameter of the PSNs was 5–30 mm; 3) non-enhanced CT scans were performed within 2 weeks prior to surgery; 4) lesions were completely removed, and the pathological diagnosis was unambiguous. In accordance with previous studies, the CT threshold of the solid components in PSNs was set to >-188 HU (27). Three-dimensional (3D) Slicer software (version 4.12; National Institutes of Health; https://www.slicer.org) can automatically identify solid components in PSNs. According to the automatic segment results of the solid components in PSNs in 3D Slicer software, authors #1 and #3 (with 10 and 7 years' experience in chest CT imaging, respectively) separately measured the maximum diameter of the solid components from the axial, coronal, and sagittal images in Picture Archiving and Communication System (PACS) (Supplementary Figure S1). PSNs with solid components greater than 6 mm measured from any direction will be excluded from the study. Author #9, a radiologist with 16 years' experience, reviewed the measurement results of authors #1 and #3 to reach a consensus.

Finally, the dataset from our hospital was randomly divided into a derivation set (735 cases: 67 AIS, 316 MIA, and 352 IAC) and an internal validation set (315 cases: 29 AIS, 135 MIA, and 151 IAC) at a ratio of 7:3.

Authors #1, #3, and #9 used the same methods andreviewed patients who underwent surgery in another hospital and were pathologically diagnosed with AIS, MIA, or IAC from January 2020 to December 2020. Finally, 198 cases (25 AIS, 79 MIA, and 94 IAC) were enrolled in our study as the external validation set. The workflow was illustrated in Figure 1.

None of the 597 IAC cases included in our study had lymph node metastases. Only 16 IAC cases were classified as pT2 stage because of the presence of pleural invasion and the rest were all classified as pT1 stage according to the eighth edition of the tumor, node, and metastasis (TNM) classification of lung cancer.

Diagnostic criteria

The resection specimens were fixed with formalin, embedded in paraffin, sectioned, and stained with hematoxylin and eosin. Author #6, a pathologist with 26 years' work experience, reviewed the pathological classification of all specimens based on the 2011 classification criteria for lung adenocarcinoma proposed by International Association for the

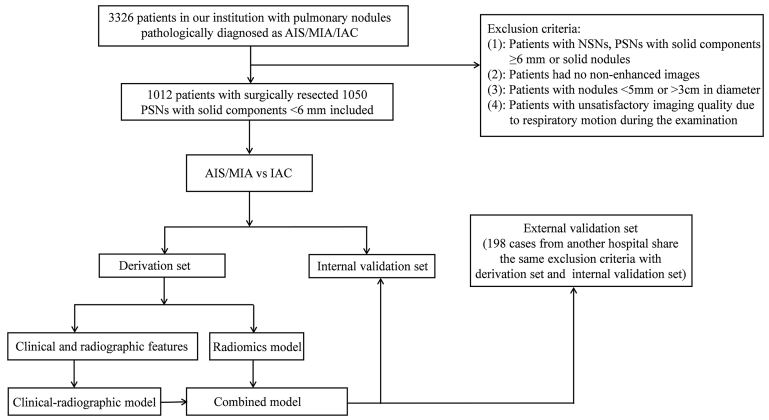


FIGURE 1
The workflow of our study.

Study of Lung Cancer/American Thoracic Society/European Respiratory Society (IASLC/ATS/ERS).

CT examination methods

All patients underwent CT examination using one of the four CT scanners: SOMATOM Definition AS+, SOMATOM Sensation 16, and GE Discovery CT750 HD of our hospital and GE LightSpeed VCT of another hospital. The detailed scan and reconstruction parameters and the number of patients performed by each scanner were shown in Supplementary Table S1.

Establishing the clinical-radiographic model

The patient's clinical information was collected from electronic medical records, including: 1) age, 2) sex, 3) smoking history, 4) carcinoembryonic antigen (CEA) level, 5) history of chronic obstructive pulmonary disease (COPD), 6) history of other cancers, and 7) family history of lung cancer. Authors #1 and #3, who did not know the patients' pathological diagnosis, interpreted the CT images individually under lung window settings (width, 1,200 Hu; level, -600 Hu). The radiographic features of all lesions evaluated in this study include the following: 1) lesion location; 2) maximal axial diameter; 3) maximum axial diameter of the solid component; 4) consolidation-to-tumor ratio (CTR); 5) border (undefined or defined); 6) shape (round/oval or irregular); 7) vacuole sign (lesions with cystic cavities with the diameter <5 mm); 8) air bronchogram sign (dilated bronchioles observed in lesions); 9) microvascular sign (lesions with convergent, dilated, or tortuous supplying vessels); and 10) pleural indentation sign (pleura adjacent to lesions showed thickening or contraction). The CTR (%) was calculated as 100×(maximum axial diameter of the solid component/maximum axial diameter of the lesion) referring to previous studies (28, 29). Kappa values and intraclass correlation coefficients (ICCs) were calculated to assess the consistency of the two authors' evaluations. To reach a consensus, author #9 rechecked the image interpretation results.

Nodule segmentation and radiomics feature extraction

CT images (DICOM format) were retrieved from the institution archive and loaded into a personal computer for further analysis. The volume of interest (VOI) was automatically segmented using a homemade software MultiLabel (version 1.1, Shanghai Key Laboratory of Magnetic Resonance, East China Normal University, China). Manual adjustment for the precise edge of the VOI was performed by author #1 if the border of the lesion was undefined or to ensure the large vessels and bronchioles were excluded from the VOI (Supplementary Figure S2). Author #9 reviewed the VOIs to ensure accurate segmentation. Then, the whole set of CT images with VOI segmentation information was converted to NII format for further radiomics analysis.

We normalized all images with the following formula: $f(x)=1000*(x-\mu_x)/\sigma_x$, where μ_x and σ_x denote the mean and standard deviation of the image intensity, respectively, before 3D radiomics features were extracted from the original image with PyRadiomics (Ver. 3.0) (30). We extracted 105 commonly used features in radiomics analysis, including 18 gray-level histogram features (e.g., mean, kurtosis, skewness), 14 shape features (e.g., compactness, sphericity), and 73 high-order texture features [gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level size-zone matrix (GLSZM), and neighborhood gray-tone difference matrix (NGTDM)]. Most features defined by Pyradiomics follow feature definitions as described by the Imaging Biomarker Standardization Initiative (IBSI) (31).

In order to check the inter- and intra-observer reproducibility of the 105 radiomics features that we extracted, 60 cases (20 AIS, 20 MIA, and 20 IAC) were randomly chosen for analysis. Author #1 and author #3 repeated the nodule segmentation procedure for the selected 60 cases separately approximately 3 months later. The intra- and inter-observer agreement of the 105 extracted features were assessed by interclass correlation coefficients and ICCs. An ICC greater than 0.75 indicated good reproducibility of the feature extraction.

Feature selection and rad-score building

Firstly, all features were normalized with z-score (subtracted mean value and divided by standard deviation), then minimum-Redundancy Maximum Relevance (mRMR) was used to remove redundant features. The remaining 30 features were used to build the radiomics model with least absolute shrinkage and selection operator (LASSO), a classifier suitable for high-dimensional data regression. Parameter and thus the appropriate number of the most weighted predictive features was determined using a 10-fold cross-validation over the derivation set. To minimize the number of features in the final model, we chose the model with the least number of features and a binomial deviance within 1 standard deviation from the minimum binomial deviance.

After the model had been built, Rad-score, namely, the predictive probability of the radiomics model for each patient, was calculated *via* a linear combination of the selected most weighted features with their respective coefficients.

Statistical analyses

Univariable and multivariable logistic regression analyses were carried out on the clinical and radiographic features of the derivation set to determine the independent predictors for IAC and establish the clinical-radiographic model. The independent predictors and Rad-score were analyzed using multivariable logistic regression; thus, a combined model and an individual prediction nomogram were constructed. Receiver operating characteristic (ROC) curve analysis was used to evaluate the performance of the clinical-radiographic model, radiomics model, and combined model in the derivation, internal validation, and external validation sets. The optimal cutoff value was determined by Youden index in the ROC analysis. DeLong's test was also used to compare the performance of the models. Model evaluation metrics such as positive predictive value (PPV), negative predictive value (NPV), F1-score, and Matthews correlation coefficient (MCC) were calculated to identify the best prediction model. Waterfall plot was used to show the prediction probability of all patients, and calibration curves were plotted to analyze the diagnostic performance of the nomogram in each dataset. Hosmer–Lemeshow test and decision curve analysis (DCA) were used to evaluate the goodness of fit and clinical value of the nomogram.

Statistical analysis was performed using IBM SPSS software (version 26.0; https://www.ibm.com) and R software (version 4.1.0; https://www.r-project.org). Specifically, we used rms package for calibration analysis, ResourceSelection package for Hosmer–Lemeshow test, rmda for DCA, mRMRe for feature selection, and glmnet for LASSO. p-values <0.05 were considered statistically significant.

Results

Predictive performance of the clinical-radiographic model

Supplementary Table S2 showed the inter-observer agreement for the radiographic sign evaluation and measurement of PSNs. The Kappa values for lesion measurement and radiographic features evaluation were medium to high.

It can be seen from Table 1 that there were no statistically significant differences between the derivation set and internal/external validation set in the comparison of clinical and radiographic features.

Table 2 and Figure 2 present the comparison on the clinical and radiographic features between AIS/MIA and IAC in the derivation, internal validation, and external validation sets. Univariable and multivariable logistic regression analyses in the derivation set revealed the maximum diameter [odds ratio (OR) 3.62, 95% confidence interval (CI) 2.26–5.80, $p < 0.001$], lesion shape (OR 1.89, 95% CI 1.26–2.81, $p = 0.002$), vacuole sign (OR 1.89, 95% CI 1.07–3.32, $p = 0.028$), microvascular sign (OR 1.91, 95% CI 1.31–2.79, $p = 0.001$), and maximum diameter of the solid component (OR 26.83, 95% CI 6.81–105.76, $p < 0.001$) were independent predictors for IAC.

The Az value of the clinical-radiographic model was 0.779 (95% CI 0.747–0.809) in the derivation set, 0.810 (95% CI 0.762–0.852) in the internal validation set, and 0.855 (95% CI 0.799–0.901) in the external validation set (Figure 3).

Feature selection and rad-score building

The inter-observer ICCs, calculated on the basis of author #1's first-extracted 105 features and those of author #3 ranged from 0.80 to 0.99 (Supplementary Table S3). The intra-observer ICCs, calculated based on author #1's twice feature extraction ranged from 0.84 to 0.99 (Supplementary Table S4). Therefore, the 105 features we extracted proved robust and achieved satisfactory inter- and intra-observer reproducibility.

A Rad-score was calculated for each patient based on seven features with non-zero coefficients selected from the 105 robust radiomics features using a LASSO logistic regression model ($\lambda = 0.039727$) (Figures 4A, B).

Rad-score = 0.831176 * gldm_DependenceEntropy
+ 0.301168 * firstorder_RootMeanSquared
- 0.004266 * shape_SurfaceVolumeRatio
+ 0.336894 * shape_Maximum2DDiameterSlice
+ 0.155065 * firstorder_90Percentile
+ 0.107055 * glcm_JointEntropy
+ 0.039099 * shape_Maximum2DDiameterColumn - 0.101629

The bar chart of the coefficients of features used in the model is shown in Figure 4C.

The Az value of the radiomics model was 0.865 (95% CI 0.839–0.889) in the derivation set, 0.884 (95% CI 0.843–0.917) in the internal validation set, and 0.924 (95% CI 0.878–0.957) in the external validation set (Figure 3).

Prediction nomogram construction and validation

Multivariable logistic regression analysis identified the Rad-score (OR 2,232.55, 95% CI 650.95–7656.89, $p < 0.001$), maximum diameter (OR 0.40, 95% CI 0.21–0.75, $p = 0.004$), and lesion shape (OR 1.94, 95% CI 1.23–3.06, $p = 0.004$) as independent predictors for IAC. All of these parameters were used to develop a prediction nomogram. Representative examples of the nomogram to predict the invasiveness are given in Figures 5.

The Az value of the combined model was 0.872 (95% CI 0.846–0.895) in the derivation set, 0.887 (95% CI 0.847–0.920) in the internal validation set, and 0.917 (95% CI 0.869–0.951) in the

TABLE 1 Comparison of clinical and radiographic characteristics between derivation and internal/external validation sets.

Characteristics	Derivation set (n = 735)	Internal validation set (n = 315)	External validation set (n = 198)	*p* value	*p'* value
Clinical Characteristics					
Age (years)	54.27 ± 10.90	53.91 ± 11.91	54.90 ± 12.29	0.364[a]	0.512[a]
Sex (Men/Women)	227 (30.9)/508 (69.1)	101 (32.1)/214 (67.9)	58 (29.3)/140 (70.7)	0.760[b]	0.666[b]
Smoking history (Yes/No)	89 (12.1)/646 (87.9)	36 (11.4)/279 (88.6)	20 (10.1)/178 (89.9)	0.835[b]	0.435[b]
CEA (ng/ml)	1.96 ± 1.28	2.08 ± 1.60	2.02 ± 1.27	0.189[a]	0.568[a]
History of COPD (Yes/No)	50 (6.8)/685 (93.2)	18 (5.7)/297 (94.3)	9 (4.5)/189 (95.5)	0.603[b]	0.247[b]
History of other cancers (Yes/No)	53 (7.2)/682 (92.8)	21 (6.7)/294 (93.3)	16 (8.1)/182 (91.9)	0.854[b]	0.678[b]
Family history of lung cancer (Yes/No)	34 (4.6)/701 (95.4)	15 (4.8)/300 (95.2)	14 (7.1)/184 (92.9)	0.676[b]	0.167[b]
Radiographic Characteristics					
Lesion location				0.833[b]	0.224[b]
Right upper lobe	271 (36.9)	117 (37.1)	61 (30.8)		
Right middle lobe	50 (6.8)	22 (7.0)	17 (8.6)		
Right lower lobe	119 (16.2)	49 (15.6)	30 (15.2)		
Left upper lobe	182 (24.7)	86 (27.3)	63 (31.8)		
Left lower lobe	113 (15.4)	41 (13.0)	27 (13.6)		
Maximum diameter (cm)	1.25 ± 0.43	1.26 ± 0.43	1.20 ± 0.41	0.236[a]	0.215[a]
Maximum diameter of the solid component (cm)	0.31 ± 0.13	0.30 ± 0.13	0.30 ± 0.14	0.493[a]	0.738[a]
CTR (%)	25.89 ± 11.23	26.09 ± 13.95	27.12 ± 13.01	0.359[a]	0.474[a]
Lesion shape				0.717[b]	0.175[b]
Round/Oval	537 (73.1)	226 (71.7)	135 (68.2)		
Irregular	198 (26.9)	89 (28.3)	63 (31.8)		
Lesion border				0.920[b]	0.350[b]
Defined	587 (79.9)	250 (79.4)	164 (82.8)		
Undefined	148 (20.1)	65 (20.6)	34 (17.2)		
Vacuole sign (Yes/No)	79 (10.7)/656 (89.3)	32 (10.2)/283 (89.8)	17 (8.6)/181 (91.4)	0.861[b]	0.374[b]
Air bronchogram (Yes/No)	81 (11.0)/654 (89.0)	28 (8.9)/287 (91.1)	14 (7.1)/184 (92.9)	0.354[b]	0.103[b]
Microvascular sign (Yes/No)	226 (30.7)/509 (69.3)	98 (31.1)/217 (68.9)	63 (31.8)/135 (68.2)	0.850[b]	0.773[b]
Pleural indentation (Yes/No)	228 (31.0)/507 (69.0)	110 (34.9)/205 (65.1)	73 (36.9)/125 (63.1)	0.243[b]	0.118[b]

CEA, carcinoembryonic antigen; COPD, chronic obstructive pulmonary disease; CTR, consolidation-to-tumor ratio.
Values are presented as no. (%) or mean ± standard deviation.
[a]Mann–Whitney U test. [b]Pearson's chi-square test and Fisher's exact test.
p value: Derivation set vs. Internal validation set; p' value: Derivation set vs. External validation set.

external validation set (Figure 3). It can be seen from the detailed metrics listed in Table 3 that the radiomics models showed superior predictive performance than the clinical-radiographic models in all sets; however, the combined models showed comparable predictive performance to the radiomics models.

The calibration curve of the radiomics nomogram smoothed with bootstrapping also indicated good agreement between predicted probability and actual occurrence in the derivation, internal validation, and external validation sets (Figure 6A). The Hosmer–Lemeshow test indicated no significant difference between the combined model's predictions and the observed values in the derivation, internal validation, and external validation sets (p = 0.779, p = 0.580, p = 0.209, respectively), implying the model's good generalization.

Figure 6B showed the decision curves of the developed models on both the internal and external validation sets. It would be more beneficial to use a model with Rad-score for identifying the invasive lesions than that without the Rad-score if the threshold probability of a patient was in the range of 10%~85%.

Waterfall plots of the combined model on both the internal and external validation sets were shown in Figure 6C. The cutoff value was set by maximizing the Youden index in the derivation set. From the waterfall plots, it can be seen clearly that the combined model can differentiate IAC from AIS/MIA well.

Discussion

Our research objective was to predict the invasiveness of lung adenocarcinoma manifesting as PSNs with solid components <6 mm. We confirmed that radiomics was superior to the clinical-radiographic model and showed comparable predictive performance to the combined model in

TABLE 2 Comparison on the clinical and radiographic characteristics between AIS-MIA and IAC in the derivation, internal validation, and external validation sets.

Characteristics	Derivation set (n =735)		*p* value	Internal validation set (n = 315)		*p* value	External validation set (n = 198)		*p* value
	AIS-MIA group (n = 383)	IAC group (n = 352)		AIS-MIA group (n = 164)	IAC group (n = 151)		AIS-MIA group (n = 104)	IAC group (n = 94)	
Clinical Characteristics									
Age (years)	52.47 ± 11.18	56.23 ± 10.24	0.000[a]	52.31 ± 12.62	55.64 ± 10.88	0.016[a]	52.21 ± 12.89	58.99 ± 10.19	0.000[a]
Sex (Men/Women)	101 (26.4)/282 (73.6)	126 (35.8)/226 (64.2)	0.006[b]	51 (31.1)/113 (68.9)	50 (33.1)/101 (66.9)	0.702[b]	27 (26.0)/77 (74.0)	31 (33.0)/63 (67.0)	0.279[b]
Smoking history (Yes/No)	34 (8.9)/349 (91.1)	55 (15.6)/297 (84.4)	0.005[b]	11 (6.7)/153 (93.3)	25 (16.6)/126 (83.4)	0.006[b]	8 (7.7)/96 (92.3)	12 (12.8)/82 (87.2)	0.237[b]
CEA (ng/ml)	1.81 ± 1.14	2.12 ± 1.40	0.003[a]	1.97 ± 1.75	2.20 ± 1.41	0.017[a]	1.83 ± 1.04	2.23 ± 1.45	0.025[a]
History of COPD (Yes/No)	21 (5.5)/362 (94.5)	29 (8.2)/323 (91.8)	0.138[b]	6 (3.7)/158 (96.3)	12 (7.9)/139 (92.1)	0.101[b]	4 (3.8)/100 (96.2)	5 (5.3)/89 (94.7)	0.877[b]
History of other cancers (Yes/No)	20 (5.2)/363 (94.8)	33 (9.4)/319 (90.6)	0.030[b]	7 (4.3)/157 (95.7)	14 (9.3)/137 (90.7)	0.075[b]	7 (6.7)/97 (93.3)	9 (9.6)/85 (90.4)	0.463[b]
Family history of lung cancer (Yes/No)	20 (5.2)/363 (94.8)	14 (4.0)/338 (96.0)	0.422[b]	7 (4.3)/157 (95.7)	8 (5.3)/143 (94.7)	0.668[b]	6 (5.8)/98 (94.2)	8 (8.5)/86 (91.5)	0.452[b]
Radiographic Characteristics									
Lesion location			0.571[b]			0.217[b]			0.200[b]
Right upper lobe	148 (38.6)	123 (34.9)		64 (39.0)	53 (35.1)		27 (26.0)	34 (36.2)	
Right middle lobe	25 (6.5)	25 (7.1)		7 (4.3)	15 (9.9)		10 (9.6)	7 (7.4)	
Right lower lobe	58 (15.1)	61 (17.3)		28 (17.1)	21 (13.9)		21 (20.2)	9 (9.6)	
Left upper lobe	99 (25.9)	83 (23.6)		47 (28.7)	39 (25.8)		31 (29.8)	32 (34.0)	
Left lower lobe	53 (13.9)	60 (17.1)		18 (10.9)	23 (15.3)		15 (14.4)	12 (12.8)	
Maximum diameter (cm)	1.09 ± 0.38	1.42 ± 0.42	0.000[a]	1.08 ± 0.34	1.45 ± 0.44	0.000[a]	0.99 ± 0.26	1.42 ± 0.43	0.000[a]
Maximum diameter of the solid component (cm)	0.26 ± 0.11	0.35 ± 0.14	0.000[a]	0.25 ± 0.10	0.34 ± 0.13	0.000[a]	0.24 ± 0.10	0.37 ± 0.14	0.000[a]
CTR (%)	25.72 ± 11.53	26.07 ± 10.90	0.424[a]	25.67 ± 13.22	26.54 ± 14.74	0.793[a]	25.77 ± 12.79	28.62 ± 13.15	0.112[a]
Lesion shape			0.000[b]			0.000[b]			0.000[b]
Round/Oval	326 (85.1)	211 (59.9)		141 (86.0)	85 (56.3)		83 (79.8)	52 (55.3)	
Irregular	57 (14.9)	141 (40.1)		23 (14.0)	66 (43.7)		21 (20.2)	42 (44.7)	
Lesion border			0.190[b]			0.014[b]			0.118[b]
Defined	313 (81.7)	274 (77.8)		139 (84.8)	111 (73.5)		82 (78.8)	82 (87.2)	
Undefined	70 (18.3)	78 (22.2)		25 (15.2)	40 (26.5)		22 (21.2)	12 (12.8)	
Vacuole sign (Yes/No)	24 (6.3)/359 (93.7)	55 (15.6)/297 (84.4)	0.000[b]	16 (9.8)/148 (90.2)	16 (10.6)/135 (89.4)	0.805[b]	5 (4.8)/99 (95.2)	12 (12.8)/82 (87.2)	0.046[b]
Air bronchogram (Yes/No)	23 (6.0)/360 (94.0)	58 (16.5)/294 (83.5)	0.000[b]	7 (4.3)/157 (95.7)	21 (13.9)/130 (86.1)	0.003[b]	6 (5.8)/98 (94.2)	8 (8.5)/86 (91.5)	0.452[b]
Microvascular sign (Yes/No)	68 (17.8)/315 (82.2)	158 (44.9)/194 (55.1)	0.000[b]	30 (18.3)/134 (81.7)	68 (45.0)/83 (55.0)	0.000[b]	25 (24.0)/79 (76.0)	38 (40.4)/56 (59.6)	0.013[b]
Pleural indentation (Yes/No)	95 (24.8)/288 (75.2)	133 (37.8)/219 (62.2)	0.000[b]	42 (25.6)/122 (74.4)	68 (45.0)/83 (55.0)	0.000[b]	30 (28.8)/74 (71.2)	43 (45.7)/51 (54.3)	0.014[b]

AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma; IAC, invasive adenocarcinoma; CEA, carcinoembryonic antigen; COPD, chronic obstructive pulmonary disease; CTR, consolidation-to-tumor ratio.
Values are presented as no. (%) or mean ± standard deviation.
[a]Mann–Whitney U test. [b]Pearson's chi-square test and Fisher's exact test.

10.3389/fonc.2022.900049

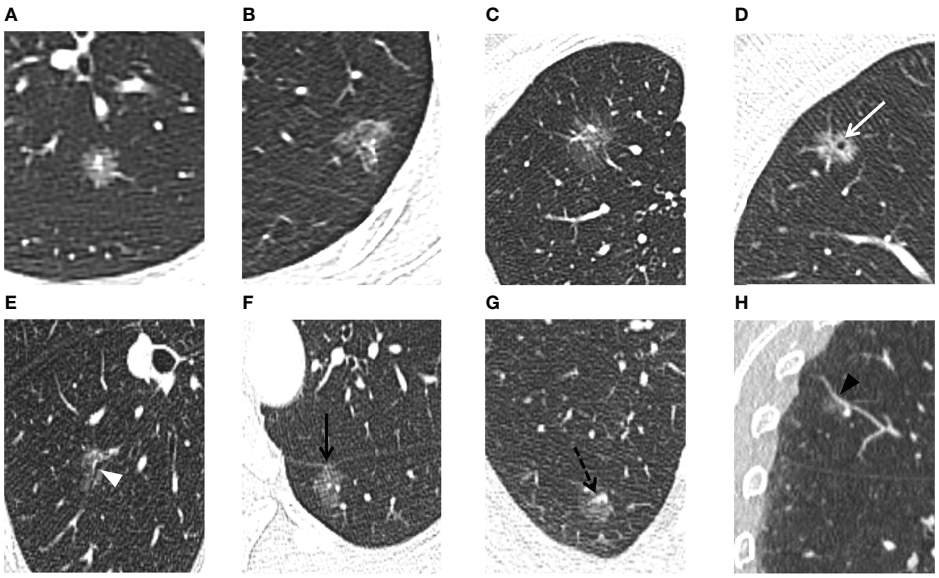


FIGURE 2
For the PSNs included in our study, thin-section CT radiographic features were evaluated. The lesion shape is evaluated as **(A)** round/oval or **(B)** irregular. The lesion border is evaluated as **(A, B)** defined and **(C)** undefined. **(D)** Vacuole sign: MIA in the right upper lobe exhibits a bubble-like lucency within the nodule (white arrow). **(E)** Air bronchogram sign: MIA in the right lower lobe shows air-filled bronchi present inside the lesion (white arrowhead). **(F)** Pleural indentation: MIA in the left lower lobe shows pleural indentation adjacent to the oblique fissure (black arrow). **(G, H)** Microvascular sign: IAC in the right upper lobe shows a small adjacent pulmonary vessel entering the lesion (black dotted arrow). Reverse tracing shows that the blood vessel is a branch of the pulmonary artery. Coronal reconstruction better shows the dilation of the supplying vessel (black arrowhead).

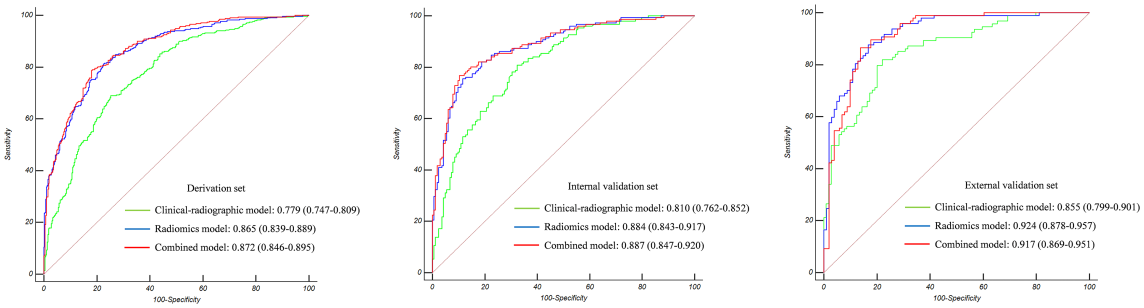


FIGURE 3
The Az value of the three models in the derivation set, internal validation set, and external validation set. Radiomics showed superior predictive performance than clinical-radiographic models in all sets; however, the combined models showed comparable predictive performance to the radiomics models.

differentiating IAC from AIS/MIA. It seems that radiomics can provide a simple and robust prediction method for accurate preoperative judgment of lesion invasiveness.

Our statistical analysis of the patients' clinical data shows that the incidence of IAC increased with age, male gender, smoking history, and high CEA level, which is consistent with the results of previous studies (32). Although some studies have found that patients with COPD, history of other cancers, and family history of lung cancer were more susceptible to lung cancer (33, 34), our data show that these features have no statistical significance in predicting the invasiveness of PSNs.

Maximum diameter, lesion shape, vacuole sign, and microvascular sign were considered independent predictors for IAC in our study. The size of PSNs is significantly related to the risk of malignancy and is the leading factor in the management of PSNs. In our study, the average maximum diameter of IAC was significantly larger than that of AIS/MIA (1.42 cm vs. 1.09 cm, $p < 0.001$), and the predicted cutoff value for IAC

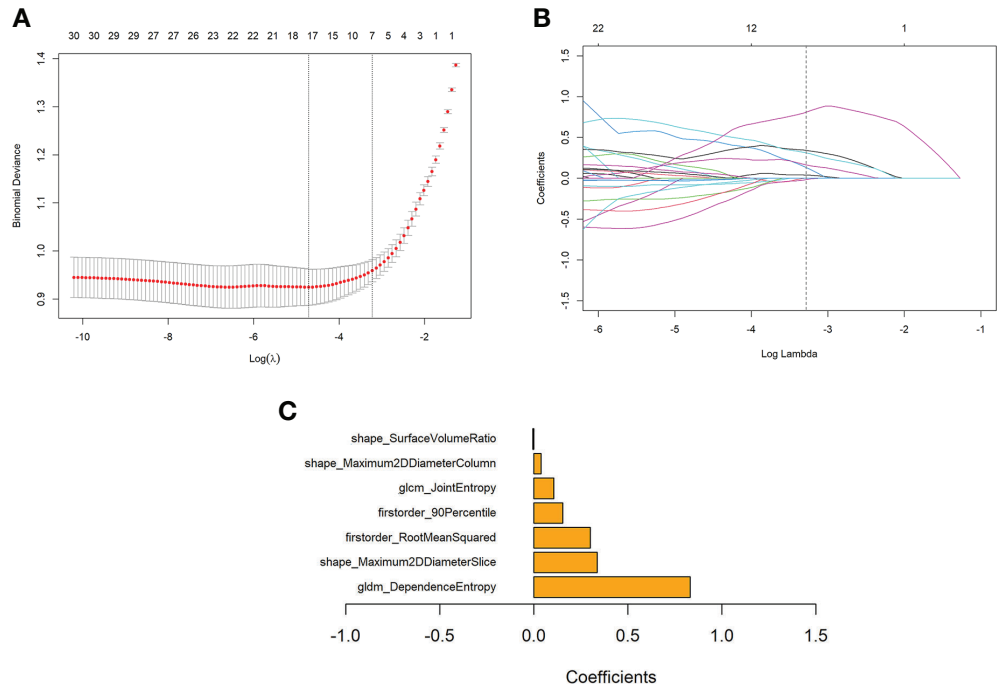


FIGURE 4
Features selection using the LASSO regression model and the selection of the tuning parameter λ. **(A)** Change of binomial deviance with log(λ). The maximum log(λ) corresponding to the binomial deviance within 1 standard error from the minimum binomial deviance was chosen for the final model. **(B)** Change of the number of features with non-zero coefficients with log(λ), as determined in a 10-fold validation. **(C)** Coefficients of the seven features retained in the model.

was 1.18 cm. Li et al. (35) and Zhou et al. (36) confirmed that the maximum diameter is an independent predictor of the invasiveness of PSNs, and the predicted cutoff values for IAC in their studies were 1.65 and 1.96 cm, respectively, which were significantly higher than ours. This difference could be due to the fact that other studies included PSNs with solid components ≥6 mm, which may be more aggressive and have larger diameters. As the invasiveness of PSNs increased, their morphology became irregular and the probability of vacuole formation increased due to the proliferation of fibroblasts, tumor cell infiltration in the lung interstitium, and heterogeneity of growth speed inside the tumor (37). Vascular remodeling and sustained angiogenesis play an important role in the early development and progression of tumors (38). When tumor cells infiltrate the pulmonary interstitium, they will create traction on the surrounding blood vessels, causing them to stiffen, twist, and aggregate. Many previous studies have confirmed that irregular shape, vacuole sign, and microvascular sign are more frequently seen in IAC manifesting as PSNs, but no studies considered any of these features as independent predictors (32, 35, 36). The main reason may be that their predictive model was based on the comprehensive analysis of both radiographic features and quantitative parameters instead of only analyzing radiographic features as we did in our study. Zhao et al. (39) found that the mean CT value showed superior predictive performance compared with irregular shape and microvascular sign and was considered to be an independent predictor for IAC. In addition, the different findings could also be due to the exclusion of PSNs with solid components ≥6 mm from our study and the difference in sample size. At present, there are a few studies on PSNs with solid components <6 mm; hence, more studies are needed to determine which radiographic features can be used as the best predictors for invasiveness, and our research results need further verification.

Previous studies have confirmed that measuring the maximum size of solid components or comparing the CTR is more accurate than measuring the maximum diameter of lesions in predicting the invasiveness of PSNs (12, 40). Our study also confirmed that the maximum size of solid components was an independent predictor for IAC. The CTR calculated in our study showed no significant statistical difference between IAC and AIS/MIA. It may be because our study excluded PSNs with solid components ≥6 mm while maintaining the same size enrollment criteria of the lesions (5–30 mm) with previous studies. Although we can clearly identify solid components smaller than 6 mm in PSNs with the aid of automatic segmentation software, errors still exist to a certain extent in the measurement of solid components. No studies focused on the solid component

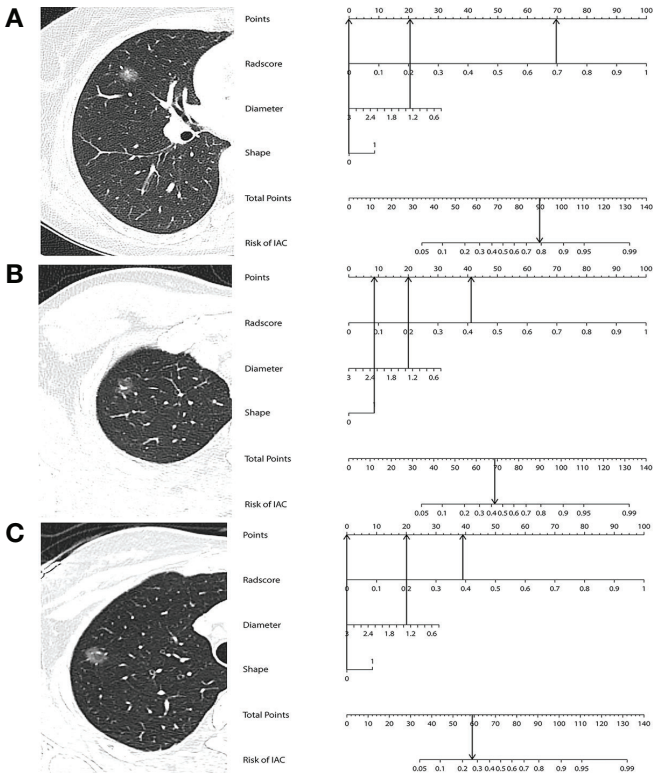


FIGURE 5
(A) A 45-year-old woman with IAC in the right middle lobe shows a regular shape and with a diameter of 1.26 cm. The nomogram shows that this case had a total of 90 points after summing all points (69 + 21 + 0), which corresponds to a 79.6% probability of IAC. **(B)** A 26-year-old woman with MIA in the right upper lobe shows an irregular shape and with a diameter of 1.32 cm. The nomogram shows that this case had a total of 69 points after summing all points (41 + 20 + 8), which corresponds to a 44.3% probability of IAC. **(C)** A 36-year-old woman with AIS in the right upper lobe shows a regular shape and with a diameter of 1.34 cm. The nomogram shows that this case had a total of 59 points after summing all points (39 + 20 + 0), which corresponds to a 25.9% probability of IAC.

analysis in PSNs with solid components <6 mm before; our research results also need further verification.

In addition to the analysis of radiographic features, there have been many studies using quantitative parameters to predict the invasiveness of PSNs. Lower kurtosis and bigger mass were confirmed as significant differentiators of IAC from AIS/MIA by Chae et al. (41). Ko et al. (27) demonstrated that the total volume and percentage solid volume measurements of PSNs helped differentiate between IAC and AIS/MIA with an accuracy of 73.2%. Therefore, although previous studies are numerous, their results vary due to differences in sample size, quantitative parameters, and analysis methods. For this reason, it is still unclear which parameters contribute the most to the prediction of invasiveness of PSNs. The classification of PSNs can still be challenging for radiologists. However, there were few pieces of research that focused on PSNs with solid component <6 mm. Qi et al. (42) illustrated that the mass of PSNs with solid component <6 mm, with the best cutoff value of 283.2 mg, was the only independent predictor for IAC, and the Az value was 0.859; sensitivity was 68.7% and specificity was 92.9%, which were lower than those achieved by the radiomics model in the external validation set of our study (Az value 0.924, sensitivity 83.5%, specificity 84.2%).

During the growth and evolution of PSNs, as the invasive components containing structural deformities of the stromal elastic fiber framework increase within a homogeneous lepidic or acinar background, the diameter and density of the lesion will increase, the shape will become irregular, and the pixel values will become inhomogeneous. In the seven selected most identifiable radiomics features in our study, gldm_DependenceEntropy and glcm_JointEntropy reflect the uniformity of texture and gray scale, higher Entropy value represents the inhomogeneous of pixels in the tumor and greater probability of IAC. Son et al. (43) reviewed 191 resected ground-glass opacity nodules with little or no solid component and identified entropy as an independent predictor for IAC, which is consistent with our results to some extent. Firstorder_RootMeanSquared and firstorder_90Percentile reflect the brightness and shape_Maximum2DdiameterSlice and shape_Maximum2DdiameterColumn reflect the maximum 2D diameter, which were all positively correlated with the

TABLE 3 Effectiveness of the three models in discriminating AIS-MIA from IAC in the derivation, internal validation, and external validation sets.

	Az (95% CI)	Cutoff value	SEN (%)	SPE (%)	PPV (%)	NPV (%)	ACC (%)	F1-score	MCC	Model-fitting information	
										AIC (%)	R^2 value
Derivation set											
Clinical-radiographic model	0.779 (0.747-0.809)	>0.471	64.8	76.8	71.9	70.3	71.0	0.682	0.443	47.5	0.225
Radiomics model	0.865 (0.839-0.889)	>0.456	75.3	81.5	78.9	78.2	78.5	0.770	0.569	63.2	0.398
Combined model	0.872 (0.846-0.895)	>0.494	78.7	82.0	80.1	80.7	80.4	0.794	0.607	64.5	0.415
Internal validation set											
Clinical-radiographic model	0.810 (0.762-0.852)	>0.416	66.9	76.8	72.7	71.6	72.1	0.697	0.439	53.0	0.279
Radiomics model	0.884 (0.843-0.917)	>0.531	78.1	83.5	81.4	80.6	81.0	0.797	0.618	66.8	0.444
Combined model	0.887 (0.847-0.920)	>0.560	80.1	84.8	82.9	82.2	82.5	0.815	0.650	68.4	0.466
External validation set											
Clinical-radiographic model	0.855 (0.799-0.901)	>0.419	71.3	80.8	77.0	75.7	76.3	0.740	0.523	60.9	0.367
Radiomics model	0.924 (0.878-0.957)	>0.444	83.5	84.2	83.5	84.2	83.8	0.835	0.613	73.3	0.535
Combined model	0.917 (0.869-0.951)	>0.472	84.5	86.1	85.4	85.3	85.4	0.850	0.707	73.8	0.542

Az, area under the receiver operating curve; CI, confidence interval; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; ACC, accuracy; MCC, Matthews correlation coefficient; AIC, Akaike information criterion; AIS, adenocarcinoma in situ; MIA, minimally invasive adenocarcinoma; IAC, invasive adenocarcinoma.

significant density and diameter difference between IAC and AIS/MIA. As the invasiveness of PSNs increase, their shape will become irregular and with higher surface/volume ratio, so it is reasonable that shape_SurfaceVolumeRatio is retained in our radiomics model.

In our study, radiomics outperformed the clinical-radiographic model in predicting the invasiveness of PSNs with solid components <6 mm. In the research of Sun et al. (25), radiomics also showed superior predictive performance compared with the clinical-radiographic model in predicting the invasiveness of NSNs. However, compared with radiomics, our combined model did not demonstrate any significant improvement in predictive ability, which was different from the study of Sun et al. This difference can be explained by the fact that we can only extract finite subjective and semiquantitative information from CT images by the naked eyes, while radiomics can analyze conventional descriptive signs and transform image data into spatial data that can be mined in depth and quantitatively analyzed. Some radiographic features we analyzed are probably included in the features extracted by radiomics. Which surprised us was that both the radiomics and the combined model achieved better results in the external validation set than in the internal validation set in our study. We studied the distribution of contributing features in these datasets, and the distributions of the two most weighted features (gldm_DependenceEntropy, shape_Maximum2DDiameter Slice) were visualized with violin-box plot in Supplementary Figure S3. It can be seen that the differences between the distributions of these two features in the positive and negative samples are larger in the external validation set. Therefore, it is understandable that models using these features achieved better results over those of the external validation set.

Deep learning has also been used to predict the invasiveness of PSNs. Kim et al. (44) developed a deep learning model using 2.5D CT images and confirmed that it performed better than the size-based logistic model in distinguishing between IAC and AIS/MIA. However, a deep learning model based on 3D convolutional neural networks in the study by Park et al. (45) showed comparable classification performance with the radiologists' measurements of solid component size in PSNs. The deep learning method learns features from data and avoids the burden of identifying the effective features manually in images without lesion segmentation. However, deep learning has its own limitations: a large number of cases are needed to train the established model and the interpretability is very limited, users cannot get an effective explanation of the classification results (46, 47). Nevertheless, after expanding the number of cases, we will try to apply the deep learning method to our studies.

Although yearly surveillance CT is recommended for PSNs with solid components <6 mm by Fleischner Society guidelines, the treatment of an individual should also be guided by the probability that the nodule is an IAC and patient preferences.

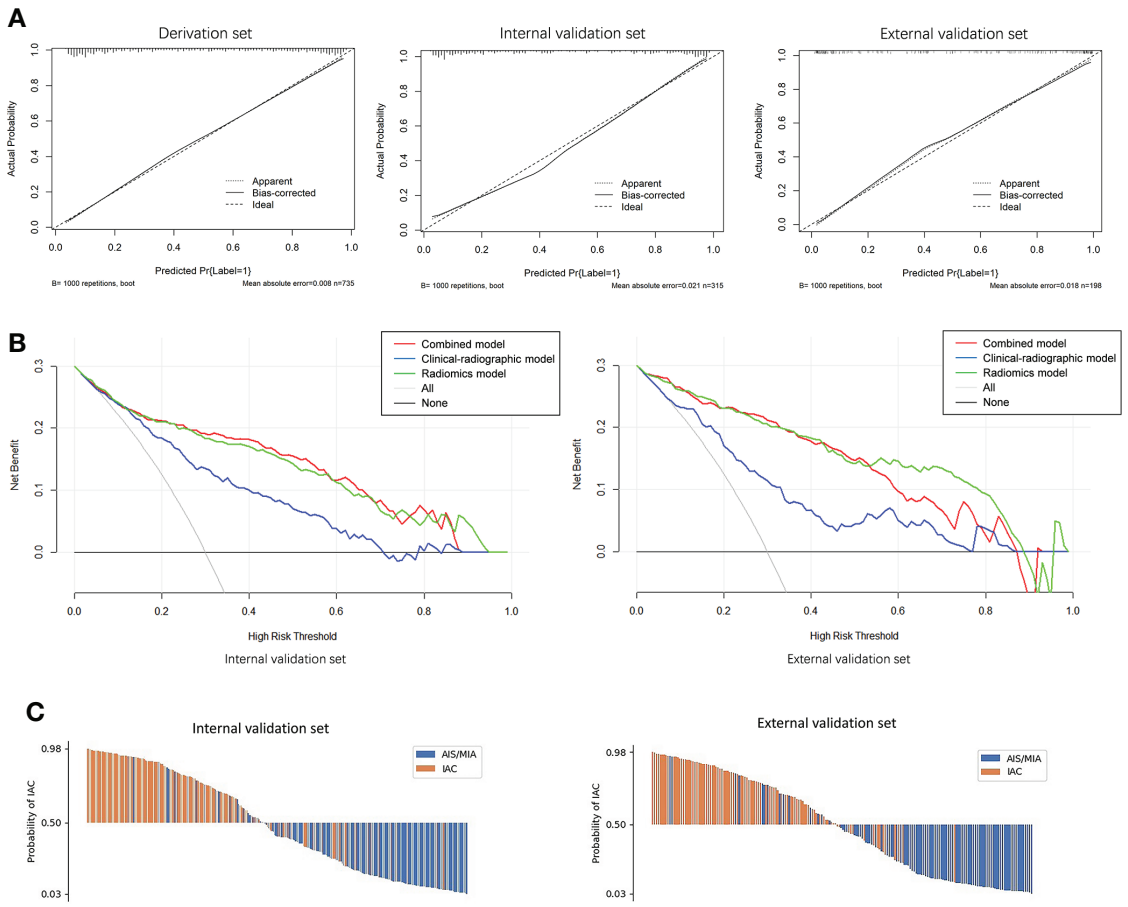


FIGURE 6
(A) Calibration curves of the combined model indicated good agreement between the predicted probability and actual occurrence in the derivation, internal validation, and external validation sets. **(B)** Decision curve analysis on both the internal and external validation sets for the models with and without Rad-score. It can be seen that if the threshold probability of a patient is in the range of 0.10~0.85, using a model with the Rad-score to predict the invasive lesion would be more beneficial than using one without the Rad-score. **(C)** Waterfall plot of the combined model showing the predicted probabilities of internal and external validation sets. It can be seen clearly that the combined model can differentiate IAC from AIS/MIA well.

Patients without risk factors (COPD history, family history of lung cancer, etc.) and lower predicted IAC possibility can be followed up routinely and avoid unnecessary surgery. But closer follow-up should be recommended for patients with risk factors and higher predicted IAC possibility. In our study, 16 IAC cases with solid components <6 mm were classified as pT2 stage because of the presence of pleural invasion. So, the accurate IAC prediction in PSNs with solid components <6 mm has important clinical value in the individualized clinical management (routine or closer follow-up, even resection for patients desire surgery) and selection of surgical methods. Our research may provide an individualized clinical management supplement to Fleischner Society guidelines to improve the diagnostic accuracy of radiologists and avoid unnecessary surgery for some patients or provide a clinical basis for some necessary surgeries.

Our study has several limitations. First, all VOIs of the lesion were automatically segmented using software, but as to the lesions with an unclear tumor–lung interface, the segment will be inaccurate and affects the result of data analysis. Manual segmentation is necessary at this time, but the process is tedious and vulnerable to readers' subjectivity. Second, the clinical-radiographic model in our study was established by the authors #1, #3, and #9 (with 10, 7, and 16 years' experience in chest CT imaging, respectively); it cannot represent the best diagnostic performance of all radiologists. More radiologists should be enrolled and the "reader performance study" can be performed in our future study. Lastly, numerous radiomics investigations have been published, but the robustness and generalizability of radiomics models still remain questionable and need to be verified by clinical practice. Although we have confirmed that the

radiomics features we extracted achieved satisfactory inter- and intra-observer reproducibility and an external validation set from another hospital was also introduced in our study, multicenter research is still necessary in the future study.

Conclusions

Radiomics has been proven to achieve outstanding classification performance in classifying PSNs with solid components <6 mm. It has the potential as an independent diagnostic tool to improve the classification ability of radiologists or thoracic surgeons and save their time and effort without compromising diagnostic accuracy.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

This study was reviewed and approved by the ethical committee of the first affiliated hospital of Nanjing medical university approved this retrospective study and waived the informed consent for the patients (approval number: 2021-SR-053). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

TZ, CZ and MY: conceptualization and methodology. TZ, CZ and GY: data curation and formal analysis. MY and QZ: funding acquisition, project administration and supervision. TZ, CZ, YZ, YS and MY: investigation. TZ, CZ, HW and GY: software. TZ, CZ, HL, MY and GY: validation. TZ and CZ: writing-original draft. MY and GY: writing-review and editing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by grants from the National Nature Science Foundation of China (81801693).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.900049/full#supplementary-material

References

1. Lin KF, Wu HF, Huang WC, Tang PL, Wu MT, Wu FZ. Propensity score analysis of lung cancer risk in a population with high prevalence of non-smoking related lung cancer. *BMC Pulm Med* (2017) 17:120. doi: 10.1186/s12890-017-0465-8

2. Wu FZ, Huang YL, Wu CC, Tang EK, Chen CS, Mar GY, et al. Assessment of selection criteria for low-dose lung screening CT among Asian ethnic groups in Taiwan: from mass screening to specific risk-based screening for non-smoker lung cancer. *Clin Lung Canc* (2016) 17:e45–56. doi: 10.1016/j.cllc.2016.03.004

3. Wang YH, Chen CF, Lin YK, Chiang C, Tzao C, Yen Y. Predicting malignancy: subsolid nodules detected on LDCT in a surgical cohort of East Asian patients. *J Thorac Dis* (2020) 12:4315–26. doi: 10.21037/jtd-20-659

4. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: from the fleischner society. *Radiology* (2017) 284:228–43. doi: 10.1148/radiol.2017161659

5. Kobayashi Y, Sakao Y, Deshpande GA, Fukui T, Mizuno T, Kuroda H, et al. The association between baseline clinical-radiological characteristics and growth of pulmonary nodules with ground-glass opacity. *Lung Canc* (2014) 83:61–6. doi: 10.1016/j.lungcan.2013.10.017

6. Kim H, Park CM, Koh JM, Lee SM, Goo JM. Pulmonary subsolid nodules: what radiologists need to know about the imaging features and management strategy. *Diagn Interv Radiol* (2014) 20:47–57. doi: 10.5152/dir.2013.13223

7. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 world health organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. *J Thorac Oncol* (2015) 10:1243–60. doi: 10.1097/JTO.0000000000000630

8. Tsutani Y, Miyata Y, Nakayama H, Okumura S, Adachi S, Yoshimura M, et al. Appropriate sublobar resection choice for ground glass opacity-dominant clinical stage IA lung adenocarcinoma: wedge resection or segmentectomy. *Chest* (2014) 145:66–71. doi: 10.1378/chest.13-1094

9. Migliore M, Palmucci S, Nardini M, Basile A. Imaging patterns of early stage lung cancer for the thoracic surgeon. *J Thorac Dis* (2020) 12:3349–56. doi: 10.21037/jtd.2020.02.61

10. Lee JH, Park CM, Kim H, Hwang EJ, Park J, Goo JM. Persistent part-solid nodules with solid part of 5 mm or smaller: can the 'follow-up and surgical resection after interval growth' policy have a negative effect on patient prognosis? *Eur Radiol* (2017) 27:195–202. doi: 10.1007/s00330-016-4364-9

11. Zhang Y, Fu F, Chen H. Management of ground-glass opacities in the lung cancer spectrum. *Ann Thorac Surg* (2020) 110:1796–804. doi: 10.1016/j.athoracsur.2020.04.094

12. Saji H, Matsubayashi J, Akata S, Shimada Y, Kato Y, Kudo Y, et al. Correlation between whole tumor size and solid component size on high-resolution computed tomography in the prediction of the degree of pathologic malignancy and the prognostic outcome in primary lung adenocarcinoma. *Acta Radiol* (2015) 56:1187–95. doi: 10.1177/0284185114554823

13. Liu Y, Sun H, Zhou F, Su C, Gao G, Ren S, et al. Imaging features of TSCT predict the classification of pulmonary preinvasive lesion, minimally and invasive adenocarcinoma presented as ground glass nodules. *Lung Canc* (2017) 108:192–7. doi: 10.1016/j.lungcan.2017.03.011

14. Ahn SY, Park CM, Jeon YK, Kim H, Lee JH, Hwang EJ, et al. Predictive CT features of visceral pleural invasion by t1-sized peripheral pulmonary adenocarcinomas manifesting as subsolid nodules. *AJR Am J Roentgenol* (2017) 209:561–6. doi: 10.2214/AJR.16.17280

15. Travis WD, Asamura H, Bankier AA, Beasley MB, Detterbeck F, Flieder DB, et al. The IASLC lung cancer staging project: proposals for coding T categories for subsolid nodules and assessment of tumor size in part-solid tumors in the forthcoming eighth edition of the TNM classification of lung cancer. *J Thorac Oncol* (2016) 11:1204–23. doi: 10.1016/j.jtho.2016.03.025

16. Aokage K, Miyoshi T, Ishii G, Kusumoto M, Nomura S, Katsumata S, et al. Clinical and pathological staging validation in the eighth edition of the TNM classification for lung cancer: correlation between solid size on thin-section computed tomography and invasive size in pathological findings in the new T classification. *J Thorac Oncol* (2017) 12:1403–12. doi: 10.1016/j.jtho.2017.06.003

17. Naidich DP, Bankier AA, MacMahon H, Schaefer-Prokop CM, Pistolesi M, Goo JM, et al. Recommendations for the management of subsolid pulmonary nodules detected at CT: a statement from the fleischner society. *Radiology* (2013) 266:304–17. doi: 10.1148/radiol.12120628

18. Lee HY, Choi YL, Lee KS, Han J, Zo JI, Shim YM, et al. Ure ground-glass opacity neoplastic lung nodules: histopathology, imaging, management. *AJR Am J Roentgenol* (2014) 202:W224–33. doi: 10.2214/AJR.13.11819

19. Ahn H, Lee KH, Kim J, Kim J, Kim J, Lee KW. Diameter of the solid component in subsolid nodules on low-dose unenhanced chest computed tomography: measurement accuracy for the prediction of invasive component in lung adenocarcinoma. *Korean J Radiol* (2018) 19:508–15. doi: 10.3348/kjr.2018.19.3.508

20. Scholten ET, Jacobs C, van Ginneken B, van Riel S, Vliegenthart R, Oudkerk M, et al. Detection and quantification of the solid component in pulmonary subsolid nodules by semiautomatic segmentation. *Eur Radiol* (2015) 25:488–96. doi: 10.1007/s00330-014-3427-z

21. Oikonomou A, Salazar P, Zhang Y, Hwang DM, Petersen A, Dmytriw AA, et al. Histogram-based models on non-thin section chest CT predict invasiveness of primary lung adenocarcinoma subsolid nodules. *Sci Rep* (2019) 9:6009. doi: 10.1038/s41598-019-42340-5

22. Lee JH, Park CM. Differentiation of persistent pulmonary subsolid nodules with a solid component smaller than 6 mm: to be invasive adenocarcinoma or not to be? *J Thorac Dis* (2020) 12:1754–7. doi: 10.21037/jtd-20-1645

23. Thawani R, McLane M, Beig N, Ghose S, Prasanna P, Velcheti V, et al. Radiomics and radiogenomics in lung cancer: a review for the clinician. *Lung Canc* (2018) 115:34–41. doi: 10.1016/j.lungcan.2017.10.015

24. Khorrami M, Bera K, Leo P, Vaidya P, Patil P, Thawani R, et al. Stable and discriminating radiomic predictor of recurrence in early stage non-small cell lung cancer: multi-site study. *Lung Canc* (2020) 142:90–7. doi: 10.1016/j.lungcan.2020.02.018

25. Sun Y, Li C, Jin L, Gao P, Zhao W, Ma W, et al. Radiomics for lung adenocarcinoma manifesting as pure ground-glass nodules: invasive prediction. *Eur Radiol* (2020) 30:3650–9. doi: 10.1007/s00330-020-06776-y

26. Yuan M, Zhang YD, Pu XH, Zhong Y, Li H, Wu JF, et al. Comparison of a radiomic biomarker with volumetric analysis for decoding tumour phenotypes of lung adenocarcinoma with different disease-specific survival. *Eur Radiol* (2017) 27:4857–65. doi: 10.1007/s00330-017-4855-3

27. Ko JP, Suh J, Ibidapo O, Escalon JG, Li J, Pass H, et al. Lung adenocarcinoma: correlation of quantitative ct findings with pathologic findings. *Radiology* (2016) 280:931–9. doi: 10.1148/radiol.2016142975

28. Kim H, Goo JM, Kim YT, Park CM. Consolidation-to-tumor ratio and tumor disappearance ratio are not independent prognostic factors for the patients with resected lung adenocarcinomas. *Lung Canc* (2019) 137:123–8. doi: 10.1016/j.lungcan.2019.09.014

29. Huang TW, Lin KH, Huang HK, Chen YI, Ko KH, Chang CK, et al. The role of the ground-glass opacity ratio in resected lung adenocarcinoma. *Eur J Cardiothorac Surg* (2018) 54:229–34. doi: 10.1093/ejcts/ezy040

30. Van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339

31. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* (2020) 295:328–38. doi: 10.1148/radiol.2020191145

32. Yue X, Liu S, Liu S, Yang G, Li Z, Wang B, et al. HRCT morphological characteristics distinguishing minimally invasive pulmonary adenocarcinoma from invasive pulmonary adenocarcinoma appearing as subsolid nodules with a diameter of ≤3 cm. *Clin Radiol* (2018) 73:411.e7–15. doi: 10.1016/j.crad.2017.11.014

33. McWilliams A, Tammemagi MC, Mayo JR, Roberts H, Liu G, Soghrati K, et al. Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* (2013) 369:910–9. doi: 10.1056/NEJMoa1214726

34. Donin NM, Kwan L, Lenis AT, Drakaki A, Chamie K. Second primary lung cancer in united states cancer survivors, 1992-2008. *Cancer Causes Control* (2019) 30:465–75. doi: 10.1007/s10552-019-01161-7

35. Li X, Zhang W, Yu Y, Zhang G, Zhou L, Wu Z, et al. CT features and quantitative analysis of subsolid nodule lung adenocarcinoma for pathological classification prediction. *BMC Canc* (2020) 20:60. doi: 10.1186/s12885-020-6556-6

36. Zhou QJ, Zheng ZC, Zhu YQ, Lu PJ, Huang J, Ye JD, et al. Tumor invasiveness defined by IASLC/ATS/ERS classification of ground-glass nodules can be predicted by quantitative CT parameters. *J Thorac Dis* (2017) 9:1190–200. doi: 10.21037/jtd.2017.03.170

37. Park CM, Goo JM, Lee HJ, Lee CH, Chun EJ, Im JG. Nodular ground-glass opacity at thin-section CT: histologic correlation and evaluation of change at follow-up. *Radiographics* (2007) 27:391–408. doi: 10.1148/rg.272065061

38. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. *Cell* (2011) 144:646–74. doi: 10.1016/j.cell.2011.02.013

39. Zhao W, Xu Y, Yang Z, Sun Y, Li C, Jin L, et al. Development and validation of a radiomics nomogram for identifying invasiveness of pulmonary adenocarcinomas appearing as subcentimeter ground-glass opacity nodules. *Eur J Radiol* (2019) 112:161–8. doi: 10.1016/j.ejrad.2019.01.021

40. Hattori A, Matsunaga T, Takamochi K, Oh S, Suzuki K. Importance of ground glass opacity component in clinical stage IA radiologic invasive lung cancer. *Ann Thorac Surg* (2017) 104:313–20. doi: 10.1016/j.athoracsur.2017.01.076

41. Chae HD, Park CM, Park SJ, Lee SM, Kim KG, Goo JM. Computerized texture analysis of persistent part-solid ground-glass nodules: differentiation of preinvasive lesions from invasive pulmonary adenocarcinomas. *Radiology* (2014) 273:285–93. doi: 10.1148/radiol.14132187

42. Qi L, Lu W, Yang L, Tang W, Zhao S, Huang Y, et al. Qualitative and quantitative imaging features of pulmonary subsolid nodules: differentiating invasive adenocarcinoma from minimally invasive adenocarcinoma and preinvasive lesions. *J Thorac Dis* (2019) 11:4835–46. doi: 10.21037/jtd.2019.11.35

43. Son JY, Lee HY, Lee KS, Kim JH, Han JH, Jeong JY, et al. Quantitative CT analysis of pulmonary ground-glass opacity nodules for the distinction of invasive adenocarcinoma from pre-invasive or minimally invasive adenocarcinoma. *PloS One* (2014) 9:e104066. doi: 10.1371/journal.pone.0104066

44. Kim H, Lee D, Cho WS, Lee JC, Goo JM, Kim HC, et al. CT-based deep learning model to differentiate invasive pulmonary adenocarcinomas appearing as subsolid nodules among surgical candidates: comparison of the diagnostic performance with a size-based logistic model and radiologists. *Eur Radiol* (2020) 30:3295–305. doi: 10.1007/s00330-019-06628-4

45. Park S, Park G, Lee SM, Kim W, Park H, Jung K, et al. Deep learning-based differentiation of invasive adenocarcinomas from preinvasive or minimally invasive lesions among pulmonary subsolid nodules. *Eur Radiol* (2021) 31:6239–47. doi: 10.1007/s00330-020-07620-z

46. Kang G, Liu K, Hou B, Zhang N. 3D multi-view convolutional neural networks for lung nodule classification. *PloS One* (2017) 12:e0188290. doi: 10.1371/journal.pone.0188290

47. Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? radiologists again at the forefront of innovation in medicine. *Eur Radiol Exp* (2018) 2:35. doi: 10.1186/s41747-018-0061-6

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 17 August 2022
DOI 10.3389/fonc.2022.840886

OPEN ACCESS

EDITED BY
Lizza E.L. Hendriks,
Maastricht University Medical Centre,
Netherlands

REVIEWED BY
Jelena Stojsic,
University of Belgrade, Serbia
Ahmad Antar,
Almoosa Specialist Hospital,
Saudi Arabia

*CORRESPONDENCE
Lei Zhang
13855282388@139.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 21 December 2021
ACCEPTED 28 July 2022
PUBLISHED 17 August 2022

CITATION
Zhu X, Chen W-B, Xing F-B, Zhou S, Tang Z, Li X-J, Zhang L and Huang Y-C (2022) Treatment, pathological characteristics, and prognosis of pulmonary inflammatory myofibroblastic tumor–a retrospective study of 8 cases.
Front. Oncol. 12:840886.
doi: 10.3389/fonc.2022.840886

Treatment, pathological characteristics, and prognosis of pulmonary inflammatory myofibroblastic tumor–a retrospective study of 8 cases

Xiao Zhu[1†], Wen-Bang Chen[1†], Fu-Bao Xing[1], Shao Zhou[1], Zhen Tang[1], Xiao-Jun Li[1], Lei Zhang[1*] and Yu-Chen Huang[2]

[1]Department of Cardiothoracic Surgery, The First Affiliated Hospital of Bengbu Medical College, Bengbu, China, [2]Department of Pathology, The First Affiliated Hospital of Bengbu Medical College, Bengbu, China

Objective: Inflammatory myofibroblastic tumor (IMT) is a rare disease. We reviewed data from eight patients diagnosed with pulmonary IMT (PIMT) at our hospital with the aim of summarizing and analyzing the characteristics of PIMT to improve our understanding of the disease.

Methods: From January 2012 to December 2019, eight patients underwent surgical intervention for PIMT at The First Affiliated Hospital of Bengbu Medical College. Resected tumors were subjected to pathological and immunohistochemical analyses. The follow-up duration for all patients ranged from 2 years and 3 months to 9 years and 9 months (median: 6 years and 9 months).

Results: The male:female ratio was 5:3, and the mean age was 48.50 years (21–74 years). Two patients (25%) with lung disease discovered *via* chest computed tomography during physical examinations had not experienced any symptoms. Six patients (75%) presented at the hospital because of cough, expectoration, blood in sputum, and chest tightness. Lesions from all eight patients were surgically removed, and PIMT was confirmed based on pathological examinations and immunohistochemical results. No patient received additional treatment after discharge. All cases have been followed up to the time of writing, without any tumor recurrence or distant metastasis.

Conclusion: The age of onset of PIMT is usually over 40 years, and its clinical symptoms are easily confused with those of lung cancer. PIMT can only be diagnosed by histopathology and immunohistochemistry. Complete surgical resection is the preferred treatment, as patients undergoing surgery require no additional treatment, such as chemotherapy, and the survival rate is good.

KEYWORDS

lung tumor, lung cancer, pathological characteristics, pulmonary inflammatory myofibroblastic tumor, prognosis

Introduction

Inflammatory myofibroblastic tumor (IMT) is a rare disease. In 2020, the World Health Organization (WHO) identified IMT as a borderline tumor with potentially recurrent and rare metastatic properties (1). The most common site of IMT is the lung, although it can also occur in the abdomen, pelvis, head and neck, upper respiratory tract, limbs, lumbar tube, uterus, and other sites (2–6). To date, the etiology of IMT remains inconclusive, although recent studies have shown that IMT is associated with anaplastic lymphoma kinase (*ALK*) gene rearrangements, as well as overexpression of ALK protein, and the disease also involves fusion of genes such as *ROS1*, *NTRK3*, and *RET* (7–10). Furthermore, a fraction of IMT cases are associated with chromosomal abnormalities (11). The diagnosis of pulmonary IMT (PIMT) is rare and such cases account for only 0.04–0.7% of all lung masses (2, 12–14). Globally, there are very few studies related to PIMT. Therefore, to analyze and improve the understanding of the pathological characteristics, treatment modalities, and prognosis of PIMT, we aimed to review the data collected from a group of patients diagnosed with PIMT at our hospital and to review the literature related to such cases published in recent years.

Material and methods

Subjects

We reviewed the data of eight patients diagnosed with PIMT by histopathology and immunohistochemistry at The First Affiliated Hospital of Bengbu Medical College between January 2012 and December 2019, including information on clinical symptoms, treatment modalities, pathological features, and prognosis. Examinations conducted prior to surgery included laboratory examinations and computed tomography (CT) scans.In addition, to determine whether there is distant metastases of the tumor, we also performed cranial magnetic resonance imaging (MRI), bone emission computed tomography (ECT), and abdominal ultrasound on patients. None of these examination findings were abnormal.

Histology and immunohistochemistry

After surgery, tumor specimens from each case were fixed with 10% neutral formaldehyde solution, extracted, paraffin-embedded, cut into 4-μm continuous sections, stained with hematoxylin and eosin (HE), and subjected to immunohistochemistry (EnVision; Agilent Technologies, Santa Clara, CA). The follow-up time ranged from 2 years and 3 months to 9 years and 9 months.

Ethical approval

This study was conducted with approval from the Ethics Committee of The First Affiliated Hospital of Bengbu Medical College. Written informed consent was obtained from the participants.

Results

The clinical data of the eight patients included in the study are presented in Table 1. Among the eight patients, the male: female ratio was 5:3, and the mean age was 48.50 years (range: 21–74 years). Two patients (25%) with lung disease discovered by chest CT during a physical examination had experienced no symptoms. Six patients (75%) presented to the hospital because of cough, expectoration, blood in sputum, and chest tightness. Chest CT of the patients revealed a tumor diameter ranging from 2–4 cm (Figures 1 and 2). None of the patients had a long-term history of respiratory tract infections.

A lobectomy was performed on all patients, resulting in successful tumor removal. None of the patients received additional treatment after discharge. There has been no tumor recurrence or distant metastasis in any patient to date.

Among these patients, the largest tumor was 4 cm in diameter and the smallest was 2 cm in diameter, with an average of 2.95 cm. The tumors were mainly composed of spindle cells, with surrounding chronic inflammatory cell infiltration (Figure 3). In terms of immunohistochemical results, the following characteristics were observed: ALK labeling was positive in five cases (Figure 4), vimentin (VIM) labeling was positive in all eight cases (Figure 5), seven cases exhibited positive smooth muscle actin (SMA) labeling (Figure 6), one case demonstrated positive calponin labeling (Figure 7), and six cases showed positive Ki67 labeling (Figure 8). All eight patients were diagnosed with PIMT according to pathological features and immunohistochemistry results, and there was no lymph node metastasis.

Discussion

IMT can occur at any age, although it most commonly occurs in young people, and there is no significant sex bias (2, 15–17). However, the mean age of our patients was 48.5 years, which is quite different from that reported in the literature. We believe this may be because the literature reports the mean age of patients with IMT at all body sites, whereas the present study only evaluated the characteristics of patients with PIMT specifically. In addition, in this study, there was no sex bias, which is consistent with previous reports.

10.3389/fonc.2022.840886

TABLE 1 Clinical data and immunohistochemical labeling of the eight patients.

Case	Sex/age (years)	Symptoms	Location/size	Surgery	Prognosis after surgery	Results of positive immunohistochemistry
1	M/38	Cough/ Hemoptysis	RLL/3.0 cm	Lobectomy	9 years 9 months, alive	ALK(++), Vim(+) SMA(+), Calponin(+), Ki67(+, 20~30%)
2	F/42	Cough/ Hemoptysis	RUL/2.0 cm	Lobectomy	9 years, alive	ALK(+), Vim(+), SMA(-), Calponin(-), Ki67(+, 10%)
3	F/46	None	RML/2.5 cm	Lobectomy	8 years, alive	ALK(+), Vim(+), SMA(+), Calponin(-), Ki67(+, 8%)
4	F/50	Cough/ Hemoptysis	RLL/4 cm	Lobectomy	7 years 7 months, alive	ALK(-), Vim(++), SMA(+), calponin(-), Ki67(-), CD68(-)
5	M/21	Cough/ Chest tightness	RUL/2.8 cm	Lobectomy	7 years 3 months, alive	ALK(+), Vim(++), SMA(+), calponin(-), Ki67(+, <5%)
6	M/50	None	LLL/2.3 cm	Lobectomy	6 years 7 months, alive	ALK(-), Vim(+++), SMA(-/+), calponin(-), Ki67(+/-, 10%)
7	M/67	Cough/ Chest tightness	LLL/3.0 cm	Lobectomy	4 years, alive	ALK(+), Vim(+), SMA(+), calponin(-), Ki67(-)
8	M/74	Thoracalgia	RUL/4.0 cm	Lobectomy+ lymph node dissection	2 years 3 months, alive	ALK(-), Vim(+++), SMA(++-++),calponin(-), Ki67(+, 20%)

F, female; M, male; RLL, right lower lobe; RUL, right upper lobe; RML, right middle lobe; LLL, left upper lobe; ALK, anaplastic lymphoma kinase; Vim, vimentin; SMA, smooth muscle actin.

The presence of cough, hemoptysis, and other symptoms reported by patients in this study support the view that the clinical manifestations of PIMT can be easily confused with lung cancer and tuberculosis (7). In X-ray and CT examinations, most cases of PIMT manifest as individual masses in the lung. The density inside the mass is basically uniform, and its boundaries tend to be clear, although lobulation and burr signs are occasionally seen (18), which is consistent with the CT characteristics of the patients in this study. Therefore, CT cannot distinguish PIMT from lung cancer. Although PIMT has also been reported in studies involving positron emission tomography (PET)/CT, spectral CT, magnetic resonance imaging (MRI), and other imaging methods (19, 20), imaging modalities alone cannot diagnose PIMT. Some scholars have reported that IMT has been misdiagnosed as lymphoma by PET/CT (21).

Pathology and immunohistochemistry are the most accurate methods for diagnosing IMT (20). Under the microscope, IMT is characterized by a series of myofibroblast proliferations and different types of inflammatory cell infiltration (14), which is consistent with the findings of this study. Most IMT immunohistochemistry has shown that spindle tumor cells were positively labeled by VIM, SMA, and ALK antibodies; more specifically, VIM labeling is usually strongly positive and diffusely observed in the cytoplasm of spindle cells, SMA labeling is mostly focal or diffusely positive, and ALK protein is expressed in 50–60% of cells (8). These immunohistochemical characteristics were also observed in the eight patients in this study. Therefore, we believe that VIM, SMA, and ALK are the three most important markers for the diagnosis of IMT by immunohistochemistry.

Although there are many ways to treat PIMT, surgical resection remains the first choice (14, 22). Although intraoperative frozen sections were collected in all cases, whether the tumors were benign or malignant could not be completely determined. At the same time, considering the clinical characteristics, CT results, and deep locations of the tumors, lobectomy was performed for all patients. In seven cases, the result of rapid frozen-section pathology was indicative of inflammatory myofibroblastoma. Therefore, only lobectomy was performed. In one case, the result of rapid frozen-section pathology was indicative of IMT, but the possibility of a malignant tumor was not ruled out, so a lobectomy with lymph node dissection was performed. Casanova and colleagues (23) believe that the prognosis of patients with PIMT who undergo early surgery is usually ideal, and there is no need for adjuvant treatment such as chemotherapy. Our results are consistent with this assertion, and none of the patients received additional treatment after discharge. To date, there have been no recurrences or distant metastases.

We performed lymph node dissection in one patient and the result was negative. There is no definitive conclusion regarding whether lymph node dissection should be conducted in patients with PIMT. Some researchers have found the existence of cancer stem cells in PIMT tissues (24), and some studies have reported

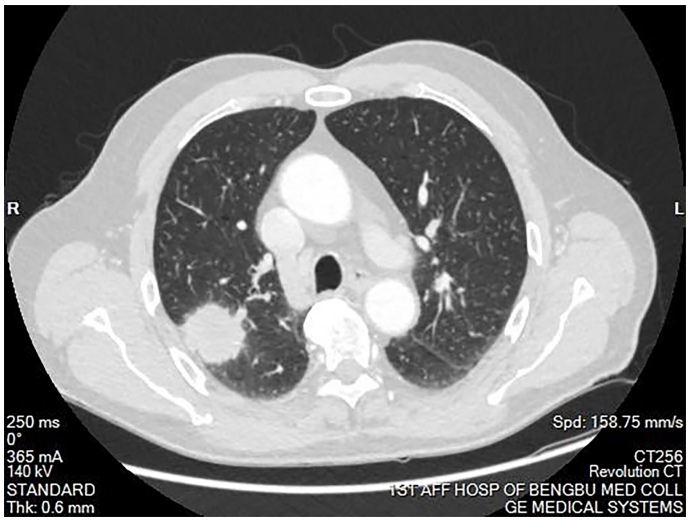


FIGURE 1
Computed tomography of the chest showing a mass in the right upper lobe of the lung. The mass was approximately 4cm in diameter. Burrs were present on the edge of the mass. A portion of the mass was connected to the pleura. The mass had uneven enhancement after the enhancement scan.

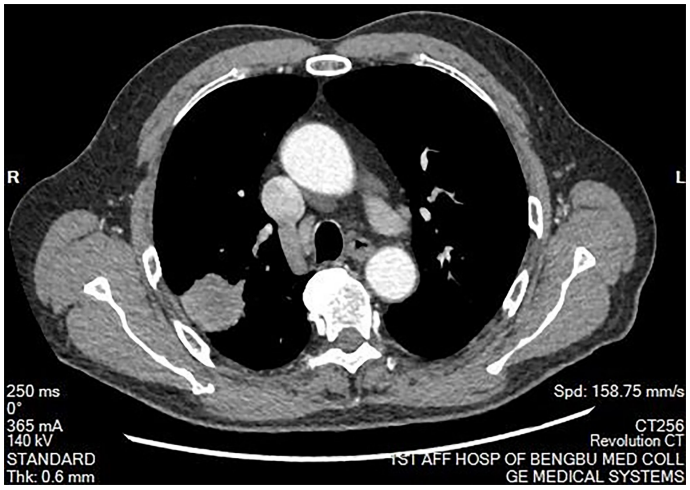


FIGURE 2
Mediastinal window of chest computed tomography. There was no abnormal mass in the mediastinum. The hilum on both sides is normal. There was no pleural effusion on both sides.

that the lymph nodes removed during surgery in patients with PIMT were positive (23). In addition, cervical lymph node metastasis has occurred 3.5 years post-operation (25). Moreover, the WHO points out that IMT is a borderline tumor with the potential for recurrence and rare metastasis. Based on the above aspects, we suggest that lymph node dissection should be performed if the pathological results of intraoperative frozen sectioning suggest PIMT. However, the current sample size is small, and there is still a lack of research in this field. More studies must be conducted in the future to verify the necessity of lymph node dissection in these patients.

The prognosis of patients with PIMT is good (22, 23), with a 5-year survival rate of 91.3% and a 10-year survival rate of 77% (2). Studies have found that all metastatic IMTs are ALK-negative, and ALK positivity may be a good prognostic indicator of IMT (26). However, the latest research by Casanova and colleagues (23) shows that even patients who cannot undergo surgery and those who are ALK-negative have

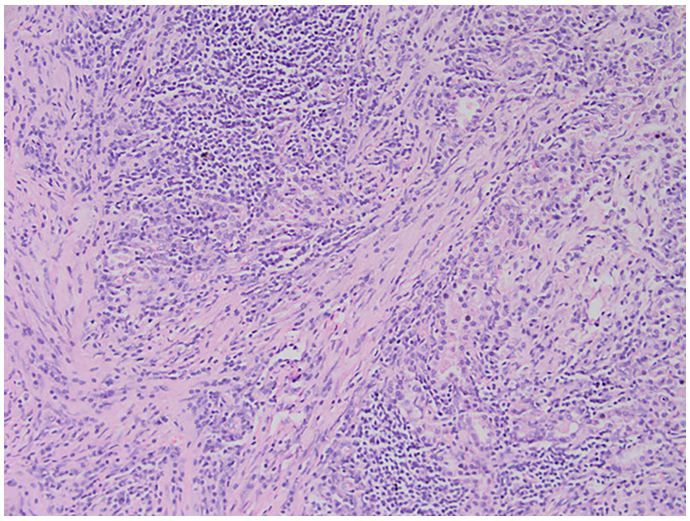


FIGURE 3
After staining of a pneumonia myofibroma, optical microscopy of the tumor reveals that it is composed of many spindle tumor cells, with some instances of plasma cell and lymphocyte infiltration into the stroma (magnification, ×200).

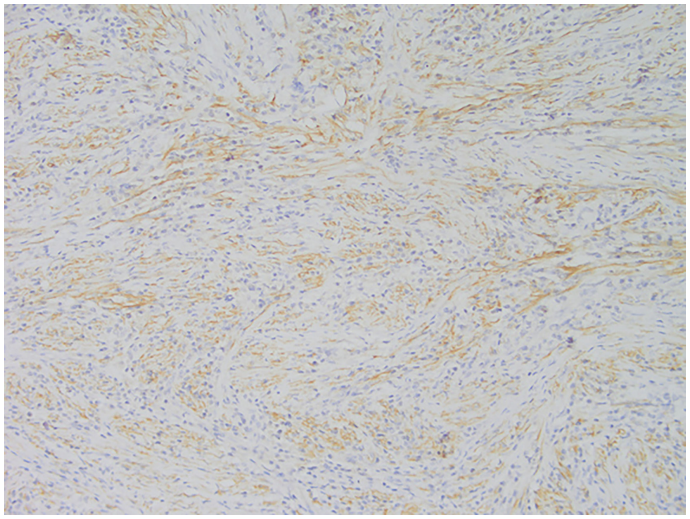


FIGURE 4
Representative image of anaplastic lymphoma kinase (ALK) expression (magnification, ×200).

a good prognosis. Regardless of whether ALK labeling was positive or not, the eight patients whose data we reviewed experienced no recurrence or metastasis after surgery, which was consistent with the results of Casanova et al. (23). Some patients with IMT who have tumor tissue removed still experience relapse and distant metastasis (25–27). In one study, researchers followed up 23 patients after PIMT excision for 2–127 months; in those cases, recurrence only occurred twice after operation, and there was no recurrence after reoperation (28). If only local recurrence occurs and the patient's physical condition is good, reoperation is still recommended (7). Systemic therapy is reserved for patients with unresectable, progressive or metastatic disease and whose body is unable to withstand lobectomy. There is controversy regarding the treatment of PIMT with steroids. On the one hand, as early as 1991, the treatment of PIMT with steroids was reported (29), and on the other hand there are reports that steroids may have an enhancing effect on IMT cell proliferation (30). In addition,

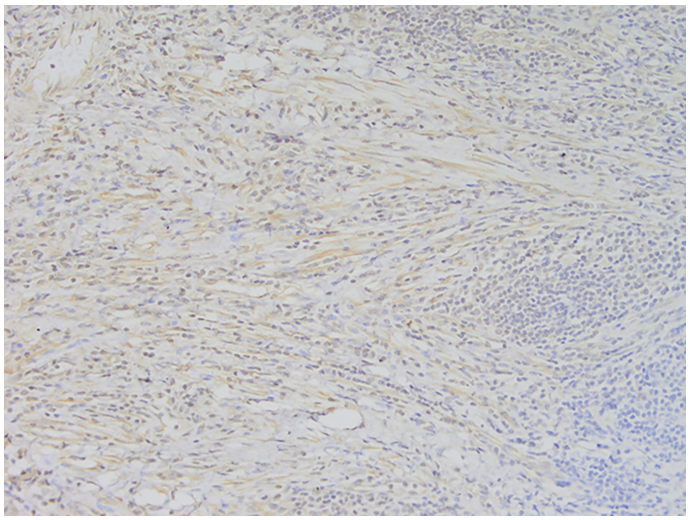


FIGURE 5
Representative image of vimentin expression(magnification, ×200).

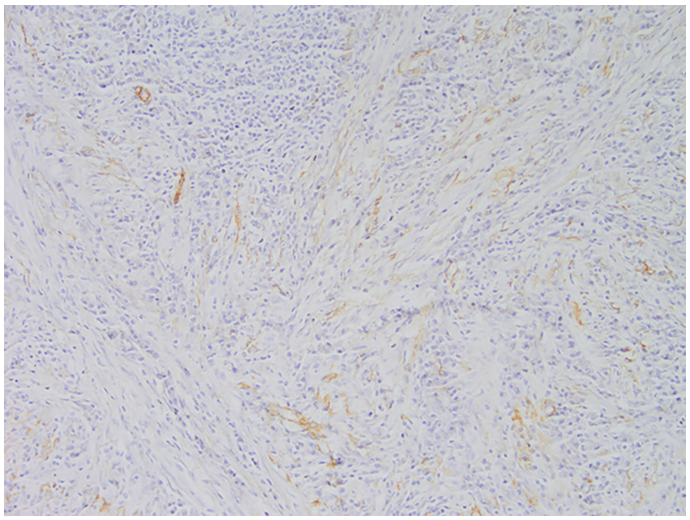


FIGURE 6
Representative image of smooth muscle actin (SMA) expression (magnification, ×200).

there are also some reports on non-steroidal anti-inflammatory drugs(NSAID) treatment for PIMT that are ALK-negative (31). Chemotherapy is a valid option for advanced IMT (23). One study has confirmed that Anthracycline-based and methotrexate plus/minus vinorelbine/vinblastine (MTX-V) regimens are very effective in IMT (17). Studies have reported the cases of using radiotherapy for recurrence of surgical resection (25), but currently there is no sufficient evidence to prove the efficacy of radiotherapy. There are reports that gene-targeted drugs are used to treat IMT (32–40). The US National Comprehensive Cancer Network recommends the use of crizotinib as the standard of care for IMT with ALK-positive (36). Ceritinib, a second-generation ALK inhibitor, has also been shown to be effective in IMT (41).

Conclusion

PIMT is a rare tumor type. Due to the lack of specificity in clinical and imaging manifestations, the diagnosis

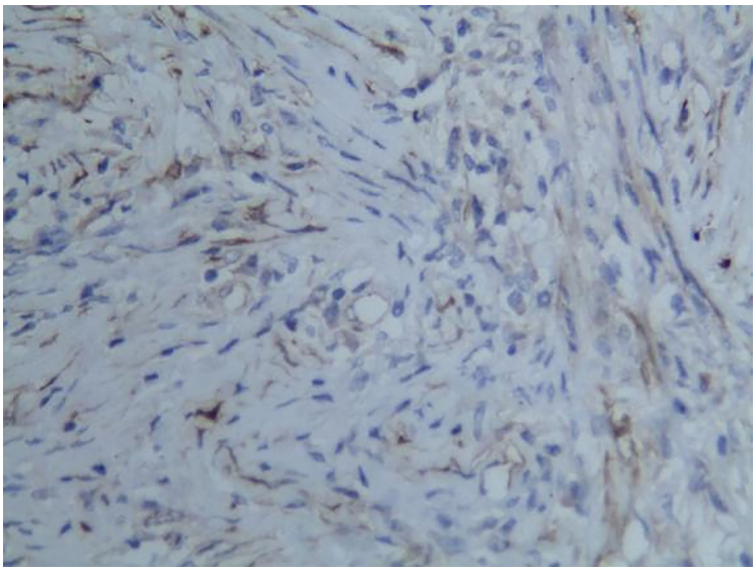


FIGURE 7
Representative image of calponin expression (magnification, ×200).

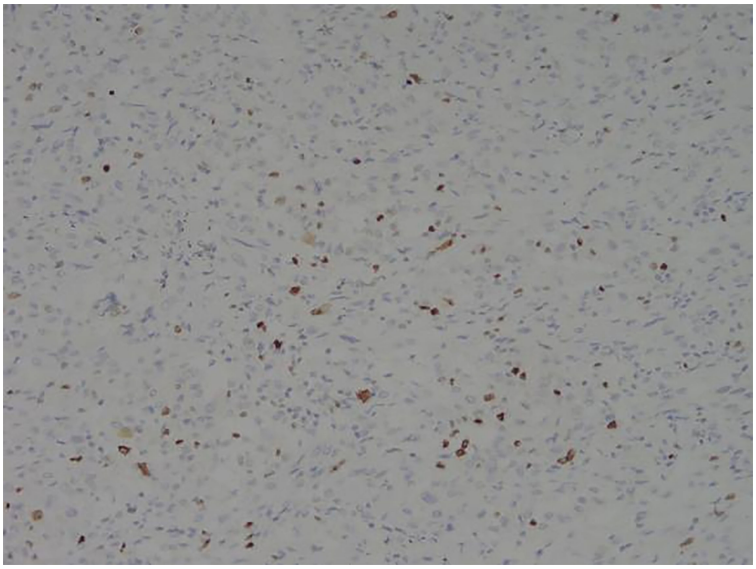


FIGURE 8
Representative image of Ki67 expression (magnification, ×200).

of PIMT can only be made based on pathology and immunohistochemistry results. Complete surgical resection is the preferred treatment in such cases and usually results in satisfactory outcomes. Because local recurrence and metastasis is possible in some cases of IMT, we recommend close, long-term follow-up. Because the disease is rare and the sample size of this study is small, the views described in this paper need to be supported by more research, such as multicenter and large-sample studies to provide more concrete clinical recommendations.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of The First Affiliated Hospital of Bengbu Medical College. The patients/participants provided their written informed consent to participate in this study.

Author contributions

XZ, WBC and LZ performed the surgeries, reviewed the literature, and contributed to manuscript drafting; FBX, SZ, and ZT reviewed the literature and contributed to manuscript drafting; and XJL and YCH were responsible for the revision of the manuscript for important intellectual content. All authors issued final approval for the version to be submitted.

Funding

The study was supported by the Scientific Research Foundation of Education Department of Anhui Province of China (KJ2019A0340).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Choi JH, Ro JY. The 2020 WHO classification of tumors of soft tissue: selected changes and new entities. *Adv Anat Pathol* (2021) 28(1):44–58. doi: 10.1097/PAP.0000000000000284

2. Hussain SF, Salahuddin N, Khan A, Memon SSJ, Fatimi SH, Ahmed R. The insidious onset of dyspnea and right lung collapse in a 35-year-old man. *Chest* (2005) 127(5):1844–7. doi: 10.1378/chest.127.5.1844

3. Dehner LP. The enigmatic inflammatory pseudotumours: the current state of our understanding, or misunderstanding. *J Pathol* (2000) 192(3):277–9. doi: 10.1002/1096-9896(200011)192:3<277::AID-PATH749>3.0.CO;2-E

4. Coffin CM, Watterson J, Priest JR, Dehner LP. Extrapulmonary inflammatory myofibroblastic tumor (inflammatory pseudotumor). a clinicopathologic and immunohistochemical study of 84 cases. *Am J Surg Pathol* (1995) 19(8):859–72. doi: 10.1097/00000478-199508000-00001

5. Wang S, Chen L, Cao Z, Mao X, Zhang L, Wang B. Inflammatory myofibroblastic tumor of the lumbar spinal canal: a case report with literature review. *Med (Baltim)* (2017) 96(26):e6488. doi: 10.1097/MD.0000000000006488

6. Bennett JA, Croce S, Pesci A, Niu N, Van de Vijver K, Burks EJ, et al. Inflammatory myofibroblastic tumor of the uterus: an immunohistochemical study of 23 cases. *Am J Surg Pathol* (2020) 44(11):1441–9. doi: 10.1097/PAS.0000000000001525

7. Preobrazhenskaya EV, Iyevleva AG, Suleymanova AM, Tiurin VI, Mitiushkina NV, Bizin IV, et al. Gene rearrangements in consecutive series of pediatric inflammatory myofibroblastic tumors. *Pediatr Blood Cancer* (2020) 67(5):e28220. doi: 10.1002/pbc.28220

8. Chang JC, Zhang L, Drilon AE, Chi P, Alaggio R, Borsu L, et al. Expanding the molecular characterization of thoracic inflammatory myofibroblastic tumors beyond ALK gene rearrangements. *J Thorac Oncol* (2019) 14(5):825–34. doi: 10.1016/j.jtho.2018.12.003

9. Yamamoto H, Yoshida A, Taguchi K, Kohashi K, Hatanaka Y, Yamashita A, et al. ROS1 and NTRK3 gene rearrangements in inflammatory myofibroblastic tumours. *Histopathology* (2016) 69(1):72–83. doi: 10.1111/his.12910

10. Alassiri AH, Ali RH, Shen Y, Lum A, Strahlendorf C, Deyell R, et al. ETV6-NTRK3 is expressed in a subset of ALK-negative inflammatory myofibroblastic tumors. *Am J Surg Pathol* (2016) 40(8):1051–61. doi: 10.1097/PAS.0000000000000677

11. Khatri A, Agrawal A, Sikachi RR, Mehta D, Sahni S, Meena N. Inflammatory myofibroblastic tumor of the lung. *Adv Respir Med* (2018) 86(1):27–35. doi: 10.5603/ARM.2018.0007

12. Cerfolio RJ, Allen MS, Nascimento AG, Deschamps C, Trastek VF, Miller DL, et al. Inflammatory pseudotumors of the lung. *Ann Thorac Surg* (1999) 67(4):933–6. doi: 10.1016/s0003-4975(99)00155-1

13. Sakurai H, Hasegawa T, Watanabe S, Suzuki K, Asamura H, Tsuchiya R. Inflammatory myofibroblastic tumor of the lung. *Eur J Cardiothorac Surg* (2004) 25(2):155–9. doi: 10.1016/s1010-7940(03)00678-x

14. Takeda S, Onishi Y, Kawamura T, Maeda H. Clinical spectrum of pulmonary inflammatory myofibroblastic tumor. *Interact Cardiovasc Thorac Surg* (2008) 7(4):629–33. doi: 10.1510/icvts.2007.173476

15. Cassivi SD, Wylam ME. Pulmonary inflammatory myofibroblastic tumor associated with histoplasmosis. *Interact Cardiovasc Thorac Surg* (2006) 5(4):514–6. doi: 10.1510/icvts.2006.129809

16. Meis JM, Enzinger FM. Inflammatory fibrosarcoma of the mesentery and retroperitoneum. a tumor closely simulating inflammatory pseudotumor. *Am J Surg Pathol* (1991) 15(12):1146–56. doi: 10.1097/00000478-199112000-00005

17. Baldi GG, Brahmi M, Lo Vullo S, Cojocaru E, Mir O, Casanova M, et al. The activity of chemotherapy in inflammatory myofibroblastic tumors: a multicenter, European retrospective case series analysis. *Oncologist* (2020) 25(11):e1777–84. doi: 10.1634/theoncologist.2020-0352

18. Surabhi VR, Chua S, Patel RP, Takahashi N, Lalwani N, Prasad SR. Inflammatory myofibroblastic tumors: current update. *Radiol Clin North Am* (2016) 54(3):553–63. doi: 10.1016/j.rcl.2015.12.005

19. Yu Y, Wang X, Shi C, Hu S, Zhu H, Hu C. Spectral computed tomography imaging in the differential diagnosis of lung cancer and inflammatory myofibroblastic tumor. *J Comput Assist Tomogr* (2019) 43(2):338–44. doi: 10.1097/RCT.0000000000000840

20. Panagiotopoulos N, Patrini D, Gvinianidze L, Woo WL, Borg E, Lawrence D. Inflammatory myofibroblastic tumour of the lung: a reactive lesion or a true neoplasm? *J Thorac Dis* (2015) 7(5):908–11. doi: 10.3978/j.issn.2072-1439.2015.04.60

21. Ma C, Lu J, Chen G, Wang W, Su F, Su X. Inflammatory myofibroblastic tumor mimicking lymphoma on 18F-FDG PET/CT. report of a case and review of the literature. *Hell J Nucl Med* (2018) 21(1):77–80. doi: 10.1967/s002449910710

22. Soyer T, Talim B, Karnak İ, Ekinci S, Andiran F, Çiftçi AÖ, et al. Surgical treatment of childhood inflammatory myofibroblastic tumors. *Eur J Pediatr Surg* (2017) 27(4):319–23. doi: 10.1055/s-0036-1593380

23. Casanova M, Brennan B, Alaggio R, Kelsey A, Orbach D, van Noesel MM, et al. Inflammatory myofibroblastic tumor: the experience of the European pediatric soft tissue sarcoma study group (EpSSG). *Eur J Cancer* (2020) 127:123–9. doi: 10.1016/j.ejca.2019.12.021

24. Masciale V, Grisendi G, Banchelli F, D'Amico R, Maiorana A, Sighinolfi P, et al. Cancer stem-like cells in a case of an inflammatory myofibroblastic tumor of the lung. *Front Oncol* (2020) 10:673. doi: 10.3389/fonc.2020.00673

25. Hou TC, Wu PS, Huang WY, Yang YT, Tan KT, Liu SH, et al. Over expression of CDK4 and MDM2 in a patient with recurrent ALK-negative mediastinal inflammatory myofibroblastic tumor: a case report. *Med (Baltim)* (2020) 99(12):e19577. doi: 10.1097/MD.0000000000019577

26. Coffin CM, Hornick JL, Fletcher CDM. Inflammatory myofibroblastic tumor: comparison of clinicopathologic, histologic, and immunohistochemical features including ALK expression in atypical and aggressive cases. *Am J Surg Pathol* (2007) 31(4):509–20. doi: 10.1097/01.pas.0000213393.57322.c7

27. Chuah YY, Tashi T, Shy CG, Shyu JS, Dong MJ, Hsueh EJ. Intracranial inflammatory myofibroblastic tumor with sarcomatous local recurrence. *World Neurosurg* (2016) 93:484.e1–4. doi: 10.1016/j.wneu.2016.07.060

28. Zhang N, Zeng Q, Chen CH, Yu J, Yan D, Xu C, et al. Clinical characteristics and prognosis of pulmonary inflammatory myofibroblastic tumor: an over 10-year retrospective analysis. *Pediatr Investig* (2020) 4(3):192–7. doi: 10.1002/ped4.12218

29. Doski JJ, Priebe CJ, Driessnack M, Smith T, Kane P, Romero J. Corticosteroids in the management of unresected plasma cell granuloma (inflammatory pseudotumor) of the lung. *J Pediatr Surg* (1991) 26(9):1064–6. doi: 10.1016/0022-3468(91)90674-i

30. Panigada S, Sacco O, Girosi D, Magnano GM, Tuo P, Tarantino V, et al. Corticosteroids may favor proliferation of thoracic inflammatory myofibroblastic tumors. *Pediatr Pulmonol* (2014) 49(3):E109–111. doi: 10.1002/ppul.22977

31. Ghani S, Desai A, Pokharel S, Demmy T, Dy GK. Pneumonectomy-sparing NSAID therapy for pulmonary inflammatory myofibroblastic tumor. *J Thorac Oncol* (2015) 10(9):e89–90. doi: 10.1097/JTO.0000000000000574

32. Watanabe H, Yamasaki N, Miyazaki T, Matsumoto K, Tsuchiya T, Abe K, et al. Successful treatment based on molecular biological assessment of invasive anaplastic lymphoma kinase-positive inflammatory myofibroblastic tumor of the lung. *Surg Case Rep* (2019) 5(1):118. doi: 10.1186/s40792-019-0674-x

33. Schöffski P, Sufliarsky J, Gelderblom H, Blay JY, Strauss SJ, Stacchiotti S, et al. Crizotinib in patients with advanced, inoperable inflammatory myofibroblastic tumours with and without anaplastic lymphoma kinase gene alterations (European organisation for research and treatment of cancer 90101 CREATE): a multicentre, single-drug, prospective, non-randomised phase 2 trial. *Lancet Respir Med Eur Organ Res Treat Cancer* (2018) 6(6):431–41. doi: 10.1016/S2213-2600(18)30116-4

34. Saiki M, Ohyanagi F, Ariyasu R, Koyama J, Sonoda T, Nishikawa S, et al. Dramatic response to alectinib in inflammatory myofibroblastic tumor with anaplastic lymphoma kinase fusion gene. *Jpn J Clin Oncol* (2017) 47(12):1189–92. doi: 10.1093/jjco/hyx133

35. Péron J, Marreaud S, Staelens D, Raveloarivahy T, Nzokirantevye A, Flament J, et al. A multinational, multi-tumour basket study in very rare cancer types: the European organization for research and treatment of cancer phase II 90101 'CREATE' trial. *Eur J Cancer* (2019) 109:192–5. doi: 10.1016/j.ejca.2018.12.013

36. Trahair T, Gifford AJ, Fordham A, Mayoh C, Fadia M, Lukeis R, et al. Crizotinib and surgery for long-term disease control in children and adolescents with ALK-positive inflammatory myofibroblastic tumors. *JCO Precis Oncol* (2019) 3:PO.18.00297. doi: 10.1200/PO.18.00297

37. Mai S, Xiong G, Diao D, Wang W, Zhou Y, Cai R. Case report: crizotinib is effective in a patient with ROS1-rearranged pulmonary inflammatory myofibroblastic tumor. *Lung Cancer* (2019) 128:101–4. doi: 10.1016/j.lungcan.2018.12.016

38. Brivio E, Zwaan CM. ALK inhibition in two emblematic cases of pediatric inflammatory myofibroblastic tumor: efficacy and side effects. *Pediatr Blood Cancer* (2019) 66(5):e27645. doi: 10.1002/pbc.27645

39. Mossé YP, Voss SD, Lim MS, Rolland D, Minard CG, Fox E, et al. Targeting ALK with crizotinib in pediatric anaplastic large cell lymphoma and inflammatory myofibroblastic tumor: a children's oncology group study. *J Clin Oncol* (2017) 35(28):3215–21. doi: 10.1200/JCO.2017.73.4830

40. Farris N, Sampson M. Single-agent rituximab for treatment of multifocal and multiple relapsed pulmonary inflammatory myofibroblastic tumor in an adolescent patient. *Pediatr Blood Cancer* (2021) 68(9):e29131. doi: 10.1002/pbc.29131

41. Mahajan P, Casanova M, Ferrari A, Fordham A, Trahair T, Venkatramani R. Inflammatory myofibroblastic tumor: molecular landscape, targeted therapeutics, and remaining challenges. *Curr Probl Cancer* (2021) 45(4):100768. doi: 10.1016/j.currproblcancer.2021.100768

frontiers | Frontiers in Oncology

TYPE Brief Research Report
PUBLISHED 27 September 2022
DOI 10.3389/fonc.2022.902955

OPEN ACCESS

EDITED BY
Mohamed Rahouma,
NewYork-Presbyterian, United States

REVIEWED BY
Janaki Deepak,
University of Maryland, Baltimore, United States
Zhixin Wang,
Affiliated Hospital of Qinghai University, China

*CORRESPONDENCE
Qun Liu
470718338@qq.com
Jinhan Chen
xmjinhan@aliyun.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 23 March 2022
ACCEPTED 31 August 2022
PUBLISHED 27 September 2022

CITATION
Ma A, Wang G, Du Y, Guo W, Guo J, Hu Y, Bai D, Huang H, Zhuang L, Chen J and Liu Q (2022) The clinical relevance of neutrophil-to-lymphocyte ratio and platelet-to-lymphocyte ratio in chronic obstructive pulmonary disease with lung cancer.
Front. Oncol. 12:902955.
doi: 10.3389/fonc.2022.902955

The clinical relevance of neutrophil-to-lymphocyte ratio and platelet-to-lymphocyte ratio in chronic obstructive pulmonary disease with lung cancer

Aiping Ma[1†], Guangdong Wang[1†], Yan Du[1†], Weixi Guo[2†], Jiaxi Guo[1], Yi Hu[3], Dongyu Bai[4], Huiping Huang[5], Lianjin Zhuang[6], Jinhan Chen[7*] and Qun Liu[1*]

[1]Department of Respiratory and Critical Medicine, The First Affiliated Hospital, School of Medicine, Xiamen University, Xiamen, China, [2]Department of Thoracic Surgery, The First Affiliated Hospital, School of Medicine, Xiamen University, Xiamen, China, [3]Department of Clinical Laboratory, The First Affiliated Hospital, School of Medicine, Xiamen University, Xiamen, China, [4]Department of Pathology, The First Affiliated Hospital, School of Medicine, Xiamen University, Xiamen, China, [5]Department of Infection Control, The First Affiliated Hospital, School of Medicine, Xiamen University, Xiamen, China, [6]Division of Quality Management, The First Affiliated Hospital, School of Medicine, Xiamen University, Xiamen, China, [7]Department of Geriatrics, The First Affiliated Hospital, School of Medicine, Xiamen University, Xiamen, China

Background: Chronic obstructive pulmonary disease (COPD) coexisting with lung cancer is associated with severe mortality and a worse prognosis. Inflammation plays an important role in common pathogenic pathways and disease progression. However, a few studies have identified the clinical value of the neutrophil-to-lymphocyte ratio (NLR) and platelet-to-lymphocyte ratio (PLR) in COPD with lung cancer, which are systemic inflammatory response markers in the blood. This study aimed to determine the association of the NLR or PLR with clinical characteristics and whether NLR or PLR can be diagnostic markers for COPD with lung cancer.

Methods: Between 2015 and 2021, we conducted a retrospective analysis of 236 COPD patients with lung cancer and 500 patients without lung cancer (control group). Clinical information, blood routine examination, and spirometry results were collected and analyzed. The receiver operating characteristic (ROC) curve was used to identify the best cutoff point of NLR or PLR. Multivariate logistic regression analysis was performed to evaluate the association of NLR or PLR with the diagnosis and prognosis of COPD with lung cancer.

Results: Compared to patients in the COPD-only group, patients in the lung cancer group had a higher percentage of current smoking and emphysema, and it was found that NLR or PLR was significantly higher in the lung cancer group. Multivariate analysis showed that age, smoking status, FEV1%pred,

emphysema, NLR, and PLR were independent risk factors for lung cancer development in COPD. Furthermore, the high level of NLR or PLR was associated with age over 70 years old, current smoking status, and ineligible surgery treatment. The level of PLR or NLR markedly increased with hypercoagulation status, the severity of airflow limitation, and advanced progression of lung cancer. Additionally, the ROC analysis also revealed that elevated NLR or PLR was an independent predictor of COPD in lung cancer patients, TNM stages IIIB–IV at first diagnosis in lung cancer, and ineligible surgery in lung cancer patients.

Conclusion: Increased NLR or PLR values might be an important and easily measurable inflammation biomarker to predict the diagnosis and severity of lung cancer with COPD.

KEYWORDS

chronic obstructive pulmonary disease, lung cancer, neutrophil-to-lymphocyte ratio, platelet-to-lymphocyte ratio, inflammation biomarkers

Introduction

Chronic obstructive pulmonary disease (COPD) and lung cancer comprise the major causes of lung disease-related deaths worldwide (1). It was estimated that the total number of newly diagnosed cases of lung cancer in China in 2015 was about 787,000 (2). Furthermore, COPD is one of the leading causes of morbidity and mortality worldwide (3). In 2016, COPD was the fifth leading cause of death in China (4). Overall, lung cancer and COPD present a major public health issue and an enormous burden on society in China (5–7). Several studies describe that there is an association between COPD and the development of lung cancer. An increase in the incidence and mortality of lung cancer was observed in patients with COPD (8). Machida and colleagues investigated the incidence of lung cancer in the stable COPD population which was 1.85% per year during a median follow-up period of 4.58 years (9). Moreover, the prevalence of COPD in lung cancer patients was 32.6%, and with the increased severity of COPD, the prognosis of lung cancer gradually worsened (10). Current research has indicated that lung cancer accounts for 12% to 14% of COPD-related deaths (11, 12). Additionally, never-smokers of non-small cell lung cancer (NSCLC) patients with COPD had shorter overall survival (OS) times, compared to non-COPD never-smoker NSCLC patients (13). It has been also found that even in patients with early-stage COPD, the prevalence of postoperative pulmonary complications (PPCs) was higher than in patients with NSCLC with normal spirometry (14). Therefore, patients with lung cancer and coexisting COPD may have a much worse prognosis than those with COPD only.

The common pathogenic pathways contributing to both diseases share underlying etiologies, such as tobacco, gene expressions, susceptibility to DNA damage, environmental factors, and inflammation (1, 15, 16). However, more details of the relationship between COPD and lung cancer remain uncertain. Some studies have demonstrated that chronic inflammation could be one of the underlying mechanisms, which has been associated with carcinogenesis in lung cancer, airway remodeling, and severe comorbidities in COPD (17–20). Although some studies have demonstrated that high levels of several inflammation markers were related to progression-free survival (PFS) and OS in lung cancer patients with and without COPD, there is still a lack of effective markers to evaluate COPD with lung cancer (21).

The neutrophil-to-lymphocyte ratio (NLR) and platelet-to-lymphocyte ratio (PLR) are the available markers of systemic inflammation and routine clinical laboratory tests (22–24). Recent studies have indicated that NLR and PLR were higher in acute exacerbation of COPD and used for predicting hospitalization (3, 25). Furthermore, they were related to short-term mortality in patients hospitalized with acute exacerbation of COPD (AECOPD) (26, 27). NLR and PLR values were also found to be higher in advanced lung cancer (28). It was reported that NLR has been associated with worse survival and recurrence (29). The combination of NLR and PLR has increased the diagnostic value and distinguished lung cancer patients from healthy subjects (30). However, few studies have reported whether NLR and PLR could better evaluate COPD

with lung cancer. In the study, we aimed to identify the baseline clinical characteristics of COPD coexisting with lung cancer and investigate whether NLR or PLR can be diagnostic markers for COPD with lung cancer.

Materials and methods

Study population

We retrospectively reviewed the medical records of 1,264 patients diagnosed with COPD between January 2015 and August 2021 at the inpatient clinic of our respiratory department at The First Affiliated Hospital of Xiamen University, Fujian, People's Republic of China. Patients who have incomplete medical records, with contraindications, or are unable to perform spirometry correctly and those who were diagnosed with asthma and interstitial lung disease were excluded. Finally, a total of 736 patients were eligible for this study. All participants underwent spirometry to measure pulmonary function, and their complete medical records were retrieved for analysis. Spirometry was done according to the American Thoracic Society recommendations with central quality assurance of spirometry tracings (19). Among the 1,264 patients, a total of 236 COPD patients with lung cancer were diagnosed through histological and/or cytological specimens.

Data collection

Basic clinical information on age, sex, smoking status, body mass index (BMI), lung cancer diagnosis, histologic type, routine blood, coagulation, and cancer stage was taken from the medical records for all patients. As a retrospective study, all data were anonymous. We declared that the patients' data were confidential and did not compromise the patients' interests. Data collection was completed by two independent authors. The ethics approval of the study was obtained from the Clinical Research Ethics Committee of the First Affiliated Hospital of Xiamen University (2021064).

Lung cancer group

Histologic lung cancer was categorized as small-cell lung cancer (SCLC) and non-SCLC (NSCLC) according to the 2015 WHO classification of lung tumors (31). NSCLC was then further categorized as adenocarcinoma, squamous cell carcinoma (SCC), or other, which consisted primarily of large-cell carcinoma and undifferentiated carcinoma of the lungs. Cancer was staged in accordance with the Union for International Cancer Control, Tumour, Node, Metastasis 7 (TNM7) classification and staging I–IV. All lung cancer patients with COPD were diagnosed through CT-guided lung tissue biopsy, thoracic surgery, or bronchoscopic biopsy.

Routine blood

Peripheral venous blood was collected from all patients into Vacutainer tubes in the morning on the first day of admission and processed immediately. The differential blood counts (white blood cell, neutrophil, lymphocyte) and platelet count were measured using the hematology analyzer Sysmex XN-9000 (XN, Sysmex, Kobe, Japan).

Assessment of COPD

Spirometry tests were performed by qualified technicians on a computerized spirometer MasterScreen, (Jaeger, Freistaat Bayern, Germany), following the criteria for the standardization of pulmonary function tests recommended by the American Thoracic Society/European Respiratory Society (ATS/ERS) Task Force (32). According to The Global Initiative for Chronic Obstructive Lung Disease (GOLD) (http://goldcopd.org) guidelines, forced vital capacity (FVC) and forced expiratory volume in 1 s (FEV1) were recorded in medical records in liters and percentage predicted values, as well as FEV1/FVC%. The post-bronchodilator fixed criteria FEV1/FVC <0.7 was applied to define COPD. The GOLD criteria were also used to assign a grade of clinical severity to COPD based on FEV1%pred and FEV1/FVC: patients with an FEV1/FVC <0.7 were classified as having COPD in all grades; GOLD 1 was defined as having an FEV1%pred >80; GOLD 2 as 50 ≤ FEV1% pred <80; GOLD 3 as 30% ≤ FEV1%pred <50%; and GOLD 4 as FEV1%pred <30% (33).

Statistical analysis

All patient data were presented in terms of frequencies and mean ± standard deviations as appropriate. Normally distributed continuous variables were evaluated using the unpaired Student's *t*-test. Continuous variables with a skewed distribution were analyzed using the Mann–Whitney *U* test. Categorical variables were assessed using the chi-square test. Correlations were done using Pearson's correlation, and "*r*" is the correlation coefficient. It ranged from −1 to +1. Receiver operating characteristic (ROC) curve analysis was used to evaluate the sensitivity and specificity of NLR and PLR for predicting the diagnosis of COPD with lung cancer. We performed a univariate analysis with age, BMI, smoking status, FEV1%pred, NLR, PLR, and emphysema. The variables with P <0.05 in the univariate analysis were entered into a multivariate logistic regression analysis. P-value <0.05 was

assumed for statistical significance in all the tests. Data were analyzed with SPSS 26.0 (SPSS Inc., Chicago, IL, USA).

Results

General characteristics of the patients

A total of 736 patients were eligible for this study: 236 patients were diagnosed as having COPD with lung cancer, and 500 patients were categorized as COPD (Figure 1). Patients' baseline characteristics included median age, gender, BMI, smoking history, pulmonary function, and inflammatory markers (Table 1). As shown in the table, there were no significant differences between BMI and gender. The median onset age was earlier in the COPD with lung cancer group than in the COPD group (66.44 ± 8.37 vs. 69.87 ± 9.37, $P < 0.001$). The proportion of current smoking was significantly higher in the COPD with lung cancer group [178 (75.4%)] than that in the COPD-only group [254 (50.8%)]. Moreover, the proportion of ever smoker was significantly lower in the COPD with lung cancer group [58 (24.6%)] compared with that in the COPD-only group [246 (49.2%)] ($P < 0.001$). In addition, there were significant differences in the degree of airflow obstruction in lung function. The values for FEV1/FVC and FEV1% predicted were all significantly higher in the COPD with lung cancer group than in the COPD group (0.59 ± 0.10 vs. 0.52 ± 0.11, $P < 0.001$; 66.46 ± 20.54 vs. 47.39 ± 21.01, $P < 0.001$, respectively).

Next, to verify the predictive value of inflammatory markers in COPD with lung cancer, peripheral blood analyses of total white blood cell, neutrophil counts, lymphocyte, NLR, and PLR were measured. The results indicated that COPD with lung cancer patients had significantly higher levels of white blood cell and neutrophil counts ($P < 0.001$). Moreover, a statistical significance was found in NLR or PLR between COPD and COPD coexisting with lung cancer. The levels of NLR and PLR were significantly increased in the COPD with lung cancer group compared with those in the COPD group (3.51 ± 0.72 vs. 2.45 ± 0.81, 172.72 ± 24.96 vs. 142.64 ± 37.56, $P < 0.001$). Finally, the results showed a significant increase in the proportion of emphysema in the COPD with lung cancer group than in the COPD group.

Baseline characteristics of COPD patients with lung cancer

Of the total population, 236 cases were diagnosed with lung cancer. Histologic type analyses showed that adenocarcinoma was the most frequent histological subtype in the COPD with lung cancer group (51.3%), and 30.9% were diagnosed with squamous cell carcinoma, 3.4% with large cell lung cancer, and 13.1% with SCC (Table 2). Among the 236 patients, 132 (55.9%) had lung cancer located in the right lobe of the lungs. Moreover, the TNM stage of lung cancer in COPD patients tended to be advanced, with stages III–VI accounting for 63.6% (Supplementary Table 1).

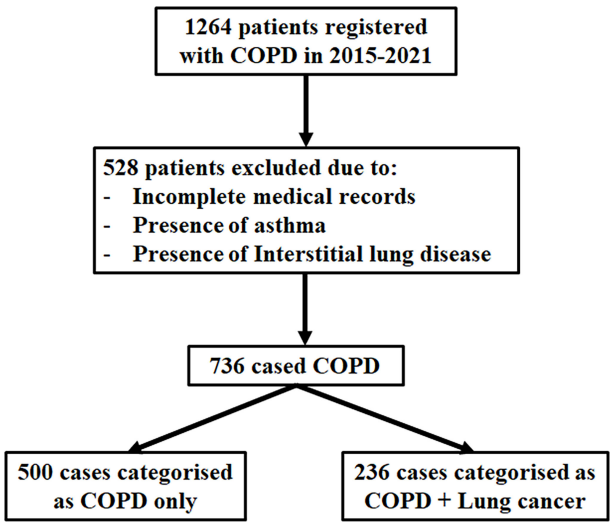


FIGURE 1
Flowchart of the population included in the study. Between 2015 and 2021, a total of 1,264 patients were retrospectively reviewed; 528 were excluded from the analysis due to missing medical records, asthma, and interstitial lung disease. The number of patients with COPD (FEV1/FVC < 70%) was 736, which were eligible for this study. Of the 736 patients included, 236 cases had COPD with lung cancer, and 500 controls had COPD only.

TABLE 1 Baseline characteristics of the study population.

Clinical characteristics	COPD with lung cancer (n = 236)	COPD only (n = 500)	P-value
Age (years)	66.44 ± 8.37	69.87 ± 9.37	<0.001
Gender, *n* (%)			
Male	223 (94.5)	470 (94.0)	0.791
Female	13 (5.5)	30 (6.0)	
BMI (kg/m^2)	21.18 ± 2.58	20.79 ± 3.29	0.114
Smoking status, *n* (%)			
Current smoking	178 (75.4)	254 (50.8)	<0.001
Never/ever smoking	58 (24.6)	246 (49.2)	
Pulmonary functions			
FEV1/FVC	0.59 ± 0.10	0.52 ± 0.11	<0.001
FEV1%pred	66.46 ± 20.54	47.39 ± 21.01	<0.001
Inflammatory markers			
WBC (10^9/L)	8.68 ± 1.60	6.69 ± 1.60	<0.001
Neutrophil (10^9/L)	6.03 ± 1.29	4.13 ± 1.24	<0.001
Lymphocyte (10^9/L)	1.75 ± 0.37	1.78 ± 0.51	0.641
NLR	3.51 ± 0.72	2.45 ± 0.81	<0.001
PLR	172.72 ± 24.96	142.64 ± 37.56	<0.001
Emphysema, *n* (%)			
Yes	205 (86.9)	341 (68.2)	<0.001
No	31 (13.1)	159 (31.8)	

BMI, body mass index; FEV1, forced expiratory volume in 1 s; FVC, forced vital capacity; WBC, white blood cell; NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio.

The levels of NLR or PLR in COPD with lung cancer

In order to confirm whether NLR or PLR served as an independent inflammation biomarker for predicting the progression of lung cancer in COPD patients, the level of NLR or PLR was compared among these groups with different clinical features. The results showed that there were no significant differences in NLR or PLR values between gender and BMI (Supplementary Table 2). NLR or PLR levels could markedly increase with advanced age ($P < 0.001$). In addition, the level of NLR or PLR in current smoker was higher than in never/ever smoker ($P < 0.05$), and the NLR or PLR was significantly higher in GOLD stages 3–4 than 1–2 ($P < 0.001$). Moreover, lung cancer patients at TNM stages IIIB–IV had higher levels of NLR and PLR than those at stages I–IIIA ($P < 0.001$). Finally, the ineligible surgery group had higher levels of NLR and PLR than the eligible for surgery lung cancer patients ($P < 0.001$).

The correlation between NLR or PLR with clinical characteristics in COPD with lung cancer

To further evaluate the correlation between NLR or PLR and other clinical parameters, Pearson's correlation coefficients were analyzed. It was indicated that a positive correlation was observed between NLR or PLR value and age, PT, FIB, and D-dimer ($P < 0.05$) (Supplementary Figure 1). Conversely, NLR or PLR value was negatively correlated with FEV1%pred ($r = -0.299$, $P < 0.001$; $r = -0.300$, $P < 0.001$, respectively). We also found a negative correlation between PLR and BMI ($r = -0.146$, $P = 0.025$). In addition, PLR was positively correlated with PLT and APTT ($r = 0.167$, $P < 0.01$; $r = 0.148$, $P = 0.023$, respectively).

Predictive role of NLR or PLR in COPD with lung cancer

ROC curve analysis was also performed for the NLR and PLR values to detect the diagnosis of lung cancer in COPD patients. The best NLR cutoff value was defined as 2.91, the sensitivity was 84.3%, and the specificity was 74.4%, with the best AUC of 0.84. The optimal PLR cutoff value was 156.53 and the AUC was 0.74, with 81.8% sensitivity and 62.8% specificity (Figure 2A). Moreover, the ROC curve analysis using NLR and PLR to predict TNM stages IIIB–IV at first diagnosis in the COPD with lung cancer patients indicated an optimal cutoff NLR of 3.53, AUC of 0.74, sensitivity of 60.9%, and specificity of 75.9%. The best cutoff of PLR was 172.10 and the AUC was 0.73, with 70.3% sensitivity and 70.4% specificity (Figure 2B). Finally, the predictive accuracy of NLR and PLR in ineligible surgery at first diagnosis in the COPD with lung cancer patients was also

10.3389/fonc.2022.902955

TABLE 2 Multivariate analysis of independent predictors of (A) lung cancer diagnosis in COPD patients, (B) TNM stages IIIB–IV at first diagnosis in the COPD with lung cancer patients, and (C) ineligible surgery at first diagnosis in the COPD with lung cancer patients.

Risk factors	OR	95% CI	*P*-value
A			
Age (≥70)	0.291	0.179–0.474	<0.001
Smoking status			
Current smoking	1.868	1.143–3.053	0.013
Never/ever smoking	–	–	–
FEV1%pred	1.049	1.038–1.061	<0.001
NLR			
NLR ≥ 2.91	12.731	7.493–21.631	<0.001
NLR < 2.91	–	–	–
PLR			
PLR ≥ 156.53	4.389	2.609–7.382	<0.001
PLR < 156.53	–	–	–
Emphysema	2.815	1.560–5.080	0.001
B			
Smoking status			
Current smoking	6.583	2.941–14.736	<0.001
Never/ever smoking	–	–	–
NLR			
NLR ≥ 3.53	3.788	1.811–7.923	<0.001
NLR < 3.53	–	–	–
PLR			
PLR ≥ 172.10	4.775	2.321–9.823	<0.001
PLR < 172.10	–	–	–
C			
BMI			
≥25 (kg/m^2)	0.219	0.061–0.792	0.021
<25(kg/m^2)	–	–	–
Smoking status			
Current smoking	4.267	2.051–8.875	<0.001
Never/ever smoking	–	–	–
PLR			
PLR ≥ 172.10	3.268	1.531–6.976	0.002
PLR < 172.10	–	–	–

BMI, body mass index; FEV1, forced expiratory volume in 1 s; NLR, neutrophil-to-lymphocyte ratio; PLR, platelet-to-lymphocyte ratio.

analyzed with the ROC curve, the best cutoff of NLR was 3.49, and the AUC was 0.68, with 55.6% sensitivity and 74.3% specificity. The ROC curve for PLR had an AUC of 0.69 with a cutoff value of 172.10, a sensitivity of 63.0%, and s specificity of 73.0% (Figure 2C).

Multivariate analyses of risk factors of COPD with lung cancer

To identify the risk factors related to lung cancer in patients with COPD, univariate and stepwise multivariate logistic regression analyses were performed. In the subgroup analysis, the results showed that age (≥70) (OR: 0.291; 95% CI: 0.179–0.474; $P < 0.001$), current smoker (OR: 1.868; 95% CI: 1.143–3.053; $P = 0.013$), FEV1%pred (OR: 1.049; 95% CI: 1.038–1.061; $P < 0.001$), NLR ≥2.91 (OR: 12.731; 95% CI: 7.493–21.631; $P < 0.001$), PLR ≥156.53 (OR: 4.389; 95% CI: 2.609–7.382; $P < 0.001$), and emphysema (OR: 2.815; 95% CI: 1.560–5.080; $P = 0.001$) were independent risk factors related to the COPD with lung cancer patients (Table 2A). Then, we evaluated the risk factors associated with TNM stages IIIB–IV at first diagnosis in the COPD with lung cancer patients. On multivariate analysis, current smoker (OR: 6.583; 95% CI: 2.941–14.736; $P < 0.001$), NLR ≥3.53 (OR: 3.788; 95% CI: 1.811–7.923; $P < 0.001$), and PLR ≥172.10 (OR: 4.775; 95% CI: 2.321–9.823; $P < 0.001$) were independent predictors of TNM stages IIIB–IV at first diagnosis in the COPD with lung cancer patients (Table 2B). Finally, we extended our analysis to

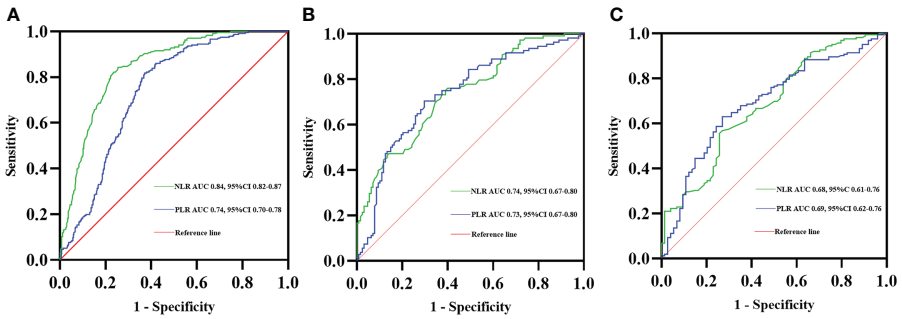


FIGURE 2
Receiver operating characteristic (ROC) curves of NLR and PLR predicting lung cancer diagnosis in COPD patients **(A)**. ROC curves of NLR and PLR predicting TNM stages IIIB–IV at first diagnosis in the COPD with lung cancer patients **(B)**. ROC curves of NLR and PLR predicting ineligible surgery at first diagnosis in the COPD with lung cancer patients **(C)**.

identify the factors that contributed to ineligible surgery at first diagnosis in the COPD with lung cancer patients. BMI ≥25 (kg/m^2) (OR: 0.219; 95% CI: 0.061–0.792; $P = 0.021$), current smoker (OR: 4.267; 95% CI: 2.051–8.875; $P < 0.001$), and PLR ≥172.10 (OR: 3.268; 95% CI: 1.531–6.976; $P = 0.002$) were the statistically significant risk factors for ineligible surgery treatment in the COPD with lung cancer patients (Table 2C).

Discussion

In the present study, we investigated the clinical features of COPD-associated lung cancer in a large cohort of COPD patients and identified the value of NLR or PLR that served as an independent predictive factor in COPD with lung cancer. The results indicated that age, smoking status, pulmonary function, serum inflammation biomarkers, and emphysema were associated with COPD coexisting with lung cancer. In terms of median age, the lung cancer group was younger than the COPD group. Airflow limitation assessed by FEV1%pred in COPD was more severe than in lung cancer. The possible reason is that low-dose spiral computed tomography (LDCT) screening was initiated in the year 2010 in China, based on the development of the National Lung Cancer Screening Program (NLST) (34). LDCT was considered as a prevalent screening tool for the early detection of lung cancer in high-risk asymptomatic individuals (35, 36). Another possible reason is that spirometry screening for airflow obstruction in asymptomatic smokers is still controversial (37). Many smokers never experienced lung function decline during the early stage (38). Therefore, based on CT, screening in smoking patients resulted in a higher detection rate of early-stage lung cancer, especially for mild-to-moderate COPD. It has been previously reported that smokers with COPD have a two- to fourfold higher risk of developing lung cancer compared with smokers without COPD (39, 40). It was elucidated that the prevalence of lung cancer was significantly higher in patients with COPD than in the average non-smoking population, reflecting the impact of cigarette smoking on both diseases (41). It has been well acknowledged that cigarette smoking exposure which caused lung barrier dysfunction and inflammation-associated COPD should be considered as an important risk factor for lung cancer (42). In our research, current smoking was an independent risk factor for lung cancer in the COPD group and related to TNM stages IIIB–IV and ineligible surgery treatment. Furthermore, there have been several studies showing that emphysema was an independent poor prognostic factor for tumor recurrence in completely resected NSCLC patients (43). Emphysema severity and centrilobular subtype were associated with a greater risk of lung cancer (44). It was considered that emphysema is an independent predictor of lung cancer diagnosis (45). Other researchers have shown that quantitative emphysema severity of the whole lung and stage III were independent predictors of lung cancer recurrence after adjusting for age, gender, smoking status, and FEV1 (46). Lung adenocarcinomas in the emphysema group have a more aggressive pathologic grade and a higher prevalence of solid lesions (47). Consistent with previous studies, our study showed that emphysema was also independently associated with lung cancer coexisting with COPD.

On the other hand, COPD is characterized as a chronic systemic inflammatory disease, which also has been found to be related to exacerbation (48). Several reports have shown that peripheral blood neutrophils and T lymphocytes were activated and there were increased levels of proinflammatory cytokines in the plasma (49–51). A recent study also demonstrated that increased baseline neutrophil counts were significantly associated with worse OS but not for lymphocyte counts in NSCLC (52). Wong et al. reported positive associations between white blood counts and lung cancer risk among never-smoking women (53). Furthermore, another study found that lymphocyte percentage exhibited a high correlation with the clinical characteristics and metastasis of lung cancer patients (54). In this study, we observed that circulating white blood counts and

neutrophil counts were significantly higher in the lung cancer group than in the COPD-only group. It was likely that neutrophils were quickly recruited from the circulation during pulmonary inflammation (55). Thus, we have considered that persistent inflammation could mediate the progression of COPD with lung cancer. Several reports have shown that the levels of NLR or PLR significantly increased in the lung cancer and COPD groups, respectively (56, 57). However, few studies have reported the levels of NLR or PLR in lung cancer with COPD. Our data revealed that NLR and PLR levels were significantly higher in the COPD with lung cancer group than in the COPD group. Then, to further confirm the clinical relevance of inflammation biomarkers, we analyzed the correlation between NLR, PLR, and clinical data from lung cancer patients with COPD. Some studies showed that elevated PLR was associated with shorter OS and poor PFS in NSCLC (58, 59). Another study suggested that elevated PLR might be a predictive factor of poor prognosis for NSCLC patients (60). Yao et al. found that the levels of NLR and PLR were significantly higher among non-survivors compared to survivors of AECOPD, and NLR may be a simple and useful prognostic marker for hospital mortality in patients with AECOPD (61). Elevated NLR could be used as a marker similar to CRP, WBC, and ESR in the determination of increased inflammation in acutely exacerbated COPD and could be beneficial for the early detection of potential acute exacerbations in patients with COPD who have normal levels of traditional markers (62). In the present study, the levels of NLR and PLR had statistical significance among the multiple stratifications in clinical parameters, especially those with a high NLR or PLR were associated with age over 70 years old, current smoking status, and ineligible surgery treatment. The level of PLR or NLR markedly increased with the severity of airflow limitation accessed with GOLD stages 3–4 and advanced progression of lung cancer evaluated with TNM stages IIIB–IV. Additionally, NLR or PLR was positively associated with age, FIB, and D-dimer. We also found a negative association between NLR and FEV1%pred. PLR was negatively associated with BMI and FEV1%pred. The results revealed that PLR or NLR values were inversely associated with the severity of airflow limitation. Also, it was suggested that higher PLR or NLR was positively associated with hypercoagulation status. In our study, the applicable thresholds for NLR and PLR were observed using the ROC curve. Multivariate analysis revealed that NLR ≥2.91 and PLR ≥156.53 were independent risk factors for lung cancer development in COPD patients; the COPD with lung cancer patients with NLR ≥3.53 and PLR ≥172.10 are more likely to be diagnosed in TNM stages IIIB–IV at first diagnosis, while patients with PLR ≥172.10 had fewer surgical opportunities. We believe that NLR or PLR could be part of an important and easily measurable inflammation biomarker system to predict the diagnosis and severity of COPD with lung cancer.

However, this study also had several limitations. Because the pulmonary function tests cannot be properly performed on a number of elderly people in this study, this may lead to the possibility of selection bias in the COPD population. Another limitation is the lack of data on follow-up survival information. We did not evaluate the outcomes of the surgical treatment and did not perform a prognostic analysis due to the high rate of missing follow-up data. Longitudinal studies are needed to evaluate the long-term clinical prognosis of COPD with lung cancer. In addition, the study has a retrospective design, in which a number of patients with severe COPD were unable to undergo lung tissue pathologic diagnosis. Thus, it may have caused bias in the selection of the study group. Moreover, this study could not provide information about the use of inhaled corticosteroids, which is a potential influencing factor for the prevention of lung cancer (63, 64).

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Ethical Committee of the Clinical Research Ethics Committee of the First Affiliated Hospital of Xiamen University. The patients/participants provided their written informed consent to participate in this study.

Author contributions

AM and WG drafted the manuscript. GW and YD collected the associated clinical data. HH and LZ contributed to the data statistical analyses. DB and YH prepared the figures and tables. JC and QL wrote the manuscript and conceptualized the framework for this research. All authors read and approved the final manuscript.

Funding

This work was supported the Xiamen Science and Technology Bureau (3502Z20194004) and Xiamen Science and Health Joint Project of Fujian Natural Science Foundation (2020J011229).

Acknowledgments

The authors would like to thank all the patients who participated in this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.902955/full#supplementary-material

References

1. Parris BA, O'Farrell HE, Fong KM, Yang IA. Chronic obstructive pulmonary disease (Copd) and lung cancer: Common pathways for pathogenesis. *J Thorac Dis* (2019) 11(Suppl 17):S2155–S72. doi: 10.21037/jtd.2019.10.54

2. Gao S, Li N, Wang S, Zhang F, Wei W, Li N, et al. Lung cancer in people's republic of China. *J Thorac Oncol* (2020) 15(10):1567–76. doi: 10.1016/j.jtho.2020.04.028

3. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. *Gold Exec Summary. Am J Respir Crit Care Med* (2017) 195(5):557–82. doi: 10.1164/rccm.201701-0218PP

4. Collaborators GBDCoD. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: A systematic analysis for the global burden of disease study 2016. *Lancet* (2017) 390(10100):1151–210. doi: 10.1016/S0140-6736(17)32152-9

5. Zhu B, Wang Y, Ming J, Chen W, Zhang L. Disease burden of copd in China: A systematic review. *Int J Chron Obstruct Pulmon Dis* (2018) 13:1353–64. doi: 10.2147/COPD.S161555

6. Yang L, Xue T, Wang N, Yuan Y, Liu S, Li H, et al. Burden of lung cancer attributable to ambient fine particles and potential benefits from air quality improvements in Beijing, China: A population-based study. *Sci Total Environ* (2020) 738:140313. doi: 10.1016/j.scitotenv.2020.140313

7. Wu F, Wang L, Zhou C. Lung cancer in China: Current and prospect. *Curr Opin Oncol* (2021) 33(1):40–6. doi: 10.1097/CCO.0000000000000703

8. Wasswa-Kintu S, Gan WQ, Man SF, Pare PD, Sin DD. Relationship between reduced forced expiratory volume in one second and the risk of lung cancer: A systematic review and meta-analysis. *Thorax* (2005) 60(7):570–5. doi: 10.1136/thx.2004.037135

9. Machida H, Inoue S, Shibata Y, Kimura T, Ota T, Ishibashi Y, et al. The incidence and risk analysis of lung cancer development in patients with chronic obstructive pulmonary disease: Possible effectiveness of annual ct-screening. *Int J Chron Obstruct Pulmon Dis* (2021) 16:739–49. doi: 10.2147/COPD.S287492

10. Wang W, Dou S, Dong W, Xie M, Cui L, Zheng C, et al. Impact of copd on prognosis of lung cancer: From a perspective on disease heterogeneity. *Int J Chron Obstruct Pulmon Dis* (2018) 13:3767–76. doi: 10.2147/COPD.S168048

11. Jenkins CR, Jones PW, Calverley PM, Celli B, Anderson JA, Ferguson GT, et al. Efficacy of Salmeterol/Fluticasone propionate by gold stage of chronic obstructive pulmonary disease: Analysis from the randomised, placebo-controlled torch study. *Respir Res* (2009) 10:59. doi: 10.1186/1465-9921-10-59

12. Bale G, Martinez-Camblor P, Burge PS, Soriano JB. Long-term mortality follow-up of the Isolde participants: Causes of death during 13 years after trial completion. *Respir Med* (2008) 102(10):1468–72. doi: 10.1016/j.rmed.2008.04.001

13. Lim JU, Yeo CD, Rhee CK, Kang HS, Park CK, Kim JS, et al. Comparison of clinical characteristics and overall survival between spirometrically diagnosed chronic obstructive pulmonary disease (Copd) and non-copd never-smoking stage I-iv non-small cell lung cancer patients. *Int J Chron Obstruct Pulmon Dis* (2019) 14:929–38. doi: 10.2147/COPD.S190244

14. Kim ES, Kim YT, Kang CH, Park IK, Bae W, Choi SM, et al. Prevalence of and risk factors for postoperative pulmonary complications after lung cancer surgery in patients with early-stage copd. *Int J Chron Obstruct Pulmon Dis* (2016) 11:1317–26. doi: 10.2147/COPD.S105206

15. Mouronte-Roibas C, Ruano-Ravina A, Fernandez-Villar A. Lung cancer and chronic obstructive pulmonary disease: Understanding the complexity of carcinogenesis. *Transl Lung Cancer Res* (2018) 7(Suppl 3):S214–S7. doi: 10.21037/tlcr.2018.08.11

16. Eapen MS, Hansbro PM, Larsson-Callerfelt AK, Jolly MK, Myers S, Sharma P, et al. Chronic obstructive pulmonary disease and lung cancer: Underlying pathophysiology and new therapeutic modalities. *Drugs* (2018) 78(16):1717–40. doi: 10.1007/s40265-018-1001-8

17. Tan Z, Xue H, Sun Y, Zhang C, Song Y, Qi Y. The role of tumor inflammatory microenvironment in lung cancer. *Front Pharmacol* (2021) 12:688625. doi: 10.3389/fphar.2021.688625

18. Nakano-Narusawa Y, Yokohira M, Yamakawa K, Ye J, Tanimoto M, Wu L, et al. Relationship between lung carcinogenesis and chronic inflammation in rodents. *Cancers (Basel)* (2021) 13(12): 2910. doi: 10.3390/cancers13122910

19. Barnes PJ. Inflammatory mechanisms in patients with chronic obstructive pulmonary disease. *J Allergy Clin Immunol* (2016) 138(1):16–27. doi: 10.1016/j.jaci.2016.05.011

20. Wang Y, Xu J, Meng Y, Adcock IM, Yao X. Role of inflammatory cells in airway remodeling in copd. *Int J Chron Obstruct Pulmon Dis* (2018) 13:3341–8. doi: 10.2147/COPD.S176122

21. Berg J, Halvorsen AR, Bengtson MB, Tasken KA, Maelandsmo GM, Yndestad A, et al. Levels and prognostic impact of circulating markers of inflammation, endothelial activation and extracellular matrix remodelling in patients with lung cancer and chronic obstructive pulmonary disease. *BMC Cancer* (2018) 18(1):739. doi: 10.1186/s12885-018-4659-0

22. Kurtipek E, Bekci TT, Kesli R, Sami SS, Terzi Y. The role of neutrophil-lymphocyte ratio and platelet-lymphocyte ratio in exacerbation of chronic obstructive pulmonary disease. *J Pak Med Assoc* (2015) 65(12):1283–7.

23. Bozan N, Alpayci M, Aslan M, Cankaya H, Kiroglu AF, Turan M, et al. Mean platelet volume, red cell distribution width, platelet-to-Lymphocyte and neutrophil-to-Lymphocyte ratios in patients with ankylosing spondylitis and their relationships with high-frequency hearing thresholds. *Eur Arch Otorhinolaryngol* (2016) 273(11):3663–72. doi: 10.1007/s00405-016-3980-y

24. Mercan R, Bitik B, Tufan A, Bozbulut UB, Atas N, Ozturk MA, et al. The association between Neutrophil/Lymphocyte ratio and disease activity in rheumatoid arthritis and ankylosing spondylitis. *J Clin Lab Anal* (2016) 30 (5):597–601. doi: 10.1002/jcla.21908

25. Sahin F, Kosar AF, Aslan AF, Yigitbas B, Uslu B. Serum biomarkers in patients with stable and acute exacerbation of chronic obstructive pulmonary disease: A comparative study. *J Med Biochem* (2019) 38(4):503–11. doi: 10.2478/jomb-2018-0050

26. Luo Z, Zhang W, Chen L, Xu N. Prognostic value of Neutrophil:Lymphocyte and Platelet:Lymphocyte ratios for 28-day mortality of patients with aecopd. *Int J Gen Med* (2021) 14:2839–48. doi: 10.2147/IJGM.S312045

27. Kumar P, Law S, Sriram KB. Evaluation of platelet lymphocyte ratio and 90-day mortality in patients with acute exacerbation of chronic obstructive pulmonary disease. *J Thorac Dis* (2017) 9(6):1509–16. doi: 10.21037/jtd.2017.05.77

28. Goksel S, Ozcelik N, Telatar G, Ardic C. The role of hematological inflammatory biomarkers in the diagnosis of lung cancer and in predicting tnm stage. *Cancer Invest* (2021) 39(6-7):514–20. doi: 10.1080/07357907.2021.1938110

29. Sebastian NT, Raj R, Prasad R, Barney C, Brownstein J, Grecula J, et al. Association of pre- and posttreatment neutrophil-lymphocyte ratio with recurrence and mortality in locally advanced non-small cell lung cancer. *Front Oncol* (2020) 10:598873. doi: 10.3389/fonc.2020.598873

30. Zhu X, Song H, Chen Y, Han F, Wang Q, Cui Y. Neutrophil-to-Lymphocyte ratio and platelet-to-Lymphocyte ratio in blood to distinguish lung cancer patients from healthy subjects. *Dis Markers* (2020) 2020:8844698. doi: 10.1155/2020/8844698

31. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 world health organization classification of lung tumors: Impact of genetic, clinical and radiologic advances since the 2004 classification. *J Thorac Oncol* (2015) 10(9):1243–60. doi: 10.1097/JTO.0000000000000630

32. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, et al. Interpretative strategies for lung function tests. *Eur Respir J* (2005) 26(5):948–68. doi: 10.1183/09031936.05.00035205

33. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, et al. Global strategy for the diagnosis, management and prevention of chronic obstructive lung disease 2017 report: Gold executive summary. *Respirology* (2017) 22(3):575–601. doi: 10.1111/resp.13012

34. National Lung Screening Trial Research T, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* (2011) 365(5):395–409. doi: 10.1056/NEJMoa1102873

35. Zhou Q, Fan Y, Wu N, Huang Y, Wang Y, Li L, et al. Demonstration program of population-based lung cancer screening in China: Rationale and study design. *Thorac Cancer* (2014) 5(3):197–203. doi: 10.1111/1759-7714.12078

36. Zhao SJ, Wu N. Early detection of lung cancer: Low-dose computed tomography screening in China. *Thorac Cancer* (2015) 6(4):385–9. doi: 10.1111/1759-7714.12253

37. Young RP, Hopkins RJ. A clinical practice guideline update on the diagnosis and management of stable chronic obstructive pulmonary disease. *Ann Intern Med* (2012) 156(1 Pt 1):68–9. doi: 10.7326/0003-4819-156-1-201201030-00021

38. Walters JA, Hansen EC, Walters EH, Wood-Baker R. Under-diagnosis of chronic obstructive pulmonary disease: A qualitative study in primary care. *Respir Med* (2008) 102(5):738–43. doi: 10.1016/j.rmed.2007.12.008

39. Nakayama M, Satoh H, Sekizawa K. Risk of cancers in copd patients. *Chest* (2003) 123(5):1775–6. doi: 10.1378/chest.123.5.1775-a

40. Papi A, Casoni G, Caramori G, Guzzinati I, Boschetto P, Ravenna F, et al. Copd increases the risk of squamous histological subtype in smokers who develop non-small cell lung carcinoma. *Thorax* (2004) 59(8):679–81. doi: 10.1136/thx.2003.018291

41. Kim TH, Oh DK, Oh YM, Lee SW, Do Lee S, Lee JS. Fibrinogen as a potential biomarker for clinical phenotype in patients with chronic obstructive pulmonary disease. *J Thorac Dis* (2018) 10(9):5260–8. doi: 10.21037/jtd.2018.08.52

42. Hou W, Hu S, Li C, Ma H, Wang Q, Meng G, et al. Cigarette smoke induced lung barrier dysfunction, emt, and tissue remodeling: A possible link between copd and lung cancer. *BioMed Res Int* (2019) 2019:2025636. doi: 10.1155/2019/2025636

43. Heo JW, Kang HS, Park CK, Kim SK, Kim JS, Kim JW, et al. Regional emphysema score is associated with tumor location and poor prognosis in completely resected nsclc patients. *BMC Pulm Med* (2020) 20(1):242. doi: 10.1186/s12890-020-01268-7

44. Gonzalez J, Henschke CI, Yankelevitz DF, Seijo LM, Reeves AP, Yip R, et al. Emphysema phenotypes and lung cancer risk. *PloS One* (2019) 14(7):e0219187. doi: 10.1371/journal.pone.0219187

45. Yong PC, Sigel K, de-Torres JP, Mhango G, Kale M, Kong CY, et al. The effect of radiographic emphysema in assessing lung cancer risk. *Thorax* (2019) 74(9):858–64. doi: 10.1136/thoraxjnl-2018-212457

46. Lee SJ, Yoo JW, Ju S, Cho YJ, Kim JD, Kim SH, et al. Quantitative severity of pulmonary emphysema as a prognostic factor for recurrence in patients with surgically resected non-small cell lung cancer. *Thorac Cancer* (2019) 10(3):421–7. doi: 10.1111/1759-7714.12920

47. Lim CG, Shin KM, Lim JK, Kim HJ, Kim WH, Cho SH, et al. Emphysema is associated with the aggressiveness of copd-related adenocarcinomas. *Clin Respir J* (2020) 14(4):405–12. doi: 10.1111/crj.13146

48. Agustí A, Edwards LD, Rennard SI, MacNee W, Tal-Singer R, Miller BE, et al. Persistent systemic inflammation is associated with poor clinical outcomes in copd: A novel phenotype. *PloS One* (2012) 7(5):e37483. doi: 10.1371/journal.pone.0037483

49. Agustí AG, Noguera A, Sauleda J, Sala E, Pons J, Busquets X. Systemic effects of chronic obstructive pulmonary disease. *Eur Respir J* (2003) 21(2):347–60. doi: 10.1183/09031936.03.00405703

50. Barnes PJ. New concepts in chronic obstructive pulmonary disease. *Annu Rev Med* (2003) 54:113–29. doi: 10.1146/annurev.med.54.101601.152209

51. Chung KF. Cytokines in chronic obstructive pulmonary disease. *Eur Respir J Suppl.* (2001) 34:50s–9s. doi: 10.1183/09031936.01.00229701

52. Biswas T, Gawdi R, Jindal C, Iyer S, Kang KH, Bajor D, et al. Pretreatment neutrophil-to-Lymphocyte ratio as an important prognostic marker in stage iii locally advanced non-small cell lung cancer: Confirmatory results from the proclaim phase iii clinical trial. *J Thorac Dis* (2021) 13(10):5617–26. doi: 10.21037/jtd-21-1018

53. Wong JYY, Bassig BA, Loftfield E, Hu W, Freedman ND, Ji BT, et al. White blood cell count and risk of incident lung cancer in the uk biobank. *JNCI Cancer Spectr* (2020) 4(2):pkz102. doi: 10.1093/jncics/pkz102

54. Huang H, Li L, Luo W, Yang Y, Ni Y, Song T, et al. Lymphocyte percentage as a valuable predictor of prognosis in lung cancer. *J Cell Mol Med* (2022) 26(7)1918–31. doi: 10.1111/jcmm.17214

55. de Oliveira S, Rosowski EE, Huttenlocher A. Neutrophil migration in infection and wound repair: Going forward in reverse. *Nat Rev Immunol* (2016) 16(6):378–91. doi: 10.1038/nri.2016.49

56. Sanchez-Salcedo P, de-Torres JP, Martinez-Urbistondo D, Gonzalez-Gutierrez J, Berto J, Campo A, et al. The neutrophil to lymphocyte and platelet to lymphocyte ratios as biomarkers for lung cancer development. *Lung Cancer (Amsterdam Netherlands)* (2016) 97:28–34. doi: 10.1016/j.lungcan.2016.04.010

57. Günay E, Sarınç Ulaşlı S, Akar O, Ahsen A, Günay S, Koyuncu T, et al. Neutrophil-to-Lymphocyte ratio in chronic obstructive pulmonary disease: A retrospective study. *Inflammation* (2014) 37(2):374–80. doi: 10.1007/s10753-013-9749-1

58. Gu X, Sun S, Gao XS, Xiong W, Qin S, Qi X, et al. Prognostic value of platelet to lymphocyte ratio in non-small cell lung cancer: Evidence from 3,430 patients. *Sci Rep* (2016) 6:23893. doi: 10.1038/srep23893

59. Zhang H, Gao L, Zhang B, Zhang L, Wang C. Prognostic value of platelet to lymphocyte ratio in non-small cell lung cancer: A systematic review and meta-analysis. *Sci Rep* (2016) 6:22618. doi: 10.1038/srep22618

60. Zhao QT, Yuan Z, Zhang H, Zhang XP, Wang HE, Wang ZK, et al. Prognostic role of platelet to lymphocyte ratio in non-small cell lung cancers: A meta-analysis including 3,720 patients. *Int J Cancer* (2016) 139(1):164–70. doi: 10.1002/ijc.30060

61. Yao C, Liu X, Tang Z. Prognostic role of neutrophil-lymphocyte ratio and platelet-lymphocyte ratio for hospital mortality in patients with aecopd. *Int J Chron. Obstruct Pulmon Dis* (2017) 12:2285–90. doi: 10.2147/copd.S141760

62. Taylan M, Demir M, Kaya H, Selimoglu Sen H, Abakay O, Carkanat AI, et al. Alterations of the neutrophil-lymphocyte ratio during the period of stable and acute exacerbation of chronic obstructive pulmonary disease patients. *Clin Respir J* (2017) 11(3):311–7. doi: 10.1111/crj.12336

63. Liu SF, Kuo HC, Lin MC, Ho SC, Tu ML, Chen YM, et al. Inhaled corticosteroids have a protective effect against lung cancer in female patients with chronic obstructive pulmonary disease: A nationwide population-based cohort study. *Oncotarget* (2017) 8(18):29711–21. doi: 10.18632/oncotarget.15386

64. Raymakers AJ, McCormick N, Marra CA, Fitzgerald JM, Sin D, Lynd LD. Do inhaled corticosteroids protect against lung cancer in patients with copd? a systematic review. *Respirology* (2017) 22(1):61–70. doi: 10.1111/resp.12919

frontiers | Frontiers in Oncology

TYPE Review
PUBLISHED 19 October 2022
DOI 10.3389/fonc.2022.862602

OPEN ACCESS

EDITED BY
Ying Liu,
Qingdao University, China

REVIEWED BY
Huaichao Luo,
Sichuan Cancer Hospital, China
Jie Song,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

*CORRESPONDENCE
Xiulin Jiang
jiangxiulin@mail.kiz.ac.cn
Luciano Mutti
luciano.mutti@temple.edu
Jun Peng
pengjun_fphyp@126.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 26 January 2022
ACCEPTED 26 August 2022
PUBLISHED 19 October 2022

CITATION
Ren W, Yuan Y, Peng J, Mutti L and Jiang X (2022) The function and clinical implication of circular RNAs in lung cancer.
Front. Oncol. 12:862602.
doi: 10.3389/fonc.2022.862602

The function and clinical implication of circular RNAs in lung cancer

Wenjun Ren[1,2†], Yixiao Yuan[3†], Jun Peng[4*], Luciano Mutti[5*] and Xiulin Jiang[6*]

[1]Department of Cardiovascular Surgery, The First People's Hospital of Yunnan Province, Kunming, Yunnan, China, [2]Department of Thoracic Surgery, The Second Affiliated Hospital of Kunming Medical University, Kunming, Yunnan, China, [3]Key Laboratory of Molecular Oncology and Epigenetics, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China, [4]Department of Thoracic Surgery, The First People's Hospital of Yunnan Province, Kunming, Yunnan, China, [5]The Sbarro Institute for Cancer Research and Molecular Medicine, Center for Biotechnology, College of Science and Technology, Temple University, Philadelphia, PA, United States, [6]Kunming College of Life Science, University of Chinese Academy of Sciences, Beijing, China

Lung cancer is the leading cause of cancer-related deaths worldwide. Despite the recent advent of promising new targeted therapies, lung cancer diagnostic strategies still have difficulty in identifying the disease at an early stage. Therefore, the characterizations of more sensible and specific cancer biomarkers have become an important goal for clinicians. Circular RNAs are covalently close, endogenous RNAs without 5′ end caps or 3′poly (A) tails and have been characterized by high stability, abundance, and conservation as well as display cell/tissue/developmental stage-specific expressions. Numerous studies have confirmed that circRNAs act as microRNA (miRNA) sponges, RNA-binding protein, and transcriptional regulators; some circRNAs even act as translation templates that participate in multiple pathophysiological processes. Growing evidence have confirmed that circRNAs are involved in the pathogenesis of lung cancers through the regulation of proliferation and invasion, cell cycle, autophagy, apoptosis, stemness, tumor microenvironment, and chemotherapy resistance. Moreover, circRNAs have emerged as potential biomarkers for lung cancer diagnosis and prognosis and targets for developing new treatments. In this review, we will summarize recent progresses in identifying the biogenesis, biological functions, potential mechanisms, and clinical applications of these molecules for lung cancer diagnosis, prognosis, and targeted therapy.

KEYWORDS

circRNA, lung cancer, clinical significance, diagnostic, mechanism

Introduction

According to the cancer statistics of 2020, lung cancer is the second most frequent human tumor and the leading cause of cancer-related deaths worldwide (1). Lung cancer includes small-cell lung carcinoma (SCLC) and non-small-cell lung carcinoma (NSCLC). NSCLC accounts for approximately 85% of all cases (2) and is mostly represented by lung adenocarcinoma (ADC), lung squamous cell carcinoma (SCC), and large-cell lung carcinoma.

Despite the recent rapid advances in diagnostic tumor biomarkers and therapeutic approaches such as surgery, chemotherapy, radiotherapy, targeted therapy, or immunotherapy, the 5-year overall survival for lung cancer still remains poor (3). Therefore, finding new biomarkers and therapeutic targets for this cancer is paramount.

The initiation and progression of lung cancer is an extremely complex process that involves genetic mutations, the role of the tumor microenvironment, and the dysregulation of epigenetic pathways (4–6). Epigenetic changes in lung cancer such as histone modifications (7), DNA methylation (8), and noncoding RNAs (ncRNAs) (9) have been extensively studied. Cellular ribonucleic acids (RNAs) can be divided into coding and non-coding RNAs; these latter kind include small interfering RNAs (siRNAs), small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs), PIWI-interacting RNAs (piRNAs), transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), microRNAs (miRNAs), long non-coding RNAs (lncRNAs), and circular RNAs (circRNAs) (10–12).

Several evidence have revealed that numerous ncRNAs are dysregulated and involved in lung cancer initiation and progression (13–15). In particular, miRNAs and lncRNAs have been shown to be involved in lung cancer progression—for example, microRNA-485-5p suppresses growth and metastasis in non-small-cell lung cancer cells by targeting IGF2BP2 (16), whereas miRNA-124 suppresses cell proliferation in NSCLC by the down-regulation of SOX8 expression (17), and lncRNA H19 promotes lung cancer proliferation and metastasis by inhibiting the miR-200a function (18). lncRNA TUC338 promotes the invasion of lung cancer by activating the MAPK pathway (19). While the role of miRNAs and lncRNAs in cancer development is currently being progressively unraveled, the functions of circRNA in the initiation and progression of lung cancers need further investigation.

As a new type of ncRNAs with distinct properties and diverse functions, unlike lncRNA and miRNA, circRNA is an endogenous RNA molecule with covalently closed loop structures without termination at 5′ caps and 3′ poly(A) tails. The first circRNA was observed in viroid (20). Next, circRNAs were discovered in the cytoplasm of eukaryotic cells by electron microscopy (21) and were earlier mainly considered to be "junk RNAs" produced by aberrant splicing events (22), until it was hypothesized as a possible function of the testis-specific circRNA, expressed by Mus spretus Sry genes (23). Recently, with the development of high-throughput RNA sequencing and bioinformatics algorithms, thousands of circRNAs in eukaryotes have been identified, including plants, fungi, protists, fish, worms, and mammals (24–26). circRNAs are composed of exonic and/or intronic sequences and are primarily generated by back-splicing, a non-canonical alternative RNA splicing event mediated by the spliceosome and regulated by a combination of cis-elements and trans-factors (27, 28). According to biogenesis from a different mechanism, circRNA can be divided into three types: exonic circRNA (ecircRNNAs) (29), intronic RNAs (ciRNAs) (30), and exon–intron circRNAs (EIcircRNA) characterized by the co-presence of both exons and introns (30).

Despite the lack of polyadenylation [poly(A)] and capping, circRNAs generally localize to the cytoplasm (29), while the exon–intron circRNAs and ciRNAs generally localize in the nucleus and promote the transcription of their parental genes *via* interaction with the U1 small nuclear ribonucleoprotein (31). circRNAs are characterized by abundance (32), stability (33), translational capacity (34), and cell specificity/tissue specificity/developmental stage specificity (29, 35). Numerous studies have demonstrated that circRNAs play a critical role in gene expression (36), act as miRNA sponge (37), interact with RNA binding protein (38), and translate to small peptides (39). Moreover, circRNAs could play paramount roles in physiological processes, including aging (40), myogenesis (41), male reproductive function (42), adipogenesis (43), innate immune response (28), synaptic function (44), insulin secretion (45), and mitochondrial ROS (mROS) output (46). On the other hand, dysregulation circRNAs are also involved in various human diseases, including neurological disorders (47), cardiovascular diseases (32), chronic inflammatory diseases (48), diabetes mellitus (49), and especially human cancers (50, 51).

Liquid biopsy is a biopsy that uses body liquids as the sample source to diagnose, predict the outcome of, or monitor the

Abbreviations: LUAC, lung adenocarcinoma; NSCLC, non-small-cell lung cancer; LUSC, lung squamous cell carcinoma; LC, lung cancer; SCLC, small-cell lung cancer; pre-mRNA, premessenger RNA; m6A, N6-methyladinosinek; ORF, open reading frame; hnRNPs, heterogeneous nuclear ribonucleoproteins; DHX9, ATP-dependent RNA helicase A; AGO2, argonaute 2; RNAi, RNA interference; circRNAs, circular RNAs; ncRNAs, non-coding RNAs; miRNAs, microRNAs; ecircRNAs, exonic circRNAs; ElciRNAs, exon–intron circRNAs; ciRNAs, intronic circRNAs; Pol II, RNA polymerase II; snRNP, small nuclear ribonucleoprotein; RBPs, RNA-binding proteins; ceRNA, competing endogenous RNA; EMT, epithelial–mesenchymal transition; TNM, tumor–node–metastasis; OS, overall survival; ROC, receiver operating characteristic; ASOs, anti-sense oligonucleotides; CRISPR/Cas9, clustered regularly interspaced short palindromic; repeats-associated nuclease Cas9; circ-AS, circRNA alternative splicing; ICSs, flanking intronic complementary sequences; CSC, cancer stem cell.

development of human diseases (52), and because of their high stability, abundant expression, and high specificity, circRNAs detected with liquid biopsy could become promising biomarkers for human diseases (53). circRNAs are involved in the pathogenesis of lung cancer through the regulation of proliferation and invasion (54), cell cycle (55), stemness of lung cancer stem cell (56), chemotherapy resistance (57), and tumor microenvironment (58, 59). Therefore, we will summarize recent progresses in identifying the biological and prognostic role of circRNAs and their potential exploitation as actionable targets in lung cancer.

Biogenesis and properties of circRNAs

The classification and biogenesis of circRNA

circRNAs are generated by RNA polymerase II from pre-mRNA (60). These circRNAs are distinct from their linear RNA counterparts because they lack the 5′–3′ ends and poly-adenylated tail due to their closed covalent structure, which usually decides the fate of many RNA transcripts (29). According to the generation mechanism, the circRNAs can be divided into exonic circRNAs (ecircRNAs) originated from one or more exonic sequences (61) and intronic circRNAs (ciRNAs) originated from intronic sequences (30), while the exon–intron circRNAs (EIciRNAs) can be produced from intron-containing exons (62). Exonic circRNAs are cytoplasmic and result from pre-mRNA splicing where the 3′ splice donor attaches to the 5′ splice acceptor forming an exonic circRNA (61). Exon–intron circRNAs are predominantly nuclear and composed of introns and exons that interact with U1 snRNP and promote the transcription of their parental genes (31). Intronic RNA (ciRNA) formation depends on the 7-nt GU-rich element near the 5′ splice site and an 11-nt C-rich element close to the branchpoint site (30).

At the moment, there are several hypotheses to explain circRNA biogenesis, including (1) intron-pairing-driven circularization (2), RNA-binding protein (RBP)-driven circularization (3), exon skipping, and (4) intron lariat circularization (63).

The intron-pairing-driven circularization model suggests that introns flanking the exon/exons of a pre-mRNA have a structure capable of joining each other. The flanking introns approach each other, creating a secondary conformation that makes the splice sites possibly carry on back-splicing and generate exonic circRNA. Adenosine deaminase 1 acting on RNA (ADAR1) is involved in the intron pairing process of circRNA formation (64).

In the RBP-driven circularization model, RBPs bind to pre-mRNAs to connect the flanking introns. This process is induced by protein dimerization, which forms an RNA loop. Muscle blind-like splicing regulator 1 (MBNL1) protein is the most frequent RBP responsible for circRNA biogenesis (65). In the exon skipping model, one or multiple exons of the mature mRNA will be missing, whereas the lariat-driven circularization hypothesis is based on the binding of adjacent exons, leading to the formation of linear mRNA and exon–intron or multiple-exon circRNA transcript with lariat structure. Finally, the fourth proposed mechanism is the intron lariats, which can form ciRNAs (66).

Role of cis-elements and trans-factors in circular RNA formation

Mounting evidence have shown that back-splicing requires spliceosomal machinery and that the regulation of circRNA formation depends on both cis-regulatory elements and trans-acting factors. In this review, we shall focus on how cis-regulatory elements and trans-acting factors influence the biogenesis of circRNA (Figure 1).

Cis elements regulate the formation of circRNAs

Back-splicing often requires regulatory elements residing in introns flanking circularized exons. Flanking intronic complementary sequences (ICSs) can promote exon circularization and then give rise to the formation of circRNAs (27). Most circRNAs in mammals are produced from internal exons with long flanking introns usually containing ICSs. RNA pairing formed across introns that flank back-spliced exons is expected to bring the distal single strand into close proximity to facilitate circRNA biogenesis (61). Consistently, the elimination of RNA pairs significantly reduces and, sometimes, even removes circRNA production, as revealed by mutagenesis analysis in circRNA expression vectors (67, 68)—for example, in cultured mouse cells, a previous study showed that circSry, derived from Sry gene, is flanked by many intronic complementary sequences, and the presence of long inverted repeats (IR) flanking of Sry gene results in the formation of the Sry circular transcript (69). Furthermore, the fusion-circRNAs (f-circRNAs) that originated from aberrant chromosomal translocations in cancers also supported the view that intronic RNA pairing across circle-forming exons is critical for enhanced circRNA formation (70). F-circM9 was identified to be produced by the MLL/AF9 fusion gene that contains the MLL gene exon 1–8 and AF9 gene exon 6–11. After translocation, the trans intronic sequences of the MLL and AF9 genes are juxtaposed in cis, which subsequently facilitates circRNA generation by forming newly paired ICSs flanking the translocation breakpoint (70). ICSs even have 23 nucleotides flanking intronic regions and also promote the formation of natural

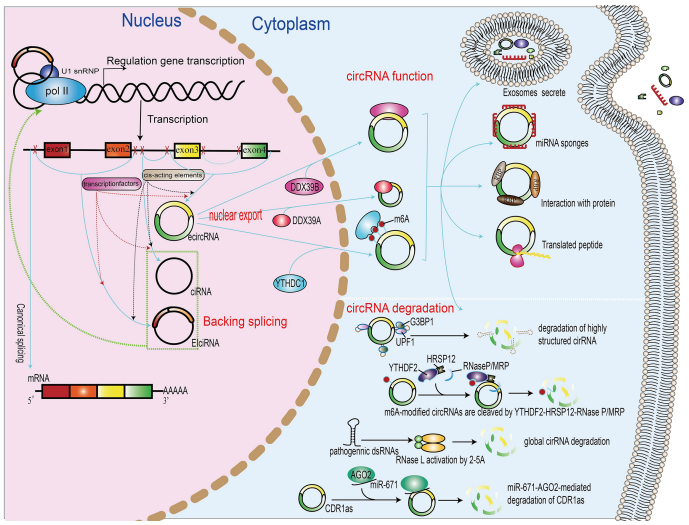


FIGURE 1
Biogenesis, nuclear export, functions, and degradation of circRNAs. In the nucleus of eukaryotic cells, DNA is transcribed to form precursor mRNA (pre-mRNA), which contains coding exons and introns. Differently from linear mRNAs, which are formed regulated by cis-acting elements and transcription factors, according to the generation mechanism, the circRNAs can be divided into exonic–intronic circRNAs (ecircRNAs), ciRNAs, and circRNAs (ElciRNAs). The proteins UAP56, URH49, and YTHDC1 can promote the nuclear export of circRNAs. The functions of circRNAs include the regulation of gene transcription or splicing, acting as miRNA sponge, binding with RNA-binding protein, translation into peptide or protein, and packing in exosomes. The degradation of circRNA includes miRNA-directed circRNA decay, m6A-mediated circRNA decay, RNase L-mediated circRNA decay, and overall structure-mediated circRNA decay.

circRNA in platelets (27). It has been shown that Alu elements are abundant and constitute 11% of the reference human genome, and nearly half of them is located in human introns (71). A recent study demonstrated that complementary inverted-repeat Alu elements drive the RNA pairs, which enhances the exon circularization efficiency and leads to multiple circular RNA transcripts produced from a single gene (27). RNA pairing formed across flanking introns generally promotes circRNA formation (72). Furthermore, RNA pairing within individual introns generally facilitates canonical splicing in linear RNA, competing with RNA pairing across flanking introns and leading to reduced circRNA formation from the same gene locus (36). These recent findings together reveal that cis elements play a key role in circRNA biogenesis.

Trans-acting factor regulate the formation of circRNAs

Previous studies have demonstrated that RNA-binding proteins are also involved in spliceosome function *via* the modulation of alternative splicing which control the information of circRNAs. In Drosophila cells, using RNAi depletion or the pharmacological inhibition of spliceosome-related gene expression inhibits canonical pre-mRNA processing and increases the output of circular RNAs (73). Another study likewise shows that the pharmacological inhibition of the spliceosome promotes the biogenesis of circRNAs in mouse brain (74). Therefore, one can assume that spliceosome regulates the activity of back-splicing and canonical splicing and then affects the biogenesis of circRNAs and mRNA. The abundance of core spliceosomal factors promotes the biogenesis of circRNAs and reduces the product of linear mRNA. The RNA-binding proteins regulate the back-splicing process through an interaction with ICSs, whereas the double-stranded RNA-binding domains in RBPs bind to ICSs and stabilize the transiently formed intronic RNA pair flanking—for example, the NF90/NF110 encoded by interleukin enhancer binding factor 3 gene, which includes a double-stranded RNA-binding domain, *via* binding to intronic inverted-repeat Alu elements, promotes circRNA production (28). The immune factors NF90 and/or NF110, each containing two dsRNA-binding domains (dsRBDs), promote circRNA formation by direct binding to IRAlus formed in nascent pre-mRNA. In the endogenous NF90 knockdown condition, the re-introduction of wild-type NF90, but not NF90 mutants with dsRBD truncations, rescued circRNA expression, further supporting that both dsRBDs of NF90 are required for circRNA production (28). However, dsRBPs can also inhibit circRNA formation by destabilizing the RNA pairing—for example, it was reported that the depletion of DHX9 from nuclear RNA helicase, known to bind to inverted-repeat Alu elements, thus regulating the biogenesis of circRNAs, leads to the double-stranded RNA accumulation defects and increased the circular RNA

production (75). ADAR1, a double-strand RNA-editing enzyme, edits A-to-I of inverted-repeat Alu elements flanking circRNA-forming exons and significantly increases circRNA expression (24). RNA-binding proteins without a dsRBD domain also participate in the regulation of circRNA levels by direct binding to specific RNA motifs—for instance, Quaking (QKI), which is an alternative splicing factor, promotes the interaction between QKI motifs and introns flanking circRNA-forming exons, thus leading to the biogenesis of hundreds of circRNAs with a paramount role in human epithelial–mesenchymal transition (EMT) (76). Other splicing factors have been found to regulate back-splicing in different biological settings: RBM20 promotes the biogenesis of circRNAs by regulating the exclusion of specific exons (77) and heterogeneous nuclear ribonucleoprotein L (HNRNPL) regulates back-splicing and promotes the biogenesis of circRNAs and tumor growth in prostate cancer (76). Splicing factor-regulated back-splicing events also occur in other species. In Drosophila melanogaster, the splicing factor Muscle blind (Mbl) regulates circRNA production from its own pre-mRNA by binding to multiple Mbl-binding sites in introns flanking the circularized exon in Mbl pre-mRNA (36). These recent findings together revealed that circRNA production is highly dependent on biological background and tightly regulated in cells using different cis elements and trans factors that are back-splicing specific.

The properties of circRNAs

circRNAs have the characteristics of abundance (32), specificity (25), stability (33), and translational capacity (34). It has also been shown that circRNAs usually have higher expression levels in low-proliferating cells, and developing tissues such as the heart, lung, and brain usually show higher levels of circRNA (78). These evidence show that the expression patterns of circRNAs are tissue and development specific. While some studies found that the expression of circRNAs was not correlated with their host mRNAs, only in some special circumstances were the expression levels of circRNAs higher than those of their host mRNAs (79). Most of the circular RNAs are cytoplasmic due to the lack of 5′ caps and 3′ poly-A tails that confer resistance to ribonuclease R (RNase R)-dependent degradation (33). The half-life of circRNAs far exceeds that of their linear RNA counterparts (48 and 10 h, respectively (61)). These characteristics make circRNAs ideal biomarkers of cancer.

Numerous results have shown that the expression of circRNAs exhibit specificity to certain tissues and developmental stage—for example, one study found that circRNAs were extraordinarily enriched in the mammalian brain compared to other tissues (80).

Although circRNAs belong to non-coding RNAs, recent studies unraveled that circRNAs can be translated into proteins or peptides —for example, circ-FBXW7 encodes a novel 21-kDa protein, called FBXW7-185aa, and the depletion of FBXW7-185aa promotes malignant phenotypes in glioma (81).

The nuclear export of circular RNAs

In spite of the fact that most circRNAs are primarily cytoplasmic while circRNAs containing retained introns are nuclear (31), some circRNAs are formed in the nucleus and then transported to the cytoplasm.

Their elaborated control of transport is essential for circular RNAs to exert their functions properly in eukaryotic cells. A recent study has demonstrated that m6A modification and other proteins modulate the circRNA export and that knock-down of the DExH/D-box helicase Hel25E leads to the nuclear accumulation of circRNAs longer than 800 nucleotides. Moreover, the depletion of UAP56 or URH49 results in long and short circRNA enrichment in the nucleus, respectively (82). These evidence indicate that the lengths of mature circRNAs may be a decisive factor for their nuclear export. In addition, a more recent study suggest that N6-methyladenosine RNA modification plays a significant role in the regulation of the nuclear export YTHDC1, a circRNA m6A reader that mediates the export of methylated mRNA from the nucleus to the cytoplasm as demonstrated by the depletion of YTHDC1 that leads to the accumulation of m6A-containing transcripts in the nucleus. Another study demonstrated the interaction of YTHDC1 with the splicing factor and nuclear export adaptor protein SRSF3, leading to the increased nuclear export of m6A-containing transcripts (83).

Very similarly, in colorectal carcinoma patients, Chen et al. recently identified the pro-metastatic effect of the upregulation of the m6A-modified circRNA circNSUN2 that enhances the stability of high-mobility group AT-hook protein 2 mRNA. Furthermore, the m6A modification increases the export of circNSUN2 to the cytoplasm (84) (Figure 1).

To conclude, it can be assumed that the mechanisms for extracellular transportation are not yet fully understood, and further studies on this function are certainly warranted.

The degradation of circular RNAs

The circular structure confers resistance of circRNAs to degradation by RNA decay nuclease and other proteins. Therefore, circRNAs have much higher stability than their cognate linear transcripts (85). Nevertheless, under stress conditions, circRNAs can also be degraded through different mechanisms. Thomas Hansen et al. revealed that miRNAs directly bind circRNAs and promote their degradation process. miR-671 directly cleaves a circular antisense transcript of the CDR1 locus in an Ago2-slicer-dependent manner, whereas the miR-671 target site of the circRNA CDR1as/ciRS-7 recruits miR-

671-loading Ago2 to CDR1as/ciRS-7, causing endonucleolytic cleavage by Ago2 and subsequent exonucleolytic RNA degradation (86). In addition, many endonucleases participate in the circRNA degradation process. Under viral infection, double-stranded RNA (dsRNA) activates the RNase L that, in turn, promotes the degradation of circRNAs (87). Fischer et al. have revealed that, during this novel structure-mediated circRNA decay, UPF1 interacts with G3BP1 and accelerates the structure-mediated circRNA decay degradation (88). Interestingly, N6-methyladenosine RNA modifications also affect the degradation of circRNAs; m6A-containing circRNAs use m6A as a recruiter of the m6A reader protein YTHDF2 and adaptor protein HRSP12. Eventually, HRSP12 directly binds to a GGUUC motif on circRNAs and serves as a bridge to bring YTHDF2 and endoribonuclease RNase P/MRP together, thus enabling RNase P/MRP to initiate circRNA degradation (89) (Figure 1).

The roles of circRNAs in physiological conditions

Previous studies have demonstrated that circRNAs contribute to gene regulation through a variety of actions, including regulating transcription, sponging microRNAs, interacting with RNA-binding proteins, and protein translation. The potential functions of circRNAs have been extensively studied (Figure 1). A number of studies have begun to reveal that at least some circRNAs could play crucial roles in physiological conditions by distinct modes of action at the molecular level.

circRNAs regulate transcription and splicing

ciRNAs and EIciRNAs are the products of processed intron lariats and back-splicing with retained introns, respectively. Both of them are mostly nuclear and regulate gene expression at the transcriptional and post-transcriptional levels (31). EIciRNAs interact with U1 small nuclear ribonucleoproteins (snRNP) and RNA PoII in the promoter region of the parental gene to enhance gene expression in cis—for instance, the blockage of circEIF3J and circPAIP2 reduces the transcriptional level of host genes (31). As expected, the processing of circRNAs affects the alternative splicing of their linear cognates and competes with pre-mRNA splicing, suggesting a negative relationship between circRNA and their linear isoforms (36)—for example, circMbl derived from the circularization of the second exon of the MBL competes with linear MBL mRNA to maintain the balance between canonical splicing and circRNA production (36). However, in addition to exon skipping, whether there are additional regulators that affect exon circularization remains to be explored. Recently, a study showed that circSEP3, derived from exon 6 of the SEPALLATA3 gene, concurs to form a RNA–DNA complex that participates in the transcription pausing and alternatively splicing process of the SEP3 gene (31). The above-mentioned research shows that some localized nuclear circRNAs can modulate gene transcription and splicing *via* different mechanisms

circRNAs act as miRNA sponges

A scenario in which competitive endogenous RNA (ceRNA) transcripts with shared microRNA (miRNA) binding sites compete for the post-transcriptional control expression of mRNA is hypothesized (90). In the case of cytoplasmic circRNAs, it acts as miRNA sponge and stabilizes mRNA stability, preventing miRNA-mediated degradation. There are many evidences showing that circRNAs act as miRNA sponges: as a sponge for miR-7, ciRS-7 contains more than 70 selectively conserved miR-7 target sites, is highly and widely connected with AGO proteins, strongly suppresses miR-7 activity, and increases the expression of miR-7 targets (37). circRNA Sry, the testis-specific circRNA in sex-determining region Y (Sry), is a potent miRNA 138 sponge that contains 16 putative target sites for miR-138 and directly binds miRNA 138 to participate in the development of mouse testis (37). Similarly, a recent study indicated that circHIPK3, derived from Exon2 of the HIPK3 gene, contains many potential binding sites of miR-124 and acts as a sponge-inhibiting miR-124, an activity with subsequent pivotal biological effects in human cells (91). It has been shown that circZNF91, sponge for miR-23b-3p, modulates human epidermal stem cell differentiation (92), whereas circBIRC6 directly binds miR-34a and miR-145 and modulates the pluripotency and differentiation of human embryonic stem cells (93). These evidence suggest that circRNAs play a crucial role in controlling the stability and quantity of miRNAs.

circRNAs interact with RNA-binding proteins

Although circRNAs mostly act as miRNA sponges, many research studies have been carried out to explore the functional relevance of circRNA–protein interactions. The most well-known proteins interacting with RNA molecules are the RNA-binding proteins (RBPs). RBPs contain a large class of over 2,000 proteins that interact with transcripts and play an important role in RNA metabolic processes (94). Previous evidence have shown that circRNAs interact with RBPs to form specific circRNA–protein complexes and affect protein localization and function.

Exon–intron circRNAs interact with U1 snRNP and regulate the expression of their related parental genes (31). Circular RNA circ-Foxo3 is mostly cytoplasmic, where it interacts with the anti-senescent protein ID-1, the transcription factor E2F1, FAK, and HIF1α, resulting in increased cellular senescence (95). circ-Foxo3 interacts with cell cycle protein cyclin-dependent kinase 2 and cyclin-dependent kinase inhibitor 1 and leads to the formation of a ternary complex, thus promoting the cell cycle progression of non-cancer cells (96). Although circ-Foxo3 is minimally expressed in tumor specimens, the expression of circ-Foxo3 gradually increases during cancer cell apoptosis. Other studies demonstrate that circ-Foxo3 and MDM2/p53 form the complex circ-Foxo3/MDM2/p53 that promotes MDM2-induced p53 ubiquitination and subsequent degradation. circ-Foxo3 plays an important regulatory role in cell apoptosis because it prevents MDM2 from inducing Foxo3 ubiquitination and degradation, resulting in elevated levels of Foxo3 protein (97). It has also been reported that circACC1 interacts with AMPK and elevates the stability of AMPK holoenzyme, promotes the enzymatic activity of the AMPK holoenzyme by forming a ternary complex with the regulatory β and γ subunits of AMPK holoenzyme, and plays a critical role in both fatty acid β-oxidation and glycolysis (98). Abdelmohsen K and colleagues provided evidence that CircPABPN1 prevents HuR binding to PABPN1 mRNA that results in the suppression of PABPN1 translation (99). Another typical example is circ-AMOTL1, which directly binds to PDK1 and AKT1, leading to AKT1 phosphorylation and nuclear translocation, and plays a cardio-protective role in cardiovascular disease (100). Furthermore, in GBM, circSMARCA5 acts as sponge for SRSF1 as well as participates in the process of VEGFA mRNA splicing regulation and angiogenesis (101). In summary, these evidence indicated that circRNAs interact with RNA-binding proteins and play a significant role in protein function regulation.

CircANRI binds pescadillo homologue 1 (PES1) and is an essential 60S-preribosomal assembly factor that impairs exonuclease-mediated pre-rRNA processing and ribosome biogenesis in vascular smooth muscle cells and macrophages (48). CircHECTD1 promotes fibroblast proliferation and migration *via* interaction with HECTD1/ZC3H12A (102). CircZKSCAN1 competes with the binding between FMRP and β-catenin-binding protein cell cycle and apoptosis regulator 1 (CCAR1) mRNA, with subsequent inhibition of WNT signaling and HCC stemness (103). CircARSP9 increases the susceptibility of HCC cells to NK cell cytotoxicity in HCC cells by the upregulation of UL16-binding protein (ULBP1) expression (104). AR suppresses CircARSP91 expression by upregulating ADAR1, while CircARSP91 plays an inhibitory role in HCC tumor growth (105). Circular RNA FECR1 controls breast cancer tumor growth by the recruitment of TET1 to FLI1 promoter, determining the over-expression of FLI1 (106). circRNA –MTO1 modulates the Eg5 protein expression, suppresses cell viability, promotes monastrol-induced cell cytotoxicity, and reverses monastrol resistance (107) via tumor necrosis factor receptor-associated factor 4 (TRAF4). Circ-Dnmt1 interacts with both p53 and AUF1, promotes the nuclear translocation of both proteins, and plays an oncogenic role in breast cancer cell autophagy (108). CircECE1 was reported to interact with c-Myc to prevent speckle-type POZ-mediated c-Myc ubiquitination and degradation and exert the proliferation and metastatic capability of osteosarcoma cells (109). circRNAs interact with proteins to influence their cellular functions, thereby regulating gene transcription and inhibiting cell cycle progression but promoting cardiac senescence, apoptosis, and cell proliferation.

circRNAs can be translated

Due to lack of 5–3′ polarity, polyadenylated tails, and internal ribosome entry sites (IRES), circRNAs were initially defined as a noncoding RNA. Notwithstanding, a recent study found that, under certain conditions, circRNAs possess translational ability and can code functional peptides—for example, Circ-ZNF609 contains an open reading frame span that can be translated into a protein that controls myoblast proliferation (39). Moreover, m6A modification regulates circRNA translation and METTL3 and YTHDC1 regulate back-splicing, while the knockdown of METTL3 downregulates Circ-ZNF609.

Another study shows that YTHDF3 and eIF4G2 recognize the m6A modifications in Circ-ZNF609 and modulate its translation (110). It was suggested that circular RNA derived from the long intergenic non-protein-coding RNA p53-induced transcript (LINC-PINT) can code an 87-amino-acid peptide. This peptide directly interacts with PAF1c, inhibits the transcriptional elongation of multiple oncogenes, and plays a suppressive role in glioblastoma cells (111). Circ-FBXW7, highly expressed in normal human brain, has been reported to encode a novel 21-kDa protein, named FBXW7-185aa, that reduces the half-life of c-Myc by antagonizing USP28-induced c-Myc stabilization, thus acting as tumor suppressor in glioblastoma cells (34). In addition, circ-SHPRH contains an open reading frame.

The IRES translate a functional protein named SHPRH-146aa and play an additional inhibitory effect in GBM, protecting the full-length SHPRH from ubiquitination (112). Similarly, circβ-catenin is highly expressed in liver cancer tissues where it promotes liver cancer cell growth. Another study shows that Circβ-catenin is predominantly localized in the cytoplasm and encodes a novel 370-amino-acid β-catenin isoform *via* a linear β-catenin mRNA. This β-catenin isoform increases the stability of full-length β-catenin by antagonizing GSK3β-induced β-catenin phosphorylation and degradation and eventually results in the activation of the Wnt pathway (113).

Circ PPP1R12A contains an open reading frame encoding the functional protein circPPP1R12A-73aa that promotes the growth and metastasis of colon cancer *via* activating the Hippo-YAP signaling pathway (114).

We named AKT3-174aa as circ AKT3 encoding a 174-amino-acid novel protein that competes with phosphorylated PDK1, reduces AKT-thr308 phosphorylation, and plays an inhibitory role in the tumorigenicity of GBM cells (115).

CircE7 is derived from human papillomaviruses and, *via* its interactionwith polyribosomes, is translated into the E7 oncoprotein. E7 oncoprotein depletion significantly suppresses cancer cell growth and tumor xenografts (116). CircLgr4 is highly expressed in colorectal tumors and colorectal cancer stem cells. The knockdown of circLgr4 inhibits colorectal cancer stem cell (CSC) self-renewal, colorectal tumorigenesis, and invasion. CircLgr4 encodes the circLgr4-peptide, which interacts with LGR4 to activate the LGR4-Wnt signaling pathway that, in turn, drives the self-renewal of colorectal CSCs (117). In colon cancer cell lines and tissue, circ FNDC3B is mostly localized in the cytoplasm, and its over-expression inhibits the proliferation, invasion, and migration of colon cancer cells. CircFNDC3B also encodes a novel protein circFNDC3B-218aa that suppresses colon cancer progression (118). These evidence pave the avenue to future studies on the translational function of circRNA.

Other physiological functions

A growing body of evidence show that circRNAs play a critical role in the physiological processes, including exosomes secretion (119), aging (40), myogenesis (41), male reproductive function (42), adipogenesis (43), innate immune response (28), synaptic function (44), insulin secretion (45), and mROS output (46). Exosomes are extracellular membranous micro-vesicles with a diameter of 40–160 nm and secreted by various cell types (120). Exosomes derived from host cells can be taken up and exert biological effects both on adjacent and distant cells (121). Exosomes have been implicated in the occurrence and development of many diseases, including cancer (122). Exosomes may include biological molecules such as lipids, proteins, DNA, and non-coding RNAs (ncRNAs), and their content and biological effects depend on the host cells.

ncRNA-containing exosomes play a role in disease progression, including cancer (123). A recent study by Zhang N et al. reported that circSATB2 is highly expressed in NSCLC cancer exosomes and promotes the proliferation, migration, and invasion of NLCSC cells. SATB2-containing exosome can be taken up by NLCSC cells to promote cell–cell communication, progression of NSCLC cells, proliferation of normal bronchial epithelial cells, and lymphatic spreading. The receiver operating characteristic (ROC) analysis curve of exosomial circSATB2 shows an area under the ROC curve (AUC) value of 0.660 and 0.797 in serum from patients with lung cancer and metastatic lung cancer patients, respectively. These results indicate that exosomial circSATB2 could act as a blood detection index for the diagnosis of lung cancer and lung cancer metastasis with high sensitivity and specificity (124). Similarly, it has also been reported that circRNA 002178 exosomes from the plasma of LUAC patients were highly expressed in LUAC tissues and LUAC cancer cells. Functional studies have demonstrated that circRNA-002178 enhances PDL1 and induces T cell exhaustion *via* sponging miR-34 in cancer cells. Interestingly, circRNA-002178 could be exploited as a novel diagnosis biomarker for lung cancer because the AUCs have been demonstrated to be higher (0.9956) in the exosomes derived from cancer cells than in those derived from normal human bronchial epithelial cells (125).

Other evidence indicate that circRNAs play a crucial role in the mitochondria, including non-alcoholic steatohepatitis pathogenesis. The steatohepatitis-associated circRNA ATP5B Regulator (SCAR) is primarily localized in the mitochondria and alleviates meta-inflammation by reducing the mROS output. Mechanistic analyses have demonstrated that circRNA SCAR binds to ATP5B and inhibits mPTP by blocking the CypD–mPTP interaction with subsequent reduction of both mROS generation and fibroblast activation (46). Therefore, circRNA SCAR is a potential therapeutic target for nonalcoholic steatohepatitis. CircSamd4 is upregulated during the differentiation of mouse C2C12 myoblasts into myotube, whereas the overexpression of circSamd4 interferes with the binding of PUR proteins to the Mhc promoter and promotes myogenesis (41). CircArhgap5-2 promotes adipogenesis through maintaining the global adipocyte transcriptional program involved in lipid biosynthesis and metabolism (43). Flies missing circular Boule (circBoule) RNAs have decreased male fertility, under heat stress conditions, and the knockdown of circBoule decreases the fertilization capacity. Moreover, circBoule RNAs inhibits\ the spermatogenesis process by interacting with heat shock proteins (HSPs) and promoting their ubiquitination (42). CircSfl is upregulated in the brain and muscle, and its overexpression significantly extends the lifespan of fruit flies, whereas mechanistic analyses have demonstrated that circSfl is translated into a protein sharing some functions with the full-length Sfl protein encoded by the host gene and extending the lifespan of cells (39). On the whole, these studies demonstrate that circRNAs play an important role in physiological processes.

The roles of circRNAs in lung cancer pathological conditions

Numerous circRNAs have been found to be dysregulated in lung cancer tissues, playing oncogenic or tumor-suppressor

roles. Extensive evidence has demonstrated that there are many differentially expressed circRNAs in the tissues and plasma of lung cancer patients. Chen et al. analyzed the circular RNA expression by high-throughput sequencing of plasmatic exosomes from patients with early-stage lung adenocarcinoma and detected 182 differentially expressed exosomal circRNAs, which included 105 upregulated and 78 downregulated compared with the controls (126). Using next-generation sequencing, Zhang et al. found that 35 circRNAs were aberrantly expressed in small-cell lung cancer tissue. Among these, five circRNAs were significantly upregulated, and 30 circRNAs were significantly downregulated (127). The roles of dysregulated circRNAs in lung cancer include proliferation, migration, invasion, apoptosis, cell cycle, stemness of lung cancer stem cell, chemotherapy resistance, tumor metabolism, tumor microenvironment (TME), and immune evasion of lung cancer cells. Here we discuss biological activities and pathogenic mechanisms of cirRNAs in lung cancer (Figure 2).

circRNAs regulate the proliferation and cell cycle of lung cancer cells

As the most important hallmarks of lung cancer, the proliferation of tumor cells accounts for over 50% of the cases (128). There are many evidence that circRNAs promote carcinogenesis and the progression of lung cancer, inducing cell cycle progression and proliferation through different mechanisms. Previous studies have demonstrated that endogenous F-circEA, which is derived from the EML4-ALK fusion gene, elevates the expression of F-circEA and promotes the migration and invasion ability of lung cells. F-circEA could be a novel liquid biopsy biomarker for the diagnosis of lung cancer because it is detectable in the plasma of patients with EML4-ALK translocation (129). In addition, circSATB2 is highly expressed in tumor cells and tissues and detectable in serum exosomes from patients with NSCLC. It acts as a sponge for miR-326, resulting in the upregulation of the actin-bundling protein 1 (FSCN1) expression and promoting the proliferation, migration, and invasion of NSCLC cells (124). It has been shown that circBIRC6 is upregulated in primary human NSCLC tissues and cells and, when knocked down, inhibits growth, proliferation, migration, and invasion of these tumor cells by sponging the tumor suppressor miR-145 (130). Circ NT5E promotes human NSCLC cell progression by sponging miR-134 (131), whereas circ-ARHGAP10 promotes human NSCLC cell progression *via* the miR-150-5p/GLUT1 axis (132). Circ-FOXM1 promotes the proliferation of NSCLC cells acting as sponge for miR-614, thus upregulating the expression of FAM83D (133). The high expression of Circ-0000326 in LUAC is correlated with tumor size, regional lymph node status, and differentiation. Mechanistic studies show that circ-0000326 acts as a miR-338-3p sponge and upregulates the expression of the downstream target RAB14, promoting the proliferation, migration, and apoptosis of LUAC cells (134). Circ PTPRA suppresses EMT and the metastasis of NSCLC cell lines by sponging miR-96-5p and upregulation of the downstream tumor suppressor RASSF8 (135). It has been shown that circ-0102231 is mainly localized to the cytoplasm, significantly upregulating the expression of RBBP4 (136) and promoting NSCLC cell proliferation and invasion by sponging miR-145. Circular RNA cMras inhibits LUAC progression *via* modulation

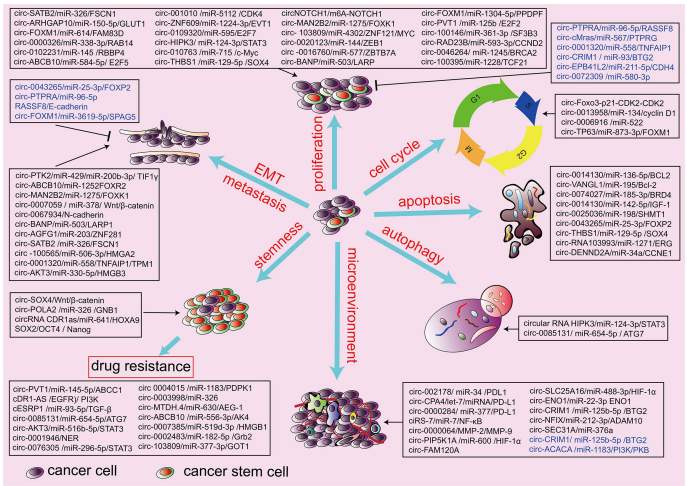


FIGURE 2
Biological function of circRNAs in the hallmarks of lung cancers. circRNAs are involved in the pathogenesis of lung cancers through the regulation of proliferation and invasion, cell cycle, cell autophagy, cell apoptosis, stemness, chemotherapy resistance, and tumor microenvironment. Black indicates an oncogenic role, and blue indicates a tumor-suppressive role.

HTTof the miR-567/PTPRG regulatory pathway (137). Circ-100565 modulates the expression of HMGA2 by sponging its regulatory Mir (miR-506-3p) (138). Reduction of circ-100565 significantly suppresses *in vitro* proliferation, migration, and invasion and the *in vivo* tumor growth of NSCLC cells, whereas in patients with NSCLC high levels of circ-100565 are associated with poor overall survival.

Circ-ABCB10 promotes NSCLC progression *via* modulating the miR-584-5p/E2F5 regulatory pathway (139). Cytoplasmic circ-0001320 is downregulated in lung cancer cells and inhibits the growth and invasion of lung cancer cells through the miR-558/TNFAIP1 and TPM1 pathways (126). circRNA-001010 is highly expressed in NSCLC patients and acts as a molecular sponge of miR-5112, leading to the increased expression of the oncogene CDK4 (127) with subsequent proliferation, migration, and invasion of NSCLC cells.

Knockdown of circ-0087862 significantly reduces NSCLC cell viability, migration, and invasion and enhances apoptosis, and a high expression of circ-0087862 is related with poor clinical outcome in NSCLC patients. It is hypothesized that circ-0087862 increases the expression of RAB3D by sponging miR-1253 (140). A decreased expression of Circ-CRIM1 in LUAC cancer is significantly correlated with lymphatic metastasis and more advanced TNM stage and is an independent negative biomarker for the overall survival of patients with LUAC. CircCRIM1 suppresses the invasion and metastasis of LUAC through sponging miR-93 and miR-182 and increases the expression of leukemia inhibitory factor receptor, a well-known tumor suppressor (141). circ-RNA EPB41L2 plays an inhibitory role in LUAC through regulating the miR-211-5p/CDH4 axis (142). Circ-ZNF609 is significantly upregulated in LUAC tissues and cell lines and promotes cell proliferation of LUAC cells through sponge miR-1224-3p to promote EVT1 expression (143). Circ-0072309 is lowly expressed in NSCLC tissues and cell lines and plays an inhibitory role in LUAC through blocking the expression of miR-580-3p (144). Overexpression of circ-11780 inhibits the proliferation, migration, and invasion of NSCLC cells *in vitro* and tumor growth *in vivo via* the miR-544a overexpression and reduced the protein concentration of F-Box and WD repeat domain containing 7 (FBXW7) (145).

Circ-0072088 acts as sponge for miR-377-5p, leading to the upregulation of NOVA2, and promotes the proliferation and metastasis of NSCLC cells (146). Circ-0109320 is significantly more expressed in NSCLC and associated with more advanced staging and lymph node metastasis of NSCLC *via* sponging of miR-595 and subsequently upregulates E2F7 expression (147). A novel circRNA (circHIPK3) increases STAT3 expression by inhibiting miR-124-3p in STK11 mutant lung cancer cells (148). Moreover, silencing circHIPK3 results in the reduction of cell proliferation, migration invasion, and promotion of macroautophagy/autophagy *via* MIR124-3p-STAT3-PRKAA/AMPKα signaling in STK11 mutant lung cancer (148). Overexpression of circAKT3 significantly promotes proliferation, migration, and invasion by sponging miR-330-5p, resulting in increased HMGB3 expression in non-small-cell lung cancer cells (149). circRNA-010763 promotes the growth and invasion of lung cancer through the regulation of the circRNA-010763/miR-715/c-Myc axis (150). Circ-THBS1 is highly expressed in patients with metastatic NSCLC and promotes the cell proliferation of LUAC cells by sponging miR-129-5p and SOX4 overexpression (151). There is also evidence that circ- 0012673 facilitates lung cancer cell proliferation and invasion *via* the miR-320a/LIMK18521axis (152). Several other regulatory cascades comprising circRNA, miRNA, and mRNA have been reported in lung cancer, including circ-MAN2B2/miR-1275/FOXK1 (153), circ-103809/miR-4302/ZNF121/MYC (154), circ-0020123/miR-144/ZEB1 and EZH2 (155), circ-FADS2/miR-498 (156), circ-0026134/miR-1256/miR-1287 (157), circ -0016760/miR-577/ZBTB7A (158), circ-0016760/miR-1287/GAGE1 (159), circ-BANP/miR-503/LARP (160), circ-001569/Wnt/β-catenin (161), circ -FOXM1/miR-1304-5p/PPDPF/MACC1 (162), Circ-FGFR3/miR-22-3p/Gal-1,p-AKT, p-ERK1/2 (163), circ-PVT1/miR-125b/E2F2 (164), circ-PVT1/miR-497/Bcl-2 (165), circ-0000735/miR-1179/miR-1182 (166), circ-100146/miR-361-3p/SF3B3 (167), circ-0043278/miR-520f/ROCK1, CDKN1B and AKT3 (168), circ-RAD23B/miR-593-3p/CCND2 and circ -RAD23B/miR-653-5p/TIAM1 (169), circ-0003645/miR-1179/TMEM14A (170), circ-P4HB/miR-133a-5p/vimentin (171), circ-0046264/miR-1245/BRCA2 (172), and circ- 100395/miR-1228/TCF21 (54).

Extensive evidence has demonstrated that circRNAs play a significant regulatory role in cell cycle transition. The dysregulation expression of circRNAs leads to tumor cell cycle progression.

circRNAs are involved in the regulation of G1/S checkpoint and participate in the development and progression of lung cancer. p21 is a cyclin-dependent kinase inhibitor that binds and inhibits the catalytic activity of Cdk2, leading to G1-phase cell cycle arrest (173). circ-Foxo3 binds to CDK2 and p21, giving origin to a circ-Foxo3-p21-CDK2 ternary complex that inhibits CDK2-dependent G1/S cell cycle progression (96). In addition, circ-0013958 acts as a sponge of miR-134, with subsequent upregulation of oncogenic cyclin D1 that plays a pivotal in the development of non-small-cell lung cancer (174). Circ-0006916 inhibits miR-522, promoting the G0/G1 progression of NSCLC (175). Circ-TP63 is highly expressed in LUSC tissues and correlates with a more advanced TNM stage. circ-TP63 competes for binding with miR-873-3p and then upregulates FOXM that, in turn, upregulates CENPA and CENP, thus promoting cell cycle progression (176).

circRNAs regulate the invasion and metastasis of lung cancer cells

Mounting evidence revealed that the aberrant expression of circRNAs is implicated in the invasion and metastasis of lung cancer —for example, circPTK2 acts a sponge for miR-429/miR-200b-3p, thus promoting TGF-β-induced EMT *via* TIF1γ and the invasiveness of NSCLC cell (15).

Circ-ABCB10 is also increased in NSCLC cell lines, and the knockdown of circ-ABCB10 suppresses NSCLC cell migration by promoting microRNA miR-1252 expression and suppressing Forkhead box 2 (FOXR2) (177). Circ-MAN2B2 promotes lung cancer cell invasion *via* the miR-1275/FOXK1 axis (153). The upregulated expression of circ-0067934 is associated with the invasiveness and migration of tumor cells and the overexpression of some EMT-associated markers such as N-cadherin (178). Circular RNA cESRP1 acts as a sponge for miR-93-5p and targets Smad 7/p21(CDKN1A), causing the inhibition of TGF-β-mediated EMT progress in lung cancer (179).

Circ-BANP promotes the migration, invasiveness, and increased expression of LARP1 in lung cancer cells *via* miR-93-5p and the inhibition of miR-503, respectively (160).

CircAGFG1 is elevated in NSCLC tissues and promotes invasion, migration, and epithelial–mesenchymal transition *via* the circ-AGFG1/miR-203/ZNF281 axis (180).

Circ-PTPRA suppresses EMT in NSCLC cell lines through the circ-PTPRA/miR-96-5p/RASSF8/E-cadherin axis and is downregulated in NSCLC tumor (135). Naringenin inhibits cell migration and invasion by regulating the circ-FOXM1/miR-3619-5p/SPAG5 axis in lung cancer (181). Several other regulatory cascades including circRNA, miRNA, and mRNA have been reported in lung cancer: circ-SATB2/miR-326/FSCN1 (124), circ-BIRC6/miR-145 (130), circ-NT5E/miR-134 (131), circ-ARHGAP10/miR-150-5p/GLUT1 (132), circ-FOXM1/miR-338-3p/RAB14 (134), circ-PTPRA/miR-96-5p/RASSF8 (135), circ-0102231/miR-145/RBBP4 (136), circ -100565/miR-506-3p/HMGA2 (138), circ-0001320/miR-558/TNFAIP1/TPM1 (126), circ-001010/miR-5112/CDK4 (126), circ-0087862/miR-1253/RAB3D (140), circ-CRIM1/miR-93 and miR-182 (141), circ-11780/miR-544a/FBXW7 (145), circ -0072088/miR-377-5p/NOVA2 (146), circ-0109320/miR-595/E2F7 (147), circ-HIPK3/miR124-3p-STAT3 (148), circ-AKT3/miR-330-5p/HMGB3 (149), circ- 010763miR-715/c-Myc (150), and circ-0012673/miR-320a/LIMK18521 axis (152).

circRNAs regulate the stemness and resistance of lung cancer stem cells

NSCLC is often a stubborn disease characterized by chemoresistance that represents a significant clinical challenge and contributes largely to disease progression, recurrence, and mortality (182). Despite the recent therapeutic advances (i.e., targeted therapy and immunotherapy), chemotherapy is still the backbone treatment for patients with advanced NSCLC cancer. Either cisplatin or carboplatin is used in combination with gemcitabine, vinorelbine, pemetrexed, or taxanes (docetaxel or paclitaxel) (183), while most patients with NSCLC are sensitive to chemotherapy at the early and the late stage. They show drug resistance that requires the identification of novel targets and development of more personalized medicine in the future (184). Despite intense efforts to overcome such resistance in lung cancer and other cancer types using novel agents (alone or in combination with chemo- and radiotherapy), the underlying mechanisms conferring this resistant phenotype in lung cancer remain largely unknown (185). It is now well established that CSCs constitute a unique subset of cells which are distinct from the bulk of tumor cells by their exclusive ability to perpetuate the growth of a malignant population of cells. This may explain the ineffectiveness of many conventional therapies and patient relapse (186). The ability of CSC to self-renew and differentiate results in tumor growth, progression, metastasis, and drug cancer treatment failure (187). This multi-drug resistance is mostly due to channel proteins that expel anticancer drugs, leading to decreased drug cell concentration (187). The stemness markers of lung cancer stem cells mainly include ALDH1, ABCG2, CD44, CD133, NANOG, and SRY-box transcription factor 2 (SOX2) (188). However, a recent study showed that circRNAs underpin cancer cell stemness by the upregulation of these stemness markers (189).On the contrary, circ-SOX4 suppresses cell proliferation, self-renewal, migration, and the invasiveness of lung tumor-initiating cells. Additionally, circ-SOX4 activates the Wnt/β-catenin pathway to maintain the stemness of lung cancer stem cells (190). Depletion of circPOLA that acts as a sponge for miR-326 with the subsequent upregulation of the G protein subunit beta 1 (GNB1) expression (56) leads to reduction of sphere formation ability, ALDH1 activity, and stemness marker expression of lung cancer cells. Consistently, the high expression of circPOLA in lung cancer tissues is associated with poor prognosis. CDR1as is upregulated in CDDP-resistant NSCLC cells, whereas its overexpression enhances the stemness signatures (SOX2, OCT4, and Nanog) *via* the miR-641/HOXA9 axis and confers resistance to CDDP-sensitive NSCLC cells (191).

Circ-PVT1 is highly expressed in LUAC cell lines and tissues and related to CDDP and MTA. Circ PVT1 mediates CDDP and MTA resistance *via* the miR-145-5p/ABCC1 axis, and Circ PVT1 knockout sensitizes tumor cells to CDDP and MTA (192). Mao et al. reported that CDR1-AS is also upregulated in cell lines and LUAC tissues and cell lines and confers resistance of patients to paclitaxel (PTX) and CDDP *via* the epidermal growth factor receptor (EGFR)/phosphatidylinositol 3-kinase (PI3K) signaling pathway (193).

cESRP1 is significantly downregulated in the chemoresistant cells and augments drug sensitivity by sponging miR-93-5p in SCLC, thereby upregulating CDKN1A, and subsequently

inhibits transforming growth factor-β-mediated epithelial–mesenchymal transition *via* miR-93-5p (179). Circ-0085131 acts as a ceRNA of miR-654-5p and upregulates the expression of autophagy-associated factor ATG7, thereby modulating cell chemoresistance. Circ-0085131 is more expressed in NSCLC tumor tissue than in the adjacent normal tissue, and the higher expression is associated with the recurrence and poorer survival of NSCLC (194). He and colleagues have shown that circ-0000567 and circ-0006867 are upregulated and downregulated, respectively, in two gefitinib-resistant cell lines. The Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway analysis indicates that dysregulation of circRNAs might play an important role in the development of acquired resistance to gefitinib in NSCLC (195). Circ AKT3 inhibits the cisplatin sensitivity of lung cancer cells *via* the miR-516b-5p/STAT3 axis-mediated glycolysis balance (196). Circ-AKT3 is upregulated in lung cancer tissues and cells, and knockdown of circAKT3 improves the cell sensitivity to CDDP and reduces glycolysis. Consistently, the inhibition of HIF-1α-dependent glycolysis attenuates the circAKT3-induced increase of chemoresistance in A549 cells. Circ-0001946 promotes the viability, proliferation, migration, and invasion of NSCLC cells and inhibits cell apoptosis. Circ-0001946 has also been proven to reduce the sensitivity of NSCLC cells to CDDP by regulating the nucleotide excision repair (NER) signaling pathway (197). Circ-0076305 is elevated in CDDP-resistant NSCLC tissues and cells and exerts this effect *via* targeting miR-296-5p and enhancing the expression of STAT3 (198). Circ 0004015 enhances the resistance of HCC827 to gefitinib by sponging miR-1183 and upregulating the expression of PDPK1 (199). Circ-0003998 targets miR-326 and inhibits the miR-326-mediated effect on chemosensitivity (200). It is highly expressed in LUAC tissues and docetaxel-resistant cell lines, and its depletion decreases chemoresistance, inhibits proliferation, and enhances apoptosis in docetaxel-resistant LUAC cells.

Circ MTDH.4 regulates AEG-1 expression *via* miR-630 and promotes chemoresistance and radio-resistance in NSCLC cells (201). Circ-ABCB10 increases the expression of AK4, promotes lung cancer progression, and sensitizes lung cancer cells to cisplatin *via* sponging miR-556-3p (202). Knockdown of circ-103762 promotes CHOP expression and inhibits multidrug resistance in NSCLC (203). Wang et al. have shown that circRNA 002178 enhances PDL1 expression and induces T cell exhaustion *via* targeting miR-34. Furthermore, circ-002178 is detectable in the exosomes of plasma from LUAC patients and is a potential biomarker for the early diagnosis of LUAC (125). Circ-FGFR1 enhances the expression of the motif chemokine receptor 4 (CXCR4) *via* miR-381-3p, thus promoting NSCLC progression and resistance to anti-programmed cell death 1 (PD-1)-based therapy (204).

Depletion of circ-0007385 suppresses cell proliferation, migration, and invasion in NSCLC cells and cisplatin (DDP) resistance. Circ-0007385 is highly expressed in NSCLC tissues and cell lines and associated with poor overall survival.

Circ-0007385 regulates HMGB1 expression and promotes the chemoresistance in NSCLC cells *via* sponging miR-519d-3p (205).

Circ-0002483 is downregulated in NSCLC cells, enhances the sensitivity of NSCLC cells by sponging miR-182-5p, and regulates its target gene growth factor receptor-bound protein2 (Grb2), forkhead box protein O1 (Foxo1), and forkhead box protein O3 (Foxo3) (57). One group investigated the expression profile of circRNAs involved in the development of early-stage lung adenocarcinoma and found that circRNA 404833 and circ 406483 might be exceptional potential candidates as early diagnostic biomarkers for lung cancer. Among them, circ-404833, through targeting miR-149-5p, regulates the cell motility and gefitinib resistance in lung adenocarcinoma (172). It has been shown that Circ-SMARCA5 is downregulated in non-small-cell lung cancer. Its low expression was negatively correlated with tumor size, lymph node metastasis, and TNM stage but positively correlated with disease-free survival and overall survival (OS). Over-expression inhibits the cell proliferation of NSCLC and enhances the chemosensitivity to cisplatin and gemcitabine (206). Overall, these studies illustrate that circRNAs have significant regulatory functions in the chemotherapy resistance of lung cancer.

circRNAs regulate the tumor metabolism and tumor microenvironment of lung cancer

Tumor metabolic reprogramming is a hallmark of cancer. In terms of energy metabolism, the glycolysis pathway often has abnormal activation in cancer cells (207). There is an increasing number of evidence that have been uncovered which show that circRNAs, through a different mechanism, can activate the glycolysis pathway and participate in the progression of cancer. One such study has shown that circ SLC25A16 can accelerate the glycolysis and proliferation of NSCLC cells. Mechanistic research demonstrated that circ-SLC25A16 can act as a sponge for miR-488-3p and elevate the expression of HIF-1α that is a target gene of miR-488-3p. In turn, HIF-1α activates lactate dehydrogenase A (LDHA) by facilitating its transcription (208). It was suggested that the overexpression of Circ-CRIM1 inhibited LUAC cell migration, invasion, EMT, glycolysis, and tumor growth through reducing the expression of miR-125b-5p and resulted in the enhance expression of BTG2 (142). The depletion of circ-ACACA inhibited the proliferation and migration of NSCLC cells and also reduced the glycolysis rate. The details of this molecular mechanism include circ-ACACA by its sponging of miR-1183 and inactivating the PI3K/PKB signaling pathway (209). Circ-ENO1, which is

derived from its host gene ENO1, was upregulated in LUAC. The overexpression of circ-ENO1 can promote glycolysis, proliferation, migration, and EMT and induce apoptosis in LUAC cells. circ-ENO1 acted as a ceRNA to interact with miR-22-3p, and an upregulated ENO1 expression promoted glycolysis and tumor progression in LUAC (58). Moreover, it was reported that Circ-CRIM1, through targeting miR-125b-5p, results in an increase of the expression of BTG2, leading to the promotion of glycolysis in lung adenocarcinoma cells (142). Circ-NFIX was highly expressed in NSCLC. The overexpression of circNFIX can promote NSCLC cell viability, migration, invasion, and glycolysis *in vitro* and hampered tumor growth *in vivo*. Mechanistic analyses have demonstrated that circ-NFIX acted as a molecular sponge of miR-212-3p and upregulated the expression of ADAM10 (210).

TME is the product of the crosstalk between different cell types and plays a crucial role in the progression, metastasis, and therapeutic treatment of cancer (211). Tumor immune escape refers to the phenomenon of tumor cells growing and metastasizing *via* various mechanisms to avoid recognition and attack by the immune system. The mechanism of tumor immune escape includes immunosuppression. Programmed death 1/programmed death-ligand 1 (PD-1/PD-L1), known as an immune checkpoint, is an important component of tumor immunosuppression (212). It has also been reported that circ-002178, from the exosomes of plasma from LUAC patients, was highly expressed in LUAC tissues and LUAC cancer cells. Functional studies have demonstrated that circ-002178 could enhance PDL1 expression *via* sponging miR-34 in cancer cells to induce T cell exhaustion (125). Circ-CPA4 was recently identified to be upregulated in NSCLC cells and cancer tissues. Compared to human bronchial epithelial cells and their paired clinical normal adjacent tissues, circ-CPA4 regulated cell growth, mobility, stemness, and drug resistance in NSCLC cells and inactivated CD8+ T cells in the tumor immune microenvironment through the let-7 miRNA/PD-L1 axis (59). Interestingly, a recent study showed that circ0000284 facilitated the progression of NSCLC by upregulating the PD-L1 expression as a competing endogenous RNA (ceRNA) of miR-377 (213). Therefore, regulating the PD-1/PD-L1 expression by targeting related circRNAs may be a direction of future immune checkpoint therapeutic research. Furthermore, some circRNAs regulate the cytokine expression to influence the function of immune cells—for example, ciRS-7 interacts with miR-7 to modulate the expression of NF-κB, modulating the activities of immune cells and affecting the development of lung cancer (214, 215). The non-cellular components of TME mainly include the extracellular matrix (ECM) and hypoxia environment. ECM, composed of matricellular proteins, elastin, collagen, and cytokines, provides the structural and functional bases for tumor cellular function (216). Circ-0000064 was reported to promote cell proliferation and inhibit apoptosis by the regulation of MMP-2 and MMP-9, participating in the destruction of the histological barrier as well as the invasion of cancer cells (217). The metabolism of cellular components in TME may result in a hypoxia environment, causing the generation of hypoxia-inducible factors (HIFs) to modulate the progression of tumor (218). A previous study has shown that upregulated circPIP5K1A inhibits miR-600 to enhance the HIF-1α expression to promote the migration and proliferation of NSCLC cells (219). Furthermore, circ-FAM120A expression was found to be downregulated in hypoxic LUAC, while HIF-1α expression increased at the same time, both promoting tumor growth (220). Therefore, circRNAs are identified to influence the hypoxia environment of TME in lung cancer.

The diagnostic and prognostic value of circRNAs

The high stability, abundance, and spatiotemporal-specific expression of circRNAs make them ideal biomarkers for liquid biopsy. The important functions of circRNAs in blood cells suggest that the dysregulation of circRNA expression in blood cells is likely to contribute to the occurrence and progression of lung cancers—for instance, in NSCLC patients, Tan et al. found that F-circEA, an f-circRNA originating from the EML4-ALK fusion gene, was exclusively expressed in the plasma of patients with the EML4-ALK fusion. Therefore, plasma F-circEA may serve as a liquid biopsy biomarker to diagnose NSCLC patients with EML4-ALK translocation and guide the targeted therapy for NSCLC patients in this subgroup (129). In a recent study, Luo et al. measured the expression levels of two circRNAs in the plasma samples of 231 lung cancer patients and 41 healthy controls using reverse transcription droplet digital PCR (221). They identified has circ-0000190 as a circRNA biomarker in human blood plasma that can predict the survival outcomes of lung cancer patients (221). Furthermore, the increased expression of circ-0000190 in plasma was also correlated with poor response to systemic therapy and immunotherapy in lung cancer patients (221). Similarly, Li et al. observed that SCLC patients with lower exosomal circFLI1 expression levels experienced longer disease remissions, indicating its prognostic power in SCLC. The authors also suggested that serum exosomal circFLI1 may be used as a biomarker that can monitor the clinical 191response to chemotherapy in SCLC patients (222). Notably, they observed that SCLC patients with lower exosomal circ-FLI1 expression levels experienced longer disease remissions, indicating its prognostic power in SCLC. The authors also suggested that serum exosomal circFLI1 may be used as a biomarker that can monitor the clinical response to chemotherapy in SCLC patients (222).

The main evaluative criterion of the diagnostic value of circRNAs is AUC in the ROC analysis. Previous studies have shown that many circRNAs may serve as potential diagnostic biomarkers. Circ 0000729 had an area under the receiver operating characteristic curve (AUC-ROC) of 0.815 for discriminating LUAC from normal controls (223). Circ-

0000064, being located in the cytoplasm circRNA, was reported to be upregulated in lung cancer tissues and lung cancer cell lines. Its aberrant expression was correlated with several clinical characteristics, including T stage, lymphatic metastasis, and TNM stage, which represents a novel potential biomarker for lung cancer diagnosis (217). It was suggested that circ-102231 was highly expressed in lung adenocarcinoma tissues and associated with advanced TNM stage, lymph node metastasis, and poor overall survival of lung cancer patients. The depletion of circ-102231 significantly restrains lung cancer cell proliferation and its invasion ability. It counts as a potential diagnostic biomarker for lung cancer patients. Its area under the ROC curve was 0.897, and circ 102231 showed good sensitivity and specificity of 81.2 and 88.7%, respectively (224). Circ-0014130 was elevated in NSCLC tissues. Its high expression correlated with TNM stage and lymphatic metastasis. The area under the receiver operating characteristic curve of circ-0014130 in NSCLC was 0.878, which showed good diagnostic potential. Furthermore, *via* gene oncology analysis and pathway analysis, it was indicated that circ-0014130 could participate in NSCLC development and could be used as a potential NSCLC diagnostic biomarker (225). Li et al. reported that the circ-0075930 expression levels were significantly higher in NSCLC cell lines and tissues. Its high expression was correlated with tumor size and lymph node metastasis. circ-0075930 had an AUC-ROC of 0.756 for discriminating NSCLC from normal controls, with sensitivity and specificity of 76.2 and 72.1%, respectively (226). Lu et al. reported that the circ0001715 levels were significantly higher in lung adenocarcinoma patients versus healthy controls. Its high expression significantly correlated with TNM stage and distant metastasis. Furthermore, the ROC curve analysis showed that the AUC of circ-0001715 was 0.871. The univariate and multivariate survival analyses showed that the plasma circ 0001715 level was an independent prognostic factor for the OS (227). The ROC curve indicated that the AUC of circ FOXO3 was 0.782, and the sensitivity and specificity of diagnosing NSLCL with circRNA-FOXO3 reached 80.0 and 73.3%, respectively (228).

circRNAs have stable circular structures which make them more stable and particularly attractive as liquid biopsy biomarkers. Circ-0013958 was elevated in lung adenocarcinoma tissues, cells, and plasma. Its high expression was associated with TNM stage and lymphatic metastasis. The area under the receiver operating characteristic curve was 0.815. In addition, circ-0013958 could be a sponge of miR-134 and thus upregulated oncogenic cyclin D1. It may be a potential non-invasive biomarker for the early detection and screening of lung adenocarcinoma (174). Circ-FARSA, derived from exon 5–7 of the FARSA gene, was elevated in cancerous tissues and patients' plasma. The elevated expression of circ-FARSA significantly promoted cell migration and invasion in NSCLC cells. Circ-FARSA could inhibit the activity of miR-330-5p and miR-326, thereby relieving their inhibitory effects on oncogene fatty acid synthase. A further analysis of the diagnostic value of plasma circ-FARSA in distinguishing NSCLC patients from non-cancer patients indicated the area under the ROC curve to be 0.71. This evidence indicated that plasma circFARSA could be a noninvasive biomarker for this NSCLC malignancy (229).

Circ-0067934 was shown to be markedly overexpressed in NSCLC tissues and cell lines. Its high expression was correlated with TNM stage, lymph node status, and distant metastasis in NSCLC patients. A higher expression of circ-0067934 results in significantly poorer survival, and the results of a multivariate Cox proportional hazard analysis indicated that circ-0067934 was an independent poor prognostic factor for patients with NSCLC. The depletion of circ-0067934 significantly suppressed NSCLC cell proliferation, migration, and invasion (178). Circ 100876 has been reported to be significantly elevated in NSCLC tissues than their adjacent nontumorous tissues. Its high expression correlated with lymph node metastasis, tumor stage, and poor overall survival in NSCLC. It might serve as a potential prognostic biomarker for NSCLC (230). It was suggested that circ-FOXM1 is highly expressed in NSCLC tissues and closely correlated with lymph node invasion, higher TNM stage, and unfavorable prognosis (162). The depletion of circ FOXM1 significantly restrains the growth, migration, and invasion of NSCLC cells. A further study has shown that Circ FOXM1 upregulated the expression of PPDPF and MACC1 *via* inhibiting the activity of miR-1304-5p (162). The above-mentioned examples demonstrated that some circRNAs could be promising biomarkers for the diagnosis and prognosis of lung cancer.

Lin et al. have shown that circ PRKCI was overexpressed in lung adenocarcinoma tissues and promoted the proliferation and tumorigenesis of lung adenocarcinoma. This group established a nude mouse xenograft model. Compared with the siRNA-transfected cell-derived tumors, the si-circPRKCI-transfected cell-derived tumors were smaller and lighter in weight (231). The therapeutic potential of circPRKCI was evaluated *via* an intratumoral injection of cholesterol-conjugated si-circPRKCI and control siRNA in patient-derived tumor xenografts (PDTX). The results indicated that si-circPRKCI significantly inhibited the growth of PDTX *in vivo*, suggesting that circPRKCI may be a promising therapeutic target for LUAC (231).

Conclusion and future direction

With the rapid development of high-throughput sequencing technologies and bioinformatics approaches, increasing evidence have uncovered that circRNAs play an important role in several human physiological and pathological processes, such as adipogenesis (43), lung inflammation (232), brain development (233), repair of ischemic muscle injury (234), mitochondrial metabolism (46), and cancer progression (235). Lung cancer is the leading form of cancer in terms of both morbidity and mortality.

Due to the lack of effective diagnostic markers, many lung cancer patients are approaching the advanced stage. Recently, with the widespread use of computed tomography and magnetic resonance imaging, incidental renal masses are increasingly being detected. However, there is an urgent need for specific biomarkers to allow for the early identification of postoperative lung cancer recurrence and metastasis. Screening and early diagnosis of lung cancer are critical for improving the treatment efficacy and reducing the mortality of patients with lung cancer and have been identified as a research priority. As a novel noncoding RNA, circRNA influences the initiation and progression of lung cancer through participation in diverse processes of lung cancer pathogenesis, including proliferation, migration, invasion, apoptosis, cell cycle, stemness of lung cancer stem cell, tumor metabolism of lung cancer, radiotherapy, and chemotherapy resistance. All of these findings suggest that circRNAs play pivotal roles in the pathological progression of cancer and may be useful as cancer biomarkers. circRNAs can be easily detected from a wide range of biological samples due to their high stability and specificity, representing as ideal diagnostic and prognostic biomarkers in lung cancer.

Previously identified lung CSC subsets have been shown to confer resistance to conventional chemotherapeutics, biological molecules, targeted therapies, and radiotherapy used in the current management of lung cancer. Specific targeting of cancer stem cells in combination with first-line chemotherapeutic agents holds great promise as a strategy to overcome chemoresistance, tumor relapse, and metastasis. A number of CSC markers have been identified and studied. These include ALDH1, ABCG2, CD44, CD133, NANOG, and SRY-box transcription factor 2 (SOX2), all of which have been linked to chemoresistance in a number of first line anti-cancer therapies—for example, axitinib is an oral, potent, small molecule ATP-competitive multitarget tyrosine kinase inhibitor. It inhibits the CSC marker ABCG2. Axitinib has also been shown to reverse multidrug resistance *via* ABCG2 inhibition both *in vitro* and *in vivo*. The axitinib–doxorubicin combination treatment promoted the intracellular accumulation of doxorubicin within the side population of CSC cells and significantly enhanced the cytotoxic effects of doxorubicin (236). In addition, a recent finding by our group was that, under normoxic conditions, YTHDF1 is highly expressed in non-small-cell lung cancer cancerous tissues and cell lines to promote cell proliferation *via* increasing cell-cycle-related factor expression. However, under hypoxic conditions or stressful chemotherapy conditions, YTHDF1 is downregulated, which leads to reduced Keap1 mRNA translational efficiency and Nrf2 protein stabilization (5).

With the deepening research on circRNAs, it is very important to choose appropriate and efficient database and analysis software for the precise analysis of circRNAs. We will also summarize recently developed database that have been developed to provide tremendous valuable information. CircBase mainly collect circRNA information of multiple species, including human, mouse, Caenorhabditis elegans, D. melanogaster, Latimeria chalumnae, and coelacanth (237). DeepBase, a database that integrate all public deep-sequencing data, provides an integrative evaluation of miRBase-annotated miRNA genes and other known ncRNAs and explores the expression patterns of miRNAs and other ncRNAs (238). CircNet database provides the following information related to circRNAs: novel circRNAs, integrated miRNA–target networks, expression profiles of circRNA isoforms, genomic annotations of circRNA isoforms, and sequences of circRNA isoforms. In addition, the CircNet database that provides tissue-specific circRNA expression profiles and circRNA–miRNAgene regulatory networks illustrates the regulation between circRNAs, miRNAs, and genes (239). circRNADb contains human circular RNAs with protein-coding annotations. The detailed information of the circRNA included genomic information, exon splicing, genome sequence, IRES, open reading frame (240). CircInteractome database is used for mapping RBP- and miRNA-binding sites on human circRNAs. circInteractome also carried out the following functions: identify potential circRNAs which can act as RBP sponges, design junction-spanning primers for the specific detection of circRNAs of interest, design siRNAs for circRNA silencing, and identify potential IRES (241). CSCD is a database for cancer-specific circular RNAs. It includes the following information: the microRNA response element sites and RNA-binding protein sites for each circRNA, the predicted potential open reading frames of circRNAs, and the splicing events in the linear transcripts of each circRNA (242). CIRCpedia v2 is a database for comprehensive circRNAs with expression features in various cell types/tissues, including disease samples. It also can perform a conservation analysis of circRNAs between humans and mice (243). TSCD is for Tissue-Specific CircRNA Database, containing information about tissue-specific circRNAs in human and mouse (244). ExoRBase is a database of circRNA, lncRNA, and mRNA derived from RNA-seq data analyses of human blood exosomes (245). circRNA Disease is a database that provides a user-friendly interface for searching disease-associated circRNAs (246). CircVAR is a database used to collect SNPs and small insertions and deletions in putative circRNA regions and to identify their potential phenotypic information (247). TransCirc is a specialized database that provides comprehensive evidences supporting the translation potential of circular RNAs (248). Circ2GO is a database that contains information on circRNAs and their function, processes, and miRNA targets (249). Although the listed databases have different advantages, there are still many problems, such as prediction overlap, lack of a unified identification standard, the results predicted by different databases that vary greatly, and the lack of a uniform identification standard. Therefore, when using these informatics methods, it is necessary to set specific conditions and improve the certainty of the data.

Exosomes can mediate cell-to-cell communication in both physiologic and pathologic processes. Exosomes contain RNAs, proteins, and lipids, all of which affect exosome functions and reflect cell characteristics. The major nucleic acids of exosomes are RNAs that contain microRNA (miRNA), tRNA, and long noncoding RNA (lncRNA) as well as circRNAs (250). Exosomes can serve as a

novel cellular communication bond by transferring their contents to target cells in a lung cancer microenvironment, thereby regulating lung cancer cell progression (251)—for example, under hypoxic conditions, miR-23a was significantly increased in exosomes from lung cancer cells, and exosomes transferred miR-23a to endothelial cells. miR-23a directly inhibited its target prolyl hydroxylase 1 and 2 (PHD1 and 2) expression, resulting in the accumulation of hypoxia-inducible factor-1α in endothelial cells. Consequently, exosomes derived from hypoxic lung cancer cells increased endothelial permeability and cancer cell transendothelial migration *in vitro* and enhanced neovascularization and tumor growth *in vivo* (252). In addition, tumor-derived exosomes carry immunosuppressive molecules, transfer these cargos to immune cells, and directly or indirectly suppress the functions of immune cells, thereby promoting tumor progression (253). circRNAs can be packaged into exosomes and transferred to receptor cells, further impacting the development of diseases. A recent study by Zhang N et al. reported that circSATB2 was highly expressed in NSCLC cancer exosomes, which can promote the proliferation, migration, and invasion of NLCSC cells. Exosomes which contained circ SATB2 can be taken up by NLCSC cells to participate in cell–cell communication, further affecting the progression of NSCLC cells and the proliferation of normal bronchial epithelial cells. Moreover, exosomal circSATB2 expression was related to lung cancer lymphatic metastasis. These results indicate that exosomal circSATB2 has the potential to act as a blood detection index for the diagnosis of lung cancer and lung cancer metastasis with high sensitivity and specificity (124). Similarly, it has also been reported that circRNA 002178 from the exosomes of plasma from LUAC patients was highly expressed in LUAC tissues and LUAC cancer cells. Functional studies have demonstrated that circRNA-002178 could enhance PDL1 expression *via* sponging miR-34 in cancer cells to induce T cell exhaustion. Interestingly, the area under the curve (AUCs) of circRNA-002178 was 0.9956, which was higher in the exosomes derived from cancer cells than the exosomes derived from normal human bronchial epithelial cells, and could serve as a novel diagnosis biomarker for lung cancer (125).

TME is the product of the crosstalk between different cell types and plays a crucial role in the progression, metastasis, and therapeutic treatment of cancer (211). The TME in both primary and secondary lung tumors is recognized as a target-rich environment for the development of novel anticancer agents. It has been reported that circRNA could modulate immune responses in the tumor microenvironment of lung cancer and be involved in lung cancer progression—for instance, ciRS-7 interacts with miR-7 to modulate the expression of NF-κB, modulating the activities of immune cells and affecting the development of lung cancer (214, 215). Mechanistic insights into the tumor-reprogrammed microenvironmental landscape in lung cancer together with the development of drugs that specifically inhibit the components of the landscape have ushered in a new era of cancer medicine.

Although there are many studies that have reported the biological function of RNA in lung cancer, there are still many problems surrounding circRNAs that remain to be clarified. First, the mechanism of expression and localization of circRNA is still unclear. Second, many new technologies, such as antisense oligonucleotide and CRISPR/Cas9-mediated knockout could be used to interfere the endogenous function of circRNA that promote cancer progression. Exosomes can transfer biologically active molecules between lung cancer cells and their microenvironment, but the mechanisms of exosome-related circRNA involved in lung cancer progression need to be illuminated. Finally, as circRNAs are stable and have unique structural conformations, additional investigations are still required to decipher the complex molecular mechanism so that circRNA could be implemented clinically through translational medicine, especially in the case of a disease such as lung cancer.

Author contributions

WR and YY contributed to the direction and guidance of this review and prepared the figure. JP, LM and XJ drafted the manuscript and revised the manuscript; All authors read and approved the final manuscript.

Funding

This work was supported by Yunnan Fundamental Research Projects (grant no. 202101AS070043), Kunming Science and Technology Plan Project (grant no. 2020-1-H-003), Kunming Medical University Graduate Student Innovation Fund (grant no. 2022B24), and The First People's Hospital of Yunnan Province Clinical Medicine Center Opening Project (grant no. 2022LCZXKF-XZ02).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA: A Cancer J Clin* (2020) 70(1):7–30.

2. Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. *Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship*. *Mayo Clinic Proc* (2008) 83(5):584–94. doi: 10.4065/83.5.584

3. Reck M, Rabe KF. Precision diagnosis and treatment for advanced non-Small-Cell lung cancer. *N Engl J Med* (2017) 377(9):849–61.

4. Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* (2012) 150(6):1121–34.

5. Shi Y, Fan S, Wu M, Zuo Z, Li X, Jiang L, et al. YTHDF1 links hypoxia adaptation and non-small cell lung cancer progression. *Nat Commun* (2019) 10 (1):4892.

6. Duruisseaux M, Esteller M. Lung cancer epigenetics: From knowledge to applications. *Semin Cancer Biol* (2018) 51:116–28.

7. Chen Y, Liu X, Li Y, Quan C, Zheng L, Huang K. *Lung cancer therapy targeting histone methylation: Opportunities and challenges*. *Comput Struct Biotechnol J* (2018) 16:211–23.

8. Hulbert A, Jusue-Torres I, Stark A, Chen C, Rodgers K, Lee B, et al. Early detection of lung cancer using DNA promoter hypermethylation in plasma and sputum. *Clin Cancer Res* (2017) 23(8):1998–2005.

9. Hua Q, Jin M, Mi B, Xu F, Li T, Zhao L, et al. LINC01123, a c-myc-activated long non-coding RNA, promotes proliferation and aerobic glycolysis of non-small cell lung cancer through miR-199a-5p/c-Myc axis. *J Hematol Oncol* (2019) 12 (1):91.

10. Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet* (2011) 12 (12):861–74.

11. Hombach S, Kretz M. Non-coding RNAs: Classification, biology and functioning. *Adv Exp Med Biol* (2016) 937:3–17.

12. Li X, Yang L, Chen LL. The biogenesis, functions, and challenges of circular RNAs. *Mol Cell* (2018) 71(3):428–42.

13. Chen Z, Li JL, Lin S, Cao C, Gimbrone NT, Yang R, et al. cAMP/CREB-regulated LINC00473 marks LKB1-inactivated lung cancer and mediates tumor growth. *J Clin Invest* (2016) 126(6):2267–79.

14. Iqbal MA, Arora S, Prakasam G, Calin GA, Syed MA. MicroRNA in lung cancer: role, mechanisms, pathways and therapeutic relevance. *Mol Aspects Med* (2019) 70:3–20.

15. Wang L, Tong X, Zhou Z, Wang S, Lei Z, Zhang T, et al. Circular RNA hsa_circ_0008305 (circPTK2) inhibits TGF-β-induced epithelial-mesenchymal transition and metastasis by controlling TIF1γ in non-small cell lung cancer. *Mol Cancer* (2018) 17(1):140. doi: 10.1186/s12943-018-0889-7

16. Huang RS, Zheng YL, Li C, Ding C, Xu C, Zhao J. MicroRNA-485-5p suppresses growth and metastasis in non-small cell lung cancer cells by targeting IGF2BP2. *Life Sci* (2018) 199:104–11.

17. Xie C, Han Y, Liu Y, Han L, Liu J. miRNA-124 down-regulates SOX8 expression and suppresses cell proliferation in non-small cell lung cancer. *Int J Clin Exp Pathol* (2014) 7(11):7518–26.

18. Zhao Y, Feng C, Li Y, Ma Y, Cai R. LncRNA H19 promotes lung cancer proliferation and metastasis by inhibiting miR-200a function. *Mol Cell Biochem* (2019) 460(1-2):1–8.

19. Zhang YX, Yuan J, Gao ZM, Zhang ZG. LncRNA TUC338 promotes invasion of lung cancer by activating MAPK pathway. *Eur Rev Med Pharmacol Sci* (2018) 22(2):443–9. doi: 10.26355/eurrev_201801_14193

20. Sanger HL, Klotz G, Riesner D, Gross HJ, Kleinschmidt AK. *Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures*. *Proc Natl Acad Sci U.S.A.* (1976) 73(11):3852–6.

21. Hsu MT, Coca-Prados M. Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells. *Nature* (1979) 280(5720):339–40.

22. Cocquerelle C, Mascrez B, Hétuin D, Bailleul B. Mis-splicing yields circular RNA molecules. *FASEB J* (1993) 7(1):155–60.

23. Capel B, Swain A, Nicolis S, Hacker A, Walter M, Koopman P, et al. Circular transcripts of the testis-determining gene sry in adult mouse testis. *Cell* (1993) 73 (5):1019–30.

24. Ivanov A, Memczak S, Wyler E, Torti F, Porath HT, Orejuela MR, et al. Analysis of intron sequences reveals hallmarks of circular RNA biogenesis in animals. *Cell Rep* (2015) 10(2):170–7.

25. Salzman J, Chen RE, Olsen MN, Wang PL, Brown PO. Cell-type specific features of circular RNA expression. *PloS Genet* (2013) 9(9):e1003777.

26. Wang PL, Bao Y, Yee MC, Barrett SP, Hogan GJ, Olsen MN, et al. Circular RNA is expressed across the eukaryotic tree of life. *PloS One* (2014) 9(6):e90859.

27. Zhang XO, Wang HB, Zhang Y, Lu X, Chen LL, Yang L. Complementary sequence-mediated exon circularization. *Cell* (2014) 159(1):134–47.

28. Li X, Liu CX, Xue W, Zhang Y, Jiang S, Yin QF, et al. Coordinated circRNA biogenesis and function with NF90/NF110 in viral infection. *Mol Cell* (2017) 67 (2):214–227.e7. doi: 10.1016/j.molcel.2017.05.023

29. Zhang Y, Zhang XO, Chen T, Xiang JF, Yin QF, Xing YH, et al. Circular intronic long noncoding RNAs. *Mol Cell* (2013) 51(6):792–806.

30. Memczak S, Jens M, Elefsinioti A, Torti F, Krueger J, Rybak A, et al. Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* (2013) 495 (7441):333–8.

31. Li Z, Huang C, Bao C, Chen L, Lin M, Wang X, et al. Exon-intron circular RNAs regulate transcription in the nucleus. *Nat Struct Mol Biol* (2015) 22(3):256–64.

32. Aufiero S, Reckman YJ, Pinto YM, Creemers EE. Circular RNAs open a new chapter in cardiovascular biology. *Nat Rev Cardiol* (2019) 16(8):503–14.

33. Suzuki H, Zuo Y, Wang J, Zhang MQ, Malhotra A, Mayeda A. Characterization of RNase r-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. *Nucleic Acids Res* (2006) 34(8):e63. doi: 10.1093/nar/gkl151

34. Yang Y, Gao X, Zhang M, Yan S, Sun C, Xiao F, et al. Novel role of FBXW7 circular RNA in repressing glioma tumorigenesis. *J Natl Cancer Inst* (2018) 110 (3):304–15.

35. Xu T, Wu J, Han P, Zhao Z, Song X. Circular RNA expression profiles and features in human tissues: a study using RNA-seq data. *BMC Genomics* (2017) 18 (Suppl 6):680.

36. Ashwal-Fluss R, Meyer M, Pamudurti NR, Ivanov A, Bartok O, Hanan M, et al. circRNA biogenesis competes with pre-mRNA splicing. *Mol Cell* (2014) 56 (1):55–66.

37. Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, et al. Natural RNA circles function as efficient microRNA sponges. *Nature* (2013) 495(7441):384–8. doi: 10.1038/nature11993

38. Ansell SM, Lesokhin AM, Borrello I, Halwani A, Scott EC, Gutierrez M, et al. PD-1 blockade with nivolumab in relapsed or refractory hodgkin's lymphoma. *N Engl J Med* (2015) 372(4):311–9.

39. Legnini I, Di Timoteo G, Rossi F, Morlando M, Briganti F, Sthandier O, et al. Circ-ZNF609 is a circular RNA that can be translated and functions in myogenesis. *Mol Cell* (2017) 66(1):22–37.e9.

40. Weigelt CM, Sehgal R, Tain LS, Cheng J, Eßer J, Pahl A, et al. An insulin-sensitive circular RNA that regulates lifespan in drosophila. *Mol Cell* (2020) 79 (2):268–279.e5.

41. Pandey PR, Yang JH, Tsitsipatis D, Panda AC, Noh JH, Kim KM, et al. circSamd4 represses myogenic transcriptional activity of PUR proteins. *Nucleic Acids Res* (2020) 48(7):3789–805.

42. Gao L, Chang S, Xia W, Wang X, Zhang C, Cheng L, et al. Circular RNAs from BOULE play conserved roles in protection against stress-induced fertility decline. *Sci Adv* (2020) 6(46).

43. Arcinas C, Tan W, Fang W, Desai TP, Teh DCS, Degirmenci U, et al. Adipose circular RNAs exhibit dynamic regulation in obesity and functional role in adipogenesis. *Nat Metab* (2019) 1(7):688–703.

44. You X, Vlatkovic I, Babic A, Will T, Epstein I, Tushev G, et al. Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nat Neurosci* (2015) 18(4):603–10.

45. Stoll L, Rodríguez-Trejo A, Guay C, Brozzi F, Bayazit MB, Gattesco S, et al. A circular RNA generated from an intron of the insulin gene controls insulin secretion. *Nat Commun* (2020) 11(1):5611.

46. Zhao Q, Liu J, Deng H, Ma R, Liao JY, Liang H, et al. Targeting mitochondria-located circRNA SCAR alleviates NASH via reducing mROS output. *Cell* (2020) 183(1):76–93.e22.

47. Hanan M, Soreq H, Kadener S. CircRNAs in the brain. *RNA Biol* (2017) 14 (8):1028–34.

48. Holdt LM, Stahringer A, Sass K, Pichler G, Kulak NA, Wilfert W, et al. Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans. *Nat Commun* (2016) 7:12429. doi: 10.1038/ncomms12429

49. Fang Y, Wang X, Li W, Han J, Jin J, Su F, et al. Screening of circular RNAs and validation of circANKRD36 associated with inflammation in patients with type 2 diabetes mellitus. *Int J Mol Med* (2018) 42(4):1865–74.

50. Hansen TB, Kjems J, Damgaard CK. Circular RNA and miR-7 in cancer. *Cancer Res* (2013) 73(18):5609–12.

51. Vo JN, Cieslik M, Zhang Y, Shukla S, Xiao L, Zhang Y, et al. The landscape of circular RNA in cancer. *Cell* (2019) 176(4):869–881.e13.

52. De Rubis G, Rajeev Krishnan S, Bebawy M. Liquid biopsies in cancer diagnosis, monitoring, and prognosis. *Trends Pharmacol Sci* (2019) 40(3):172–86.

53. Zhang Z, Yang T, Xiao J. Circular RNAs: Promising biomarkers for human diseases. *EBioMedicine* (2018) 34:267–74.

54. Chen D, Ma W, Ke Z, Xie F. CircRNA hsa_circ_100395 regulates miR-1228/TCF21 pathway to inhibit lung cancer progression. *Cell Cycle* (2018) 17 (16):2080–90.

55. Chen T, Yang Z, Liu C, Wang L, Yang J, Chen L, et al. Circ_0078767 suppresses non-small-cell lung cancer by protecting RASSF1A expression via sponging miR-330-3p. *Cell Prolif* (2019) 52(2):e12548.

56. Fan Z, Bai Y, Zhang Q, Qian P. CircRNA circ_POLA2 promotes lung cancer cell stemness via regulating the miR-326/GNB1 axis. *Environ Toxicol* (2020) 35 (10):1146–56.

57. Li X, Yang B, Ren H, Xiao T, Zhang L, Li L, et al. Hsa_circ_0002483 inhibited the progression and enhanced the taxol sensitivity of non-small cell lung cancer by targeting miR-182-5p. *Cell Death Dis* (2019) 10(12):953.

58. Zhou J, Zhang S, Chen Z, He Z, Xu Y, Li Z. CircRNA-ENO1 promoted glycolysis and tumor progression in lung adenocarcinoma through upregulating its host gene ENO1. *Cell Death Dis* (2019) 10(12):885.

59. Hong W, Xue M, Jiang J, Zhang Y, Gao X. Circular RNA circ-CPA4/ let-7 miRNA/PD-L1 axis regulates cell growth, stemness, drug resistance and immune evasion in non-small cell lung cancer (NSCLC). *J Exp Clin Cancer Res* (2020) 39 (1):149.

60. Chen LL. The biogenesis and emerging roles of circular RNAs. *Nat Rev Mol Cell Biol* (2016) 17(4):205–11.

61. Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, et al. Circular RNAs are abundant, conserved, and associated with ALU repeats. *Rna* (2013) 19 (2):141–57. doi: 10.1261/rna.035667.112

62. Jeck WR, Sharpless NE. Detecting and characterizing circular RNAs. *Nat Biotechnol* (2014) 32(5):453–61.

63. Barrett SP, Wang PL, Salzman J. Circular RNA biogenesis can proceed through an exon-containing lariat precursor. *Elife* (2015) 4:e07540.

64. Athanasiadis A, Rich A, Maas S. Widespread a-to-I RNA editing of alu-containing mRNAs in the human transcriptome. *PloS Biol* (2004) 2(12):e391.

65. Chen LL, Yang L. Regulation of circRNA biogenesis. *RNA Biol* (2015) 12 (4):381–8.

66. Kristensen LS, Andersen MS, Stagsted LVW, Ebbesen KK, Hansen TB, Kjems J. The biogenesis, biology and characterization of circular RNAs. *Nat Rev Genet* (2019) 20(11):675–91.

67. Zhang XO, Dong R, Zhang Y, Zhang JL, Luo Z, Zhang J, et al. Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Res* (2016) 26(9):1277–87.

68. Liang D, Wilusz JE. Short intronic repeat sequences facilitate circular RNA production. *Genes Dev* (2014) 28(20):2233–47. doi: 10.1101/gad.251926.114

69. Dubin RA, Kazmi MA, Ostrer H. *Inverted repeats are necessary for circularization of the mouse testis sry transcript. Gene* (1995) 167:245–8.

70. Guarnerio J, Bezzi M, Jeong JC, Paffenholz SV, Berry K, Naldini MM, et al. Oncogenic role of fusion-circRNAs derived from cancer-associated chromosomal translocations. *Cell* (2016) 166(4):1055–6.

71. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature* (2001) 409 (6822):860–921.

72. Dong R, Zhang XO, Zhang Y, Ma XK, Chen LL, Yang L. CircRNA-derived pseudogenes. *Cell Res* (2016) 26(6):747–50. doi: 10.1038/cr.2016.42

73. Liang D, Tatomer DC, Luo Z, Wu H, Yang L, Chen LL, et al. The output of protein-coding genes shifts to circular RNAs when the pre-mRNA processing machinery is limiting. *Mol Cell* (2017) 68(5):940–954.e3.

74. Wang M, Hou J, Müller-McNicoll M, Chen W, Schuman EM. *Long and repeat-rich intronic sequences favor circular RNA formation under conditions of reduced spliceosome activity. iScience* (2019) 20:237–47. doi: 10.1016/j.isci.2019.08.058

75. Aktaş T, Avşar Ilık İ, Maticzka D, Bhardwaj V, Pessoa Rodrigues C, Mittler G, et al. DHX9 suppresses RNA processing defects originating from the alu invasion of the human genome. *Nature* (2017) 544(7648):115–9. doi: 10.1038/nature21715

76. Fei T, Chen Y, Xiao T, Li W, Cato L, Zhang P, et al. Genome-wide CRISPR screen identifies HNRNPL as a prostate cancer dependency regulating RNA splicing. *Proc Natl Acad Sci United States America* (2017) 114(26):E5207–15.

77. Khan MA, Reckman YJ, Aufiero S, van den Hoogenhof MM, van der Made I, Beqqali A, et al. RBM20 regulates circular RNA production from the titin gene. *Circ Res* (2016) 119(9):996–1003. doi: 10.1161/CIRCRESAHA.116.309568

78. Bachmayr-Heyda A, Reiner AT, Auer K, Sukhbaatar N, Aust S, Bachleitner-Hofmann T, et al. Correlation of circular RNA abundance with proliferation–exemplified with colorectal and ovarian cancer, idiopathic lung fibrosis, and normal human tissues. *Sci Rep* (2015) 5:8057.

79. Guo JU, Agarwal V, Guo H, Bartel DP. *Expanded identification and characterization of mammalian circular RNAs. Genome Biol* (2014) 15(7):409. doi: 10.1186/s13059-014-0409-z

80. Circular RNAs in the mammalian brain are highly abundant, conserved, and dynamically expressed. *Mol Cell* (2015) 58(5):870–85. doi: 10.1016/j.molcel.2015.03.027

81. Jiang N, Meng X, Mi H, Chi Y, Li S, Jin Z, et al. Circulating lncRNA XLOC_009167 serves as a diagnostic biomarker to predict lung cancer. *Clin Chim Acta* (2018) 486:26–33. doi: 10.1038/s41467-019-12651-2

82. Huang C, Liang D, Tatomer DC, Wilusz JE. *A length-dependent evolutionarily conserved pathway controls nuclear export of circular RNAs. Genes Dev* (2018) 32:639–44.

83. Roundtree IA, Luo GZ, Zhang Z, Wang X, Zhou T, Cui Y, et al. YTHDC1 mediates nuclear export of n-methyladenosine methylated mRNAs. *eLife* (2017) 6.

84. Chen RX, Chen X, Xia LP, Zhang JX, Pan ZZ, Ma XD, et al. N-methyladenosine modification of circNSUN2 facilitates cytoplasmic export and stabilizes HMGA2 to promote colorectal liver metastasis. *Nat Commun* (2019) 10(1):4695.

85. Enuka Y, Lauriola M, Feldman ME, Sas-Chen A, Ulitsky I, Yarden Y, et al. *Circular RNAs are long-lived and display only minimal early alterations in response to a growth factor. Nucleic Acids Res* (2016) 44(3):1370–83. doi: 10.1093/nar/gkv1367

86. Hansen TB, Wiklund ED, Bramsen JB, Villadsen SB, Statham AL, Clark SJ, et al. miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J* (2011) 30(21):4414–22. doi: 10.1038/emboj.2011.359

87. Liu CX, Li X, Nan F, Jiang S, Gao X, Guo SK, et al. Structure and degradation of circular RNAs regulate PKR activation in innate immunity. *Cell* (2019) 177 (4):865–880.e21.

88. Guo Y, Wei X, Peng Y. *Structure-mediated degradation of CircRNAs. Trends Cell Biol* (2020) 30(7):501–3. doi: 10.1016/j.tcb.2020.04.001

89. Park OH, Ha H, Lee Y, Boo SH, Kwon DH, Song HK, et al. Endoribonucleolytic cleavage of mA-containing RNAs by RNase P/MRP complex. *Mol Cell* (2019) 74(3):494–507.e8. doi: 10.1016/j.molcel.2019.02.034

90. Thomson DW, Dinger ME. Endogenous microRNA sponges: evidence and controversy. *Nat Rev Genet* (2016) 17(5):272–83.

91. Zheng Q, Bao C, Guo W, Li S, Chen J, Chen B, et al. Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nat Commun* (2016) 7:11215. doi: 10.1038/ncomms11215

92. Kristensen LS, Okholm TLH, Venø MT, Kjems J. *Circular RNAs are abundantly expressed and upregulated during human epidermal stem cell differentiation. RNA Biol* (2018) 15(2):280–91.

93. Yu CY, Li TC, Wu YY, Yeh CH, Chiang W, Chuang CY, et al. The circular RNA circBIRC6 participates in the molecular circuitry controlling human pluripotency. *Nat Commun* (2017) 8(1):1149. doi: 1038/s41467-017-01216-w

94. Corley M, Burns MC, Yeo GW. How RNA-binding proteins interact with RNA: Molecules and mechanisms. *Mol Cell* (2020) 78(1):9–29. doi: 10.1016/j.molcel.2020.03.011

95. Du WW, Yang W, Chen Y, Wu ZK, Foster FS, Yang Z, et al. Foxo3 circular RNA promotes cardiac senescence by modulating multiple factors associated with stress and senescence responses. *Eur Heart J* (2017) 38(18):1402–12. doi: 10.1093/eurheartj/ehw001

96. Du WW, Yang W, Liu E, Yang Z, Dhaliwal P, Yang BB, et al. *Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. Nucleic Acids Res* (2016) 44(6):2846–58. doi: 10.1093/nar/gkw027

97. Du WW, Fang L, Yang W, Wu N, Awan FM, Yang Z, et al. Induction of tumor apoptosis through a circular RNA enhancing Foxo3 activity. *Cell Death differentiation* (2017) 24(2):357–70.

98. Li Q, Wang Y, Wu S, Zhou Z, Ding X, Shi R, et al. CircACC1 regulates assembly and activation of AMPK complex under metabolic stress. *Cell Metab* (2019) 30(1):157–173.e7. doi: 10.1016/j.cmet.2019.05.009

99. Abdelmohsen K, Panda AC, Munk R, Grammatikakis I, Dudekula DB, De S, et al. Identification of HuR target circular RNAs uncovers suppression of PABPN1 translation by CircPABPN1. *RNA Biol* (2017) 14(3):361–9. doi: 10.1080/15476286.2017.1279788

100. Zeng Y, Du WW, Wu Y, Yang Z, Awan FM, Li X, et al. A circular RNA binds to and activates AKT phosphorylation and nuclear localization reducing apoptosis and enhancing cardiac repair. *Theranostics* (2017) 7(16):3842–55.

101. Barbagallo D, Caponnetto A, Brex D, Mirabella F, Barbagallo C, Lauretta G, et al. CircSMARCA5 regulates VEGFA mRNA splicing and angiogenesis in glioblastoma multiforme through the binding of SRSF1. *Cancers* (2019) 11(2). doi: 10.3390/cancers11020194

102. Zhou Z, Jiang R, Yang X, Guo H, Fang S, Zhang Y, et al, et al. circRNA mediates silica-induced macrophage activation via HECTD1/ZC3H12A-dependent ubiquitination. *Theranostics* (2018) 8(2):575–92. doi: 10.7150/thno.21648

103. Zhu YJ, Zheng B, Luo GJ, Ma XK, Lu XY, Lin XM, et al. Circular RNAs negatively regulate cancer stem cells by physically binding FMRP against CCAR1 complex in hepatocellular carcinoma. *Theranostics* (2019) 9(12):3526–40. doi: 10.7150/thno.32796

104. Ma Y, Zhang C, Zhang B, Yu H, Yu Q. *circRNA of AR-suppressed PABPC1 91 bp enhances the cytotoxicity of natural killer cells against hepatocellular carcinoma via upregulating UL16 binding protein 1. Oncol Lett* (2019) 17(1):388–97. doi: 10.3892/ol.2018.9606

105. Shi L, Yan P, Liang Y, Sun Y, Shen J, Zhou S, et al. Circular RNA expression is suppressed by androgen receptor (AR)-regulated adenosine deaminase that acts on RNA (ADAR1) in human hepatocellular carcinoma. *Cell Death Dis* (2017) 8 (11):e3171. doi: 10.1038/cddis.2017.556

106. Chen N, Zhao G, Yan X, Lv Z, Yin H, Zhang S, et al. A novel FLI1 exonic circular RNA promotes metastasis in breast cancer by coordinately regulating TET1 and DNMT1. *Genome Biol* (2018) 19(1):218. doi: 10.1186/s13059-018-1594-y

107. Liu Y, Dong Y, Zhao L, Su L, Luo J. *Circular RNA–MTO1 suppresses breast cancer cell viability and reverses monastrol resistance through regulating the TRAF4/Eg5 axis. Int J Oncol* (2018) 53(4):1752–62.

108. Du WW, Yang W, Li X, Awan FM, Yang Z, Fang L, et al. A circular RNA circ-DNMT1 enhances breast cancer progression by activating autophagy. *Oncogene* (2018) 37(44):5829–42. doi: 10.1038/s41388-018-0369-y

109. Shen S, Yao T, Xu Y, Zhang D, Fan S, Ma J. CircECE1 activates energy metabolism in osteosarcoma by stabilizing c-myc. *Mol Cancer* (2020) 19(1):151.

110. Di Timoteo G, Dattilo D, Centrón-Broco A, Colantoni A, Guarnacci M, Rossi F, et al. Modulation of circRNA metabolism by m(6)A modification. *Cell Rep* (2020) 31(6):107641.

111. Zhang M, Zhao K, Xu X, Yang Y, Yan S, Wei P, et al. A peptide encoded by circular form of LINC-PINT suppresses oncogenic transcriptional elongation in glioblastoma. *Nat Commun* (2018) 9(1):4475.

112. Zhang M, Huang N, Yang X, Luo J, Yan S, Xiao F, et al. A novel protein encoded by the circular form of the SHPRH gene suppresses glioma tumorigenesis. *Oncogene* (2018) 37(13):1805–14.

113. Liang WC, Wong CW, Liang PP, Shi M, Cao Y, Rao ST, et al. Translation of the circular RNA circβ-catenin promotes liver cancer cell growth through activation of the wnt pathway. *Genome Biol* (2019) 20(1):84.

114. Zheng X, Chen L, Zhou Y, Wang Q, Zheng Z, Xu B, et al. A novel protein encoded by a circular RNA circPPP1R12A promotes tumor pathogenesis and metastasis of colon cancer via hippo-YAP signaling. *Mol Cancer* (2019) 18(1):47.

115. Xia X, Li X, Li F, Wu X, Zhang M, Zhou H, et al. A novel tumor suppressor protein encoded by circular AKT3 RNA inhibits glioblastoma tumorigenicity by competing with active phosphoinositide-dependent kinase-1. *Mol Cancer* (2019) 18(1):131. doi: 10.1186/s12943-019-1056-5

116. Zhao J, Lee EE, Kim J, Yang R, Chamseddin B, Ni C, et al. Transforming activity of an oncoprotein-encoding circular RNA from human papillomavirus. *Nat Commun* (2019) 10(1):2300.

117. Zhi X, Zhang J, Cheng Z, Bian L, Qin J. *circLgr4 drives colorectal tumorigenesis and invasion through Lgr4-targeting peptide. Int J Cancer* (2019).

118. Pan Z, Cai J, Lin J, Zhou H, Peng J, Liang J, et al. A novel protein encoded by circFNDC3B inhibits tumor progression and EMT through regulating snail in colon cancer. *Mol Cancer* (2020) 19(1):71. doi: 10.1186/s12943-020-01179-5

119. Wang Y, Liu J, Ma J, Sun T, Zhou Q, Wang W, et al. Exosomal circRNAs: biogenesis, effect and application in human diseases. *Mol Cancer* (2019) 18(1):116. doi: 10.1186/s12943-019-1041-z

120. Pegtel DM, Gould SJ. Exosomes. *Annu Rev Biochem* (2019) 88:487–514.

121. Qu L, Ding J, Chen C, Wu ZJ, Liu B, Gao Y, et al. Exosome-transmitted lncARSR promotes sunitinib resistance in renal cancer by acting as a competing endogenous RNA. *Cancer Cell* (2016) 29(5):653–68. doi: 10.1016/j.ccell.2016.03.004

122. Wang X, Luo G, Zhang K, Cao J, Huang C, Jiang T, et al. Hypoxic tumor-derived exosomal miR-301a mediates M2 macrophage polarization via PTEN/PI3Kγ to promote pancreatic cancer metastasis. *Cancer Res* (2018) 78 (16):4586–98.

123. Chen G, Huang AC, Zhang W, Zhang G, Wu M, Xu W, et al. Exosomal PD-L1 contributes to immunosuppression and is associated with anti-PD-1 response. *Nature* (2018) 560(7718):382–6. doi: 10.1158/0008-5472.CAN-17-3841

124. Zhang N, Nan A, Chen L, Li X, Jia Y, Qiu M, et al. Circular RNA circSATB2 promotes progression of non-small cell lung cancer cells. *Mol Cancer* (2020) 19(1):101. doi: 10.1038/s41419-020-2230-9

125. Wang J, Zhao X, Wang Y, Ren F, Sun D, Yan Y, et al. circRNA-002178 act as a ceRNA to promote PDL1/PD1 expression in lung adenocarcinoma. *Cell Death Dis* (2020) 11(1):32.

126. Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJJr., Wu YL, et al. Lung cancer: current therapies and new targeted treatments. *Lancet* (2017) 389(10066):299–311.

127. Tan S, Gou Q, Pu W, Guo C, Yang Y, Wu K, et al. Circular RNA f-circEA produced from EML4-ALK fusion gene as a novel liquid biopsy biomarker for non-small cell lung cancer. *Cell Res* (2018) 28(6):693–5. doi: 10.1038/s41422-018-0033-7

128. Thompson JC, Hwang WT, Davis C, Deshpande C, Jeffries S, Rajpurohit Y, et al. Gene signatures of tumor inflammation and epithelial-to-mesenchymal transition (EMT) predict responses to immune checkpoint blockade in lung cancer with high accuracy. *Lung Cancer* (2020) 139:1–8. doi: 10.1016/j.lungcan.2019.10.012

129. Yuan Z, Yuan Z, Hasnat M, Zhang H, Liang P, Sun L, et al. *A new perspective of triptolide-associated hepatotoxicity: the relevance of NF- κb and NF- κb-mediated cellular FLICE-inhibitory protein. Acta Pharm Sin B* (2020) 10(5):861–77. doi: 10.1016/j.apsb.2020.02.009

130. Jin M, Shi C, Yang C, Liu J, Huang G. *Upregulated circRNA ARHGAP10 predicts an unfavorable prognosis in NSCLC through regulation of the miR-150-5p/GLUT-1 axis. Mol Ther Nucleic Acids* (2019) 18:219–31.

131. Yu C, Cheng Z, Cui S, Mao X, Li B, Fu Y, et al. circFOXM1 promotes proliferation of non-small cell lung carcinoma cells by acting as a ceRNA to upregulate FAM83D. *J Exp Clin Cancer Res CR* (2020) 39(1):55.

132. Xiong Q, Liu B, Ding M, Zhou J, Yang C, Chen Y. Hypoxia and cancer related pathology. *Cancer Lett* (2020) 486:1–7.

133. Cao X, Li F, Shao J, Lv J, Chang A, Dong W, et al. Circular RNA hsa_circ_0102231 sponges miR-145 to promote non-small cell lung cancer cell proliferation by up-regulating the expression of RBBP4. *J Biochem* (2020). doi: 10.1093/jb/mvaa093

134. Yu C, Tian F, Liu J, Su M, Wu M, Zhu X, et al. Circular RNA cMras inhibits lung adenocarcinoma progression via modulating miR-567/PTPRG regulatory pathway. *Cell proliferation* (2019) 52(3):e12610.

135. Li L, Wei H, Zhang H, Xu F, Che G. *Circ_100565 promotes proliferation, migration and invasion in non-small cell lung cancer through upregulating HMGA2 via sponging miR-506-3p. Cancer Cell Int* (2020) 20:160.

136. Ma D, Qin Y, Huang C, Chen Y, Han Z, Zhou X, et al. Circular RNA ABCB10 promotes non-small cell lung cancer progression by increasing E2F5 expression through sponging miR-584-5p. *Cell Cycle* (2020), 1–10. doi: 10.1080/15384101.2020.1761617

137. Mao Y, He JX, Zhu M, Dong YQ, He JX. Circ0001320 inhibits lung cancer cell growth and invasion by regulating TNFAIP1 and TPM1 expression through sponging miR-558. *Hum Cell* (2020). doi: 10.1007/s13577-020-00453-4

138. Wang Q, Kang PM. CircRNA_001010 adsorbs miR-5112 in a sponge form to promote proliferation and metastasis of non-small cell lung cancer (NSCLC). *Eur Rev Med Pharmacol Sci* (2020) 24(8):4271–80. doi: 10.26355/eurrev_202004_21007

139. Li L, Wan K, Xiong L, Liang S, Tou F, Guo S, et al. *CircRNA hsa_circ_0087862 acts as an oncogene in non-small cell lung cancer by targeting miR-1253/RAB3D axis.* OncoTargets and therapy. (2020) 13:2873–86.

140. Wang Liang L, Mao Y, Xia Q, Chen W, Shen B, H, et al. Circular RNA circCRIM1 inhibits invasion and metastasis in lung adenocarcinoma through the microRNA (miR)-182/miR-93-leukemia inhibitory factor receptor pathway. *Cancer Sci* (2019) 110(9):2960–72.

141. Zuo Y, Shen W, Wang C, Niu N, Pu J. *Circular RNA circ-ZNF609 promotes lung adenocarcinoma proliferation by modulating miR-1224-3p/ETV1 signaling. Cancer Manage Res* (2020) 12:2471–9.

142. Ruiz-Cordero R, Devine WP. Targeted therapy and checkpoint immunotherapy in lung cancer. *Surg Pathol Clin* (2020) 13(1):17–33.

143. Liu Y, Yang C, Cao C, Li Q, Jin X, Shi H, et al. *Hsa_circ_RNA_0011780 represses the proliferation and metastasis of non-small cell lung cancer by decreasing FBXW7 via targeting miR-544a. OncoTargets Ther* (2020) 13:745–55.

144. Tan Z, Cao F, Jia B, Xia L. *Circ_0072088 promotes the development of non-small cell lung cancer via the miR-377-5p/NOVA2 axis. Thorac Cancer* (2020).

145. Bai Q, Li L, Chen F, Zhu J, Cao L, Yang Y, et al. Suppression of circular RNA Hsa_circ_0109320 attenuates non-small cell lung cancer progression via MiR-595/E2F7 axis. *Med Sci monitor Int Med J Exp Clin Res* (2020) 26:e921200.

146. Chen X, Mao R, Su W, Yang X, Geng Q, Guo C, et al. Circular RNA circHIPK3 modulates autophagy via MIR124-3p-STAT3-PRKAA/AMPKα signaling in STK11 mutant lung cancer. *Autophagy* (2020) 16(4):659–71.

147. Zhou ZF, Wei Z, Yao JC, Liu SY, Wang F, Wang Z, et al. CircRNA_102179 promotes the proliferation, migration and invasion in non-small cell lung cancer cells by regulating miR-330-5p/HMGB3 axis. *Pathol Res Pract* (2020) 216 (11):153144.

148. Zhang P, Xue XF, Ling XY, Yang Q, Yu Y, Xiao J, et al. CircRNA_010763 promotes growth and invasion of lung cancer through serving as a molecular sponge of miR-715 to induce c-myc expression. *Eur Rev Med Pharmacol Sci* (2020) 24(13):7310–9.

149. Chen H, Wu G, Chen G, Wang W, Ruan F. Circular RNA THBS1 promotes proliferation and apoptosis of non-small cell lung cancer cells by sponging miR-129-5p and regulating SOX4 expression. *J buon* (2020) 25 (4):1721–7.

150. Qin H, Liu J, Du ZH, Hu R, Yu YK, Wang QA. Circular RNA hsa_circ_0012673 facilitates lung cancer cell proliferation and invasion via miR-320a/LIMK18521 axis. *Eur Rev Med Pharmacol Sci* (2020) 24(4):1841–52.

151. Ma X, Yang X, Bao W, Li S, Liang S, Sun Y, et al. Circular RNA circMAN2B2 facilitates lung cancer cell proliferation and invasion via miR-1275/FOXK1 axis. *Biochem Biophys Res Commun* (2018) 498(4):1009–15.

152. Liu W, Ma W, Yuan Y, Zhang Y, Sun S. Circular RNA hsa_circRNA_103809 promotes lung cancer progression via facilitating ZNF121-dependent MYC expression by sequestering miR-4302. *Biochem Biophys Res Commun* (2018) 500(4):846–51.

153. Qu D, Yan B, Xin R, Ma T. A novel circular RNA hsa_circ_0020123 exerts oncogenic properties through suppression of miR-144 in non-small cell lung cancer. *Am J Cancer Res* (2018) 8(8):1387–402.

154. Zhao F, Han Y, Liu Z, Zhao Z, Li Z, Jia K. circFADS2 regulates lung cancer cells proliferation and invasion via acting as a sponge of miR-498. *Biosci Rep* (2018) 38(4).

155. Chang H, Qu J, Wang J, Liang X, Sun W. Circular RNA circ_0026134 regulates non-small cell lung cancer cell proliferation and invasion via sponging miR-1256 and miR-1287. *BioMed Pharmacother* (2019) 112:108743.

156. Hao Y, Xi J, Peng Y, Bian B, Hao G, Xi Y, et al. Circular RNA Circ_0016760 modulates non-Small-Cell lung cancer growth through the miR-577/ZBTB7A axis. *Cancer Manag Res* (2020) 12:5561–74.

157. Li Y, Hu J, Li L, Cai S, Zhang H, Zhu X, et al. Upregulated circular RNA circ_0016760 indicates unfavorable prognosis in NSCLC and promotes cell progression through miR-1287/GAGE1 axis. *Biochem Biophys Res Commun* (2018) 503(3):2089–94.

158. Han J, Zhao G, Ma X, Dong Q, Zhang H, Wang Y, et al. CircRNA circ-BANP-mediated miR-503/LARP1 signaling contributes to lung cancer progression. *Biochem Biophys Res Commun* (2018) 503(4):2429–35. doi: 10.1016/j.bbrc.2018.06.172

159. Ding L, Yao W, Lu J, Gong J, Zhang X. Upregulation of circ_001569 predicts poor prognosis and promotes cell proliferation in non-small cell lung cancer by regulating the wnt/β-catenin pathway. *Oncol Lett* (2018) 16(1):453–8. doi: 10.3892/ol.2018.8673

160. Liu G, Shi H, Deng L, Zheng H, Kong W, Wen X, et al. Circular RNA circ-FOXM1 facilitates cell progression as ceRNA to target PPDPF and MACC1 by sponging miR-1304-5p in non-small cell lung cancer. *Biochem Biophys Res Commun* (2019) 513(1):207–12.

161. Qiu BQ, Zhang PF, Xiong D, Xu JJ, Long X, Zhu SQ, et al. CircRNA fibroblast growth factor receptor 3 promotes tumor progression in non-small cell lung cancer by regulating galectin-1-AKT/ERK1/2 signaling. *J Cell Physiol* (2019) 234(7):11256–64.

162. Li X, Zhang Z, Jiang H, Li Q, Wang R, Pan H, et al. Circular RNA circPVT1 promotes proliferation and invasion through sponging miR-125b and activating E2F2 signaling in non-small cell lung cancer. *Cell Physiol Biochem* (2018) 51 (5):2324–40.

163. Qin S, Zhao Y, Lim G, Lin H, Zhang X, Zhang X. Circular RNA PVT1 acts as a competing endogenous RNA for miR-497 in promoting non-small cell lung cancer progression. *BioMed Pharmacother* (2019) 111:244–50. doi: 10.1016/j.biopha.2018.12.007

164. Li W, Jiang W, Liu T, Lv J, Guan J. Enhanced expression of circ_0000735 forecasts clinical severity in NSCLC and promotes cell progression via sponging miR-1179 and miR-1182. *Biochem Biophys Res Commun* (2019) 510(3):467–71.

165. Chen L, Nan A, Zhang N, Jia Y, Li X, Ling Y, et al. Circular RNA 100146 functions as an oncogene through direct binding to miR-361-3p and miR-615-5p in non-small cell lung cancer. *Mol Cancer* (2019) 18(1):13.

166. Cui J, Li W, Liu G, Chen X, Gao X, Lu H, et al. A novel circular RNA, hsa_circ_0043278, acts as a potential biomarker and promotes non-small cell lung cancer cell proliferation and migration by regulating miR-520f. *Artif Cells Nanomed Biotechnol* (2019) 47(1):810–21.

167. Han W, Wang L, Zhang L, Wang Y, Li Y. Circular RNA circ-RAD23B promotes cell growth and invasion by miR-593-3p/CCND2 and miR-653-5p/TIAM1 pathways in non-small cell lung cancer. *Biochem Biophys Res Commun* (2019) 510(3):462–6.

168. An J, Shi H, Zhang N, Song S. Elevation of circular RNA circ_0003645 forecasts unfavorable prognosis and facilitates cell progression via miR-1179/TMEM14A pathway in non-small cell lung cancer. *Biochem Biophys Res Commun* (2019) 511(4):921–5.

169. Wang T, Wang X, Du Q, Wu N, Liu X, Chen Y, et al. The circRNA circP4HB promotes NSCLC aggressiveness and metastasis by sponging miR-133a-5p. *Biochem Biophys Res Commun* (2019) 513(4):904–11.

170. Yang L, Wang J, Fan Y, Yu K, Jiao B, Su X. Hsa_circ_0046264 up-regulated BRCA2 to suppress lung cancer through targeting hsa-miR-1245. *Respir Res* (2018) 19(1):115.

171. Abbas T, Dutta A. p21 in cancer: intricate networks and multiple activities. *Nat Rev Cancer* (2009) 9(6):400–14.

172. Zhu X, Wang X, Wei S, Chen Y, C. Y, Fan X, et al. hsa_circ_0013958: a circular RNA and potential novel biomarker for lung adenocarcinoma. *FEBS J* (2017) 284(14):2170–82. doi: 10.1111/febs.14132

173. Dai X, Zhang N, Cheng Y, Yang T, Chen Y, Liu Z, et al. RNA-Binding protein trinucleotide repeat-containing 6A regulates the formation of circular RNA circ0006916, with important functions in lung cancer cells. *Carcinogenesis* (2018) 39(8):981–92. doi: 10.1093/carcin/bgy061

174. Cheng Z, Yu C, Cui S, Wang H, Jin H, Wang C, et al. circTP63 functions as a ceRNA to promote lung squamous cell carcinoma progression by upregulating FOXM1. *Nat Commun* (2019) 10(1):3200.

175. Tian X, Zhang L, Jiao Y, Chen J, Shan Y, Yang W. CircABCB10 promotes nonsmall cell lung cancer cell proliferation and migration by regulating the miR-1252/FOXR2 axis. *J Cell Biochem* (2019) 120(3):3765–72. doi: 10.1002/jcb.27657

176. Wang J, Li H. CircRNA circ_0067934 silencing inhibits the proliferation, migration and invasion of NSCLC cells and correlates with unfavorable prognosis in NSCLC. *Eur Rev Med Pharmacol Sci* (2018) 22(10):3053–60.

177. Huang W, Yang Y, Wu J, Niu Y, Yao Y, Zhang J, et al. Circular RNA cESRP1 sensitises small cell lung cancer cells to chemotherapy by sponging miR-93-5p to inhibit TGF-β signalling. *Cell Death Differ* (2020) 27(5):1709–27. doi: 10.1038/s41418-019-0455-x

178. Xue YB, Ding MQ, Xue L, Luo JH. CircAGFG1 sponges miR-203 to promote EMT and metastasis of non-small-cell lung cancer by upregulating ZNF281 expression. *Thorac Cancer* (2019) 10(8):1692–701.

179. Wei S, Zheng Y, Jiang Y, Li X, Geng J, Shen Y, et al. The circRNA circPTPRA suppresses epithelial-mesenchymal transitioning and metastasis of NSCLC cells by sponging miR-96-5p. *EBioMedicine* (2019) 44:182–93. doi: 10.1016/j.ebiom.2019.05.032

180. Tan Z, Sun Y, Liu M, Xia L, Cao F, Qi Y, et al. Naringenin inhibits cell migration, invasion, and tumor growth by regulating circFOXM1/miR-3619-5p/SPAG5 axis in lung cancer. *Cancer Biother Radiopharm* (2020). doi: 10.1089/cbr.2019.3520

181. Sarkar FH, Li Y, Wang Z, Kong D, Ali S. Implication of microRNAs in drug resistance for designing novel cancer therapy. *Drug Resist Update* (2010) 13(3):57–66. doi: 10.1016/j.drup.2010.02.001

182. Suresh R, Ali S, Ahmad A, Philip PA, Sarkar FH. The role of cancer stem cells in recurrent and drug-resistant lung cancer. *Adv Exp Med Biol* (2016) 890:57–74.

183. Gower A, Wang Y, Giaccone G. Oncogenic drivers, targeted therapies, and acquired resistance in non-small-cell lung cancer. *J Mol Med (Berl)* (2014) 92 (7):697–707.

184. Szakács G, Paterson JK, Ludwig JA, Booth-Genthe C, Gottesman MM. Targeting multidrug resistance in cancer. *Nat Rev Drug Discovery* (2006) 5 (3):219–34.

185. Nguyen LV, Vanner R, Dirks P, Eaves CJ. Cancer stem cells: an evolving concept. *Nat Rev Cancer* (2012) 12(2):133–43.

186. Yadav AK, Desai NS. Cancer stem cells: Acquisition, characteristics, therapeutic implications, targeting strategies and future prospects. *Stem Cell Rev Rep* (2019) 15(3):331–55. doi: :10.1007/s12015-019-09887-2

187. Moro M, Bertolini G, Pastorino U, Roz L, Sozzi G. Combination treatment with all-trans retinoic acid prevents cisplatin-induced enrichment of CD133+ tumor-initiating cells and reveals heterogeneity of cancer stem cell compartment in lung cancer. *J Thorac Oncol* (2015) 10(7):1027–36.

188. Szabo L, Morey R, Palpant NJ, Wang PL, Afari N, Jiang C, et al. Statistically based splicing detection reveals neural enrichment and tissue-specific induction of circular RNA during human fetal development. *Genome Biol* (2015) 16:126.

189. Wang L, Zheng C, Wu X, Zhang Y, Yan S, Ruan L, et al. Circ-SOX4 promotes non-small cell lung cancer progression by activating the wnt/β-catenin pathway. *Mol Oncol* (2020). doi: 10.1002/1878-0261.12656

190. Zhao Y, Zheng R, Chen J, Ning D. CircRNA CDR1as/miR-641/HOXA9 pathway regulated stemness contributes to cisplatin resistance in non-small cell lung cancer (NSCLC). *Cancer Cell Int* (2020) 20:289.

191. Zheng F, Xu R. CircPVT1 contributes to chemotherapy resistance of lung adenocarcinoma through miR-145-5p/ABCC1 axis. *BioMed Pharmacother* (2020) 124:109828.

192. Mao Y, Xu R. Circular RNA CDR1-AS contributes to pemetrexed and cisplatin chemoresistance through EGFR/PI3K signaling pathway in lung adenocarcinoma. *BioMed Pharmacother* (2020) 123:109771. doi: 10.1016/j.biopha.2019.109771

193. R K. Circular RNA hsa_circ_0085131 is involved in cisplatin-resistance of non-small cell lung cancer cells by regulating autophagy. *Cell Biol Int* (2020).

194. Wen C, Xu G, He S, Huang Y, Shi J, Wu L, et al. Screening circular RNAs related to acquired gefitinib resistance in non-small cell lung cancer cell lines. *J Cancer* (2020) 11(13):3816–26.

195. Xu Y, Jiang T, Wu C, Zhang Y. CircAKT3 inhibits glycolysis balance in lung cancer cells by regulating miR-516b-5p/STAT3 to inhibit cisplatin sensitivity. *Biotechnol Lett* (2020) 42(7):1123–35.

196. Huang MS, Liu JY, Xia XB, Liu YZ, Li X, Yin JY, et al. Hsa_circ_0001946 inhibits lung cancer progression and mediates cisplatin sensitivity in non-small cell lung cancer via the nucleotide excision repair signaling pathway. *Front Oncol* (2019) 9:508.

197. Dong Y, Xu T, Zhong S, Wang B, Zhang H, Wang X, et al. Circ_0076305 regulates cisplatin resistance of non-small cell lung cancer via positively modulating STAT3 by sponging miR-296-5p. *Life Sci* (2019) 239:116984.

198. Zhou Y, Zheng X, Xu B, Chen L, Wang Q, Deng H, et al. Circular RNA hsa_circ_0004015 regulates the proliferation, invasion, and TKI drug resistance of non-small cell lung cancer by miR-1183/PDPK1 signaling pathway. *Biochem Biophys Res Commun* (2019) 508(2):527–35.

199. Yu W, Peng W, Sha H, Li J. Hsa_circ_0003998 promotes chemoresistance via modulation of miR-326 in lung adenocarcinoma cells. *Oncol Res* (2019) 27 (5):623–8.

200. Li YH, Xu CL, He CJ, Pu HH, Liu JL, Wang Y. circMTDH.4/miR-630/AEG-1 axis participates in the regulation of proliferation, migration, invasion, chemoresistance, and radioresistance of NSCLC. *Mol Carcinog* (2020) 59(2):141–53.

201. Wu Z, Gong Q, Yu Y, Zhu J, Li W. Knockdown of circ-ABCB10 promotes sensitivity of lung cancer cells to cisplatin via miR-556-3p/AK4 axis. *BMC Pulm Med* (2020) 20(1):10.

202. Xiao G, Huang W, Zhan Y, Li J, Tong W. CircRNA_103762 promotes multidrug resistance in NSCLC by targeting DNA damage inducible transcript 3 (CHOP). *J Clin Lab Anal* (2020):e23252.

203. Zhang PF, Pei X, Li KS, Jin LN, Wang F, Wu J, et al. Circular RNA circFGFR1 promotes progression and anti-PD-1 resistance by sponging miR-381-3p in non-small cell lung cancer cells. *Mol Cancer* (2019) 18(1):179.

204. Ye Y, Zhao L, Li Q, Xi C, Li Y, Li Z, et al. circ_0007385 served as competing endogenous RNA for miR-519d-3p to suppress malignant behaviors and cisplatin resistance of non-small cell lung cancer cells. *Thorac Cancer* (2020).

205. Tong S. Circular RNA SMARCA5 may serve as a tumor suppressor in non-small cell lung cancer. *J Clin Lab Anal* (2020) 34(5):e23195.

206. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* (2011) 144(5):646–74.

207. Shangguan H, Feng H, Lv D, Wang J, Tian T, Wang X, et al. Circular RNA circSLC25A16 contributes to the glycolysis of non-small-cell lung cancer through epigenetic modification. *Cell Death Dis* (2020) 11(6):437. doi: 10.1038/s41419-020-2635-5

208. Wu W, Xi W, Li H, Yang M, Yao X. Circular RNA circ−ACACA regulates proliferation, migration and glycolysis in non−small−cell lung carcinoma via miR −1183 and PI3K/PKB pathway. *Int J Mol Med* (2020) 45(6):1814–24.

209. Zhang SJ, Ma J, Wu JC, Hao ZZ, Zhang YA, Zhang YJ. Circular RNA circCRIM1 suppresses lung adenocarcinoma cell migration, invasion, EMT, and glycolysis through regulating miR-125b-5p/BTG2 axis. *Eur Rev Med Pharmacol Sci* (2020) 24(7):3761–74. doi: 10.26355/eurrev_202004_20841

210. Lu J, Zhu Y, Qin Y, Chen Y. CircNFIX acts as a miR-212-3p sponge to enhance the malignant progression of non-small cell lung cancer by up-regulating ADAM10. *Cancer Manag Res* (2020) 12:9577–87. doi: 10.2147/CMAR.S272309

211. Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med* (2013) 19(11):1423–37. doi: 10.1038/nm.3394

212. Keir ME, Butte MJ, Freeman GJ, Sharpe AH. PD-1 and its ligands in tolerance and immunity. *Annu Rev Immunol* (2008) 26:677–704. doi: 10.1146/annurev.immunol.26.021607.090331

213. Li L, Zhang Q, Lian K. Circular RNA circ_0000284 plays an oncogenic role in the progression of non-small cell lung cancer through the miR-377-3p-mediated PD-L1 promotion. *Cancer Cell Int* (2020) 20:247.

214. Su C, Han Y, Zhang H, Li Y, Yi L, Wang X, et al. CiRS-7 targeting miR-7 modulates the progression of non-small cell lung cancer in a manner dependent on NF-κB signalling. *J Cell Mol Med* (2018) 22(6):3097–107.

215. Kumar M, Sahu SK, Kumar R, Subuddhi A, Maji RK, Jana K, et al. MicroRNA let-7 modulates the immune response to mycobacterium tuberculosis infection via control of A20, an inhibitor of the NF-κB pathway. *Cell Host Microbe* (2015) 17(3):345–56.

216. Schaefer L, Reinhardt DP. Special issue: Extracellular matrix: Therapeutic tools and targets in cancer treatment. *Adv Drug Delivery Rev* (2016) 97:1–3.

217. Luo YH, Zhu XZ, Huang KW, Zhang Q, Fan YX, Yan PW, et al. Emerging roles of circular RNA hsa_circ_0000064 in the proliferation and metastasis of lung cancer. *BioMed Pharmacother* (2017) 96:892–8.

218. Velasco-Hernandez T, Hyrenius-Wittsten A, Rehn M, Bryder D, Cammenga J. HIF-1α can act as a tumor suppressor gene in murine acute myeloid leukemia. *Blood* (2014) 124(24):3597–607.

219. Chi Y, Luo Q, Song Y, Yang F, Wang Y, Jin M, et al. Circular RNA circPIP5K1A promotes non-small cell lung cancer proliferation and metastasis through miR-600/HIF-1α regulation. *J Cell Biochem* (2019) 120(11):19019–30.

220. Cheng X, Qiu J, Wang S, Yang Y, Guo M, Wang D, et al. Comprehensive circular RNA profiling identifies CircFAM120A as a new biomarker of hypoxic lung adenocarcinoma. *Ann Transl Med* (2019) 7(18):442.

221. Luo YH, Yang YP, Chien CS, Yarmishyn AA, Ishola AA, Chien Y, et al. Plasma level of circular RNA hsa_circ_0000190 correlates with tumor progression and poor treatment response in advanced lung cancers. *Cancers (Basel)* (2020) 12(7).

222. Chen X, Xie R, Gu P, Huang M, Han J, Dong W, et al. Long noncoding RNA LBCS inhibits self-renewal and chemoresistance of bladder cancer stem cells through epigenetic silencing of SOX2. *Clin Cancer Res* (2019) 25(4):1389–403.

223. Huang H, Weng H, Sun W, Qin X, Shi H, Wu H, et al. Recognition of RNA N(6)-methyladenosine by IGF2BP proteins enhances mRNA stability and translation. *Nat Cell Biol* (2018) 20(3):285–95.

224. Zong L, Sun Q, Zhang H, Chen Z, Deng Y, Li D, et al. Increased expression of circRNA_102231 in lung cancer and its clinical significance. *Biomedicine pharmacotherapy* (2018) 102:639–44. doi: 10.1016/j.biopha.2018.03.084

225. Zhang S, Zeng X, Ding T, Guo L, Li Y, Ou S, et al. Microarray profile of circular RNAs identifies hsa_circ_0014130 as a new circular RNA biomarker in non-small cell lung cancer. *Sci Rep* (2018) 8(1):2878.

226. Li J, Wang J, Chen Z, Chen Y, Jin M. Hsa_circ_0079530 promotes cell proliferation and invasion in non-small cell lung cancer. *Gene* (2018) 665:1–5.

227. Lu GJ, Cui J, Qian Q, Hou ZB, Xie HY, Hu W, et al. Overexpression of hsa_circ_0001715 is a potential diagnostic and prognostic biomarker in lung adenocarcinoma. *Onco Targets Ther* (2020) 13:10775–83.

228. Zhang Y, Zhao H, Zhang L. Identification of the tumor−suppressive function of circular RNA FOXO3 in non−small cell lung cancer through sponging miR−155. *Mol Med Rep* (2018) 17(6):7692–700.

229. Hang Zhou D, Qin J, Zhou N, Ma W, Jin H, G, et al. A novel plasma circular RNA circFARSA is a potential biomarker for non-small cell lung cancer. *Cancer Med* (2018) 7(6):2783–91.

230. Yao JT, Zhao SH, Liu QP, Lv MQ, Zhou DX, Liao ZJ, et al. Over-expression of CircRNA_100876 in non-small cell lung cancer and its prognostic value. *Pathology Res Pract* (2017) 213(5):453–6.

231. Qiu M, Xia W, Chen R, Wang S, Xu Y, Ma Z, et al. The circular RNA circPRKCI promotes tumor growth in lung adenocarcinoma. *Cancer Res* (2018) 78(11):2839–51.

232. Li M, Hua Q, Shao Y, Zeng H, Liu Y, Diao Q, et al. Circular RNA circBbs9 promotes PM(2.5)-induced lung inflammation in mice via NLRP3 inflammasome activation. *Environ Int* (2020) 143:105976.

233. Xu K, Zhang Y, Xiong W, Zhang Z, Wang Z, Lv L, et al. CircGRIA1 shows an age-related increase in male macaque brain and regulates synaptic plasticity and synaptogenesis. *Nat Commun* (2020) 11(1):3594.

234. Ke L, Yang DC, Wang Y, Ding Y, Gao G. AnnoLnc2: the one-stop portal to systematically annotate novel lncRNAs for human and mouse. *Nucleic Acids Res* (2020) 48(W1):W230–w238.

235. Zhang HD, Jiang LH, Sun DW, Hou JC, Ji ZL. CircRNA: a novel type of biomarker for cancer. *Breast Cancer* (2018) 25(1):1–7.

236. Wang F, Mi YJ, Chen XG, Wu XP, Liu Z, Chen SP, et al. Axitinib targeted cancer stemlike cells to enhance efficacy of chemotherapeutic drugs via inhibiting the drug transport function of ABCG2. *Mol Med* (2012) 18(1):887–98.

237. Glažar P, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs. *Rna* (2014) 20(11):1666–70.

238. Yang JH, Qu LH. DeepBase: annotation and discovery of microRNAs and other noncoding RNAs from deep-sequencing data. *Methods Mol Biol* (2012) 822:233–48.

239. Liu YC, Li JR, Sun CH, Andrews E, Chao RF, Lin FM, et al. CircNet: a database of circular RNAs derived from transcriptome sequencing data. *Nucleic Acids Res* (2016) 44(D1):D209–15.

240. Chen X, Han P, Zhou T, Guo X, Song X, Li Y. circRNADb: A comprehensive database for human circular RNAs with protein-coding annotations. *Sci Rep* (2016) 6:34985.

241. Dudekula DB, Panda AC, Grammatikakis I, De S, Abdelmohsen K, Gorospe M. CircInteractome: A web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biol* (2016) 13(1):34–42.

242. Xia S, Feng J, Chen K, Ma Y, Gong J, Cai F, et al. CSCD: a database for cancer-specific circular RNAs. *Nucleic Acids Res* (2018) 46(D1):D925–d929.

243. Dong R, Ma XK, Li GW, Yang L. CIRCpedia v2: An updated database for comprehensive circular RNA annotation and expression comparison. *Genomics Proteomics Bioinf* (2018) 16(4):226–33.

244. Xia S, Feng J, Lei L, Hu J, Xia L, Wang J, et al. Comprehensive characterization of tissue-specific circular RNAs in the human and mouse genomes. *Brief Bioinform* (2017) 18(6):984–92.

245. Li S, Li Y, Chen B, Zhao J, Yu S, Tang Y, et al. exoRBase: a database of circRNA, lncRNA and mRNA in human blood exosomes. *Nucleic Acids Res* (2018) 46(D1):D106–d112.

246. Zhao Z, Wang K, Wu F, Wang W, Zhang K, Hu H, et al. circRNA disease: a manually curated database of experimentally supported circRNA-disease associations. *Cell Death Dis* (2018) 9(5):475.

247. Zhao M, Qu H. circVAR database: genome-wide archive of genetic variants for human circular RNAs. *BMC Genomics* (2020) 21(1):750.

248. Huang W, Ling Y, Zhang S, Xia Q, Cao R, Fan X, et al. TransCirc: an interactive database for translatable circular RNAs based on multi-omics evidence. *Nucleic Acids Res* (2020).

249. Lyu Y, Caudron-Herger M, Diederichs S. circ2GO: A database linking circular RNAs to gene function. *Cancers (Basel)* (2020) 12(10).

250. Chaput N, Flament C, Viaud S, Taieb J, Roux S, Spatz A, et al. Dendritic cell derived-exosomes: biology and clinical implementations. *J Leukoc Biol* (2006) 80(3):471–8.

251. Chen R, Xu X, Qian Z, Zhang C, Niu Y, Wang Z, et al. The biological functions and clinical applications of exosomes in lung cancer. *Cell Mol Life Sci* (2019) 76(23):4613–33.

252. Hsu YL, Hung JY, Chang WA, Lin YS, Pan YC, Tsai PH, et al. Hypoxic lung cancer-secreted exosomal miR-23a increased angiogenesis and vascular permeability by targeting prolyl hydroxylase and tight junction protein ZO-1. *Oncogene* (2017) 36(34):4929–42.

253. Whiteside TL. Exosomes and tumor-mediated immune suppression. *J Clin Invest* (2016) 126(4):1216–23.

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 11 November 2022
DOI 10.3389/fonc.2022.894780

OPEN ACCESS

EDITED BY
Leli Zeng,
Seventh Affiliated Hospital, Sun Yat-sen University, China

REVIEWED BY
Xiaobo Du,
Mianyang Central Hospital, China
Nguyen Quoc Khanh Le,
Taipei Medical University, Taiwan
Yuchen Liu,
Seventh Affiliated Hospital, Sun Yat-sen University, China

*CORRESPONDENCE
Xun Wang
wangxun04275@pkuph.edu.cn
Fan Yang
yangfan@pkuph.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 12 March 2022
ACCEPTED 18 October 2022
PUBLISHED 11 November 2022

CITATION
Cai J-S, Qiu M-T, Yang F and Wang X (2022) Stage IV non-small cell lung cancer among young individuals: Incidence, presentations, and survival outcomes of conventional therapies. *Front. Oncol.* 12:894780. doi: 10.3389/fonc.2022.894780

Stage IV non-small cell lung cancer among young individuals: Incidence, presentations, and survival outcomes of conventional therapies

Jing-Sheng Cai†, Man-Tang Qiu†, Fan Yang* and Xun Wang*

Department of Thoracic Surgery, Peking University People's Hospital, Beijing, China

Background: There is a paucity of data published on the clinicopathological features and prognosis of stage IV non-small cell lung cancer (NSCLC) patients aged ≤45 years. Herein, we evaluated a large clinical series in an effort to provide a clearer picture of this population.

Methods: The least absolute shrinkage and selection operator (LASSO)-penalized Cox regression model was performed to identify prognostic factors for NSCLC among individuals aged ≤45 years. The Kaplan–Meier method with log-rank test was used to compare overall survival (OS) differences between groups. Competing risk analysis with the Fine–Gray test was used to analyze cancer-specific survival (CSS) differences. Propensity score matching (PSM) was used to minimize selection bias.

Results: Incidence-rate analyses, including 588,680 NSCLC cases (stage IV, 233,881; age ≤ 45 years stage IV, 5,483; and age > 45 years stage IV, 228,398) from 2004 to 2015, showed that the incidence of stage IV NSCLC among young individuals decreased over the years. In comparative analyses of clinical features and survival outcomes, a total of 48,607 eligible stage IV cases (age ≤ 45 years stage IV, 1,390; age > 45 years stage IV, 47,217) were included. The results showed that although patients in the young cohort were more likely to be diagnosed at advanced stages, they were also more likely to receive aggressive treatments. In addition, the survival rates of the young patients were superior to those of the older patients both before and after PSM.

Conclusions: Stage IV NSCLC patients aged ≤45 years comprise a relatively small but special NSCLC subgroup. Although this population had better survival outcomes than older patients, these patients deserve more attention due to their young age and the significant socioeconomic implications.

KEYWORDS

stage IV, non-small cell lung cancer, age ≤ 45 years, incidence, outcome

Introduction

Lung cancer is a serious global pandemic (1, 2). Non-small cell lung cancer (NSCLC) accounts for approximately 85% of all lung cancer cases. Approximately 40% of NSCLC patients are initially diagnosed at stage IV (3). Because the median age at diagnosis is 70 years (4), NSCLC is often regarded as a disease among older people. However, over the past few decades, the incidence of NSCLC in young individuals has been increasing gradually (5–7). Given the substantial societal and economic effects of NSCLC, more in-depth investigations are needed to examine this disease among young patients (8).

Previous clinical series demonstrated that young NSCLC patients are more likely to be female, to be non-smokers, to have adenocarcinoma (ADC) subtypes, and to have advanced-stage diseases than older NSCLC patients (4, 5, 9–16). Furthermore, the survival rate of young patients is inferior to that of older patients (9–11, 13). However, these studies focused on the entire entity of young NSCLC patients, and no literature is available about the clinicopathological features and prognosis of young stage IV NSCLC patients.

Given the paucity of related studies, we sought to better understand stage IV NSCLC among young patients (age ≤ 45 years) by analyzing the data deposited in the Surveillance, Epidemiology, and End Results (SEER) Program with the purpose of sketching an outline of this population.

Materials and methods

Included subjects

To analyze the incidence rate, lung cancer patients from 2004 to 2015 were extracted from the SEER database (https://seer.cancer.gov/). The inclusion criterion was a diagnosis of lung malignancy. The exclusion criteria were 1) lung tumors other than NSCLC and 2) stage I–III diseases (the 8th edition of the tumor–node–metastasis [TNM] staging system (17)). The flowchart of patient selection is shown in Figure 1.

To analyze clinicopathological features and survival outcomes, lung cancer patients from 2010 to 2015 were extracted from the SEER database. The inclusion criteria were as follows: 1) pathologically diagnosed as NSCLC and 2) stage IV diseases. The exclusion criteria were as follows: 1) unavailable TNM stage information, 2) unavailable survival information, 3) previous other cancers, and (4) age < 18 years. The eligible stage IV NSCLC patients were categorized into two groups: stage IV NSCLC aged ≤45 years and stage IV NSCLC aged >45 years. The corresponding flowchart of patient selection is shown in Figure 1.

Stage IV patients from 2004 to 2015 were included in the incidence-rate analysis because of the large number of cases, which could lead to a more reliable conclusion. Patients from 2010 to 2015 were included in the clinical characteristics and survival outcomes analysis because only the information about specific metastasis sites was available for the patients in this period, which is an important prognostic factor for metastatic NSCLC patients.

Ethics

We obtained permission to access the SEER dataset (reference number 12962-Nov2019) using SEER*Stat software version 8.3.4. The study was conducted in accordance with the Declaration of Helsinki. The Ethics Board of Peking University People's Hospital approved this study. This was an open database study, and only non-identifiable information was used. Therefore, this study was dispensed with acquiring signed informed consent forms and ethical approval.

Data collection

This work was supported by the NationalNatural Science funds (grant number82173386).The following anonymized data, including age (continue), sex (male and female), ethnicity (Caucasian, African, and other), marital status (married and other), tumor location (upper lobe [UL], middle lobe [ML], low lobe [LL], and other), surgery (no and yes), chemotherapy (no and yes), radiotherapy (no and yes), grade (well differentiated, moderately differentiated, poorly differentiated, undifferentiated, and unknown), TNM stage (stage IVA, stage IVB, and stage IV), T category (T1, T2, T3, and T4), N category (N0, N1, N2, and N3), M category (M1a, M1b, M1c, and M1), bone metastasis (no and yes), brain metastasis (no and yes), liver metastasis (no and yes), intrapulmonary metastasis (no and yes), cause of death, patient status, and survival time, were retrieved. The current 8th edition of the TNM staging system (17) was used in this study.

Follow-up

The primary endpoints were overall survival (OS) and cancer-specific survival (CSS). OS was calculated as the time interval from the date of initial diagnosis to the date of death or the date of the last follow-up evaluation. CSS was calculated as the time interval from the date of initial diagnosis to the date of death attributed to NSCLC or the date of the last follow-up evaluation. The survival information, including survival time, survival status, and cause of death, is available in the SEER database. NSCLC patients with an exact survival status and survival time were included, and those with a survival time = 0 months were excluded from this study. The median follow-up time of the entire stage IV cohort, age ≤ 45 years stage IV cohort, and age > 45 years stage IV cohort were 6 months (range, 1–83

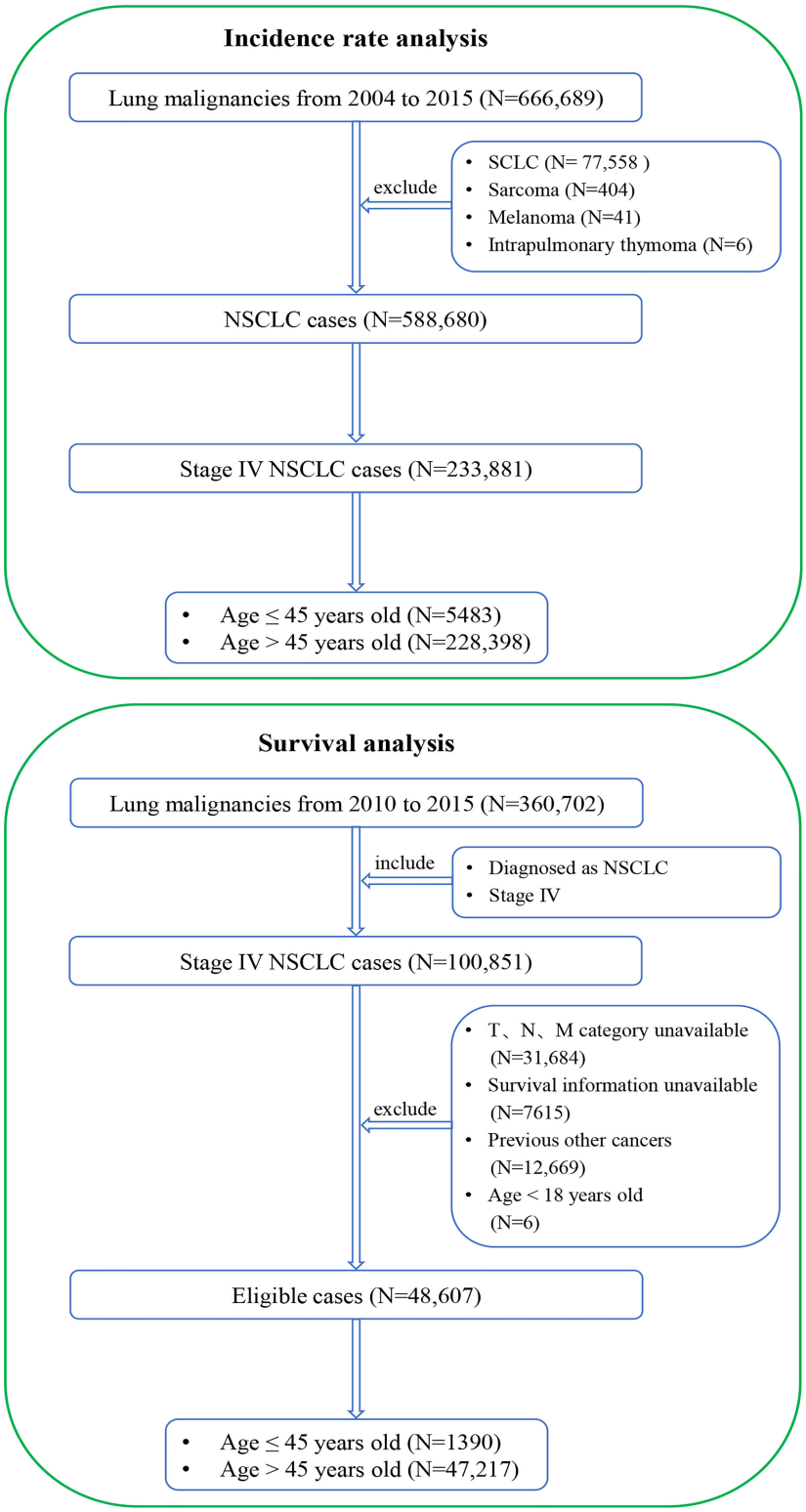


FIGURE 1
Flowcharts of patient selection. The detailed selection process of stage IV NSCLC patients from 2004 to 2015 for incidence-rate analysis (top) and the detailed selection process of stage IV NSCLC patients from 2010 to 2015 for presentations and survival outcomes analysis (bottom). NSCLC, non-small cell lung cancer; SCLC, small cell lung cancer.

months), 12 months (range, 1–83 months), and 6 months (range, 1–83 months), respectively.

Statistical analysis

R version 3.5.2 (The R Foundation for Statistical Computing, Vienna, Austria; http://www.r-project.org) and IBM SPSS Statistics (version 25.0, IBM Corp, Armonk, NY, USA) were applied to the statistical analysis. The Kaplan–Meier method with log-rank test was used to compare OS differences between groups. Competing risk analysis with the Fine–Gray test (18) was used to compare CSS differences between groups. One-to-two propensity score matching (PSM) (19) between the age ≤ 45 years cohort and the age > 45 years cohort was carried out to minimize bias. The nearest-neighbor matching method with a caliper distance of 0.0001 was used in the PSM algorithm. The variables, including sex, surgery, chemotherapy, radiotherapy, histology, grade, TNM stage, T category, N category, M category, bone metastasis, brain metastasis, liver metastasis, and intrapulmonary metastasis, were included in the PSM algorithm. The variables, including age, sex, ethnicity, marital status, tumor location, surgery, chemotherapy, radiotherapy, histology, grade, TNM stage, T category, N category, M category, bone metastasis, brain metastasis, liver metastasis, and intrapulmonary metastasis, were included in the least absolute shrinkage and selection operator (LASSO) regression model (20), which minimizes the risk of overfitting and further selects the potential prognostic factors. The selected variables were then included in a stepwise multivariable Cox regression analysis to determine the final independent prognostic factors. Statistically significant factors selected from the LASSO-penalized multivariable Cox analysis were used to develop a nomogram (21). The C-index (22) and the receiver operating characteristic (ROC) curves with an area under the curve (AUC) were used to evaluate the performances of the models. Categorical variables were expressed as numbers and percentages and were compared between groups using Pearson's χ^2 test. A two-sided $p < 0.05$ was considered statistically significant.

Results

Incidence-rate analysis

This work was supported by the NationalNatural Science funds (grant number82173386).From 2004 to 2015, 666,689 cases of lung malignancies were retrospectively reviewed. After the inclusion and exclusion criteria were applied, 233,881 eligible stage IV NSCLC cases (age ≤ 45 years, 5,483 cases; age > 45 years, 228,398 cases) were selected. In the entire cohort, the crude incidence of stage IV diseases was stable between 2004 and 2009 (approximately 40.00%). However, it increased by approximately 6.00% between 2010 and 2015 (Figure 2A). Considering the incidence of age ≤ 45 years stage IV NSCLC, a decreasing tendency was observed between 2004 and 2015 (1.33% in 2004 and 0.79% in 2015; Figure 2B). In stage IV NSCLC cohort, the incidence of age ≤ 45 years still had an uninterrupted decrease during this period (3.41% in 2004 and 1.73% in 2015; Figure 2C).

Patient characteristics

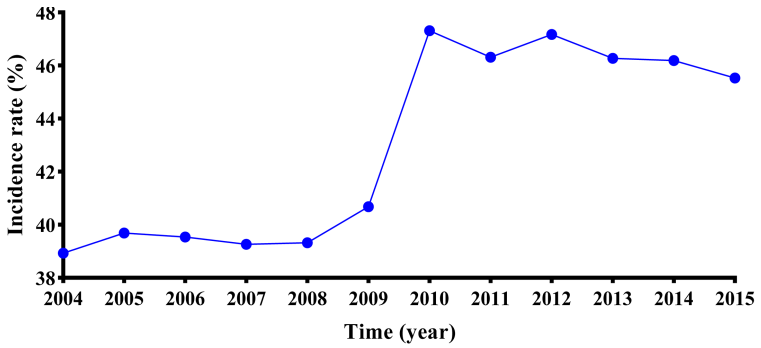
From 2010 to 2015, the data of 360,702 lung malignancy cases were reviewed. A total of 48,607 eligible stage IV NSCLC cases (age ≤ 45 years, 1,390 cases; age > 45 years, 47,217 cases) were selected. The clinicopathological features are listed in Table 1. Regarding the age ≤ 45 years in stage IV NSCLC patients, there was no sex difference (male vs. female = 48.8% vs. 51.2%). The majority of patients were Caucasian (67.9%). Only a small proportion of patients underwent surgery (9.1%) and radiotherapy (21.7%). Most patients underwent chemotherapy (81.5%). ADC was the predominant histological subtype (66.6%). Over half of the cases were diagnosed as stage IVA disease (53.5%). Most of the cases had local/distant lymph node metastasis (N2 category, 46.0%; N3 category, 27.0%). Regarding the metastatic sites, 43.9% of the patients had bone metastasis, 38.6% had brain metastasis, 19.0% had liver metastasis, and 31.3% had intrapulmonary metastasis.

Before PSM, when compared with the age > 45 years stage IV NSCLC cohort, there were more men in the age ≤ 45 years stage IV NSCLC cohort ($p < 0.001$). In addition, more patients received surgery, chemotherapy, and radiotherapy in the age ≤ 45 years stage IV NSCLC cohort (surgery, 9.1% vs. 4.4%, $p < 0.001$; chemotherapy, 81.5% vs. 61.0%, $p < 0.001$; radiotherapy, 21.7% vs. 11.8%, $p < 0.001$). More patients in the age > 45 years stage IV NSCLC cohort suffered from intrathoracic metastasis (62.1% vs. 53.5%, $p < 0.001$). However, they were unlikely to have lymph node metastasis ($p < 0.001$). After PSM, there were 1,320 and 2,589 cases in the young and older patient groups, respectively. The covariates between these two groups were well-balanced (Table 1).

Least absolute shrinkage and selection operator-penalized Cox regression analysis and nomogram

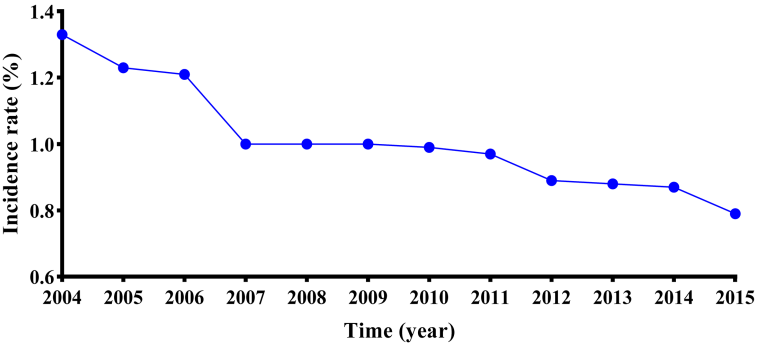
To examine OS, 14 variables, including age, sex, ethnicity, marital status, tumor location, surgery, chemotherapy, histology, grade, T category, N category, M category, bone metastasis, and liver metastasis, were selected using the LASSO model (Figures S1A, B). In further analyses, the multivariable Cox regression analysis confirmed that age, ethnicity, surgery, chemotherapy, histology, grade, T category, N category, M category, and bone metastasis were independent prognostic factors (Table 2). The

A



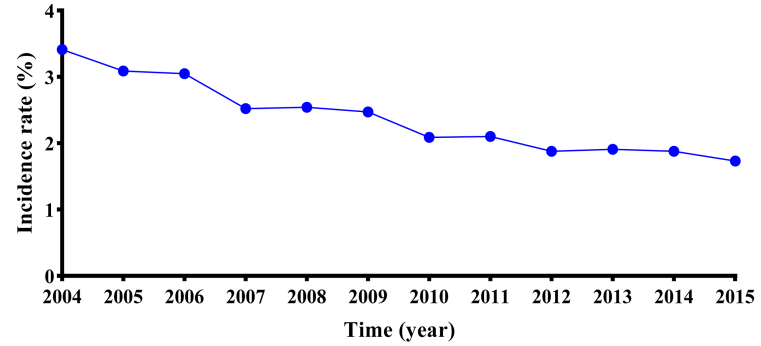
Time (Year)	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
All NSCLC	43,637	43,408	45,215	45,413	45,735	46,258	45,136	45,060	45,438	45,711	46,129	46,211
Stage IV	16,988	17,229	17,880	17,829	17,981	18,820	21,355	20,869	21,434	21,150	21,305	21,042
Rate (%)	39.83	39.69	39.54	39.26	39.32	40.68	47.31	46.31	47.17	46.27	46.19	45.53

B



Time (Year)	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
All NSCLC	43,637	43,408	45,215	45,413	45,735	46,258	45,136	45,060	45,438	45,711	46,129	46,211
⩽45 Stage IV	579	533	546	450	456	464	447	438	404	403	400	363
Rate (%)	1.33	1.23	1.21	1.00	1.00	1.00	0.99	0.97	0.89	0.88	0.87	0.79

C



Time (Year)	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
All Stage IV	16,988	17,229	17,880	17,829	17,981	18,820	21,355	20,869	21,434	21,150	21,305	21,042
⩽45 Stage IV	579	533	546	450	456	464	447	438	404	403	400	363
Rate (%)	3.41	3.09	3.05	2.52	2.54	2.47	2.09	2.10	1.88	1.91	1.88	1.73

FIGURE 2
Line charts of the incidence of stage IV NSCLC patients from 2004 to 2015. The incidence-rate analyses of stage IV NSCLC patients in the entire cohort **(A)**. The incidence-rate analyses of stage IV NSCLC patients aged ≤45 years in the entire cohort **(B)**; and the incidence-rate analyses of stage IV NSCLC patients aged ≤45 years in the stage IV NSCLC cohort. **(C)** NSCLC, non-small cell lung cancer.

TABLE 1 Baseline features of stage IV NSCLC patients ≤45 years and those >45 years before and after PSM.

Characteristic	Before PSM			After PSM		
	≤45 (N = 1,390)	>45 (N = 47,217)	*p*	≤45 (N = 1,320)	>45 (N = 2,589)	*p*
Sex			<0.001			1.000
Male	679 (48.8)	25,887 (54.8)		652 (49.4)	1,278 (49.4)	
Female	711 (51.2)	21,330 (45.2)		668 (50.6)	1,311 (50.6)	
Surgery			<0.001			0.334
No	1,264 (90.9)	45,124 (95.6)		1,230 (93.2)	2,433 (94.0)	
Yes	126 (9.1)	2,093 (4.4)		90 (6.8)	156 (6.0)	
Chemotherapy			<0.001			0.938
No	257 (18.5)	18,399 (39.0)		242 (18.3)	472 (18.2)	
Yes	1,133 (81.5)	28,818 (61.0)		1,078 (81.7)	2,117 (81.8)	
Radiotherapy			<0.001			0.618
No	1,089 (78.3)	41,643 (88.2)		1,064 (80.6)	2,104 (81.3)	
Yes	301 (21.7)	5,574 (11.8)		256 (19.4)	485 (18.7)	
Histology			<0.001			0.954
ADC	926 (66.6)	26,715 (56.6)		895 (67.8)	1,768 (68.3)	
SCC	125 (9.0)	9,999 (21.2)		116 (8.8)	224 (8.7)	
Other	339 (24.4)	10,503 (22.2)		309 (23.4)	597 (23.1)	
Grade			0.167			0.210
Well	50 (3.6)	1,344 (2.8)		47 (3.6)	95 (3.7)	
Moderately	190 (13.7)	6,618 (14.0)		176 (13.3)	285 (11.0)	
Poorly	387 (27.8)	14,058 (29.8)		364 (27.6)	725 (28.0)	
Undifferentiated	29 (2.1)	770 (1.6)		25 (1.9)	38 (1.5)	
Unknown	734 (52.8)	24,427 (51.7)		708 (53.6)	1,446 (55.9)	
TNM stage			<0.001			0.936
IVA	744 (53.5)	29,315 (62.1)		719 (54.5)	1,426 (55.1)	
IVB	321 (23.1)	7,608 (16.1)		294 (22.3)	568 (21.9)	
IV	325 (23.4)	10,294 (21.8)		307 (23.3)	595 (23.0)	
T category			0.005			0.999
T1	346 (24.9)	10,822 (22.9)		330 (25.0)	650 (25.1)	
T2	412 (29.6)	15,680 (33.2)		404 (30.6)	796 (30.7)	
T3	260 (18.7)	9,410 (19.9)		241 (18.3)	470 (18.2)	
T4	372 (26.8)	11,305 (23.9)		345 (26.1)	673 (26.0)	
N category			<0.001			0.899
N0	266 (19.1)	11,708 (24.8)		255 (19.3)	495 (19.1)	
N1	110 (7.9)	4,119 (8.7)		95 (7.2)	171 (6.6)	
N2	639 (46.0)	21,612 (45.8)		617 (46.7)	1,231 (47.5)	
N3	375 (27.0)	9,778 (20.7)		353 (26.7)	692 (26.7)	
M category			<0.001			0.988
1a	259 (18.6)	12,472 (26.4)		250 (18.9)	495 (19.1)	
1b	485 (34.9)	16,842 (35.7)		469 (35.5)	931 (36.0)	
1c	321 (23.1)	7,608 (16.1)		294 (22.3)	568 (21.9)	
1	325 (23.4)	10,295 (21.8)		307 (23.3)	595 (23.0)	
Bone metastasis			<0.001			0.926
No	780 (56.1)	28,782 (61.0)		751 (56.9)	1,477 (57.0)	
Yes	610 (43.9)	18,435 (39.0)		569 (43.1)	1,112 (43.0)	
Brain metastasis			<0.001			0.732
No	854 (61.4)	33,743 (71.5)		816 (61.8)	1,615 (62.4)	
Yes	536 (38.6)	13,474 (28.5)		504 (38.2)	974 (37.6)	

(Continued)

TABLE 1 Continued

Characteristic	Before PSM			After PSM		
Liver metastasis			0.097			0.547
No	1,126 (81.0)	39,056 (82.7)		1,085 (82.2)	2,148 (83.0)	
Yes	264 (19.0)	8,161 (17.3)		235 (17.8)	441 (17.0)	
Intrapulmonary metastasis			0.710			0.818
No	955 (68.7)	32,661 (69.2)		913 (69.2)	1,800 (69.5)	
Yes	435 (31.3)	14,556 (30.8)		407 (30.8)	789 (30.5)	

NSCLC, non-small cell lung cancer; PSM, propensity score matching; ADC, adenocarcinoma; SCC, squamous cell carcinoma; TNM, tumor–node–metastasis.

corresponding nomogram was developed (Figure S2A). The C-index of this nomogram was 0.67 (95% CI: 0.65–0.69). The AUC of the nomogram was 0.69 (Figure S3A).

To examine CSS, 15 variables, including age, sex, ethnicity, marital status, tumor location, surgery, chemotherapy, radiotherapy, histology, grade, T category, N category, M category, bone metastasis, and liver metastasis, were selected using the LASSO model (Figures S1C, D). The multivariable Cox regression analysis confirmed that age, ethnicity, surgery, chemotherapy, grade, T category, N category, and M category were independent prognostic factors (Table 2). The corresponding nomogram was established (Figure S2B). The C-index of this nomogram was 0.66 (95% CI: 0.64–0.68). The AUC of the nomogram was 0.67 (Figure S3B).

Survival analysis

Before PSM, survival analysis showed that stage IV NSCLC patients ≤45 years had better OS than patients >45 years (3-year OS rate, 20.3% vs. 9.8%, $p < 0.001$; Figure 3A). In the matched cohort, the OS rate of stage IV NSCLC patients ≤45 years was still superior to that of stage IV NSCLC patients >45 years (3-year OS rate, 20.3% vs. 14.6%, $p < 0.001$; Figure 3B).

Regarding CSS, before PSM, the 3-year cancer-specific cumulative mortality rate of stage IV NSCLC patients ≤45 years was lower than that of stage IV NSCLC patients >45 years (74.8% vs. 84.6%, $p < 0.001$; Figure 3C). After PSM, stage IV NSCLC patients ≤45 years still had better CSS than stage IV NSCLC patients >45 years (3-year cancer-specific cumulative mortality rate, 74.9% vs. 80.9%, $p < 0.001$; Figure 3D).

Discussion

This large population-based study was the first to investigate the incidence, presentations, and survival outcomes of stage IV NSCLC among young patients. The findings of this study can be summarized as follows. The incidence of stage IV NSCLC patients ≤45 years has declined over the years. Stage IV NSCLC patients ≤45 years were more likely to be diagnosed at an advanced stage, but they were also more likely to receive aggressive treatments. Multivariable Cox analyses revealed that receiving surgery and chemotherapy were important prognostic factors. Survival analyses showed that the survival of younger stage IV NSCLC patients was superior to that of the older patients. Our study focused on a special subset of stage IV NSCLC patients, which may provide comprehensive knowledge of this population.

There are conflicting data about the incidence of NSCLC among young individuals. Several studies have demonstrated that the incidence rate has been increasing gradually (5–7). Another study revealed that the incidence decreased from 1978 to 2010 (15). However, there are no studies that specifically investigated the incidence of stage IV NSCLC among young individuals. Our results showed a decreasing incidence of stage IV NSCLC among young individuals ≤ 45 years between 2004 and 2015. In our view, the reduced incidence might be attributed to economic development and improved living standards. Over the past few decades, poverty and geographical restraint have prevented some NSCLC patients from receiving standard cancer care services (9). Therefore, many NSCLC patients are diagnosed at advanced stages on initial examinations. With the advancement of diagnostic and treatment modalities, many patients are well managed with surgery and other novel efficient therapies in the early stage. In addition, less indoor air pollution and less exposure to coal dust might be partially responsible for the decline in incidence.

Consistent with previous studies (4, 5, 9–12, 14–16, 23), we also found that more young NSCLC patients were diagnosed with the ADC histology subtype. A possible explanation might be related to smoking: it is known that smoking is considered an important carcinogenesis factor for SCC (24, 25), and young patients are unlikely to be smokers (9, 13, 23). Therefore, the ADC histology subtype dominated in this population. No sex differences were observed in our study, which was similar to several other studies (4, 9, 23). Our study showed that when compared with the older cohort, younger patients were more likely to be diagnosed at advanced stages and receive aggressive treatments. This result was also confirmed by Arnold et al. (26), who demonstrated that young patients were administered more aggressive therapies than older patients at each TNM stage.

TABLE 2 LASSO-penalized multivariable Cox analysis of the stage IV NSCLC patients ≤45 years.

Characteristic	OS			CSS		
	HR	95% CI	*p*	HR	95% CI	*p*
Age, years	1.026	1.013–1.039	<0.001	1.029	1.015–1.043	<0.001
Sex			0.527			0.586
Male	1			1		
Female	1.040	0.920–1.176		1.036	0.913–1.176	
Ethnicity			<0.001			<0.001
Caucasian	1			1		
African	1.171	0.989–1.387		1.164	0.977–1.387	
Other	0.670	0.561–0.800		0.664	0.553–0.798	
Marital status			0.204			0.485
Married	1			1		
Other	1.085	0.957–1.230		1.047	0.920–1.192	
Tumor location			0.113			0.118
UL	1			1		
ML	0.785	0.594–1.038		0.779	0.583–1.041	
LL	0.863	0.743–1.003		0.867	0.742–1.013	
Other	0.995	0.844–1.174		1.016	0.857–1.204	
Surgery			<0.001			<0.001
No	1			1		
Yes	0.459	0.349–0.604		0.406	0.301–0.549	
Chemotherapy			<0.001			<0.001
No	1			1		
Yes	1.948	1.668–2.275		1.959	1.667–2.302	
Radiotherapy						0.553
No				1		
Yes				1.050	0.893–1.235	
Histology			0.029			0.055
ADC	1			1		
SCC	1.260	1.022–1.552		1.226	0.986–1.524	
Other	1.159	0.999–1.343		1.160	0.996–1.351	
Grade			<0.001			<0.001
Well	1			1		
Moderately	2.109	1.292–3.442		2.114	1.261–3.544	
Poorly	2.533	1.574–4.077		2.608	1.580–4.307	
Undifferentiated	4.025	2.198–7.369		4.194	2.234–7.873	
Unknown	2.241	1.399–3.589		2.240	1.362–3.682	
T category			<0.001			<0.001
T1	1			1		
T2	0.951	0.803–1.128		0.956	0.802–1.139	
T3	1.203	1.000–1.448		1.209	0.998–1.464	
T4	1.423	1.193–1.699		1.426	1.188–1.713	
N category			<0.001			<0.001
N0	1			1		
N1	1.395	1.074–1.811		1.347	1.024–1.772	
N2	1.409	1.179–1.685		1.420	1.180–1.709	
N3	1.625	1.336–1.977		1.633	1.333–2.000	
M category			0.004			<0.001
1a	1			1		

(Continued)

TABLE 2 Continued

Characteristic	OS			CSS		
	HR	95% CI	*p*	HR	95% CI	*p*
1b	1.110	0.912–1.351		1.347	1.024–1.772	
1c	1.158	0.870–1.540		1.420	1.180–1.709	
1	1.401	1.148–1.709		1.633	1.333–2.000	
Bone metastasis			0.039			0.061
No	1			1		
Yes	1.192	1.009–1.408		1.183	0.992–1.410	
Liver metastasis			0.134			0.116
No	1			1		
Yes	1.149	0.958–1.379		1.164	0.963–1.406	

LASSO, least absolute shrinkage and selection operator; NSCLC, non-small cell lung cancer; PSM, propensity score matching; UL, upper lobe; ML, middle lobe; LL, low lobe; ADC, adenocarcinoma; SCC, squamous cell carcinoma.

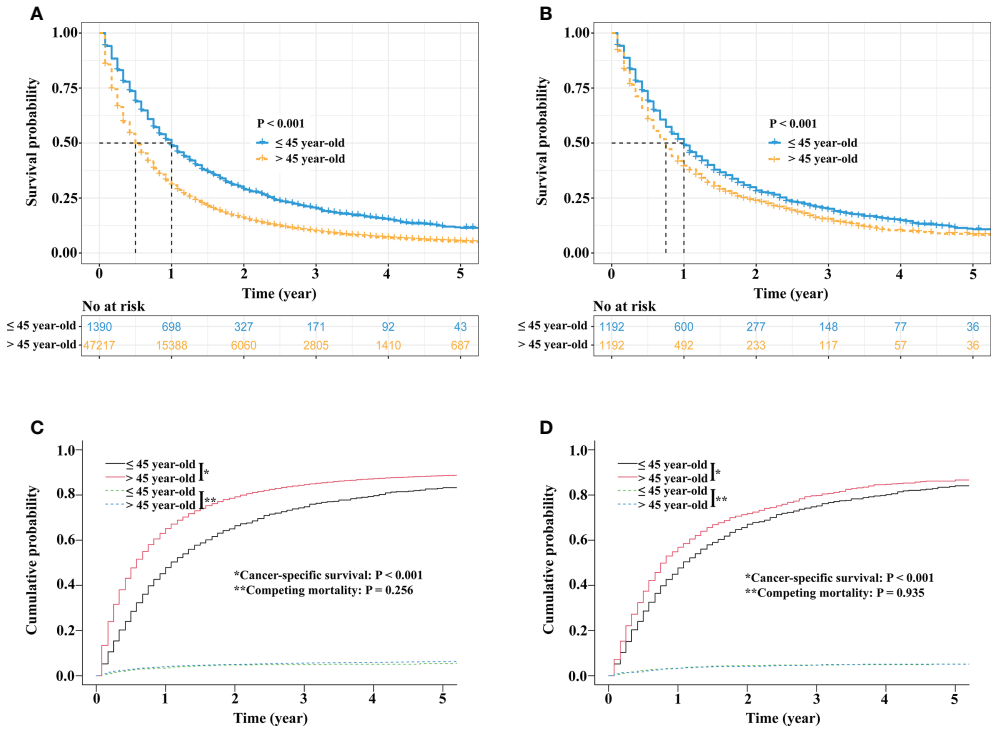


FIGURE 3
Survival comparisons between stage IV NSCLC patients aged ≤45 years and stage IV NSCLC patients aged >45 years. Kaplan–Meier survival curve comparison before PSM **(A)**. Kaplan–Meier survival curve comparison after PSM **(B)**. Competing risk analyses before PSM **(C)** and competing risk analyses after PSM **(D)**. NSCLC, non-small cell lung cancer; PSM, propensity score matching.

Better performance status and a stronger desire to live may explain this difference.

Controversy exists with regard to the survival differences between younger and older stage IV NSCLC patients. A study by Vashistha et al. demonstrated that the survival of young NSCLC patients is similar to that of older NSCLC patients in India (9). This was also confirmed in a study by Mauri et al. (14). However, Bratova et al. (10) and Bryant et al. (13) suggested that young NSCLC patients have worse survival rates than older NSCLC patients. Our study showed that the survival of stage IV

NSCLC patients ≤45 years was superior to that of stage IV NSCLC patients >45 years. Similar findings were also observed in the studies by Arnold et al. (26) and Subramanian et al. (12). One potential reason for this difference was that young patients were more likely to undergo surgery and chemotherapy than older patients. Previous clinical series suggested that surgical resection could prolong stage IV patient survival (27–32). Our nomograms also confirmed that surgery was the strongest predictor of favorable outcomes, followed by chemotherapy and the T category. Therefore, it is reasonable to observe that young patients had better survival outcomes than older patients. Additionally, a better performance status might bring survival benefits to young patients. Herein, we proposed that a multidisciplinary collaborative treatment modality, including surgery, might be the preferred option for these patients with advanced diseases.

Our study had several limitations. First, we could not evaluate the influence of performance status, smoking history, comorbidities, timing, dosage and regimens of treatments, and treatment-related side effects because they were not recorded in the SEER database. It is likely that with increasing age, more comorbidities and worse performance status might hinder older patients from receiving more aggressive treatments. Second, in the era of targeted therapy and immunotherapy, older stage IV NSCLC patients could gain more survival benefit from novel efficient therapies. However, genetic features such as epidermal growth factor receptor mutation status, anaplastic lymphoma kinase mutation status, programmed death ligand 1 expression level, and tumor mutation burden were not recorded in the database. Further efforts on broader clinicopathological features such as radiomics features (33) and tumor molecular profiles (34) recruitment are also warranted. Third, external validation was lacking in this study. Therefore, our results needed to be further validated in other clinical series and should be cautiously interpreted. Finally, inevitable bias was inherent to the retrospective design of the study, although PSM was used in the study.

Taken together, the incidence of stage IV NSCLC patients aged ≤45 years declined over the years. Although patients in this cohort were more likely to be diagnosed at advanced stages, they were also more likely to receive aggressive treatments. The prognosis of the young patients was better than that of the older patients. Although the younger cohort had better survival outcomes, they deserved more attention due to their young age and the significant socioeconomic implications of disease among this group.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving human participants were reviewed and approved by Institutional Review Board of Peking University People's Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

FY and XW contributed to conception and design of the study. J-SC and M-TQ organized the database. J-SC performed the statistical analysis. J-SC wrote the first draft of the manuscript. J-SC wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This work was supported by the National Natural Science funds (grant number 82173386).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.894780/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Prognostic predictor selection using the LASSO regression analysis. LASSO coefficient profiles of 18 variables against the log (Lambda) sequence for OS **(A)** and CSS **(C)**. Tuning parameter (Lambda) selection in the LASSO model used 10-fold cross-validation *via* minimum criteria for OS **(B)** and CSS **(D)**. LASSO, least absolute shrinkage and selection operator; OS, overall survival; CSS, cancer specific survival.

SUPPLEMENTARY FIGURE 2
Prognostic nomograms for OS **(A)** and CSS **(B)**. OS, overall survival; CSS, cancer specific survival; ADC, adenocarcinoma; SCC, squamous cell carcinoma.

SUPPLEMENTARY FIGURE 3
The receiver operating curves of the nomograms. **(A)** OS and **(B)** CSS. AUC, area under curve; OS, overall survival; CSS, cancer specific survival.

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. *CA Cancer J Clin* (2021) 71(1):7–33. doi: 10.3322/caac.21654

2. Ganti AK, Klein AB, Cotarla I, Seal B, Chou E. Update of incidence, prevalence, survival, and initial treatment in patients with non-small cell lung cancer in the us. *JAMA Oncol* (2021) 7(12):1824–32. doi: 10.1001/jamaoncol.2021.4932

3. United States, Department of Health and Human Services, National Institutes of Health, National Cancer Institute (Nci). *Non–small cell lung cancer treatment (Pdq) – health professional version: Stage information for nsclc*. Bethesda, Md: Nci (2020).

4. Galvez-Nino M, Ruiz R, Pinto JA, Roque K, Mantilla R, Raez LE, et al. Lung cancer in the young. *Lung* (2020) 198(1):195–200. doi: 10.1007/s00408-019-00294-5

5. Hsu CL, Chen KY, Shih JY, Ho CC, Yang CH, Yu CJ, et al. Advanced non-small cell lung cancer in patients aged 45 years or younger: Outcomes and prognostic factors. *BMC Cancer* (2012) 12:241. doi: 10.1186/1471-2407-12-241

6. Marugame T, Yoshimi I, Kamo K, Imamura Y, Kaneko S, Mizuno S, et al. Trends in lung cancer mortality among young adults in Japan. *Jpn J Clin Oncol* (2005) 35(4):177–80. doi: 10.1093/jjco/hyi054

7. Strand TE, Malayeri C, Eskonsipo PK, Grimsrud TK, Norstein J, Grotmol T. Adolescent smoking and trends in lung cancer incidence among young adults in Norway 1954-1998. *Cancer Causes Control* (2004) 15(1):27–33. doi: 10.1023/B:CACO.0000016575.31651.b0

8. Fidler MM, Gupta S, Soerjomataram I, Ferlay J, Steliarova-Foucher E, Bray F. Cancer incidence and mortality among young adults aged 20-39 years worldwide in 2012: A population-based study. *Lancet Oncol* (2017) 18(12):1579–89. doi: 10.1016/S1470-2045(17)30677-0

9. Vashistha V, Garg A, Iyer H, Jain D, Madan K, Hadda V, et al. A comprehensive comparison between young and older-age non-small cell lung cancer patients at a public referral centre in Delhi, India. *Ecancermedicalscience* (2021) 15:1223. doi: 10.3332/ecancer.2021.1223

10. Bratova M, Brat K, Hurdalkova K, Barinova M, Drosslerova M, Kultan J, et al. Lung cancer versus "Young cancer": Is non-small cell lung cancer in young patients a different entity? *J Adolesc Young Adult Oncol* (2021) 11(5):451–8. doi: 10.1089/jayao.2021.0069

11. Zhang J, Chen SF, Zhen Y, Xiang J, Wu C, Bao P, et al. Multicenter analysis of lung cancer patients younger than 45 years in shanghai. *Cancer* (2010) 116 (15):3656–62. doi: 10.1002/cncr.25100

12. Subramanian J, Morgensztern D, Goodgame B, Baggstrom MQ, Gao F, Piccirillo J, et al. Distinctive characteristics of non-small cell lung cancer (Nsclc) in the young: A surveillance, epidemiology, and end results (Seer) analysis. *J Thorac Oncol* (2010) 5(1):23–8. doi: 10.1097/JTO.0b013e3181c41e8d

13. Bryant AS, Cerfolio RJ. Differences in outcomes between younger and older patients with non-small cell lung cancer. *Ann Thorac Surg* (2008) 85(5):1735–9. doi: 10.1016/j.athoracsur.2008.01.031

14. Mauri D, Pentheroudakis G, Bafaloukos D, Pectasides D, Samantas E, Efstathiou E, et al. Non-small cell lung cancer in the young: A retrospective analysis of diagnosis, management and outcome data. *Anticancer Res* (2006) 26 (4B):3175–81.

15. Thomas A, Chen Y, Yu T, Jakopovic M, Giaccone G. Trends and characteristics of young non-small cell lung cancer patients in the united states. *Front Oncol* (2015) 5:113. doi: 10.3389/fonc.2015.00113

16. Bigay-Game L, Bota S, Greillier L, Monnet I, Madroszyk A, Corre R, et al. Characteristics of lung cancer in patients younger than 40 years: A prospective multicenter analysis in France. *Oncology* (2018) 95(6):337–43. doi: 10.1159/000489784

17. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, et al. The iaslc lung cancer staging project: Proposals for revision of the tnm stage groupings in the forthcoming (Eighth) edition of the tnm classification for lung cancer. *J Thorac Oncol* (2016) 11(1):39–51. doi: 10.1016/j.jtho.2015.09.009

18. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: New representations of old estimators. *Stat Med* (1999) 18(6):695–706. doi: 10.1002/(sici)1097-0258(19990330)18:6

19. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med* (2009) 28(25):3083–107. doi: 10.1002/sim.3697

20. Tibshirani R. The lasso method for variable selection in the cox model. *Stat Med* (1997) 16(4):385–95. doi: 10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3

21. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol* (2008) 26(8):1364–70. doi: 10.1200/JCO.2007.12.9791

22. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology* (2010) 21(1):128–38. doi: 10.1097/EDE.0b013e3181c30fb2

23. Li F, He H, Qiu B, Ji Y, Sun K, Xue Q, et al. Clinicopathological characteristics and prognosis of lung cancer in young patients aged 30 years and younger. *J Thorac Dis* (2019) 11(10):4282–91. doi: 10.21037/jtd.2019.09.60

24. Etzel CJ, Amos CI, Spitz MR. Risk for smoking-related cancer among relatives of lung cancer patients. *Cancer Res* (2003) 63(23):8531–5.

25. Centers for Disease Control and Prevention. The health consequences of smoking. a report of the surgeon general, in: *Https://www.cdc.gov/tobacco/sgr/e-cigarettes/* (2016) (Accessed October 8).

26. Arnold BN, Thomas DC, Rosen JE, Salazar MC, Blasberg JD, Boffa DJ, et al. Lung cancer in the very young: Treatment and survival in the national cancer data base. *J Thorac Oncol* (2016) 11(7):1121–31. doi: 10.1016/j.jtho.2016.03.023

27. Zhu S, Ge T, Hu J, Jiang G, Zhang P. Prognostic value of surgical intervention in advanced lung adenocarcinoma: A population-based study. *J Thorac Dis* (2021) 13(10):5942–53. doi: 10.21037/jtd-21-997

28. Shyr BS, Huang CS, Chen HS, Hsu PK, Chiu CH, Hsu HS, et al. Sequence for surgical resection of primary lung tumor for oligometastatic non-small cell lung cancer. *Ann Thorac Surg* (2021) 113(4):1333–40. doi: 10.1016/j.athoracsur.2021.04.057

29. Jia J, Guo B, Yang Z, Liu Y, Ga L, Xing G, et al. Outcomes of local thoracic surgery in patients with stage iv non-Small-Cell lung cancer: A seer-based analysis. *Eur J Cancer* (2021) 144:326–40. doi: 10.1016/j.ejca.2020.12.002

30. Hao Z, Liang H, Zhang Y, Wei W, Lan Y, Qiu S, et al. Surgery for advanced-stage non-small cell lung cancer: Lobectomy or Sub-lobar resection? *Transl Lung Cancer Res* (2021) 10(3):1408–23. doi: 10.21037/tlcr-21-39

31. Fisch D, Bozorgmehr F, Kazdal D, Kuon J, Klotz LV, Shah R, et al. Comprehensive dissection of treatment patterns and outcome for patients with metastatic Large-cell neuroendocrine lung carcinoma. *Front Oncol* (2021) 11:673901. doi: 10.3389/fonc.2021.673901

32. Xu J, Fan L, Yu H, Lu D, Peng W, Sun G. Survival value of primary tumour resection for stage iv non-Small-Cell lung cancer: A population-based study of 6466 patients. *Clin Respir J* (2020). online ahead of print. doi: 10.1111/crj.13194

33. Le VH, Kha QH, Hung TNK, Le NQK. Risk score generated from ct-based radiomics signatures for overall survival prediction in non-small cell lung cancer. *Cancers (Basel)* (2021) 13(14):3616. doi: 10.3390/cancers13143616

34. Dang HH, Ta HDK, Nguyen TTT, Anuraga G, Wang CY, Lee KH, et al. Prospective role and immunotherapeutic targets of sideroflexin protein family in lung adenocarcinoma: Evidence from bioinformatics validation. *Funct Integr Genomics* (2022) 22(5):1057–72. doi: 10.1007/s10142-022-00883-3

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 02 December 2022
DOI 10.3389/fonc.2022.868216

OPEN ACCESS

EDITED BY
Ryogo Minamimoto,
National Center For Global Health and Medicine, Japan

REVIEWED BY
Xiaohu Li,
First Affiliated Hospital of Anhui Medical University, China
Asli Suner,
Ege University, Turkey

*CORRESPONDENCE
Xiqi Zhu
xiqi.zhu@163.com

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Thoracic Oncology, a section of the journal Frontiers in Oncology

RECEIVED 02 February 2022
ACCEPTED 14 November 2022
PUBLISHED 02 December 2022

CITATION
Mu R, Meng Z, Guo Z, Qin X, Huang G, Yang X, Jin H, Yang P, Deng M, Zhang X and Zhu X (2022) Diagnostic value of dual-layer spectral detector CT in differentiating lung adenocarcinoma from squamous cell carcinoma. *Front. Oncol.* 12:868216. doi: 10.3389/fonc.2022.868216

Diagnostic value of dual-layer spectral detector CT in differentiating lung adenocarcinoma from squamous cell carcinoma

Ronghua Mu[1†], Zhuoni Meng[1†], Zixuan Guo[1†], Xiaoyan Qin[2], Guangyi Huang[2], Xuri Yang[2], Hui Jin[2], Peng Yang[2], Meimei Deng[2], Xiaodi Zhang[3] and Xiqi Zhu[2*]

[1]Department of Radiology, Graduate School of Guilin Medical University, Guilin, China, [2]Department of Radiology, Nanxishan Hospital of Guangxi Zhuang Autonomous Region, Guilin, China, [3]Philips (China) Investment Co., Ltd., Chengdu Branch, Chengdu, China

Background and objective: The pathological type of non–small cell lung cancer is considered to be an important factor affecting the treatment and prognosis. The purpose of this study was to investigate the diagnostic value of spectral parameters of dual-layer spectral detector computed tomography (DLCT) in determining efficacy to distinguish adenocarcinoma (AC) and squamous cell carcinoma (SC), and their combined diagnostic efficacy was also analyzed.

Methods: This is a single-center prospective study, and we collected 70 patients with lung SC and 127 patients with lung AC confirmed by histopathological examination. Morphological parameters, plain scan CT value, biphasic enhanced CT value, and spectral parameters were calculated. The diagnostic efficiency of morphological parameters, spectral parameters, and spectral parameters combined with morphological parameters was obtained by statistical analysis.

Results: In univariate analysis, seven morphological CT features differed significantly between SC and AC: tumor location (distribution), lobulation, spicule, air bronchogram, vacuole sign, lung atelectasis and/or obstructive pneumonia, and vascular involvement (all $p < 0.05$). In the arterial phase and the venous phase, the spectral parameters of AC were higher than those of SC (AP-Zeff: 8.07 ± 0.23 vs. 7.85 ± 0.16; AP-ID: 1.41 ± 0.47 vs. 0.94 ± 0.28; AP-NID: 0.13 ± 0.04 vs. 0.09 ± 0.03; AP-λ: 3.42 ± 1.10 vs. 2.33 ± 0.96; VP-Zeff: 8.26 ± 0.23 vs. 7.96 ± 0.16; VP-ID: 1.18 ± 0.51 vs. 1.16 ± 0.30; VP-NID: 0.39 ± 0.13 vs. 0.29 ± 0.08; VP-λ: 4.42 ± 1.28 vs. 2.85 ± 0.72; $p < 0.001$). When conducting multivariate analysis combining CT features and DLCT parameters with the best diagnostic efficacy, the independent predictors of AC were distribution on peripheral (OR, 4.370; 95% CI, 1.485–12.859; p = 0.007), presence of air bronchogram (OR, 5.339; 95% CI, 1.729–16.484; p = 0.004), and presence of

vacuole sign (OR, 7.330; 95% CI, 1.030–52.184; p = 0.047). Receiver operating characteristic curves of the SC and AC showed that VP-λ had the best diagnostic performance, with an area under the curve (AUC) of 0.864 and sensitivity and specificity rates of 85.8% and 74.3%, respectively; the AUC was increased to 0.946 when morphological parameters were combined, and sensitivity and specificity rates were 89.8% and 87.1%, respectively.

Conclusion: The quantitative parameters of the DLCT spectrum are of great value in the diagnosis of SC and AC, and the combination of morphological parameters and spectral parameters is helpful to distinguish SC from AC.

KEYWORDS

lung cancer, pathological classification, dual layer detector, energy spectrum, X-ray computed tomography

Introduction

Lung cancer is a disease that seriously affects human health. Approximately 85% of patients are non–small cell lung cancer (NSCLC), of which adenocarcinoma (AC) and squamous cell carcinoma (SC) are the most common subtypes (1, 2). Histological classification of lung cancer has been demonstrated as an independent prognostic indicator. Because there are significant differences in biological behavior, treatment strategies and prognostic evaluation between different pathological subtypes of lung cancer (3, 4), a reasonable choice of treatment strategies can reduce mortality, prolong the survival time, and improve the quality of life.

Chest CT is the preferred imaging examination for the diagnosis of lung cancer. Traditional CT can evaluate the benign and malignant lesions according to their morphological features, intensity, and lymph node metastasis. However, it is difficult to evaluate the pathological subtypes of lung cancer because of the lack of quantitative indicators. If morphological criteria did not help to distinguish benign from malignant lung lesions, then it is strongly dependent on invasive pathologic examination or follow-up studies (5). As a non-invasive and effective pathological classification method of lung cancer, energy spectrum CT has become a hot spot in clinical research. The basic structure of the new generation of spectral CT is similar to that of ordinary CT, but it has two spatially equivalent detectors: the upper layer only absorbs low-energy photons, whereas the lower layer absorbs high-energy photons. On the premise of a perfect match in time and space within the projected data domain, high-energy and low-energy data can be parsed to obtain both traditional CT images and spectral images (6, 7).

Previous studies on energy spectrum CT in SC and AC are insufficient and controversial. Fehrenbach et al. (8) considered that only normalized iodine density (NID) in the arterial phase (AP) was significant for the differential diagnosis of NSCLC. Wang et al. (9) showed that the slope of the spectral curve (λ) of 40 to 70 keV and iodine density (ID) in AP had diagnostic value in differentiating SC from AC, but the venous phase (VP) parameters were not included in the study. In contrast, Li et al. (10) showed that VP-ID can differentiate SC from AC by reflecting tumor microvessel density. Jia et al. (11) showed that AP-λ (40–100), AP-Zeff, AP-ID, and VP-ID were significant for their identification, but these data of SC were greater than that of AC. Problems, such as single-phase scanning, single spectral parameters, and the exclusion of morphological parameters, have caused a lot of controversies.

This study improves the above issues. The purpose of this study was to investigate the diagnostic value of spectral parameters of DLCT in determining efficacy to distinguish AC and SC, and their combined diagnostic efficacy was also analyzed.

Materials and methods

Patients

A prospective study was conducted in our hospital between 1 August 2020 and 31 March 2021; a total of 656 consecutive patients who were diagnosed with lung cancer, confirmed by histopathological examination, were enrolled for the present study. To ensure the reliability of the results, all patients were required to conduct DLCT scan within 7 days before treatment. Inclusion criteria were as follows: (1) patients who underwent dual-layer detector spectral CT before treatment; (2) imaging, clinical, and pathological data were complete; (3) the short diameter of the tumor was greater than 2.0 cm; (4) no clinical antineoplastic treatment was given before enrollment; (5)

patients can undergo the CT scan with the breath-holding; and (6) no other cancer history. Exclusion criteria were as follows: (1) patients with hyperthyroidism were not cured; (2) history of allergy to iodine contrast medium; and (3) complication of the heart, liver, kidney, and other important organ damage. The flow diagram of the present study is shown in Figure 1.

This single-center prospective study was approved by the Hospital Institution Review Committee [no. 2020(KY-E-30)], and written informed consent was obtained from all of the participants.

CT examination

All scans were performed using the IQon Spectral CT device (Philips Healthcare). The scans included routine CT plain film and AP and VP enhancement scan of the lung. The field of vision ranges from apex to diaphragm level. The contrast agent iophorol (70 ml of iopromide; Beijing Beilu Pharmaceutical Co., Ltd.) was administered *via* an antecubital vein at an intended flow rate of 3.0 ml/s with a high-pressure syringe, followed by 30 ml of saline, injected at the same flow rate. Scans were performed using the bolus chase method. The region of interest (ROI) was located in the descending aorta with a trigger threshold of 150 Hu. AP was started at 6 s after contrast agent injection, and VP was performed at 36 s after contrast agent injection. Parameters included tube voltages of 120 kVp, spectral CT adaptive current, collimator width of 64 × 0.625 mm, pitch of 1.234, rotation time of 0.27 s, and matrix of 512 × 512. After the scans were completed, the data obtained in the enhanced double phases were reconstructed by projected spatial-spectral reconstruction (spectral, level 4). The image reconstruction image thickness was 1 mm, and the image spacing was 1 mm.

Imaging analysis

The images were sent to the Philips Spectral Diagnostic Suite 9.0 (Philips Healthcare) workstation, and all data were processed and analyzed by two radiologists with more than 5 years of chest CT diagnosis experience. They were blinded to any patient's clinical data to mitigating potential cognitive biases. Three consecutive image sections containing the maximum cross section of the tumor and the adjacent upper and lower levels were chosen for measurement; the round or oval ROI was drawn as large as possible (areas close to half to two-thirds of the lesion area) to minimize the influences of noise and the partial volume effect. Two radiologists first sketched the ROI on the conventional CT enhanced and unenhanced images, and the morphological parameters (including location, diameter, margin, internal, and other features) of the tumor were evaluated. The ROI was placed on solid regions of the tumor, avoiding areas with vessels, calcification, cystic, and necrotic change. They then obtained data from other parameter images [spectral CT-mono energetic (40 and 60 keV), Z effective (Zeff) images, ID images, etc.], and copy/paste functions were used to ensure the consistency of the size, shape, and position of the ROIs. Enhancement value was defined as the difference of the

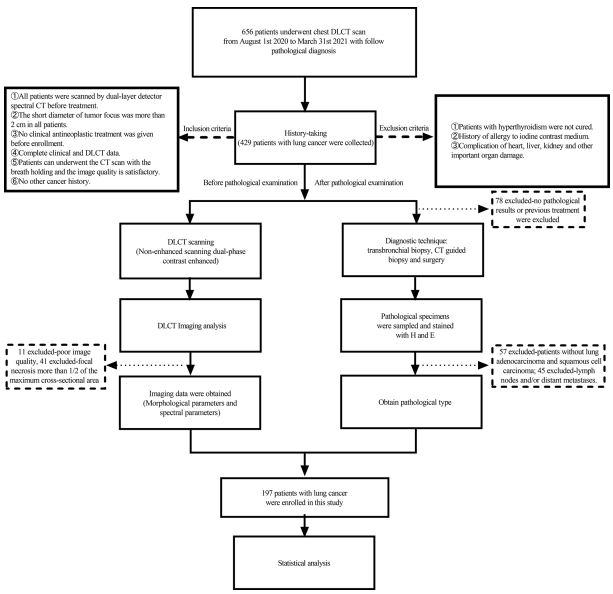


FIGURE 1
Study flowchart for the inclusion and exclusion criteria of the lung cancer patient sample.

CT value (AP/VP) and plain CT value according to the following formula: enhancement value (AP/VP) = CT value (AP/VP) − plain CT value. To minimize the influence of the individual circulation status and scanning times, the ID values of lung lesions were normalized to that of the aorta in the aortopulmonary window level to calculate the NID: NID = ID/ID_{aorta}. The slope of the spectral curve was defined as the difference of the CT value at 40 and 60 keV divided by the energy difference (60 − 40) according to the following formula: $\lambda = [CT_{(40)} - CT_{(60)}]/(60 - 40)$.

Histochemical examination

Tumor specimens were analyzed by a pathologist with 14 years of immunohistochemical staining experience. Without knowing clinical information and spectral CT results, the pathologist numbered and evaluated all sections and analyzed and recorded their pathological types. The histologic criteria used for diagnosis were in accordance with the World Health Organization Classification of Lung Tumors (12, 13).

Statistical analysis

Kappa and Intraclass correlation coefficient (ICC) tests were used to test the consistency of the measurement results of the two radiologists, with Kappa > 0.8 or ICC > 0.7 indicating satisfaction. All parameters were finally averaged over the measurements made by the two physicians. The normal distribution of all measured data was tested by the Shapiro–Wilk test. Measurements that follow a normal distribution were expressed as mean (standard deviation). In addition, the counting data were expressed by n (%). Independent sample t-test was adopted for all measurement parameters, and the chi-square test was adopted for counting data. For parameters that were significantly different, the results of the pathological examination were used as the gold standard, receiver operating characteristic (ROC) curves were drawn, and the diagnostic efficiency of each parameter is analyzed. First, univariate analysis regression model was established to analyze the correlation between morphological parameters and SC and AC. Then, multivariate logistic regression analyses were applied to identify independent predictors of SC or AC, with the model of CT features alone and CT features combined with optimal spectral parameters; factors with $p < 0.05$ on univariate analysis were used as the input variables for multiple logistic regression analysis, with the final model selected with the forward selection; model I (CT morphological parameters), model II (optimal spectral parameters), model III (morphological parameters and CT enhancement value), and model IV (morphological parameters and optimal spectral parameters) were established respectively. Finally, the ROC curves of the above models are drawn, and their diagnostic efficiency is analyzed. In the case of two-tailed p-value < 0.05, the difference was considered statistically significant.

All data were analyzed using SPSS 21.0.0 (IBM Corp., Armonk, NY, USA), and graphics were drawn using GraphPad Prism version 9.0.2 (GraphPad software Inc., San Diego, USA).

We conducted an *a priori* power analysis to test the adequacy of our sample size to independent sample t-test using G*Power (26). We specified an alpha level of 0.05, a 1-β error probability of 0.80, and an effect size (f = 0.50) for an estimated medium effect. The results of the analysis suggested a total recommended sample size of 128. A *post-hoc* power analysis revealed that a sample size of 197 (lung AC : SC = 27:70) resulted in a reported power of 0.917 to detect a medium effect (f = 0.50) with an alpha level of 0.05.

Results

Patients' information

Eleven patients with poor image quality, 45 patients with lymph nodes and/or distant metastases, 32 patients with focal necrosis greater than one-half of the maximum cross-sectional area, and 60 patients without pathological findings or with previous treatment were excluded. Finally, a total of 197 patients met the inclusion criteria, including 70 patients with SC (male patients, 91.4%) and 127 patients with AC patients (male patients, 52.8%). The clinical data are shown in Table 1.

CT characteristics of the SC and AC

Univariate analysis revealed that seven CT features differed significantly between SC and AC in Table 2: tumor location (distribution), lobulation, spicule, air bronchogram, vacuole sign, lung atelectasis and/or obstructive pneumonia, and vascular involvement (all $p < 0.05$). AC was more frequently found in the periphery, and the lobulation, spicule, air bronchogram, and vacuole sign were more likely to be observed among AC. Tumors with lung atelectasis and/or obstructive pneumonia and vascular involvement were more likely to be observed among SC. When it came to other CT features, no significant differences were noted between AC and SC. Table 3 shows the traditional CT quantitative parameters, and there was no significant difference in CT values during plain scan. AP-CT value of AC was higher than that of SC, and AP-enhancement value of AC was higher than that of SC. The VP-CT value of AC was higher than that of SC, and the VP-enhancement value of AC was higher than that of SC.

TABLE 1 Patient characteristics.

Variable	AC	SC	Statistic	*P*-value
No. of patients	127	70		
Age, mean (SD)	63.8 (9.5)	62.5 (8.4)	−0.969	0.334
Man	64.4 (9.0)	63.2 (8.1)		
Woman	63.3 (10.2)	55.0 (8.2)		
Sex, n (%)			30.294	<0.001
Male	67 (52.8)	64 (91.4)		
Female	60 (47.2)	6 (8.6)		
Diagnostic technique, n (%)			9.164	0.010
Transbronchial biopsy	11 (8.7)	17 (24.3)		
CT-guided biopsy	49 (38.6)	24 (34.3)		
Surgery	67 (52.8)	29 (41.4)		

AC, adenocarcinoma; SC, squamous cell carcinoma.

CT spectral quantitative parameters of the SC and AC

As shown in Table 3, all CT spectral parameters in the AC group in both the arterial and venous phases were higher than in the SC group (P<0.001).

Diagnostic implication

Because spectral quantitative parameters have statistically significant differences between SC and AC in the AP and the VP, diagnostic capacity was assessed using ROC curves in Table 4 and Figure 2. The spectral parameters in VP are generally better than those in AP in differentiating SC from AC. The VP-λ has the best diagnostic efficacy, with AUC = 0.864, 95% CI = 0.813~0.915, sensitivity = 85.8%, and specificity = 74.3%.

Multivariable analysis and joint diagnostic efficiency

In multivariate analysis, after adjusting for other confounding factors, models of distribution, air bronchus, vacuole sign, and VP-λ are associated with AC (Figure 3).

As shown in Table 5 and Figure 4, ROC curve analysis in diagnosing of the AC and SC using morphological parameters, optimum spectral parameter (VP-λ), combining morphological parameters with CT value, and combining morphological parameters with spectral parameters in the VP. When morphological parameters were combined with VP-λ, the diagnostic efficiency is the highest, with AUC = 0.946, 95% CI = 0.917–0.975, sensitivity = 89.8%, and specificity = 87.1%. Figure 5 shows the imaging and pathological images.

Reliability of measurements

For all the given parameters, the Kappa and ICC tests found that the measurement results were found to be consistent between the two radiologists ($p > 0.05$).

Discussion

Our results showed that the spectral parameters of the VP have a higher diagnostic value in differentiating SC from AC than those of the AP. Morphological parameters also have important diagnostic value, in which distribution, air bronchus, and vacuole sign are associated with AC. The joint diagnosis efficacy is higher than that of single index, and the joint diagnosis of spectral parameters and morphological parameters had the best efficacy. Therefore, the combined diagnosis of energy spectral parameters and morphological parameters is more effective and instructive for the pathological classification of lung cancer.

CT is widely used in the diagnosis of lung diseases. In recent years, spectral CT has become a hot spot in the study of lung cancer. However, in diagnostic studies of lung cancer (8–11), most researchers were keen to study the optimal diagnostic performance of quantitative parameters, ignoring the morphological features. Therefore, we have supplemented the study of the morphological features of tumors. Our results suggest that the morphological features of tumors can be used for histological inference of AC and SC, although some of the signs are not significant. Several previous studies have compared the clinical and imaging features of lung cancer of different pathological types. Chu et al. (14) showed that the radiologic and microscopic findings correlate well with each other and are closely associated with tumour prognosis. Understanding their morphological features is helpful for

TABLE 2 Morphological parameters in patients with AC and SC.

Characteristics	AC	SC	*P*-value	Univariate OR (95% CI)	*P*-value	Multivariable OR (95% CI)
Location, n (%)						
Distribution			**<0.001**		**0.007**	
Peripheral	83 (65.4)	12 (17.1)		9.117 (4.433, 18.751)		4.370 (1.485-12.859)
Central	44 (34.6)	58 (89.2)		Reference		Reference
Lobe location						
Right upper lobe	27 (21.3)	10 (14.3)	0.525	1.440 (0.468, 4.430)		
Right middle lobe	33 (26.0)	17 (24.3)	0.948	1.035 (0.366, 2.925)		
Right lower lobe	23 (18.1)	16 (22.9)	0.626	0.767 (0.263, 2.234)		
Left upper lobe	29 (22.8)	19 (27.1)	0.697	0.814 (0.289, 2.291)		
Left lower lobe	15(11.8)	8 (11.4)		Reference		
Diameter, mean (SD)						
Long-axis diameter (cm)	4.28 (1.67)	4.73 (1.92)	0.087	0.867 (0.736, 1.021)		
Short-axis diameter (cm)	3.00 (1.55)	3.19 (1.31)	0.394	0.918 (0.755, 1.117)		
Margin, n (%)						
Contour (Irregular)	59 (46.5)	41 (58.6)	0.105	0.614 (0.340, 1.107)		
Border definition (poorly)	65 (51.2)	46 (65.7)				
Lobulation (Y)	76 (59.8)	30 (42.9)	**0.023**	1.987 (1.100, 3.590)	0.528	1.406 (0.448–4.049)
Spicule (Y)	71 (55.9)	25 (35.7)	**0.007**	2.282 (1.251, 4.164)	0.198	2.037 (0.690–6.017)
Internal, n (%)						
Air bronchogram (Y)	73 (57.5)	22 (31.4)	**0.001**	2.949 (1.594, 5.456)	**0.004**	5.339 (1.729–16.484)
Vacuole sign (Y)	28 (22.0)	3 (4.3)	**0.003**	6.316 (1.846, 21.618)	**0.047**	7.330 (1.030–52.184)
Vessel convergence (Y)	86 (67.7)	44 (62.9)	0.491	1.239 (0.673, 2.284)		
Liquefactive necrosis (Y)	51 (40.2)	36 (51.4)	0.128	0.643 (0.352, 1.141)		
Calcification (Y)	15 (11.8)	15 (21.4)	0.076	0.491 (0.224, 1.077)		
Other, n (%)						
Pleural indentation (Y)	54 (42.5)	23 (32.9)	0.185	1.512 (0.821, 2.783)		
Mediastinal/hilar lymphadenopathy (Y)	60 (47.2)	25 (35.7)	0.119	1.612 (0.884, 2.938)		
Lung atelectasis and/or obstructive pneumonia (Y)	11 (8.7)	16 (22.9)	**0.007**	0.320 (0.139, 0.736)	0.536	1.778 (0.287–11.003)
Vascular involvement (Y)	12 (9.4)	22 (31.4)	**<0.001**	0.228 (0.104, 0.497)	**0.047**	5.121 (1.018–25.749)

AC, adenocarcinoma; SC, squamous cell carcinoma; Y, yes.
Values in bold indicate statistical significance, $p < 0.05$.

early identification and diagnosis. Wang et al. (15) have shown that AC and SC have different clinical and imaging characteristics and that imaging features are useful for differential diagnosis. Koenigkam et al. (16) found that SC more likely to occur in male patients, more frequently found in the central distribution, and less lobulation, burr, and air bronchogram. In addition, previous studies did not analyze the contrast enhancement between AC and SC, but we did a detailed analysis. We found the differences between them and verified the diagnostic value of morphological features.

Conventional CT is used to differentiate subtypes of tumors by their size, shape, and attenuation value. Although it has some advantages, the accuracy of diagnosis is still limited (11, 17). The spectral parameters provide quantitative analysis for histological diagnosis of NSCLC, and its accuracy is higher than that of conventional CT, making up for the deficiency of conventional CT. There are several reasons: i. From the perspective of histology, AC is a malignant epithelial neoplasm of glandular differentiation, in which glandular tubular and glandular luminal–like structures are common, and it contains more microvessels with abundant blood supply (18). Whereas, SC is defined as keratinized malignant epithelial tumor, which mostly grows in a stacked pattern, and its internal structure is dense, with frequent tumor nests, keratinized beads, and intercellular bridges, and the internal part is mostly accompanied by liquefactive necrosis (18, 19). ii. The degree of tumor enhancement depends only on the amount of contrast medium in the tumor vessels. The central type of lung cancer is mainly SC, mainly supplied by bronchial artery, whereas the peripheral type of lung cancer is mainly AC with dual blood supply of bronchial artery and pulmonary artery (20, 21). Because the blood supply from the bronchial artery is later

TABLE 3 Conventional CT enhancement values and spectral parameters in patients with AC and SC.

	AC	SC	t-value	*P*-value
Plain CT value, mean (SD)	34.12 (12.75)	30.07 (13.10)	0.545	0.586
AP, mean (SD)				
AP-CT value (Hu)	76.37 (12.17)	67.43 (11.67)	5.006	<0.001
AP-enhancement value (Hu)	40.44 (9.02)	34.36 (15.22)	3.520	<0.001
Zeff	8.07 (0.23)	7.85 (0.16)	8.024	<0.001
ID	1.41 (0.47)	0.94 (0.28)	8.683	<0.001
NID	0.13 (0.04)	0.09 (0.03)	5.482	<0.001
λ	3.42 (1.10)	2.33 (0.69)	8.579	<0.001
VP, mean (SD)				
VP-CT value (Hu)	87.47 (14.60)	75.03 (11.55)	6.148	<0.001
VP-enhancement value (Hu)	51.54 (12.94)	41.95 (16.51)	4.504	<0.001
Zeff	8.26 (0.23)	7.96 (0.16)	10.962	<0.001
ID	1.78 (0.51)	1.16 (0.30)	10.859	<0.001
NID	0.39 (0.13)	0.29 (0.08)	7.354	<0.001
λ	4.42 (1.28)	2.85 (0.72)	11.009	<0.001

AC, adenocarcinoma; SC, squamous cell carcinoma; AP, arterial phase; VP, venous phase; Zeff, effective atomic number; ID, iodine density; NID, normalized iodine density; λ, the slope of spectral curve.

than that from the pulmonary artery, the VP better reflects the blood supply of the tumor.

ID reflects intravascular blood flow distribution and vascular status by quantitative analysis of iodine content, but it is influenced by many factors, including cardiac output and blood volume of the patient, concentration and flow rate of the contrast medium, and dose and rate of injection (20). NID was defined as the ratio of tumor ID to that of the aorta or subclavian arteries at the same level, and most researchers (8, 21–23) state that NID can reduce the effect of individual circulatory variability on tumor iodine content and thus more accurately reflect the blood supply of the lesion. However, in this comparison of AC and SC, we found that ID is more effective than NID in both the AP and the VP, which is consistent with the results of the study by Li et al. (10). This suggests that NID depends on the extent of lesion and aortic enhancement, and changes in aortic ID may cause NID to deviate.

Energy spectrum curve is the curve of material decay varying with x-ray energy, which reflects the decay characteristics of material. Theoretically, each substance has its own specific

TABLE 4 Performance of differential parameters in distinguishing the AC and SC in the receiver operating characteristic analysis.

Parameter	AUC	Thresholds	Sensitivity (%)	Specificity (%)	95% CI	*P*-value
AP						
CT value	0.704	69.47	72.4	64.3	0.692~0.779	<0.001
CT enhancement value	0.721	36.50	63.8	75.7	0.641~0.802	<0.001
Zeff	0.798	7.975	71.7	82.9	0.735~0.860	<0.001
ID	0.811	1.145	71.7	82.9	0.737~0.864	<0.001
NID	0.753	0.110	68.5	74.3	0.683~0.823	<0.001
λ	0.808	2.800	71.7	84.3	0.747~0.869	<0.001
VP						
CT value	0.760	82.13	69.3	78.6	0.691~0.828	<0.001
CT enhancement value	0.733	48.65	65.4	81.4	0.656~0.811	<0.001
Zeff	0.862	8.055	82.7	75.7	0.810~0.913	<0.001
ID	0.857	1.265	83.5	74.3	0.805~0.909	<0.001
NID	0.787	0.297	79.5	65.7	0.721~0.852	<0.001
λ	0.864	3.088	85.8	74.3	0.813~0.915	<0.001

AP, arterial phase; VP, venous phase; Zeff, effective atomic number; ID, iodine density; NID, normalized iodine density; λ, the slope of spectral curve.

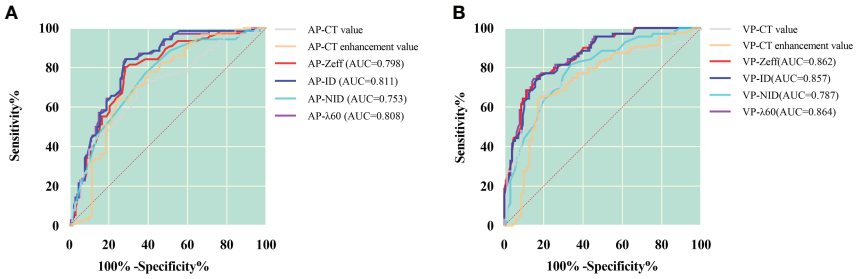


FIGURE 2
ROC curves for dual-layer spectral detector CT quantitative parameters between squamous cell carcinoma and adenocarcinoma in the arterial phase **(A)** and the venous phase **(B)**. In the arterial phase **(A)**, the ID has the best diagnostic efficiency; AUC, 0.811; sensitivity and specificity, 71.7% and 82.9%, respectively; in the venous phase **(B)**, the λ has the best diagnostic efficiency; AUC, 0.864; sensitivity and specificity, 85.8% and 74.3%, respectively. AP, arterial phase; VP, venous phase; Zeff, effective atomic number; ID, iodine density; NID, normalized iodine density; λ, slope of 40- to 60-keV spectral curve.

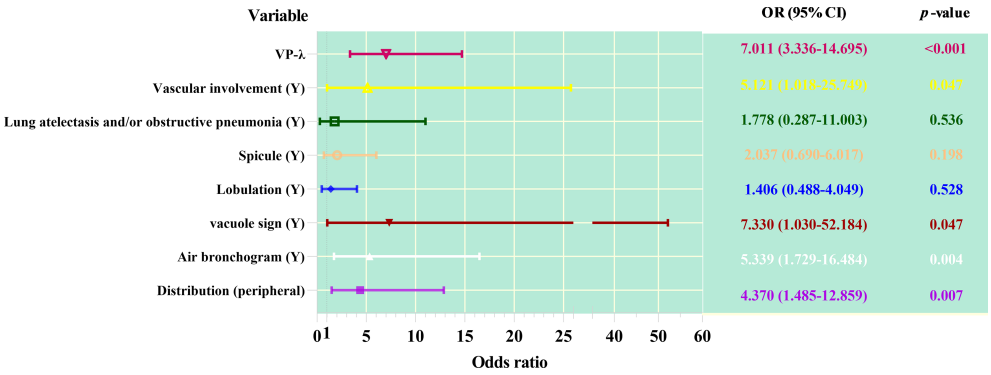


FIGURE 3
Forest plot of multivariable regression analysis of dual-layer spectral detector CT metrics combined with CT features predicting adenocarcinoma from squamous cell carcinoma.

TABLE 5 Different models in distinguishing the AC and SC in the receiver operating characteristic analysis.

Parameter	AUC	Thresholds	Sensitivity (%)	Specificity (%)	95% CI	*P*-value
Model I	0.863	—	83.5	71.4	0.813~0.914	<0.001
Model II	0.864	3.088	85.8	74.3	0.813~0.915	<0.001
Model III	0.899	—	73.2	88.6	0.857~0.941	<0.001
Model IV	0.946	—	89.8	87.1	0.917~0.975	<0.001

Model I, morphological parameters; Model II, λ in the venous phase; Model III, combination of morphology and CT value in the venous phase; Model IV, combination of morphology and λ in the venous phase.

spectrum curve, and the slope of the spectrum curve decreases with the expansion of its range (24). The K edge of iodine is 33.2 keV, so the lower the keV, the higher the CT value (25). From IQon, monochrome images of the 40- to 200-keV energy range can be obtained. However, not all images can be used to observe and diagnose lesions, because image quality varies at different energy levels. Low–kilo–electron volt images can improve tissue enhancement, increase tissue contrast, and make the display of small lesions clearer. Equivalent kilo–electron volt images can reduce image noise and improve image quality. This study shows that VP-λ is conducive to the differentiation of lung SC and AC. Jia et al. (9) reported that the spectral parameters of SC in the AP

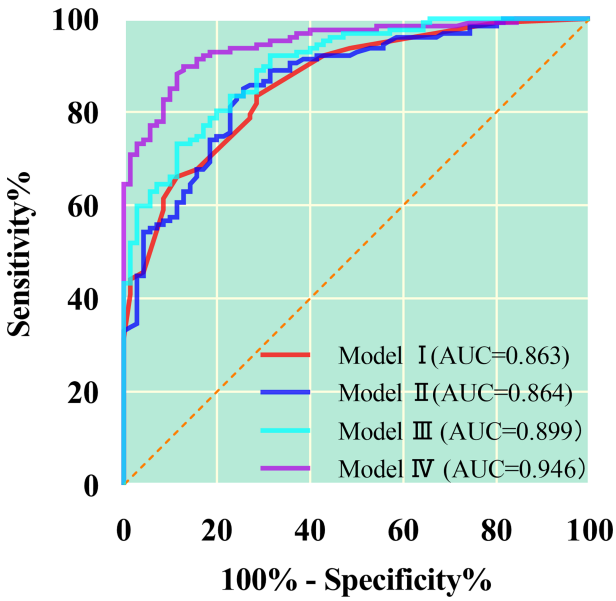


FIGURE 4
The quantitative parameters of different models were used to identify the lung adenocarcinoma from squamous cell carcinoma. The combination of morphology and λ in the venous phase has the greatest diagnostic efficiency; AUC, 0.946; sensitivity and specificity, 89.8% and 87.1%, respectively. Model I, morphological parameters; Model II;, λ in the venous phase; Model III, combination of morphology and CT value in the venous phase; Model IV, combination of morphology and λ in the venous phase.

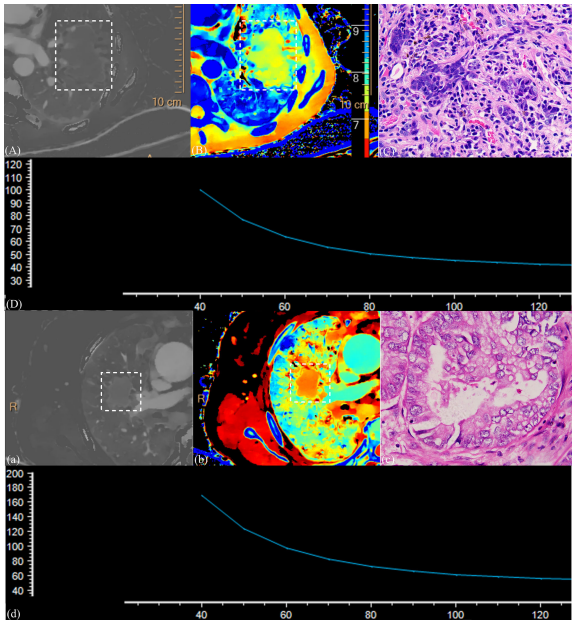


FIGURE 5
(1) CT images and the pathological section of a 58-year-old man with squamous cell carcinoma in the venous phase **(A–D)** and a 63-year-old man with adenocarcinoma in the venous phase **(a–d)**. Iodine density images **(A, a)**; Z-effective images **(B, b)**; the dashed box point to lung cancer tissue. Magnification: C, ×600; slope of spectral curve in the arterial phase; the horizontal axis represents the energy (keV), and the vertical axis represents mean attenuation (Hu) **(D, d)**. λ = 1.72, ID = 0.70, NID = 0.19, and Zeff = 7.71 for squamous cell carcinoma, and λ = 3.91, ID = 1.59, NID = 0.31, and Zeff = 8.19 for adenocarcinoma. Significant differences were observed in all parameters representing these cases.

and the VP were larger than those of AC, and this is different from our experimental results.

Zeff is the atomic number of an element that has the same decay coefficient as a compound or mixture and can be used to identify the tissue composition of a substance, especially in substances with similar CT values. It is a quantitative indicator of different substances (11, 26). Moreover, Zeff can indirectly provide information about the accumulation of contrast media (27). Previous studies have shown that Zeff can identify substances (28, 29). In the differential diagnosis between AC and SC, Zeff in VP was greater than that in AP. This is related to the blood supply of the tumors described above, their histological characteristics, the growth pattern of the tumors, and the changes of their surrounding microenvironment.

In this study, we provide not only quantitative parameters for tumor enhancement but also qualitative parameters for morphological features. The diagnostic ability of combining morphological features with quantitative parameters, especially DLCT, has been significantly improved, which confirms the additional diagnostic value of quantitative analysis of spectral parameters. Although there are more and more studies on quantitative data, CT morphology of lung cancer remains an important indicator of cancer diagnosis. On the one hand, they are closely related to the growth characteristics of tumors and easy to collect. On the other hand, in the era of precision medicine and big data, a single spectral parameter may also be considered unreliable.

The limitations of this study are as follows: (1) because of the small sample size, this study failed to stratify the degree of differentiation of tumors; (2) unenhanced spectral parameters were not included in the study; (3) not all pathological sections were matched with imaging ROI; and (4) some patients' unenhanced images were replaced by virtual unenhanced images, which may affect our results.

In summary, the combination of morphological features and DLCT spectral parameters improved the diagnostic efficiency in distinguishing SC from AC. This may help clinicians develop initial treatment strategy and prognostic predictions. Although relatively accurate pathological diagnosis can be obtained by invasive methods, complications or patient intolerance still exist.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Nanxishan Hospital of Guangxi Zhuang Autonomous Region Institution Review Committee [NO.2020 (KY-E-30)]. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

First author: RM. Co-author: ZM; ZG. Other authors: XQ; GH; XY; HJ; PY; MD; XDZ. Guarantor and correspondent: XQZ. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors thank the dedicated this project participants, their loved ones, and the devoted staff and trainees who contributed to recruitment, screening, and enrollment of the cohort.

Conflict of interest

Author XZ is employed by Philips China Investment Co.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Jia Y, Xiao X, Sun Q, Jiang H. PRR11 induces filopodia formation and promotes cell motility *via* recruiting ARP2/3 complex in non-small cell lung cancer cells. *Genes Dis* (2021) 9(1):230–44. doi: 10.1016/j.gendis.2021.02.012

2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660

3. Zheng XQ, Huang JF, Lin JL, Chen L, Zhou TT, Chen D, et al. Incidence, prognostic factors, and a nomogram of lung cancer with bone metastasis at initial diagnosis: A population-based study. *Transl Lung Cancer Res* (2019) 8(4):367–79. doi: 10.21037/tlcr.2019.08.16

4. Saijo N, Niitani H, Tominaga K, Eguchi K, Koketsu H, Fujino T, et al. Comparison of survival in nonresected well differentiated and poorly differentiated adenocarcinoma of the lung. *J Cancer Res Clin Oncol* (1980) 97(1):71–9. doi: 10.1007/BF00411280

5. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: From the fleischner society 2017. *Radiology* (2017) 284(1):228–43. doi: 10.1148/radiol.2017161659

6. Kikano EG, Rajdev M, Salem KZ, Laukamp K, Felice CD, Gilkeson RC, et al. Utility of iodine density perfusion maps from dual-energy spectral detector CT in evaluating cardiothoracic conditions: A primer for the radiologist. *AJR Am J Roentgenol* (2020) 214(4):775–85. doi: 10.2214/AJR.19.21818

7. van Hamersvelt RW, Išgum I, de Jong PA, Cramer MJM, Leenders GEH, Willemink MJ, et al. Application of speCtraL computed tomograPhy to impRove specIficity of cardiac compuTed tomographY (CLARITY study): Rationale and design. *BMJ Open* (2019) 9(3):e025793. doi: 10.1136/bmjopen-2018-025793

8. Fehrenbach U, Kahn J, Böning G, Feldhaus F, Merz K, Frost N, et al. Spectral CT and its specific values in the staging of patients with non-small cell lung cancer: Technical possibilities and clinical impact. *Clin Radiol* (2019) 74(6):456–66. doi: 10.1016/j.crad.2019.02.010

9. Wang G, Zhang C, Li M, Deng K, Li W. Preliminary application of high-definition computed tomographic gemstone spectral imaging in lung cancer. *J Comput Assist Tomogr* (2014) 38(1):77–81. doi: 10.1097/RCT.0b013e3182a21633

10. Li Q, Li X, Li XY, Huo JW, Lv FJ, Luo TY. Spectral CT in lung cancer: Usefulness of iodine concentration for evaluation of tumor angiogenesis and prognosis. *AJR Am J Roentgenol* (2020) 215(3):595–602. doi: 10.2214/AJR.19.22688

11. Jia Y, Xiao X, Sun Q, Jiang H. CT spectral parameters and serum tumour markers to differentiate histological types of cancer histology. *Clin Radiol* (2018) 73 (12):1033–40. doi: 10.1016/j.crad.2018.07.104

12. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 world health organization classification of lung tumors: Impact of genetic, clinical and radiologic advances since the 2004 classification. *J Thorac Oncol* (2015) 10(9):1243–60. doi: 10.1097/JTO.0000000000000630

13. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol* (2015) 10(9):1243–60. doi: 10.1097/JTO.0000000000000630

14. Chu ZG, Yang ZG, Shao H, Zhu ZY, Deng W, Tang SS, et al. Small peripheral lung adenocarcinoma: CT and histopathologic characteristics and prognostic implications. *Cancer Imaging* (2011) 11(1):237–46. doi: 10.1102/1470-7330.2011.0033

15. Wang Z, Li M, Huang Y, Ma L, Zhu H, Kong L, et al. Clinical and radiological characteristics of central pulmonary adenocarcinoma: A comparison with central squamous cell carcinoma and small cell lung cancer and the impact on treatment response. *Onco Targets Ther* (2018) 11:2509–17. doi: 10.2147/OTT.S154385

16. Koenigkam Santos M, Muley T, Warth A, de Paula WD, Lederlin M, Schnabel PA, et al. Morphological computed tomography features of surgically resectable pulmonary squamous cell carcinomas: Impact on prognosis and comparison with adenocarcinomas. *Eur J Radiol* (2014) 83(7):1275–81. doi: 10.1016/j.ejrad.2014.04.019

17. Xu X, Sui X, Zhong W, Xu Y, Wang Z, Jiang J, et al. Clinical utility of quantitative dual-energy CT iodine maps and CT morphological features in distinguishing small-cell from non-small-cell lung cancer. *Clin Radiol* (2019) 74 (4):268–77. doi: 10.1016/j.crad.2018.10.012

18. Mairinger T. [Histology, cytology and molecular diagnostics of lung cancer]. *Pathologe* (2019) 40(6):649–61. doi: 10.1007/s00292-019-00677-8

19. Zheng M. Classification and pathology of lung cancer. *Surg Oncol Clin N Am* (2016) 25(3):447–68. doi: 10.1016/j.soc.2016.02.003

20. Milne EN. Circulation of primary and metastatic pulmonary neoplasms. A postmortem microarteriographic study. *Am J Roentgenol Radium Ther Nucl Med* (1967) 100(3):603–19. doi: 10.2214/ajr.100.3.603

21. Cheng Z, Wang Y, Yuan M, Liang J, Feng Y, Shi Y, et al. CT perfusion imaging can detect residual lung tumor early after radiofrequency ablation: A preliminary animal study on both tumoral and peri-tumoral region assessment. *J Thorac Dis* (2022) 14(1):64–76. doi: 10.21037/jtd-21-967

22. Zhang M, Kono M. Solitary pulmonary nodules: Evaluation of blood flow patterns with dynamic CT. *Radiology* (1997) 205(2):471–8. doi: 10.1148/radiology.205.2.9356631

23. Wu F, Zhou H, Li F, Wang JT, Ai T. Spectral CT imaging of lung cancer: Quantitative analysis of spectral parameters and their correlation with tumor CharacteristIDs. *AcademID Radiol* (2018) 25(11):1398–404. doi: 10.1016/j.acra.2018.04.017

24. Lin LY, Zhang Y, Suo ST, Zhang F, Cheng JJ, Wu HW. Correlation between dual-energy spectral CT imaging parameters and pathological grades of non-small cell lung cancer. *Clin Radiol* (2018) 73(4):412.e1–7. doi: 10.1016/j.crad.2017.11.004

25. Kulpe S, Dierolf M, Günther B, Busse M, Achterhold K, Gleich B, et al. K-Edge subtraction computed tomography with a compact synchrotron X-ray source. *Sci Rep* (2019) 9(1):13332. doi: 10.1038/s41598-019-49899-z

26. Goodsitt MM, Christodoulou EG, Larson SC. Accuracies of the synthesized monochromatic CT numbers and effective atomic numbers obtained with a rapid kVp switching dual energy CT scanner. *Med Phys* (2011) 38(4):2222–32. doi: 10.1118/1.3567509

27. Deniffel D, Sauter A, Fingerle A, Rummeny EJ, Makowski MR, Pfeiffer D. Improved differentiation between primary lung cancer and pulmonary metastasis by combining dual-energy CT-derived biomarkers with conventional CT attenuation. *Eur Radiol* (2021) 31(2):1002–10. doi: 10.1007/s00330-020-07195-9

28. Ju Y, Liu A, Dong Y, Liu Y, Wang H, Sun M. The value of nonenhanced single-source dual-energy CT for differentiating metastases from adenoma in adrenal glands. *Acad Radiol* (2015) 22(7):834–9. doi: 10.1016/j.acra.2015.03.004

29. Gümüştaş S, Inan N, Akansel G, Ciftçi E, Demirci A, Ozkara SK. Differentiation of malignant and benign lung lesions with diffusion-weighted MR imaging. *Radiol Oncol* (2012) 46(2):106–13. doi: 10.2478/v10019-012-0021-3

Frontiers in Oncology

Advances knowledge of carcinogenesis and tumor progression for better treatment and management

The third most-cited oncology journal, which highlights research in carcinogenesis and tumor progression, bridging the gap between basic research and applications to imrpove diagnosis, therapeutics and management strategies.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact