

Predictive Value of Folate Receptor-Positive Circulating Tumor Cells for the Preoperative Diagnosis of Lymph Node Metastasis in Patients with Lung Adenocarcinoma

Zhao Li, Ke Xu, Lekai Xu, Jie Dai, Kaiqi Jin, Yuming Zhu,* Yang Yang,* and Gening Jiang*

Noninvasive assessments of the risk of lymph node metastasis (LNM) in patients with lung adenocarcinoma (LAD) are of great value for selecting individualized treatment options. However, the diagnostic accuracies of different preoperative LN evaluation methods in routine clinical practice are not satisfactory. Here, an assessment to detect folate receptor (FR)-positive circulating tumor cells (CTCs) based on ligand-targeted enzyme-linked polymerization is established. FR-positive CTCs have the potential to improve the specificity and sensitivity of diagnosing LNM in lung cancer patients. The addition of CTC level improved the diagnostic efficiency of the initial prediction model that comprises other clinical parameters. A nomogram for predicting preoperative LNM is established, which showed good prediction and calibration capacities and achieved an average area under the curve of 0.786. Good correlations are observed between the CTC level and nodal classifications, such as the number of positive LNs and the ratio of the number of positive LNs to removed LNs (LN ratio or LNR). The ligand-targeted enzyme-linked polymerization-assisted assessment of CTCs enables noninvasive detection and has a useful predictive value for the preoperative diagnosis of LNM in patients with LAD.

is the most common non-small cell lung cancer (NSCLC).^[2,3] A portion of cancer cells will metastasize to the lymph nodes (LNs) around the lung hilar and mediastinum, including mainly N1 station (ipsilateral parabranchial, interlobar, or hilar node involvement), N2 station (ipsilateral mediastinal node involvement), and N3 station (contralateral mediastinal, contralateral hilar, or supraclavicular node involvement) lymph node metastasis (LNM).^[4] Usually, LNM occurs first in the N1 station LNs, followed by the N2 and N3 regions. LNM is an essential indicator of LAD prognosis^[5] and scope, dictates the method of surgery^[6] and is also a significant factor associated with a high recurrence rate and low survival rate.^[7] Accurate clinical staging of LNM can help choose the proper treatment and standardize management procedures.

However, the diagnostic accuracies of different preoperative LN evaluation

1. Introduction

Lung cancer affects 2 918 200 people worldwide, and its incidence rate is ranked first among all cancers.^[1] Approximately 60% of lung cancer patients have lung adenocarcinoma (LAD), which

methods in routine clinical practice are not satisfactory. Computed tomography (CT) is the standard examination and available in most hospitals, but it has a low sensitivity of 55%.^[8] Positron emission tomography-computed tomography (PET-CT) has been reported to be superior to CT in staging mediastinal LNs.^[9] However, PET-CT is more expensive and not as popular as CT, and PET-CT has a lower negative predictive value (NPV) for suspected N1 metastasis. There is an urgent need for a safe and efficient method to predict LNM in LAD patients to guide the clinical diagnosis and treatment process.


Folate receptor (FR)-positive circulating tumor cells (CTCs) have the potential to improve the specificity and sensitivity of diagnosing LNM in lung cancer patients. CTCs exhibit great potential for early diagnosis, prognosis evaluation, treatment response monitoring, and predicting resistance to therapy in a noninvasive manner. Recent reports have illustrated that the reported diagnostic sensitivity of CTCs for lung cancer ranges from 87 to 100%. Before or after surgery, high levels of CTCs are associated with a high risk of metastases or early recurrence after surgery, and this has been supported by strong evidence. Because CTCs are from primary tumors and spread via blood or lymphatic vessels, the presence of CTCs is suggested to be an important predictor of tumor micro-metastasis and the source of metastasis.^[10,11] This real-time dynamic method has become essential to evaluating the status of lung cancer. Several studies have found an association between CTCs and LN involvement.^[12,13] Therefore, CTCs may improve the diagnostic

Dr. Z. Li, Dr. J. Dai, Dr. K. Jin, Prof. Y. Zhu, Prof. Y. Yang, Prof. G. Jiang
Department of Thoracic Surgery
Shanghai Pulmonary Hospital
Tongji University School of Medicine
No. 507 Zhengmin Road, Shanghai 200433, China
E-mail: zhuyuming@tongji.edu.cn; timyangsh@tongji.edu.cn;
geningjiang@tongji.edu.cn

K. Xu
Department of Thoracic Surgery
The First Affiliated Hospital of Guangzhou Medical University
No. 151 Yanjiang Road, Guangzhou 510120, China

L. Xu
State Key Laboratory of Biochemical Engineering
Institute of Process Engineering
Chinese Academy of Sciences
No. 1 Beiertiao, Zhongguancun, Beijing 100109, China

Prof. Y. Yang
School of Materials Science and Engineering
Tongji University
Shanghai 201804, China

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/smt.202100152>.

DOI: 10.1002/smt.202100152

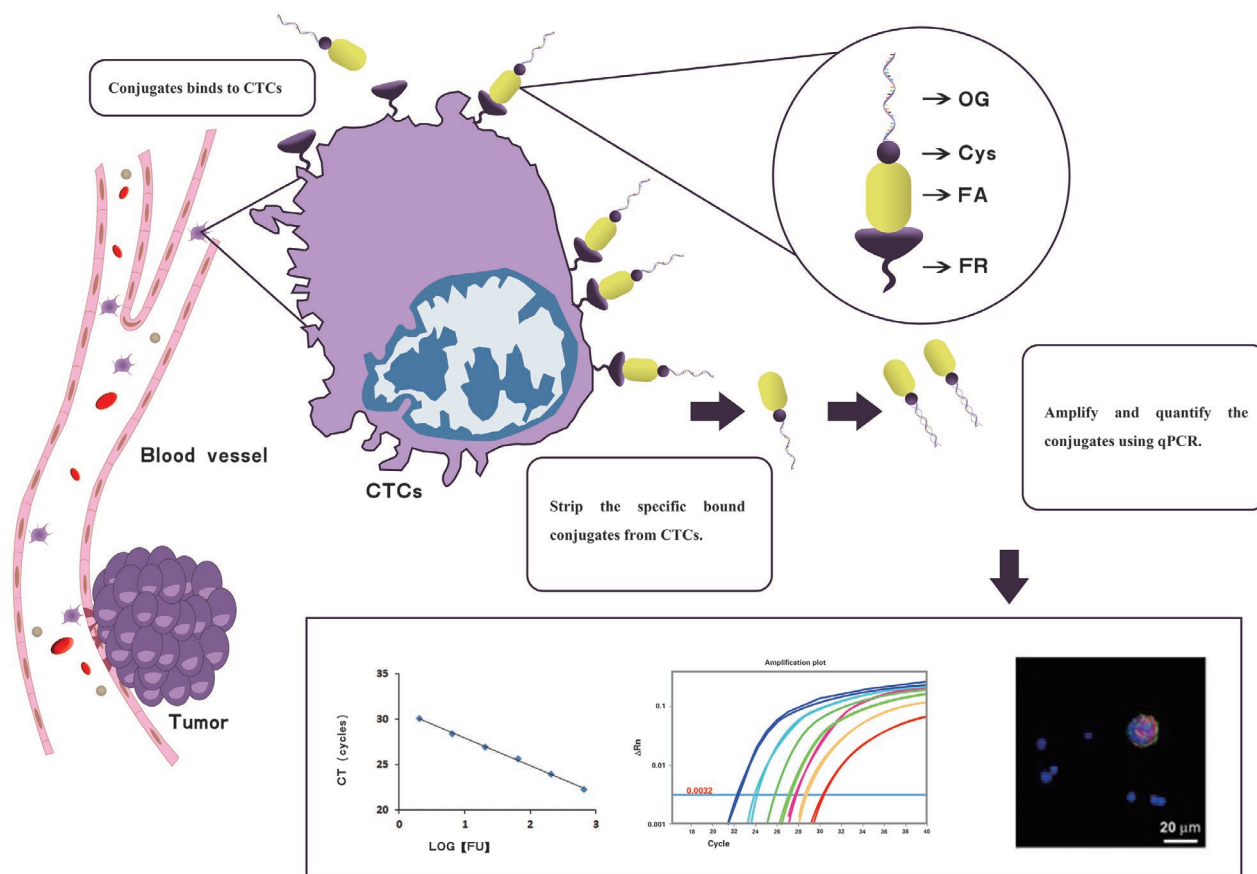


Figure 1. Schematic of the FR-positive CTC concentration analysis. Abbreviations: CTCs, circulating tumor cells; OG, oligonucleotide probe; Cys, cysteine; FA, folic acid; FR, folic acid receptor; qPCR, quantitative real-time polymerase chain reaction.

sensitivity for LNM in lung cancer patients. In addition, FR is highly expressed in epithelial-derived malignant tumor cells and is significantly upregulated in 75.7% of NSCLC patients but has no expression in normal human blood cells.^[14] Consequently, FR is an ideal detection marker for CTCs in patients with lung cancer. FR-positive CTCs can be detected by ligand-targeted enzyme-linked polymerization and are used as a novel diagnostic biomarker for lung cancer, increasing the specificity for LNM of LAD. Our team has developed a convenient and time-saving strategy based on FR-positive CTCs concentration to improve the pathological diagnosis of pulmonary nodules and showed the satisfactory performance of FR-positive CTCs.^[15]

In this study, we evaluated the diagnostic value of preoperative FR-positive CTCs as a useful tool during noninvasive liquid biopsy to predict LNM preoperatively in patients with LAD.

2. Results

2.1. CTC Detection and Patient Characteristics

In all patients scheduled for surgery, peripheral blood samples were collected from the anterior umbilical vein in ethylenediaminetetraacetic acid (EDTA)-containing tubes. The CTCs were enriched through negative enrichment with immunomagnetic beads. The enriched CTCs were incubated with labeling buffer containing the

tumor-specific locus of the ligand folic acid and synthetic oligonucleotide conjugates. Specific ligand-nucleotide conjugates were removed, collected by centrifugation, and neutralized for further reverse transcription (RT)-polymerase chain reaction (PCR) analysis. CTC detection and quantification were performed by ligand-targeted PCR. A series of standards containing oligonucleotides were used for CTC quantification. A schematic of the FR-positive CTC concentration analysis is summarized in **Figure 1**.

For this patient cohort study, the institutional database was searched for medical records from January 2013 to December 2019 to identify patients with pathologically confirmed LAD who underwent surgical resection with curative intent. In total, 2,821 patients were identified: 1,131 were males, and 1,690 were females. A total of 1,551 patients were over 60 years old, while 1,270 were younger than 60 years old. LNM was present in 4.0% (N1 disease) and 9.8% (N2 disease) of all patients. The clinical characteristics of the overall patient population are summarized in Table S1, Supporting Information.

2.2. Correlation Analysis between FR-Positive CTCs and Clinicopathologic Parameters

Next, we determined the correlation between the FR-positive CTC level and other clinicopathologic parameters (age, sex, smoking status, serum tumor biomarkers, other laboratory

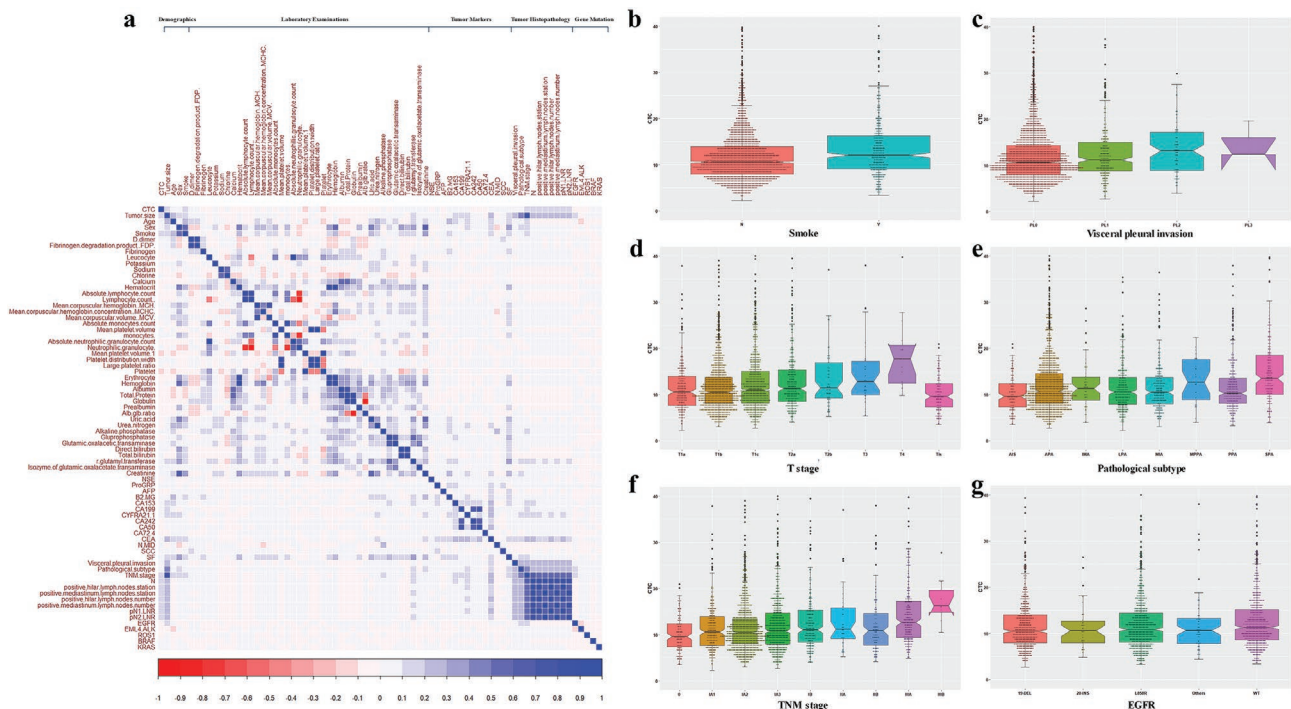


Figure 2. The correlation between FR-positive CTCs and clinicopathologic parameters and CTC levels stratified by clinicopathologic parameters. a) A heatmap (74 × 74) was constructed to show the correlation between FR-positive CTCs and clinicopathologic parameters, consisting of demographics, laboratory examinations, tumor markers, tumor histopathology, and gene mutations on the X-axis. The negative correlation is shown in red (the correlation coefficient [R] is less than 0), while the positive correlation is shown in blue (the correlation coefficient [R] is greater than 0). Univariate analysis regarding the association of CTC level to b) smoking status; c) visceral pleural invasion; d) T stage; e) pathological subtypes; f) TNM stage; and g) EGFR mutation subtype.

examinations, tumor size, adenocarcinoma subtypes, visceral pleural invasion, LNM status (N stage), and gene mutation), as shown in **Figure 2a**. We found that age, sex, smoking status, LAD subtype, pathological TNM stage, and LNM status (nodal stage) were significantly associated with the CTC level. Notably, the positive LN station, positive nodal number (nN) and the combination of pN and lymph node ratio (pN-LNR) were associated with the CTC level.

Univariate analysis was conducted to individually assess the differences in CTC level among patients with different clinicopathologic variables, such as smoking status, visceral pleural invasion, T stage, pathological LAD subtype, TNM stage, and EGFR mutation subtype, as shown in **Figure 2b,g** and **Table S2**, Supporting Information. Smokers had significantly higher CTC levels than nonsmokers. Regarding the pathologic variables, as the tumor invasiveness and stage increased in terms of visceral pleural invasion, pathological LAD subtype, T stage, and TNM stage, a creeping trend in CTC levels was observed. In addition, for EGFR gene mutations, we did not note significant differences in exon-19 deletion, exon 20-ins, L858R, wild type, or other mutation subtypes.

2.3. Feature Selection and Development of an Individualized Prediction Model

The least absolute shrinkage and selection operator (LASSO) method, which is suitable for the regression of

high-dimensional data, was used to select the most useful predictive features from the training data set. The LASSO process for feature selection is shown in **Figure S1**, Supporting Information. Among 52 clinicopathologic parameters, the following six parameters were included as candidate clinical predictors: tumor size, absolute leucocyte count, absolute monocyte count, levels of the tumor markers CA153 and CEA, and CTC level. A total of 2,821 patients were randomly assigned to the training cohort ($N = 1,854$, 66.7%) and the validation cohort ($N = 967$, 33.3%). No significant differences were observed between the training cohort and validation cohort in any clinical factor (**Table S1**, Supporting Information). A logistic regression analysis identified the CTC level, CA153 and CEA level, tumor size, absolute leucocyte count, and absolute monocyte count as independent predictors (**Table 1**).

2.4. Nomogram Development and Validation

A model incorporating the above independent predictors was developed and presented as a nomogram (**Figure 3**). The distinguishing prognostic ability of the model (model 2) with patient clinical parameters only (tumor size, absolute leucocyte count, absolute monocyte count, and CA153 and CEA levels) presented an area under the receiver operating characteristic curve (AUC) of 0.780 for diagnosing LNM. In an attempt to improve the AUC, CTC level was added (model 1). A slightly higher AUC was observed for the model that integrated CTC level in the training

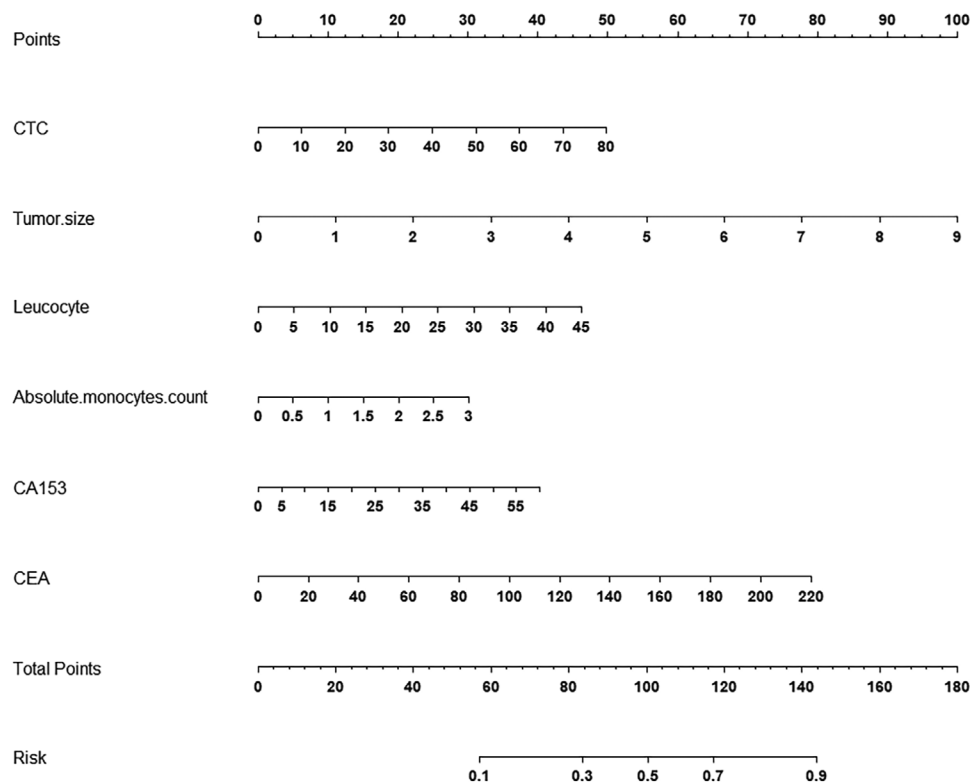


Figure 3. Nomogram developed based on CTC level. The nomogram was developed with the training cohort and incorporated the CTC level, tumor size, absolute leucocyte count, absolute monocyte count, CA153 level, and CEA level. According to the location of the value on the 2nd to the 6th axes, we can obtain the vertically corresponding points on the first axis. By summing all points, we obtain a total number of points, and the vertically corresponding predicted value on the last axis shows the predicted possibility of LNM.

cohort, indicating an improved prediction performance. The prediction model based on the clinical parameters and with the addition of CTC level is shown in **Table 1**. The AUCs of model 1 and model 2 are shown in Figure S2, Supporting Information.

Furthermore, a good internal calibration curve was observed for the probability of LNM in the validation cohort (Figure S3, Supporting Information). The Hosmer–Lemeshow test yielded a nonsignificant statistic ($p > 0.05$), and the C-index of the

nomogram for the prediction of LNM status was 0.786. The performance of the internally validated nomogram was tested in the validation cohort. The logistic regression formula formed in the training cohort was applied to all patients in the validation cohort, and total points for each patient were calculated. Logistic regression in this cohort was then performed by using the total points as a factor. Finally, the C-index and calibration curve were derived based on the regression analysis.

Table 1. Multivariate logistic analysis of clinicopathologic characteristics as predictors of LNM.

Intercept and variable	Model 1			Model 2		
	Formula	Std. error	P value	Formula	Std. error	P value
	-5.079	0.379	<0.001	-4.722	0.353	<0.001
Tumor size	0.617	0.063	<0.001	0.631	0.062	<0.001
Absolute leucocyte count	0.065	0.025	0.009	0.060	0.025	0.016
Absolute monocyte count	0.362	0.264	0.169	0.369	0.263	0.159
CA153	0.025	0.012	0.044	0.028	0.012	0.019
CEA	0.022	0.006	<0.001	0.022	0.005	0.019
CTC	0.028	0.010	0.005	NA	NA	NA
Cox-Snell R ²	0.097			0.093		
Nagelkerke R ²	0.173			0.166		
AUC	0.786			0.780		

Std. error, standard error; CA153, carbohydrate antigen 153; CEA, carcinoembryonic antigen; CTC, circulating tumor cells; AUC, area under curve.

2.5. The Associations between FR-Positive CTCs and Nodal Stage

According to the previous correlation analysis between the CTC level and clinicopathologic parameters, not only was nodal stage found to have a strong association with FR-positive CTC levels but also LNM in the hilar (N1 disease) and mediastinum (N2 disease) stations correlated with FR-positive CTC levels. The TNM classification for breast, gastric, and colorectal cancer have been updated to include the number of positive LNs (nN) in nodal staging. As nN is confounded by the number of LNs sampled and LNR was defined as the ratio of the number of positive LNs to the number of removed LNs, both were investigated and found to be better independent prognostic factors for NSCLC than the proposed 8th TNM subclassification for pathological N staging (pN), which may limit the interpretation of prognosis prediction. For the first time, we comprehensively compared the correlation between FR-positive CTCs and the nodal stage subgroups (N1a, N1b, N2a1, N2a2, and N2b), the combination of pN and positive LNs (nN) (pN-nN) and the combination of pN and LNR (pN-LNR), as shown in **Figure 4**.

In terms of nodal stage, we found that CTCs had different distributions in N1 (N1a and N1b) and N2 (N2a1, N2a2 and N2b) diseases and showed good ability to distinguish between stages in intergroup comparisons. As the nodal stage increased, the crest of the distribution curves of FR-positive CTCs showed a

posterior shift, which is shown in Figure S4, Supporting Information. We then classified patients with pN1 or pN2 NSCLC into seven categories: pN1-nN1 (1 positive LN), pN1-nN2 (2 positive LNs), pN1-nN3 (3 positive LNs), pN1-nN4-5 (4-5 positive LNs), pN2-nN1-3 (1-3 positive LNs), pN2-nN4-6 (4-6 positive LNs) and pN2-nN7-14 (7-14 positive LNs). Similar distribution curves were observed in the different subgroups of positive LNs (nN) (Figure 4a,d). In the correlation analysis, the pN1a-nN (single N1 station), pN2a-nN (single N2 station), and pN2b-nN (multiple N2 station) classifications showed significant positive correlations with CTCs, but not the pN1b subtype (Figure 4b,e). Moreover, regarding LNR, we found that pN-LNR showed a positive correlation with the CTC level regardless of pathological N1 (pN1) or N2 (pN2) status. To discriminate the associations of single-station LNM and multiple-station LNM to the CTC level, we also conducted subgroup analyses of N1a versus N1b and N2a versus N2b. Positive correlations of single- and multiple-station LNM with CTC level were still observed and achieved good consistency (Figure 4c,f).

3. Discussion

In this study, we found that FR-positive CTC levels were correlated with smoking ($P < 0.001$), adenocarcinoma subtype

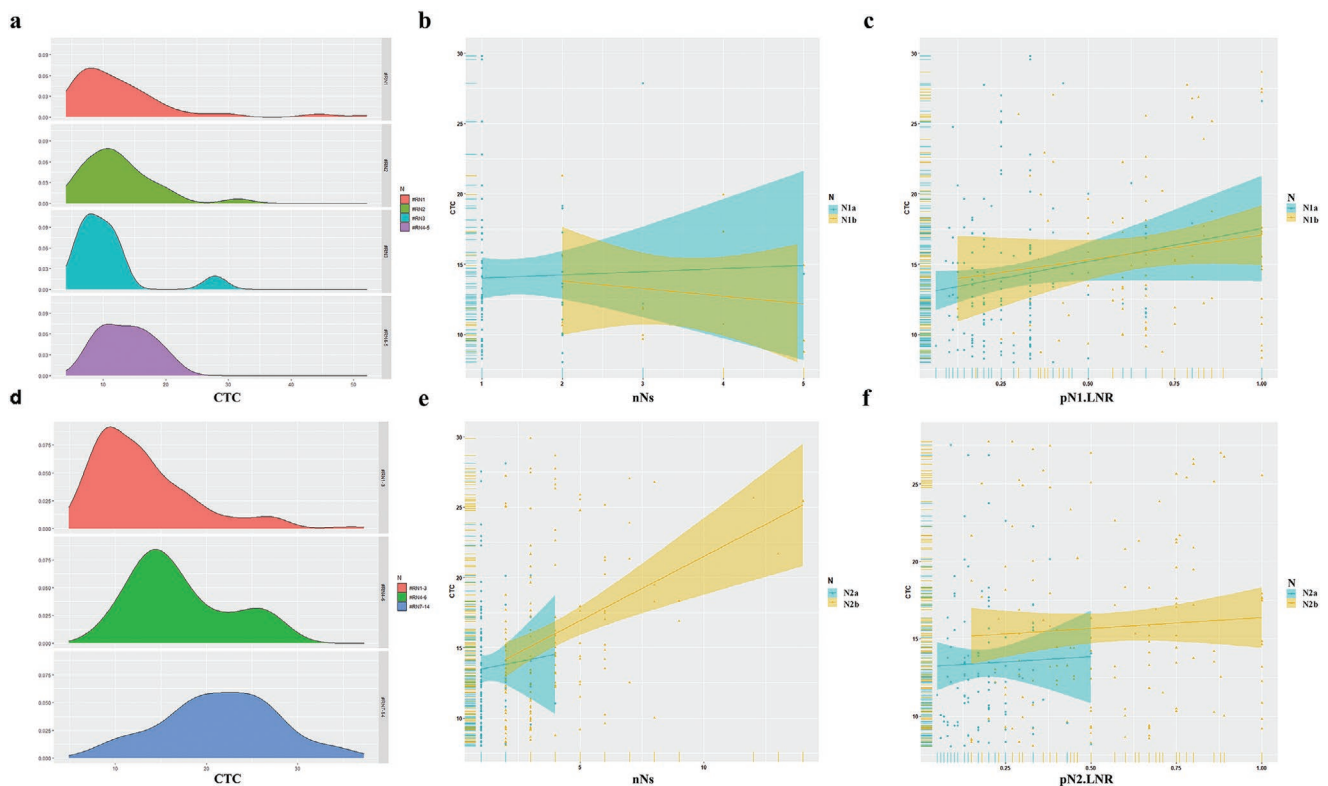


Figure 4. The correlation between FR-positive CTCs and nodal stage. a) The distribution of FR-positive CTC level in different groups stratified by the number of positive LNs in patients with N1 disease. b) The correlation between the FR-positive CTC level and pN1-nN stratified by the involvement of a single (N1a) or multiple (N1b) N1 station LNs. c) The correlation between the FR-positive CTC level and the pN1-LNR stratified by the involvement of a single (N1a) or multiple (N1b) N1 station LNs. d) The distribution of FR-positive CTC levels among different groups stratified by the number of positive LNs in patients with N2 disease. e) The correlation between the FR-positive CTC level and pN2-nN stratified by the involvement of a single (N2a) or multiple (N2b) N2 station LNs. f) The correlation between the FR-positive CTC level and the pN2-LNR stratified the involvement of a single (N2a) or multiple (N2b) N2 station LNs.

($P < 0.001$) and TNM stage ($P < 0.001$). We developed and validated a diagnostic, FR-positive CTC-based model for the individualized preoperative prediction of LNM in patients with LAD, confirming that FR-positive CTCs have value for diagnosing LNM preoperatively. The nomogram incorporated preoperative CTC level, tumor size, absolute leucocyte count, absolute monocyte count, and levels of the tumor markers CA153 and CEA and successfully stratified patients according to LNM status and has clinical value. Furthermore, we also found that the FR-positive CTC level was associated with nodal stage, nodal subgroup (N1a, N1b, N2a1, N2a2 and N2b), number of positive LNs (nN), and the combination of pN and LNR (pN-LNR) ($P < 0.05$).

We constructed a diagnostic model based on FR-positive CTCs because CTCs have a potential diagnostic and prognostic association with LNM. CTC analyses have become a diagnostic tool in the lung cancer field. Tong and his colleagues reported a prospective study^[16] and were the first to show that CTC testing could detect early-stage lung cancer in clinically undetectable patients 1–4 years earlier. Zhou and his colleagues used the CytoploRar kit platform to enrich CTCs, which led to the a high diagnostic efficacy (AUC: 0.823), with a 73.2% sensitivity and 84.1% specificity,^[10] suggesting that CTCs could be a potential biomarker for the early diagnosis of lung cancer. Moreover, CTCs have been positively associated with LNM, both CTCs and LNM are prognostic indicators of lung cancer. Two meta-analyses^[12,13] on the prognostic relevance of CTCs in patients with metastatic lung cancer have indicated that the CTC count is a reliable prognostic factor in patients with NSCLC, revealing that the CTC level was statistically associated with primary tumor size, clinical stage, lymph node involvement, and distant metastases. CTC analyses offer much promise as a real-time “liquid biopsy” for the preoperative diagnosis of LNM in LAD.

Considering strong evidence indicating the association between CTC level and LNM and that both are associated with tumor metastasis, CTCs could be used as a sensitive biomarker to isolate early-stage lung cancer patients with a negative/positive LNM status.^[17] Based the correlation analysis shown in Figure 2a, CTCs were strongly related to smoking status and tumor size, and the Mann–Whitney U test conducted between CTC level and smoking status (Figure 2b) and the Kruskal–Wallis test conducted between CTC level and tumor stage (Figure 2d) showed significant associations, further corroborating the former correlated analysis. Smoking is considered a significant risk factor for cancer, while tumor size may reflect invasiveness, as was also found in other research.^[18,19] In the subgroup analysis of LNM status, the average CTC level in the LNM-positive subgroups was 12.10 (8.26–15.84) units, which was significantly higher than the average of 10.43 (7.91–13.37) units in the LNM-negative subgroups. Surprisingly, nodal stage and the novel nodal classifications had strong correlations to CTC level, which supported our previous hypothesis and the results of the meta-analysis above.^[8,12,13] In addition, the CTC level was included in the logistic regression model following LASSO regression analysis, indicating that CTC level was an independent risk factor for LNM. The coefficient of CTCs was larger than that of tumor markers, suggesting that CTCs are more sensitive than conventional serum oncological indexes,

which was consistent with other evidence confirming that FR-positive CTCs, as an effective auxiliary diagnostic marker for NSCLC, are superior to conventional tumor markers such as CEA and CA153.^[20]

In our study, 52 preoperative clinical variables were included in the feature selection process, for which histopathologic tumor information was not considered. There are two major reasons for this. First, our study was designed based on clinical parameters in the preoperative diagnostic setting. Second, most of the patients in our cohort had early-stage LAD, which manifested as peripheral nodules on radiological images, making it impossible for clinical physicians to perform an invasive preoperative tumor biopsy, endobronchial examination or direct lymph node staging, so these pathologic features were not included in the LASSO analysis and multivariate logistic regression analysis; this differed from other diagnostic model studies on the risk of LNM among lung cancer patients.^[21,22] The 52 candidate features were reduced to six potential predictors by examining the predictor–outcome association and shrinking the regression coefficients with the LASSO method, which is superior to the method of choosing predictors on the basis of the strength of their univariate association with outcome. Tumor size, absolute leucocyte count, absolute monocyte count, and CA153 and CEA levels were identified as predictors of LNM, and these factors have similarly been shown to be predictive factors in previous reports.^[21,23–25] We are the first to report CTCs as an independent predictor of LNM.

Furthermore, we explored the role of CTCs in the preoperative LNM diagnostic model in terms of diagnostic and predictive value. The FR-positive CTC-based diagnostic model was constructed with the training cohort and consisted of preoperative CTC level, tumor size, absolute leucocyte count, absolute monocyte count, and CA153 and CEA levels, while model 2 comprised the same clinical parameters but without CTC level. When compared to model 2, model 1, which included CTC level, showed an improved prediction performance and can provide additional value in the prediction of LNM, which provides the basis for the next study on the clinical application of this model as a nomogram. Thus, we established a nomogram to predict LNM preoperatively based on multivariate logistic regression model 1. The nomogram exhibited excellent prediction and calibration capacities and achieved an average AUC of 0.786 through internal and external validation, suggesting good clinical application value.

Additionally, we comprehensively assessed the distribution of preoperative FR-positive CTCs in groups stratified by different nodal stages, the combination of pN and positive lymph node number (pN-nN), and the correlation between FR-positive CTCs and the combination of pN and LNR (pN-LNR). The reason behind this is that the current pathologic nodal classification (pN) based on the anatomic location of the involved lymph nodes is unsatisfactory in distinguishing heterogeneous pN1 or pN2 NSCLC. A more adequate nodal assessment could further stratify patients into subgroups to more precisely predict survival and individualize treatment.^[26] The number of positive LNs (nN) and the ratio of positive LNs to removed LNs (LNR) are regarded as more robust relevant prognostic factors for N+ NSCLC than conventional anatomic location-based pathological nodal stage (pN).^[26–28] Based on the present evidence,

our results, which are illustrated in Figure 4a,d, showed good discrimination among different pN-nN subgroups, especially in discriminating pN1-nN ≥ 3 from pN1-nN 1–2 and pN2-nN1-6 from pN2-nN7-14. The distribution curve and range of CTC levels had noted variations. This result was consistent with a previous study that reported differences in prognosis across pN–nN subgroups, suggesting that CTCs have the same prognostic risk stratification function as LN stage.^[26] Regarding the correlation between CTCs and pN–nN, not only the number of positive LNs but also the number of involved LN stations affected CTC levels. Thus, as seen in Figure 4b,e, the patients were stratified by N1a, N1b, N2a, and N2b. Only N1b was not positively correlated with CTCs, meaning that for patients with N1 disease, when multiple N1 LN stations were involved, the CTC level did not change significantly as the number of positive N1 LNs detected increased. This result was in contrast to that in the N2b subgroup. These results may be attributed to the difficulty in separating N1 LNs from the resected lung tissue, leading to an underestimation of the removed LNs. Moreover, Figure 4c,d indicates that the preoperative FR-positive CTC level had a good correlation with the LNR; even when patients were stratified by the number of LN stations involved, the correlation still existed after controlling for potential confounders. The correlation analysis of CTC level and pN-LNR showed more consistent results than that of CTC level and pN–nN, indicating the superior prognostic value of pN-LNR for the disease burden. To the best of knowledge, this is the first investigation on distribution and correlation between FR-positive CTCs and novel nodal classifications such as pN–nN and pN-LNR. We propose that CTCs, a potential biomarker for the revised nodal classification, can help stratify patients into pN1 or pN2 subgroups to systemically predict their LNM status. Further studies should investigate the predictive and prognostic value of FR-positive CTCs with pN–nN and pN-LNR.

The study limitations include the fact that preoperative histologic diagnoses through tissue biopsy were not available for some early-stage lung cancer patients in our cohorts; although a histologic diagnosis has been proven to be an independent risk factor for LNM in other studies, obtaining this variable was not suitable for our study design. Second, this was a retrospective study, but we applied internal and external validation to eliminate the intrinsic bias inherent to the study design. Future multicenter studies, ideally with prospective data obtained via population-based screening, are warranted to confirm our findings.

4. Conclusion

This study presents a diagnostic model constructed with internal and external validation that incorporates both FR-positive CTC level and clinical risk factors and revealed that FR-positive CTCs serve as a predictive biomarker that can be conveniently used to facilitate individualized preoperative predictions of LNM in patients with LAD.

5. Experimental Section

Patients: Patients who underwent radical lung resection with systemic LN dissection by thoracotomy or video-assisted thoracoscopic

surgery (VATS) in the Department of Thoracic Surgery at the Shanghai Pulmonary Hospital from January 2013 to December 2019 were eligible for this retrospective study. The inclusion criteria were as follows: 1) pathological diagnosis of LAD based on hematoxylin-eosin (HE) and immunohistochemistry (IHC) staining performed by pathologists from the same hospital and classified according to the 2015 World Health Organization (WHO) recommendations and TNM staging performed in accordance with the International Association for the Study of Lung Cancer (IASLC) staging system (the 8th edition); 2) undergoing radical lung resection (lobectomy, bi-lobectomy, pneumonectomy, and segmentectomy) with systemic LN dissection by thoracotomy or VATS; at least 3 stations of mediastinum LNs and 3 stations of intrapulmonary LNs should be dissected, meanwhile at least twelve LNs should be dissected totally in accordance with the European Society of Thoracic Surgeons Guidelines^[29] and Chinese Medical Association Guidelines for Clinical Diagnosis and Treatment of Lung Cancer (2019 edition,^[30]) 3) sufficient amount of peripheral blood collected before surgery. The exclusion criteria were as follows: 1) history of cancer prior to the diagnosis of lung cancer; 2) recent pulmonary infection; 3) previous treatment with anticancer agents, corticosteroids, or other nonsteroid anti-inflammatory drugs; and 4) multiple lung nodules suspected to be multiple primary lung cancers. Patient demographic and clinical characteristics, including age, sex, smoking status, operative procedure, pathological diagnosis, and stage, were recorded. Patients were grouped as LN-negative (N0) versus LN-positive (N1+N2). Subgroup analyses were stratified by age, sex, smoking history, histopathologic subtype, TNM stage, nodal stage, and gene mutation status. The patients' selection criteria for the training and validation group were illustrated in Figure S5, Supporting Information. This study was approved by the Ethics Committee of Shanghai Pulmonary Hospital.

FR-Positive CTCs Analysis: Processing of Blood Samples: A catheter was used to prevent contaminating the epithelial cells, and the first 5 mL of blood was discarded. Next, peripheral blood samples were collected in 6-mL EDTA-containing tubes (BD Diagnostics, Sparks, MD). The samples were stored at 4 °C and processed within 24 h of blood collection. Blood samples were taken from an antecubital vein in all patients who were scheduled for surgery. The technologists who performed CTC detection were blinded to the sample groups. All samples were deidentified before the technologists received them.

Detection of CTCs: The CytoploRare Circulating Lung Cancer Cell Kit was provided by GenoSaber Biotech Co Ltd. (Shanghai, China). The fundamental design of this kit was previously described and was modified to detect CTCs in patients with lung cancer.

In short, the method comprised two components: one was for CTC enrichment based on negative enrichment by immunomagnetic beads, and the other was for CTC detection and quantification based on ligand-targeted PCR. The primer sequences were as follows: reverse transcription (RT) primer (an oligonucleotide that is conjugated to the tumor-specific ligand folic acid), 5'-CTCAACTGGTTCGTGGAGTCCGGCAATTGAGTTGAGGGTTCTAA-3'; forward primer, 5'-TATGATTATGAGGCA-TGA-3'; reverse primer, 5'-GGTGTCTGGAGTTCG-3'; TaqMan probe, 5'-FAM-CAGTTGAGGGTTC-MGB-3'.

Following the manufacturer's instruction manual, red cell lysis buffer was added to 3 mL of whole blood to lyse erythrocytes (vol/vol, 1:4) for 15 min at 4 °C and then depleted of leukocytes with 150 μ L of anti-CD45 magnetic beads and macrophages with 50 μ L of anti-CD14 beads for 30 min at 4 °C. Next, the enriched CTCs were incubated with 10 μ L of labeling buffer that contained conjugates of the tumor-specific ligand folic acid and a synthesized oligonucleotide for 40 min at room temperature. The samples were then washed three times with 1 mL of wash buffer and centrifuged at 500 rpm for 10 min at 4 °C to remove the unbound conjugates. The specific ligand-oligonucleotide conjugates were removed with 120 μ L of stripping buffer for 2 min at 4 °C, collected by centrifugation, and neutralized by 24 μ L of neutralization buffer for further RT-PCR analysis.

For the PCR analysis, RT-PCR and data collection were performed with an ABI 7300 StepOne System (Life Technologies, Carlsbad, CA). The following reaction conditions were used on the ABI 7300 instrument:

denaturation at 95 °C for 2 min, annealing at 40 °C for 30 sec, extension at 72 °C for 30 sec, and then cooling at 8 °C for 5 min; and 40 cycles of denaturation at 95 °C for 10 sec, annealing at 35 °C for 30 sec, and extension at 72 °C for 10 sec.

A self-defined measurement, named the CTC unit, which was defined as the number of CTCs detected in 3 mL of blood, was used in our study. If one CTC was detected in 3 mL of blood, it was defined as one CTC unit. A series of standards containing oligonucleotides (from 10^{-14} to 10^{-9} M, corresponding from 2 to 2×10^5 CTC units) were used for CTC quantification. All blood samples were tested in duplicate with six standards and three quality controls.

Immunofluorescence Staining of Enriched CTCs: The enriched CTCs from 3 mL of blood were fixed with methanol for 10 min at -20 °C. Then, these cells were washed with phosphate-buffered saline for 1 min and stained with phycoerythrin-conjugated pancytokeratin monoclonal antibody (sc-8018PE; Santa Cruz Biotechnology, Dallas, TX) and Alexa Fluor 488-conjugated folic acid that was synthesized by WuXi AppTec. Co (Shanghai, China) for 1 h at room temperature. After washing with phosphate-buffered saline, the samples were finally mounted with 4'-6-diamidino-2-phenylindole-containing mounting media (D3571; Life Technologies). Imaging was carried out with an inverted Nikon ECLIPSE TE2000-S microscope (Nikon, Tokyo, Japan).

Clinicopathologic Data Collection: Demographic data, such as age, sex and smoking status, of the patients were retrospectively collected from a prospectively organized database, and clinicopathologic data were retrospectively obtained from clinical records. A series of laboratory examination indexes such as D-dimer, fibrinogen degradation product, fibrinogen, absolute leucocyte count, potassium, sodium, chlorine, calcium, hematocrit, absolute lymphocyte count, lymphocyte count%, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration (MCHC), mean corpuscular volume (MCV), absolute monocyte count, monocyte%, absolute neutrophilic granulocyte count, neutrophilic granulocyte%, mean platelet volume, platelet distribution width, large platelet ratio, platelet, erythrocyte, hemoglobin, albumin (alb), total protein, globulin (glb), prealbumin, alb/glb ratio, uric acid, urea nitrogen, alkaline phosphatase, glutamylphosphatase, glutamic-oxalacetic transaminase, direct bilirubin, total bilirubin, r-glutamyl transferase, isozyme of glutamic oxalacetate transaminase, creatinine and tumor markers such as NSE, AFP, B2-MG, CA-153, CA-199, CYFRA21-1, CA242, CA50, CA72-4, CEA, N-MID, SCC, SF and pro-GRP were retrieved without filtration. Tumor size was measured on contrast-enhanced chest CT scans acquired 2 weeks before the scheduled surgery if the patient underwent multiple chest CT examinations. A radiologist retrospectively reviewed the CT images and collected data regarding tumor size and tumor location and excluded the presence of multiple primary lung lesions. We reviewed the postoperative pathologic reports to determine the LAD subtype, LNM status and presence of visceral pleural invasion. The demographic data, preoperative CTC level, and TNM stage were used for baseline comparisons of clinicopathologic features between the training and validation cohorts. In addition, pathologic data were excluded because we aimed to develop the model in a preoperative setting.

Feature Selection and Development of Prediction Model: We used the LASSO method in logistic regression to select the most useful predictive features for preoperative LNM from the training cohort. The LASSO method is a penalized technique for variable selection and is suitable for the regression of high-dimensional data. Finally, backward elimination was applied to decline the number of remaining final features. All the selected key features of each sequence were combined and were introduced to the multivariate logistic regression to build the prediction model. The backward stepwise selection was applied with Akaike's information criterion (AIC) as the stopping rule. The predictive accuracy, sensitivity, and specificity were calculated using the area under the receiver operating characteristic curve (AUC) in the training group.

Development, Performance, and Validation of Radiomics Nomogram: Similarly, the aforementioned clinical candidate predictors were tested in the stepwise multivariate logistic regression model to develop a nomogram for predicting LNM in the training set and AIC as the

stopping rule. To provide a more understandable outcome measure, a nomogram was then constructed by the selected predictors. The AUCs quantified the discrimination performance of established models. The AUCs of models were compared using a DeLong test.^[31] The nomogram's calibration was assessed with a calibration curve by plotting via bootstrapping with 1000 resamples, and the goodness of fit was assessed with the Hosmer-Lemeshow test.^[32] The nomogram's performance was then tested in the validation set using the formula derived from the training set.

Statistical Analysis: The CTC units are presented as the median and interquartile range. We compared the CTC units between two groups using the Mann-Whitney U test and among three groups or more using the Kruskal-Wallis test. LASSO binary logistic regression was performed using the "glmnet" package. The "rms" package was used to perform multivariate binary logistic regression and create the nomogram and calibration plots. The C-index was calculated using the "Hmisc" package. Internal validation was performed with the "rms" package. Statistical analysis was conducted with R software (version 3.0.1; <http://www.Rproject.org>). The reported statistical significance levels were all two-sided, with statistical significance set at 0.05.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

Z.L, K.X., and L.X. contributed equally to this work. This work was supported by the National Natural Science Foundation of China (NSFC) (No. 51872205, 51922077), National Program on Key Basic Research Project of China (Youth 973 Program) (NO. 2020YFA0211100), and Cultivation Project of Shanghai Pulmonary Hospital (No Fkcx1905, fk1911).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

Research data are not shared.

Keywords

folate receptor-positive circulating tumor cells, lymph node metastasis, non-small cell lung cancer, prediction models

Received: February 8, 2021

Revised: March 18, 2021

Published online:

- [1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, A. Jemal, *Ca-Cancer J. Clin.* **2018**, *68*, 394.
- [2] W. Chen, R. Zheng, P. D. Baade, S. Zhang, H. Zeng, F. Bray, A. Jemal, X. Q. Yu, J. He, *Ca-Cancer J. Clin.* **2016**, *66*, 115.
- [3] R. L. Siegel, K. D. Miller, A. Jemal, *Ca-Cancer J. Clin.* **2018**, *68*, 7.
- [4] P. Goldstraw, K. Chansky, J. Crowley, R. Rami-Porta, H. Asamura, W. E. Eberhardt, A. G. Nicholson, P. Groome, A. Mitchell, V. Bolejack, *J. Thorac. Oncol.* **2016**, *11*, 39.

- [5] H. Asamura, K. Chansky, J. Crowley, P. Goldstraw, V. W. Rusch, J. F. Vansteenkiste, H. Watanabe, Y.-L. Wu, M. Zielinski, D. Ball, R. Rami-Porta, *J Thorac Oncol* **2015**, *10*, 1675.
- [6] S. Call, C. Obiols, R. Rami-Porta, *J Thorac. Dis.* **2018**, *10*, S2601.
- [7] Y. Zhao, D. Chen, Y. Chen, *Zhongguo Fei Ai Za Zhi* **2018**, *21*, 547.
- [8] B. Prabhakar, P. Shende, S. Augustine, *Biomed. Pharmacother.* **2018**, *106*, 1586.
- [9] S. Sheikhbahaei, E. Mena, A. Yanamadala, S. Reddy, L. B. Solnes, J. Wachsmann, R. M. Subramaniam, *AJR Am. J. Roentgenol.* **2017**, *208*, 420.
- [10] L. Zhou, D. T. Dicker, E. Matthew, W. S. El-Deiry, A. RK, *F1000Research* **2017**, *6*, 1445.
- [11] F. Castro-Giner, N. Aceto, *Genome Med.* **2020**, *12*, 31.
- [12] X. L. Ma, Z. L. Xiao, L. Liu, X. X. Liu, W. Nie, P. Li, N.-Y. Chen, Y.-Q. Wei, *Asian Pac. J. Cancer Prev.* **2012**, *13*, 1137.
- [13] J. Wang, K. Wang, J. Xu, J. Huang, T. Zhang, *PLoS One* **2013**, *8*, e78070.
- [14] B. Tong, M. Wang, *Future Oncol.* **2019**, *15*, 2531.
- [15] W. Yin, J. Zhu, B. Ma, G. Jiang, Y. Zhu, W. He, Y. Yang, Z. Zhang, *Small* **2020**, *16*, 2001695.
- [16] M. Ilie, V. Hofman, E. Long-Mira, E. Selva, J. M. Vignaud, B. Padovani, J. Mouroux, C.-H. Marquette, P. Hofman, *PLoS One* **2014**, *9*, e111597.
- [17] K. Kawada, M. M. Taketo, *Cancer Res.* **2011**, *71*, 1214.
- [18] A. Yannoutsos, M. Fontaine, A. Galloula, D. Damotte, G. Chatellier, P. Paterlini-Bréchet, G. Meyer, J. Pastre, V. Duchatelle, V. Marini, K.-L. Schwering, I. Lazareth, P. Ghaffari, A. Stansal, H. Sanson, C. Labrousse, H. Beaussier, N. B. Nasr, M. Zins, S. Salmeron, E. Messas, J.-P. Lajonchère, J. Emmerich, P. Priollet, J. Trédaniel, *BMC Cardiovasc. Disord.* **2019**, *19*, 212.
- [19] J. Yu, Y. Deng, T. Liu, J. Zhou, X. Jia, T. Xiao, S. Zhou, J. Li, Y. Guo, Y. Wang, J. Zhou, C. Chang, *Nat. Commun.* **2020**, *11*, 4807.
- [20] Y. Yu, Z. Chen, J. Dong, P. Wei, R. Hu, C. Zhou, N. Sun, M. Luo, W. Yang, R. Yao, Y. Gao, J. Li, G. Yang, W. He, J. He, *Transl. Oncol.* **2013**, *6*, 697.
- [21] N. Ding, Y. Mao, S. Gao, Q. Xue, D. Wang, J. Zhao, Y. Gao, J. Huang, K. Shao, F. Feng, Y. Zhao, L. Yuan, *J. Thorac. Dis.* **2018**, *10*, 4061.
- [22] R. G. Vaghjiani, Y. Takahashi, T. Eguchi, S. Lu, K. Kameda, Z. Tano, J. Dozier, K. S. Tan, D. R. Jones, W. D. Travis, P. S. Adusumilli, *J. Thorac. Oncol.* **2020**, *15*, 792.
- [23] C. Huang, J. Yue, Z. Li, N. Li, J. Zhao, D. Qi, *Tumour Biol.* **2015**, *36*, 7581.
- [24] C. Zhang, G. Pang, C. Ma, J. Wu, P. Wang, K. Wang, *J Immunol. Res.* **2019**, *2019*, 1.
- [25] J. Wang, Z. Hui, Y. Men, J. Kang, X. Sun, L. Deng, Y. Zhai, W. Wang, N. Bi, J. Liang, J. Lv, Z. Zhou, Q. Feng, Z. Xiao, D. Chen, L. Wang, J. Zhao, *Ann. Thorac. Surg.* **2019**, *108*, 1701.
- [26] X. Ding, Z. Hui, H. Dai, C. Fan, Y. Men, W. Ji, J. Liang, J. Lv, Z. Zhou, Q. Feng, Z. Xiao, D. Chen, H. Zhang, W. Yin, N. Lu, J. He, L. Wang, *J. Thorac. Oncol.* **2016**, *11*, 1565.
- [27] J. P. Wisnivesky, J. Arciniega, G. Mhango, J. Mandeli, E. A. Halm, *Thorax* **2011**, *66*, 287.
- [28] S. Jonnalagadda, J. Arciniega, C. Smith, J. P. Wisnivesky, *Cancer* **2011**, *117*, 4724.
- [29] D. Lardinois, P. De Leyn, P. Van Schil, R. R. Porta, D. Waller, B. Passlick, M. Zielinski, T. Lerut, W. Weder, *Eur. J. Cardiothorac. Surg.* **2006**, *30*, 787.
- [30] Chinese Medical Association Lung Cancer Clinical Diagnosis and Treatment Guidelines (2019 Edition) *Chin. J. Oncol.* **2020**, *42*, 257.
- [31] E. R. DeLong, D. M. DeLong, D. L. Clarke-Pearson, *Biometrics* **1988**, *44*, 837.
- [32] A. A. Kramer, J. E. Zimmerman, *Crit. Care Med.* **2007**, *35*, 2052.