doi:10.4149/neo_2021_210425N567

Development and validation of a multivariable predictive model for EGFR gene mutation status in patients with lung adenocarcinoma

Yiping SHI[1,#], Qiongfang ZHA[2,#], Qing YE[2], Shan XUE[2], Huawei WU[3], Daoqiang TANG[4], Hui QIN[2,*], Jing ZOU[2,*]

[1]Department of Nuclear Medicine, Renji Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, China; [2]Department of Respirology, Renji Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, China; [3]Department of Radiology, Renji Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, China; [4]Department of Pathology, Renji Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, China

**Correspondence: drqinhui@126.com; drzouj@163.com*
[#]Contributed equally to this work.

Received April 25, 2021 / Accepted July 20, 2021

Detection of epidermal growth factor receptor (EGFR) is one real dilemma owing to the non-sufficient tissue for testing EGFR mutations in lung adenocarcinoma. A model for predicting EGFR mutations would be helpful for clinical decisions in those patients. A retrospective cohort of 1,196 patients diagnosed with lung adenocarcinoma was investigated between December 1, 2017, and December 31, 2019, in Renji Hospital, Shanghai, China. All patients were tested for EGFR mutations (amplification refractory mutation system, n=1,144; next-generation sequencing, n=52). Of 1,196 patients with lung adenocarcinoma, 944 met the inclusion criteria. A nomogram model was developed based on 567 patients and validated in 377 patients. Variables associated with EGFR mutations were age, sex, smoking history, lepidic predominant subtype, solid predominant subtype, mucinous adenocarcinoma, Ki67 expression, lobulation, solid texture in radiology, and pleural retraction. The nomogram based on the model performed well in the development group (c-index 0.789, 95% CI: 0.751–0.827), and the validation group (c-index 0.809, 95% CI: 0.771–0.847). At the probability cut-point of 0.7, the diagnostic efficiency was 82.7% in patients with NGS liquid biopsy. Decision curve analysis further confirmed the clinical usefulness of the nomogram, which showed that predicting the EGFR mutations probability applying this nomogram would be better than having all patients or none patients use this nomogram. A high probability group (>0.7) by nomogram model may suggest a high possibility of EGFR mutation, if tissue is limited, NGS-based ctDNA with liquid biopsy could be implemented effectively.

Key words: adenocarcinoma, lung cancer, epidermal growth factor receptor, mutation, predictive model

The incidence and mortality of lung cancer are on the top ranking of cancer reported by Global cancer statistics 2018 [1]. Lung adenocarcinoma which accounts for more than 40% of lung cancer [2] is the most common histologic subtype of non-small cell lung cancer (NSCLC) [3, 4]. It has been confirmed that NSCLC patients with a somatic mutation in the epidermal growth factor receptor (EGFR) after treatment with EGFR tyrosine kinase inhibitors (TKIs) present higher responsive rate, prolonged progression-free survival (PFS), and health-related quality of life improvement, in comparison with those who received standard chemotherapy [5–10]. The prevalence of EGFR mutations in Asians is 38.4%, which is the highest prevalence in the world [11]. Due to the fact that lung adenocarcinoma has a high frequency of EGFR mutations compared with other histologic subtypes of NSCLC, clinical practice guidelines recommended EGFR mutation testing once lung adenocarcinoma was diagnosed [12, 13–15]. However, a systematic review recently reported that only 31% of over 50,000 patients from 18 eligible studies were tested for EGFR mutations [16]. There are two major causes that may directly restrict the application of EGFR mutation testing; one is the tissue sample availability and adequacy, the other is the costs of testing. The use of pretest probability of EGFR mutations from universally available factors has been suggested as a potential accessory for situations when EGFR mutation results cannot be obtained because of limited testing resources or limited tissue [17]. EGFR mutation status has been found to have substantial correlations with histological subtypes of lung adenocarcinoma [18–21]. And likewise, the computer tomographic (CT) characteristics had a relationship with EGFR mutations [22–26]. A model with both clinical characteristics and CT features appeared to be a more reliable tool for predicting EGFR mutations probability. However, until now, no nomograms which combined

clinical characteristics, lung adenocarcinoma histological characteristics and CT features have been developed, and no proper methods for the assessment of the clinical utility for a risk model have been reported.

The aim of this study was to generate a multivariate logistic regression model and an associated nomogram, based on clinical characteristics, histological characteristics, and CT features to predict the probability of EGFR mutations in lung adenocarcinoma, which would assist clinical decisions for those who have limited tissue for further EGFR mutation testing.

Patients and methods

Source of data. The Ethics Committee of Renji Hospital approved the study (KY2021-003). The Ethics Committee waived the requirement for informed consent of the patients because the study had a non-interventional retrospective design and all data were analyzed anonymously. We followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) [27].

Participants. We included patients age 18 years and older who met the inclusion criteria: 1) available clinical data, including age, sex, smoke history (patients were classified as nonsmokers if they had never smoked, or smokers if they were former or current smokers), and TNM stage [28]; 2) preoperative thin-section CT images were acquired with our picture archiving and communication system (the interval between CT and subsequent surgery was less than one month); 3) available pathology reports with a diagnosis of lung adenocarcinoma; 4) available amplification refractory mutation system (ARMS) test results for EGFR mutation status.

Outcome. EGFR mutations were examined with ARMS using the Human EGFR Gene Mutation Detection Kit (Amoy Diagnostics Co. Ltd., Xiamen, China). Molecular analyses of the mutation status of EGFR exons 18, 19, 20, and 21 were performed. The positive EGFR mutation in this study refers to any positive mutation in the EGFR test (including exon 18 G719X, exon 19 deletions, exon 20 T790M, exon 20 insertions, exon 21 L858R, exon 21 L861Q). Our goal was to develop a nomogram model based on the clinical characteristic, histological characteristics, and CT features for clinicians when EGFR mutation testing resources are limited or limited tissue is available, and thus the visualization model may be helpful to make a clinical decision.

Predictors. We chose to focus on predictors that were objective, readily available, and required little computation so that the nomogram model could be calculated by hand at the point of evaluation. We extracted age, sex, smoking history (patients were classified as nonsmokers if they had never smoked, or smokers if they were former or current smokers), histological results, tumor marker, and TNM stage. Laboratory analysis of tumor markers including carcinoembryonic antigen (CEA), carbohydrate antigen199 (CA199), carbohydrate antigen125 (CA125), and cytokeratin-19 fragment (Cyfra 21-1) were done via routine blood tests within 1 week before surgery.

We extracted the patient's chest CT parameters including tumor distribution, lobe location, size category, long axis diameter, short axis diameter, lobulation, spiculation, texture (including pure GGO, mixed GGO, solid texture), calcification, CT value, air bronchogram, bubblelike lucency, peripheral emphysema, vascular convergence and pleural retraction. Chest CT examinations were performed by using one of two multidetector CT (Lightspeed16, GE Healthcare, Milwaukee, WI, USA; Discovery CT750 HD, GE Healthcare). The images were reviewed in random order by two independent radiologists with 15 and 20 years of experience. Both of them were blinded to clinical and histologic findings. The majority class was used as the final CT feature value in case of disagreement. Mean values were used for continuous variables. CT images were read with both mediastinal (width, 350 HU; level, 40 HU) and lung (width, 1500 HU; level, -600 HU) window settings. And findings were agreed upon by consensus between the two radiologists.

Sample size. Our dataset comprised of a development group to derive a nomogram model and a validation group to validate the model. To ensure that the sample size in the development group was sufficient for the estimates for p-values to be valid, we applied the rule that the number of events (positive for the outcome of EGFR mutation) and non-events (negative for the composite) per covariate in the model should be at least 10 [29–31]. We aimed for a nomogram model with no more than 10 predictors (and thus no more than 10 "covariates"). Thus, our development group would require at least 100 events and 100 non-events. Within 567 patients in the development group (341 events and 226 non-events), we had sufficient patients to validly estimate the beta coefficients in the logistic regression model.

Statistical analysis. We report descriptive statistics on all variables, reporting median and interquartile range or frequencies and percentages. The study sample was divided into a 60% group for developing and a 40% group for validation by time. Data from 567 patients diagnosed from 1 Dec 2017 to 28 Feb 2019 were used for model development (development group). And the model was further validated in another 377 patients diagnosed from 1 Mar 2019 to 31 Dec 2019 (validation group).

We used Fisher's exact test to compare the categorical variables between different groups and the Mann-Whitney U test to compare differences between the two groups for continuous variables. The Fisher's exact test and Mann-Whitney U test were applied for univariate analysis. Next, we screened out the optimal variables with nonzero coefficients as potential predictors of this prediction model using the least absolute shrinkage and selection operator (LASSO) method [32]. Then multivariable logistic regression (stepwise backward logistic regression) analysis was applied to construct the predictive model based on the results of LASSO regres-

sion and a further nomogram was developed. The prediction efficiency of this predictive model was assessed by C-index and AUC as well as calibration curves in both development group and validation group. Calibration was considered poor if the p-value was less than 0.05 by Hosmer-Lesmeshow test. Decision Curve Analysis (DCA) curve was also performed to determine the clinical value of the predictive model by quantifying the net benefit at disparate threshold probabilities. All statistical analyses were conducted using R software (version 3.6.3 (http://www.Rproject.org)). Statistical significance was decided by a criterion of two-sided p<0.05.

Results

Patient characteristics. A total of 1,196 patients with lung adenocarcinoma had done the EGFR mutation test, 944 met the criteria for inclusion, and 567 were included in the development group (Figure 1). All patients were ethnically Asian. The EGFR mutation rate was 61.2% (578 of 944). The most common EGFR mutation types were exon 21 L858R (327 of 578, 56.6%) and exon 19 deletion (200 of 578, 34.6%). In total, the age range was 63 (54–69) yrs, female patients (558 of 944, 59.1%), and non-smokers (764 of 944, 80.9%) were predominant in our study.

Patients in the development group had a median age of 63 (55–68) yrs, and 59.6% female, 79.4% without smoke history, 48% acinar predominant subtype, 83.8% low Ki67 expression, 68.2% lobulation, 46.4% pleural retraction (Tables 1 and 2).

Of the patients in the development group, 60.1% were with EGFR mutation (55.1% L858R, 36.1% 19-deletion). With positive EGFR mutation, there were 69.2% female, 88.3% without smoke history, 88.9% invasive adenocarcinoma, 55.4% acinar predominant subtype, 89.7% low Ki67 expression, 73.9% lobulation, 51.9% mixed GGO, lower CT value (–237 vs –134), 67.2% airbronchogram, and 51.6% pleural retraction.

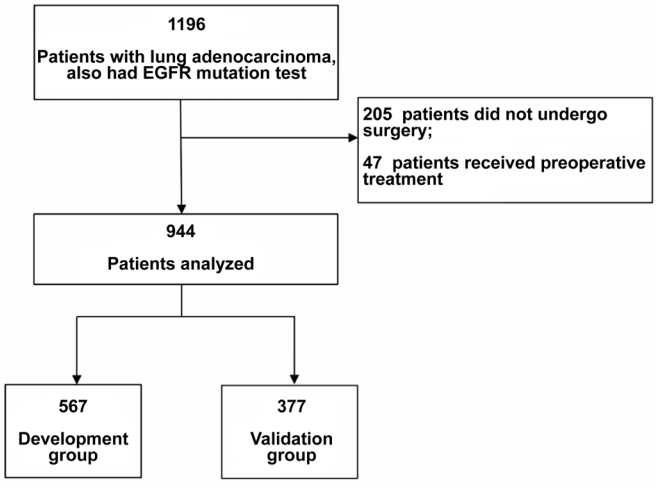

Figure 1. Sample selection schematic, lung adenocarcinoma with EGFR mutation test results.

Model development. The 16 variables were significantly correlated with the EGFR mutations via univariate analysis (Tables 1 and 2) including age, sex, smoke history, predominant subtypes (acinar, lepidic, solid, micropapillary predominant), mucinous adenocarcinoma, Ki67 expression, lobulation, texture, mixed GGO, solid texture, CT value, air bronchogram, and pleural retraction (p<0.05) in both development group and validation group. In order to avoid the influence of confounding factors, we performed a LASSO regression analysis to re-valuate the variables. Finally, we retained 12 variables with nonzero coefficients (Figures 2A, 2B). Then, we further performed a multivariable logistic regression analysis and constructed a predictive model. The results of the logistic regression analysis are shown in Table 3. The final predictors included age, sex, smoke history, lepidic predominant subtype, solid predominant subtype, mucinous adenocarcinoma, Ki67 expression, lobulation, solid texture, and pleural retraction. These 10 variables were used as potential predictors of the prediction model. The model incorporating the above independent predictors was developed and presented as the nomogram to help practice in the clinic (Figure 2C). The maximum separation was at a probability cut-point of 0.7, with 61% sensitivity and 82% specificity for the development group. A negative predictive value (NPV) was 57.9%, a positive predicted value (PPV) was 82.9%, and an informedness index was 0.43 for the development group.

Model validation. The development and validation group presented with good calibration. The Hosmer-Lemeshow test showed adequate goodness-of-fit of the model both in the development group (p=0.674) and in the validation group (p=0.412) (Figures 3A, 3B). In the development group, the C-index of the predictive model was 0.789 (95% CI: 0.751–0.827). Meanwhile, the validation group was 0.809 (95% CI: 0.771–0.847) through cross-validation. The AUC of the development group was 0.789, and 0.809 of the validation group (Figures 3C, 3D) that suggested a good prediction capability of the model. Then, a Decision Curve Analysis (DCA) was performed to evaluate the prediction model. It indicated that predicting the EGFR mutations by applying this model would be better than having all patients or none patients treated by this model with a range of the risk threshold between >3% and <75% (Figure 3E). The maximum separation was at a probability cut-point of 0.7, with 67% sensitivity and 80% specificity in the validation group. An NPV was 61.2%, a PPV was 86.8%, and an informedness index was 0.47 for the validation group.

The model performance in the NGS test group. The model was assessed in a group of 52 patients who were tested for the EGFR mutations by NGS based on a liquid biopsy (peripheral blood or pleural effusion). As the maximum separation was at a probability cut-point of 0.7, there were 23 patients were categorized as a high positive group (>0.7) applying this model, among them, 21 of 23 patients

Table 1. Clinical features and histological characteristics of the development group and validation group.

Mutation status (n/%)	Development group (n=567)			Validation group (n=377)			Total (n=944)
	EGFR mutations 341 (60.1%)	Wild-type mutations 226 (39.9%)	p-value	EGFR mutations 237 (62.9%)	Wild-type mutations 140 (37.1%)	p-value	
Mutation types							
Exon 21 L858R (n/%)	188 (55.1%)			139 (58.6%)			327(56.6%)
Exon 19 deletion (n/%)	123(36.1%)			77 (32.5%)			200 (34.6%)
Exon 18 G719X (n/%)	9 (2.6%)			4 (1.7%)			13 (2.2%)
Exon 20 Ins (n/%)	7 (2.1%)			10 (4.2%)			17 2.9%)
Exon 21 L861Q (n/%)	6 (1.7%)			1 (0.4%)			7 (1.2%)
Exon 21 L861Q+ Exon 18 G719X (n/%)	2 (0.6%)			1 (0.4%)			3 (0.5%)
Exon 20 T790M+ Exon 21 L858R (n/%)	2 (0.6%)			2 (0.8%)			4 (0.7%)
Exon 20 T790M+ Exon 18 G719X (n/%)	1 (0.3%)			/			1 (0.2%)
Exon 20 T790M+ Exon 19 deletion (n/%)	1 (0.3%)			2 (0.8%)			3 (0.5%)
Exon 19 deletion+ Exon 21 L861Q (n/%)	1 (0.3%)			/			1 (0.2%)
Exon 18 G719X+ Exon 20 S768I (n/%)	1 (0.3%)			/			1 (0.2%)
Exon 20 T790M+ Exon 20 Ins (n/%)	/			1 (0.4%)			1 (0.2%)
Age (year)	63(56–69)	61(52–67)	0.004	63 (56–69)	60 (49–69)	0.014	63 (54–69)
Sex			<0.001			<0.001	
Female (n/%)	236 (69.2%)	102 (45.1%)		160 (67.5%)	60 (42.9%)		558 (59.1%)
Male (n/%)	105 (30.7%)	124 (54.9%)		77(32.5%)	80 (57.1%)		386 (40.1%)
Smoking history			<0.001			<0.001	
Yes (n/%)	40 (11.7%)	77 (34.1%)		19 (8.0%)	44 (31.4%)		180 (19.1%)
No (n/%)	301 (88.3%)	149 (65.9%)		218 (92.0%)	96 (68.6%)		764 (80.9%)
Histology							
MIA (n/%)	38 (11.1%)	43 (19.0%)	0.009	28 (11.8%)	24 (17.1%)	0.148	133 (14.1%)
IA (n/%)	303 (88.9%)	183 (81.0%)	0.009	209 (88.2%)	116 (82.9%)	0.148	811 (85.9%)
Predominant subtypes							
Acinar (n/%)	189 (55.4%)	83 (36.7%)	<0.001	126 (53.1%)	50 (35.7%)	0.001	448 (47.5%)
Papillary (n/%)	89 (26.1%)	43 (19.0%)	0.051	63 (26.6%)	22 (15.7%)	0.015	217 (23.0%)
Lepidic (n/%)	54 (15.8%)	56 (24.8%)	0.008	44 (18.6%)	39 (27.9%)	0.036	193 (20.4%)
Solid (n/%)	5 (1.5%)	26 (11.5%)	<0.001	2 (0.8%)	18 (12.9%)	<0.001	51 (5.4%)
Micropapillary (n/%)	2 (0.6%)	6 (2.7%)	0.041	1 (0.4%)	5 (3.6%)	0.018	14 (1.5%)
Invasive adenocarcinoma variant-Mucinous adenocarcinoma (n/%)	2 (0.6%)	12 (5.3%)	<0.0001	1 (0.4%)	6 (4.3%)	0.007	21 (2.2%)
Ki67 expression			<0.001			<0.001	
Low Ki67 expression (0–20%)	306 (89.7%)	169 (74.8%)		225 (95.0%)	106 (44.7%)		806 (85.4%)
High Ki67 expression (>20%)	35 (11.3%)	57 (25.3%)		12 (5.0%)	34 (55.3%)		138 (14.6%)
Stage			0.124			0.104	
I–II (n/%)	311 (91.2%)	197 (87.2%)		220 (92.8%)	123 (87.8%)		851 (90.1%)
III–IV (n/%)	30 (8.8%)	29 (12.8%)		17 (7.2%)	17 (12.2%)		93 (9.9%)
CEA (ng/ml)	2.64 (1.63–4.21)	2.77 (1.57–4.97)	0.373	2.33 (1.60–3.55)	2.50 (1.55–3.94)	0.458	2.53 (1.58–4.08)
CA199 (U/ml)	10.58 (6.83–15.93)	10.44 (6.70–15.22)	0.722	11.00 (7.70–15.90)	10.86 (7.21–16.55)	0.802	10.65 (7.00–15.76)
CA125 (U/ml)	10.49 (7.45–15.72)	10.82 (7.83–16.39)	0.431	10.40 (7.63–14.64)	10.93 (7.67–15.85)	0.540	10.62 (7.64–15.55)
Cyfra21-1 (ng/ml)	2.46 (1.86–3.20)	2.31 (1.78–3.10)	0.416	2.67 (1.92–3.41)	2.62 (2.07–4.02)	0.302	2.53 (1.87–3.30)

Abbreviations: MIA-minimally invasive adenocarcinoma, IA-invasive adenocarcinoma, CEA-carcinoembryonic antigen, CA199-carbohydrate antigen199; CA125-carbohydrate antigen 125, Cyfra 21-1-cytokeratin-19 fragment

Table 2. CT features of the development group and validation group.

	Development group (n=567)			Validation group (n=377)			Total (n = 944)
	EGFR mutations	Wild-type mutation	p-value	EGFR mutations	Wild-type mutation	p-value	
Mutation status (n/%)	341 (60.1%)	226 (39.9%)		237 (62.9%)	140 (37.1%)		
Distribution			0.607			0.628	
Central (n/%)	62 (18.2%)	45(13.2%)		41 (17.3%)	27 (19.3%)		175 (18.5%)
Peripheral (n/%)	279 (91.8%)	181(86.8%)		196 (82.7%)	113 (80.7%)		769 (81.5%)
Lobe location			0.452			0.045	
Right upper lobe (n/%)	115 (33.7%)	82 (36.3%)		92 (38.8%)	41 (29.3%)		330 (35.0%)
Right middle lobe (n/%)	23 (6.7%)	17 (7.5%)		14 (5.9%)	9 (6.4%)		63 (6.7%)
Right lower lobe (n/%)	53 (15.3%)	38 (16.8%)		37 (15.6%)	23 (16.4%)		151 (16.0%)
Left upper lobe (n/%)	99 (29.0%)	54 (23.9%)		62 (26.2%)	40 (28.5%)		255 (27.0%)
Left lower lobe (n/%)	51 (14.9%)	35 (15.5%)		32 (13.5%)	27 (19.3%)		145 (15.4%)
Size category			0.489			0.009	
≥ 30 mm	50 (14.7%)	38 (16.8%)		17 (7.2%)	22 (15.7%)		127 (13.5%)
< 30 mm	291 (85.3%)	188 (83.2%)		220 (92.8%)	118 (84.3%)		817 (86.5%)
Long axis diameter (mm)	15.9 (12.2–24.0)	14.6 (10.5–24.0)	0.048	17.0 (12.6–23.2)	15.4 (11.1–22.1)	0.174	16.14 (11.90–23.98)
Short axis diameter (mm)	11.7 (9.1–16.9)	10.9 (7.7–16.1)	0.195	11.9 (8.6–16.2)	11.2 (7.6–16.9)	0.412	11.61 (8.40–16.54)
Lobulation (n/%)	252 (73.9%)	135 (59.7%)	<0.001	203 (85.7%)	101 (72.1%)	0.001	691 (73.2%)
Spiculation (n/%)	279 (81.8%)	176 (77.9%)	0.249	217 (91.6%)	121 (86.4%)	0.114	793 (84.0%)
Texture			0.009			0.001	
Pure GGO (n/%)	78 (22.9%)	55 (24.3%)	0.688	87 (36.7%)	42 (30.0%)	0.185	262 (27.8%)
Mixed GGO (n/%)	177 (51.9%)	70 (31.0%)	<0.001	93 (39.2%)	34 (24.3%)	0.003	374 (39.7%)
Solid (n/%)	86 (25.2%)	101 (44.7%)	<0.001	57 (24.1%)	64 (45.7%)	<0.001	308 (32.6%)
Calcification	6 (1.8%)	10 (4.4%)	0.061	4 (1.7%)	5(3.6%)	0.248	25 (2.6%)
CT value	–237 (–418 – –1)	–134(–396–19)	0.012	–205 (–455– –12)	–73 (–388–25)	0.008	–188(–407–13)
Air bronchogram (n/%)	229 (67.2%)	129 (57.1%)	0.015	129 (54.4%)	59 (42.1%)	0.021	546 (57.8%)
Bubblelike lucency (n/%)	234 (68.6%)	130 (57.5%)	0.007	102 (43.0%)	72 (51.4%)	0.115	538 (57.0%)
Peripheral emphysema (n/%)	60 (17.6%)	39 (17.3%)	0.917	30 (12.7%)	19 (13.6%)	0.277	148 (15.7%)
Vascular convergence (n/%)	260 (76.2%)	155 (68.6%)	0.044	176 (74.3%)	101 (72.1%)	0.653	692 (73.3%)
Pleural retraction (n/%)	176 (51.6%)	87 (38.5%)	0.002	107 (45.1%)	27 (19.3%)	<0.0001	397 (42.1%)

Abbreviation: GGO-grand glass opacity

were defined as EGFR mutation-positive by NGS methods. There were 29 patients categorized as a low positive group (≤0.7), among them, 22 of 29 patients were defined as EGFR mutation-negative by NGS methods. An NPV was 75.9%, a PPV was 91.3%, and the diagnostic efficiency was 82.7% for the NGS group.

Discussion

In the present study, we developed and validated a multivariable logistic regression-based model to estimate the probability of the EGFR mutations in lung adenocarcinoma patients from a single center. The predictors included age, sex, smoke history, lepidic predominant subtype, solid predominant subtype, mucinous adenocarcinoma, Ki67 expression, lobulation, solid texture, and pleural retraction.

The nomogram based on the predictive model showed good predictive performance in both the development group (C-index: 0.798) and the validation group (C-index: 0.809) with good calibration. DCA showed that predicting the EGFR mutations probability applying this nomogram would be better than having all patients or none patients use this nomogram.

The female sex and non-smoker status were highly associated with a higher prevalence of the EGFR mutation, as observed in previous studies [11, 33]. That was consistent with our study.

Previous studies suggested histological characteristics including lepidic predominant subtype, acinar predominant subtype, and micropapillary predominant subtype were related with the EGFR mutations [18–21]. While, the solid predominant subtype had a negative correlation with the EGFR mutations [34, 35]. In our study, acinar predominant subtype had the highest proportion of 47.5% (448 of 944), compared with other histological subtypes. The LASSO regression showed that acinar predominant subtype was

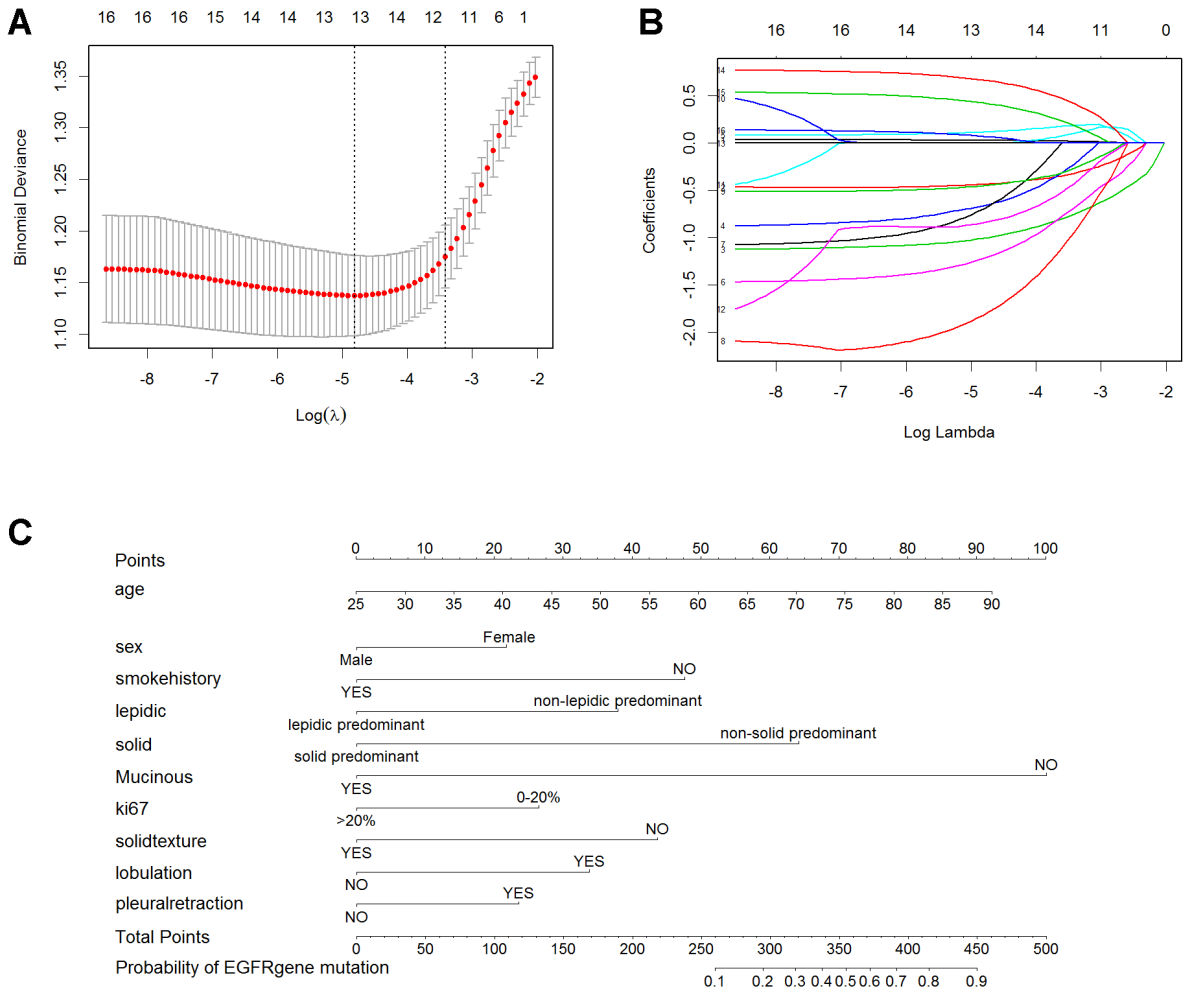

Figure 2. Texture feature selection using the least absolute shrinkage and selection operator (LASSO) binary logistic regression model. 12 variables of predictors were retained, including age, sex, smoke history, lepidic-predominant histology, acinar-predominant histology, solid-predominant histology, mucinous adenocarcinoma, Ki67 expression, lobulation, pleural retraction, mixed GGO, solid texture. A) Suitable parameter (λ) selection in the LASSO model used 10-fold cross-validation via minimum criteria. We plotted the partial likelihood deviance (binomial deviance) curve versus log (λ) were drawn at the optimal values applying the minimum criteria and the 1 standard error of the minimum criteria (the 1-SE criteria). A λ value of 0.0328, with log (λ), -3.418 was chosen (1-SE criteria) according to 10-fold cross-validation. B) LASSO coefficient profiles of the 16 texture features (The number information of each texture feature: 1. age, 2. sex, 3. smoke history, 4. lepidic predominant subtype, 5. acinar predominant subtype, 6. solid predominant subtype, 7. micropapillary predominant subtype, 8. mucinous adenocarcinoma, 9. Ki67 expression, 10. Texture, 11. MGGO, 12. solid texture, 13. CT value, 14. lobulation, 15. pleural retraction, 16. air bronchogram). A coefficient profile plot was produced against the log (λ) sequence. C) Nomogram for predicting EGFR mutations in primary lung adenocarcinoma patients.

positively correlated with EGFR mutations; while, lepidic predominant subtype and solid predominant subtype were negatively correlated with EGFR mutations.

Mucinous adenocarcinoma is classified as an invasive adenocarcinoma variant. Wakejima et al. [36] found that the frequency of EGFR mutations in mucinous adenocarcinoma patients was 20% (without female predominance), which was lower than that in lung non-mucinous adenocarcinoma (54%) [36]. In our study, there were 21 cases of mucinous adenocarcinoma, in which the frequency of the EGFR mutation was 14.3% (3 of 21), with 2 cases of exon 19 deletion and 1 case of exon 21 L858R. On the contrary, the frequency of EGFR mutation in non-mucinous adenocarcinoma was 62.3% (575 of 923).

Ki67 expression is used to assess cell proliferation and is associated with tumor growth. Li et al. [37] reported that Ki67 expression was significantly decreased (p=0.030) in NSCLCs with the EGFR mutations, which was consistent with our study. We also found that when Ki67 expression was more than 20%, it was less frequent with EGFR mutations.

Previous studies of lung cancer suggested that models combined with clinical and CT features may improve the

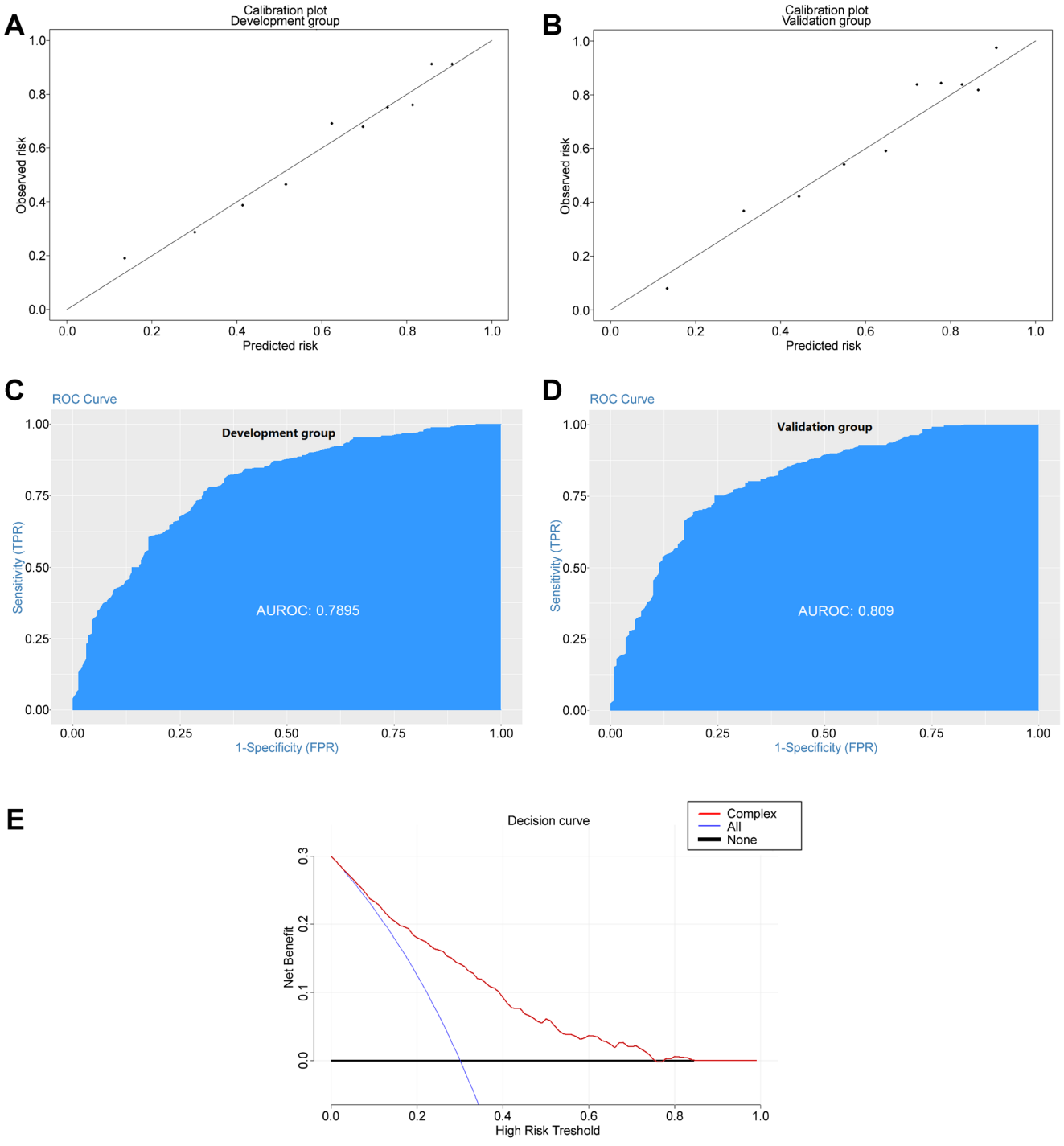

Figure 3. Calibration curves, ROC curves of the nomogram for predicting EGFR mutation in primary lung adenocarcinoma patients and decision curve analysis for EGFR mutation. A) Calibration curve of the nomogram for the development group (Hosmer-Lemeshow goodness-of-fit test: p=0.674). B) Calibration curve of the nomogram for the validation group (Hosmer-Lemeshow goodness-of-fit test: p=0.412). A p-value of >0.05 indicates good calibration. C) ROC curve for the development group. D) ROC curve for the validation group. AUROC = area under ROC curve; ROC = receiver operating characteristic. E) DCA: The y-axis represents the net benefit. The red line represents the nomogram of EGFR mutations. The grey line displays the assumption that all patients have EGFR mutations. The black line represents the assumption that no patients have EGFR mutations. The decision curve showed that predicting the EGFR mutations applying this nomogram would be better than having all patients or none patients treated by this nomogram with a range of the risk threshold between >3% and <75%.

Table 3. Multivariate logistic regression analysis for the EGFR mutations in patients with primary lung adenocarcinoma (development group).

Predictors	Multivariable analysis		
	OR	(95% CI)	p-value
Age (year)	1.033	1.014–1.053	0.001
Sex	0.605	0.381–0.961	0.033
Smoking history	0.334	0.191–0.584	<0.001
Predominant subtypes			
Acinar	1.158	0.713–1.881	0.553
Lepidic	0.417	0.248–0.702	0.001
Solid	0.228	0.076–0.678	0.008
Invasive adenocarcinoma variant Mucinous adenocarcinoma	0.099	0.021–0.481	0.004
Ki67 expression (0–20% vs. >20%)	0.544	0.296–1.000	0.050
Lobulation	2.182	1.394–3.415	0.001
Mixed GGO	0.931	0.537–1.615	0.799
Solid texture	0.365	0.220–0.607	<0.001
Pleural retraction	1.719	1.134–2.607	0.011

Abbreviation: GGO-grand glass opacity

performance in EGFR mutation prediction [22–26, 28, 38–40]. In our study, lobulation, solid texture, and pleural retraction were associated with the EGFR mutations. In addition to shallow lobulation [41], other CT features were found to have a positive correlation with EGFR mutation in lung adenocarcinoma, including pleural retraction, air bronchogram, or bubblelike lucency [22, 42]. But, GGO and spiculation didn't show a significant association with EGFR mutation status [43]. Our study showed that mixed GGO was more frequent in patients with EGFR mutation, and Zhang et al. also found that [26] part-solid GGO was more frequent in NSCLC patients with EGFR mutation when compared with nonsolid GGO; however, tumor size, cavitation, air bronchogram, lobulation, and spiculation did not significantly correlate with the EGFR mutation, respectively. The controversial outcomes from researches between EGFR mutations and histological or CT features might be associated with study sample size, ethnic difference, and a research design.

Zhou et al. [44] found that the overall response rate and overall survival were not demonstrated with significant difference between patients with a high abundance of EGFR (positive for mutation with both ARMS and direct DNA sequencing) and a low abundance of EGFR mutations (positive for mutation with ARMS but negative with direct DNA sequencing); while patients with low abundance of EGFR mutations had a longer median PFS than those with wild-type tumors (negative for mutation with both ARMS and direct DNA sequencing). This indicates that it is EGFR mutation status rather than the abundance that influences EGFR-TKI therapy response rate, and the survival time of patients with lung adenocarcinoma.

According to the 2017 molecular testing guideline for the selection of lung cancer patients for treatment with targeted TKI [45], next-generation sequencing (NGS), a massively parallel sequencing has changed the practice of molecular diagnostic in lung cancer and in other contexts. Numerous studies [46–50] have demonstrated the excellent sensitivity of NGS methods relative to single-gene targeted assays. The greater analytic sensitivity of NGS makes it suitable for very small or heterogeneous samples. While the previous capture-based methods, using hybridization to generate the library are less sensitive in highly heterogeneous or small samples.

The purpose of promoting the application of the EGFR mutation predictive model in patients with lung adenocarcinoma is to more accurately enrich patient populations for the EGFR-TKI therapy. In our study, we defined the NGS group as a high positive group (>0.7) and a low positive group (≤0.7). Our model performed well in the NGS group, with 82.7% of diagnostic efficiency. Therefore, when a patient acquires a high score of the EGFR mutation probability according to a nomogram model, but with a negative result of EGFR molecular test with capture-based methods or unavailable tissue for EGFR testing, this patient might be highly suggested to test the EGFR mutation by NGS-based circulating tumor DNA (ctDNA) method for a potential positive result of the EGFR mutation and the TKI therapy opportunity. Our EGFR mutation predictive model cannot replace the EGFR mutation testing. But in lung adenocarcinoma patients, whose tissue biopsy is not enough for further EGFR mutation examination, a high estimated probability of EGFR mutations by our model, suggests that further NGS-based ctDNA with liquid biopsy could be implemented.

Our study has several limitations. First, it was a single-center retrospective study in China, which might cause a selection bias. A large-scale validation across multiple centers would substantially strengthen the score. Second, because of the majority of EGFR mutations are occurring in lung adenocarcinoma, we focused on patients with adenocarcinoma, limiting the application of other histology. This was done because the majority of EGFR mutations are found in adenocarcinoma. Third, any prediction instrument inherently incorporates a certain degree of uncertainty [31] and individual predictions remain imperfect. Finally, there may be some differences in CT features between patients with low and high proportions of the EGFR mutation. However, we did not have data to analyze the differences in the present study. It is expected that future research with large academic centers will help to reveal the relations of EGFR mutations types, clinical characteristics, histological characteristics, CT features, and TKI therapy response.

In conclusion, we developed an effective nomogram based on clinical, histological characteristics, and CT features to estimate the probability of the EGFR mutations in primary lung adenocarcinoma. A high score on the nomogram may suggest the possibility of positive EGFR mutation. If a patient has a high score probability based on our model, while his or her tissue is limited, NGS-based ctDNA could be implemented effectively in this situation.

Acknowledgments: We like to thank all departments involved in the data collection of the patients. Hui Qin and Jing Zou made contributions to the integrity of the entire study. Yiping Shi, Qiongfang Zha, Shan Xue, Huawei Wu, Daoqiang Tang made contributions to the clinical data collection; Jing Zou, Yiping Shi, and Qing Ye made contributions to the statistical analysis; all authors made substantial contributions to the execution of the review and interpretation of the data. All authors contributed to the intellectual content of the manuscript, reviewed, and critically revised it, and approved the final submitted version.

References

[1] BRAY F, FERLAY J, SOERJOMATARAM I, SIEGEL RL, TORRE LA et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. [published correction appears in CA Cancer J Clin. 2020; 70: 313]. CA Cancer J Clin 2018; 68: 394–424. https://doi.org/10.3322/caac.21492

[2] 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: e0121323. https://doi.org/10.1371/journal.pone.0121323

[3] TORRE LA, SIEGEL RL, JEMAL A. Lung Cancer Statistics. Adv Exp Med Biol 2016; 893: 1–19. https://doi.org/10.1007/978-3-319-24223-1_1

[4] TRAVIS WD, BRAMBILLA E, NOGUCHI M, NICHOLSON AG, GEISINGER K 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: 381–385. https://doi.org/10.1513/pats.201107-042ST

[5] MAEMONDO M, INOUE A, KOBAYASHI K, SUGAWARA S, OIZUMI S et al. Gefitinib or chemotherapy for non–small-cell lung cancer with mutated EGFR. N Engl J Med 2010; 362: 2380–2388. https://doi.org/10.1056/NEJMoa0909530

[6] MITSUDOMI T, MORITA S, YATABE Y, NEGORO S, OKAMOTO I et al. Gefitinib versus cisplatin plus docetaxel in patients with non-small-cell lung cancer harbouring mutations of the epidermal growth factor receptor (WJTOG3405): an open label, randomised phase 3 trial. Lancet Oncol 2010; 11: 121–128. https://doi.org/10.1016/S1470-2045(09)70364-X

[7] SEQUIST LV, YANG JC, YAMAMOTO N, O'BYRNE K, HIRSH V et al. Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. J Clin Oncol 2013; 31: 3327–3334. https://doi.org/10.1200/JCO.2012.44.2806

[8] ROSELL R, CARCERENY E, GERVAIS R, VERGNENEGRE A, MASSUTI B 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: 239–246. https://doi.org/10.1016/S1470-2045(11)70393-X

[9] ZHOU C, WU YL, CHEN G, FENG J, LIU XQ et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. Lancet Oncol 2011; 12: 735–742. https://doi.org/10.1016/S1470-2045(11)70184-X

[10] ROSELL R, CARCERENY E, GERVAIS R, VERGNENEGRE A, MASSUTI B, 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: 239–246. https://doi.org/10.1016/S1470-2045(11)70393-X

[11] ZHANG YL, YUAN JQ, WANG KF, FU XH, HAN XR et al. The prevalence of EGFR mutation in patients with non-small cell lung cancer: a systematic review and meta-analysis. Oncotarget 2016; 7: 78985–78993. https://doi.org/10.18632/oncotarget.12587

[12] PIRKER R, HERTH FJ, KERR KM, FILIPITS M, TARON M et al. Consensus for EGFR mutation testing in non-small cell lung cancer: results from a European workshop. J Thorac Oncol 2010; 5: 1706–1713. https://doi.org/10.1097/JTO.0b013e3181f1c8de

[13] LINDEMAN NI, CAGLE PT, BEASLEY MB, CHITALE DA, DACIC S et al. Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. J Thorac Oncol 2013; 8: 823–859. https://doi.org/10.1097/JTO.0b013e318290868f

[14] ETTINGER DS, WOOD DE, AKERLEY W, BAZHENOVA LA, BORGHAEI H et al. Non-Small Cell Lung Cancer, Version 6.2015. J Natl Compr Canc Netw 2015; 13: 515–524. https://doi.org/10.6004/jnccn.2015.0071

[15] RECK M, POPAT S, REINMUTH N, DE RUYSSCHER D, KERR K, PETERS S. Metastatic non-small-cell lung cancer (NSCLC): ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol 2014; 25: iii27–39. https://doi.org/10.1093/annonc/mdu199

[16] THI AM, TIN TIN S, MCKEAGE M, ELWOOD JM. Utilisation and Determinants of Epidermal Growth Factor Receptor Mutation Testing in Patients with Nonsmall Cell Lung Cancer in Routine Clinical Practice: A Global Systematic Review. Target Oncol 2020; 15: 279–299. https://doi.org/10.1007/s11523-020-00718-w

[17] MARTIN P, LEIGHL NB. Review of the use of pretest probability for molecular testing in non-small cell lung cancer and overview of new mutations that may affect clinical practice. Ther Adv Med Oncol 2017; 9: 405–414. https://doi.org/10.1177/1758834017704329

[18] VILLA C, CAGLE PT, JOHNSON M, PATEL JD, YELDANDI AV et al. Correlation of EGFR mutation status with predominant histologic subtype of adenocarcinoma according to the new lung adenocarcinoma classification of the International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society. Arch Pathol Lab Med 2014; 138: 1353–1357. https://doi.org/10.5858/arpa.2013-0376-OA

[19] YOSHIZAWA A, SUMIYOSHI S, SONOBE M, KOBAYASHIM, FUJIMOTOM, et al. Validation of the IASLC/ATS/ERS lung adenocarcinoma classification for prognosis and association with EGFR and KRAS gene mutations: analysis of 440 Japanese patients. J Thorac Oncol 2013; 8: 52–61. https://doi.org/10.1097/JTO.0b013e3182769aa8

[20] RUSSELL PA, BARNETT SA, WALKIEWICZ M, WAINER Z, CONRON M, et al. Correlation of mutation status and survival with predominant histologic subtype according to the new IASLC/ATS/ERS lung adenocarcinoma classification in stage III (N2) patients. J Thorac Oncol 2013; 8: 461–468. https://doi.org/10.1097/JTO.0b013e3182828fb8

[21] SHIM HS, LEE H, PARK EJ, KIM SH. Histopathologic characteristics of lung adenocarcinomas with epidermal growth factor receptor mutations in the International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society lung adenocarcinoma classification. Arch Pathol Lab Med 2011; 135: 1329–1334. https://doi.org/10.5858/arpa.2010-0493-OA

[22] LIU Y, KIM J, QU F, LIU S, WANG H et al. CT Features Associated with Epidermal Growth Factor Receptor Mutation Status in Patients with Lung Adenocarcinoma. Radiology 2016; 280: 271–280. https://doi.org/10.1148/radiol.2016151455

[23] SABRI A, BATOOL M, XU Z, BETHUNE D, ABDOLELL M et al. Predicting EGFR mutation status in lung cancer: proposal for a scoring model using imaging and demographic characteristics. Eur Radiol 2016; 26: 4141–4147. https://doi.org/10.1007/s00330-016-4252-3

[24] RIZZO S, PETRELLA F, BUSCARINO V, DE MARIA F, RAIMONDI S et al. CT Radiogenomic characterization of EGFR, K-RAS, and ALK mutations in nonsmall cell lung Cancer. Eur Radiol 2016; 26: 32–42. https://doi.org/10.1007/s00330-015-3814-0

[25] HSU JS, HUANG MS, CHEN CY, LIU GC, LIU TC et al. Correlation between EGFR mutation status and computed tomography features in patients with advanced pulmonary adenocarcinoma. J Thorac Imaging 2014; 29: 357–363. https://doi.org/10.1097/RTI.0000000000000116

[26] ZHANG H, CAI W, WANG Y, LIAO M, TIAN S. CT and clinical characteristics that predict risk of EGFR mutation in non-small cell lung cancer: a systematic review and meta-analysis. Int J Clin Oncol 2019; 24: 649–659. https://doi.org/10.1007/s10147-019-01403-3

[27] COLLINS GS, REITSMA JB, ALTMAN DG, MOONS KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. Ann Intern Med 2015; 162: 55–63. https://doi.org/10.7326/M14-0697

[28] GOLDSTRAW P, CHANSKY K, CROWLEY J, RAMI-PORTA R, ASAMURA H 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: 39–51. https://doi.org/10.1016/j.jtho.2015.09.009

[29] PEDUZZI P, CONCATO J, KEMPER E, HOLFORD TR, FEINSTEIN AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996; 49: 1373–1379. https://doi.org/10.1016/s0895-4356(96)00236-3

[30] HARRELL FE, LEE KL, CALIFF RM, PRYOR DB, ROSATI RA. Regression modelling strategies for improved prognostic prediction. Stat Med 1984; 3: 143–152. https://doi.org/10.1002/sim.4780030207

[31] HARRELL FE, LEE KL, MARK DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996; 15: 361–387. https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4

[32] FRIEDMAN J, HASTIE T, TIBSHIRANI R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 2010; 33: 1–22.

[33] MIDHA A, DEARDEN S, MCCORMACK R. EGFR mutation incidence in non-smallcell lung cancer of adenocarcinoma histology: a systematic review and global map by ethnicity (mutMapII). Am J Cancer Res 2015; 5: 2892–2911.

[34] SONG Z, ZHU H, GUO Z, WU W, SUN W, ZHANG Y. Correlation of EGFR mutation and predominant histologic subtype according to the new lung adenocarcinoma classification in Chinese patients. Med Oncol 2013; 30: 645. https://doi.org/10.1007/s12032-013-0645-1

[35] ZHANG Y, SUN Y, PAN Y, LI C, SHEN L et al. Frequency of driver mutations in lung adenocarcinoma from female never-smokers varies with histologic subtypes and age at diagnosis. Clin Cancer Res 2012; 18: 1947–1953. https://doi.org/10.1158/1078-0432.CCR-11-2511

[36] WAKEJIMA R, INAMURA K, NINOMIYA H, NAGANO H, MUN M et al. Mucinous lung adenocarcinoma, particularly referring to EGFR-mutated mucinous adenocarcinoma. Pathol Int 2020; 70: 72–83. https://doi.org/10.1111/pin.12879

[37] LI M, ZHANG Q, LIU L, LU W, WEI H et al. Expression of the Mismatch Repair Gene hMLH1 Is Enhanced in Non-Small Cell Lung Cancer with EGFR Mutations. PloS One 2013; 8: E78500. https://doi.org/10.1371/journal.pone.0078500.

[38] GEVAERT O, ECHEGARAY S, KHUONG A, HOANG CD, SHRAGER JB et al. Predictive radiogenomics modeling of EGFR mutation status in lung cancer. Sci Rep 2017; 7: 41674. https://doi.org/10.1038/srep41674

[39] ZHANG L, CHEN B, LIU X, SONG J, FANG M et al. Quantitative biomarkers for prediction of epidermal growth factor receptor mutation in non-small cell lung Cancer. Transl Oncol 2018; 11: 94–101. https://doi.org/10.1016/j.tranon.2017.10.012

[40] TU W, SUN G, FAN L, WANG Y, XIA Y et al. Radiomics signature: a potential and incremental predictor for EGFR mutation status in NSCLC patients, comparison with CT morphology. Lung Cancer 2019; 132: 28–35. https://doi.org/10.1016/j.lungcan.2019.03.025

[41] CHEN Y, YANG Y, MA L, ZHU H, FENG T et al. Prediction of EGFR mutations by conventional CT-features in advanced pulmonary adenocarcinoma. Eur J Radiol 2019; 112: 44–51. https://doi.org/10.1016/j.ejrad.2019.01.005

[42] ZHOU JY, ZHENG J, YU ZF, XIAO WB, ZHAO J, et al. Comparative analysis of clinicoradiologic characteristics of lung adenocarcinomas with ALK rearrangements or EGFR mutations. Eur Radiol 2015; 25: 1257–1266. https://doi.org/10.1007/s00330-014-3516-z

[43] SUGANO M, SHIMIZU K, NAKANO T, KAKEGAWA S, MIYAMAE Y et al. Correlation between computed tomography findings and epidermal growth factor receptor and KRAS gene mutations in patients with pulmonary adenocarcinoma. Oncol Rep 2011; 26: 1205–1211. https://doi.org/10.3892/or.2011.1412

[44] ZHOU Q, ZHANG XC, CHEN ZH, YIN XL, YANG JJ, et al. Relative abundance of EGFR mutations predicts benefit from gefitinib treatment for advanced non-small-cell lung cancer. J Clin Oncol 2011; 29: 3316–3321. https://doi.org/10.1200/JCO.2010.33.3757

[45] LINDEMAN NI, CAGLE PT, AISNER DL, ARCILA ME, BEASLEY MB et al. Updated Molecular Testing Guideline for the Selection of Lung Cancer Patients for Treatment With Targeted Tyrosine Kinase Inhibitors: Guideline From the College of American Pathologists, the International Association for the Study of Lung Cancer, and the Association for Molecular Pathology. Arch Pathol Lab Med 2018; 142: 321–346. https://doi.org/10.5858/arpa.2017-0388-CP

[46] DRILON A, WANG L, ARCILA ME, BALASUBRAMANIAN S, GREENBOWE JR et al. Broad, hybrid capture-based next-generation sequencing identifies actionable genomic alterations in lung adenocarcinomas otherwise negative for such alterations by other genomic testing approaches. Clin Cancer Res 2015; 21: 3631–3639. https://doi.org/10.1158/1078-0432.CCR-14-2683

[47] SU J, ZHANG XC, AN SJ, ZHONG WZ, HUANG Y et al. Detecting the spectrum of multigene mutations in non-small cell lung cancer by Snapshot assay. Chin J Cancer 2014; 33: 346–350. https://doi.org/10.5732/cjc.013.10195

[48] HAN JY, KIM SH, LEE YS, HWANG JA, KIM JY et al. Comparison of targeted next-generation sequencing with conventional sequencing for predicting the responsiveness to epidermal growth factor receptor-tyrosine kinase inhibitor (EGFR-TKI) therapy in never-smokers with lung adenocarcinoma. Lung Cancer 2014; 85: 161–167. https://doi.org/10.1016/j.lungcan.2014.04.009

[49] TUONONEN K, MAKI-NEVALA S, SARHADI VK, WIRTANEN A, RONTY M et al. Comparison of targeted next-generation sequencing (NGS) and real-time PCR in the detection of EGFR, KRAS, and BRAF mutations on formalin-fixed, paraffin-embedded tumor material of non-small cell lung carcinoma-superiority of NGS. Genes Chromosomes Cancer 2013; 52: 503–511. https://doi.org/10.1002/gcc.22047

[50] SCARPA A, SIKORA K, FASSAN M, RACHIGLIO AM, CAPPELLESSO R, et al. Molecular typing of lung adenocarcinoma on cytological samples using a multigene next generation sequencing panel. PloS One 2013; 8: E80478. DOI: https://doi.org/10.1371/journal.pone.0080478